Measuring Sleep and Respiration with Chest-Wall Accelerometry



Fons Schipper

Measuring Sleep and Respiration with Chest-Wall Accelerometry

A catalogue record is available from the Eindhoven University of Technology Library
ISBN: 978-90-386-6637-2
Cover design: Studio 0404
Layout and printing: ProefschriftMaken.nl

©Fons Schipper, 2026 Copyright of the individual chapters containing published articles belongs to the publisher of the journal listed at the beginning of the respective chapters. All rights reserved. No part of this thesis may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the copyright owner.

Measuring Sleep and Respiration with Chest-Wall Accelerometry

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven,
op gezag van de rector magnificus prof. dr. S.K. Lenaerts,
voor een commissie aangewezen door het College voor Promoties,
in het openbaar te verdedigen op dinsdag 17 maart 2026 om 13:30 uur

door

Alphonsus Tarcisius Jozef Maria Schipper

geboren te Purmerend

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

Voorzitter:	Prof. dr. ir. P.G.M. Baltus
Promotoren:	Prof. dr. S. Overeem Dr. ir. R.J.G. van Sloun
Copromotor:	Dr. ir. P. Fonseca
Leden:	Prof. dr. J. Verbraecken (Universitair Ziekenhuis Antwerpen) Prof. dr. R.M. Aarts Dr. E. Peri
Adviseur:	Prof. dr. N. de Vries (Universiteit van Amsterdam)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Table of Contents

Summary 7

Samenvatting 9

Introduction 13
- The patient 15
- The sleep study 18
- The diagnosis 20
- Therapy for positional OSA 22
- Cardiorespiratory measurements from chest-wall accelerometry 24
- Estimation of AHI and sleep stages from cardiorespiratory measurements 27
- Research goals and thesis outline 29

1 Respiratory Effort Estimation with Principal Component Analysis 35
- 1.1 Abstract 36
- 1.2 Introduction 37
- 1.3 Methods 38
- 1.4 Results 47
- 1.5 Discussion 53

2 Respiratory Effort Estimation with Deep Learning 59
- 2.1 Abstract 60
- 2.2 Introduction 61
- 2.3 Methods 63
- 2.4 Results 74
- 2.5 Discussion 79

3 Maximum A Posteriori Detection of Heartbeats 85
- 3.1 Abstract 86
- 3.2 Introduction 87
- 3.3 Methods 89
- 3.4 Results 97
- 3.5 Discussion 100

4 Sleep Stage Estimation using Chest-Wall Accelerometry 105
4.1 Abstract 106
4.2 Introduction 107
4.3 Methods 109
4.4 Results 113
4.5 Discussion 117

5 Apnea-Hypopnea Index Estimation using Chest-Wall Accelerometry 121
5.1 Abstract 122
5.2 Introduction 123
5.3 Methods 124
5.4 Results 131
5.5 Discussion 135

6 General Discussion and Perspectives 141
6.1 Research goals 143
6.2 Interpretation of results 144
6.3 Strengths and Limitations 150
6.4 Applications 153
6.5 Future Perspectives 157
6.6 Conclusions 162

Dankwoord 165

About the author 167

Publications 169

The TU/e sleep series 171

References 173

Summary

Sleep is important for daily wellbeing and long-term health. Obstructive sleep apnea (OSA) is a common sleep disorder that deteriorates sleep quality and has various adverse health effects. OSA is caused by repetitive partial or complete obstruction of the upper airways during sleep, leading to hypopneas or obstructive apneas respectively, typically followed by oxygen desaturations and arousals.

Despite several limitations, the severity of obstructive sleep apnea (OSA) is typically measured using the apnea-hypopnea index (AHI), which counts the number of obstructive apneas and hypopneas per hour of sleep. Monitoring the AHI is useful during OSA treatment, as it provides an indication of whether respiratory events are adequately addressed, or, in case it increases, that therapy might require adjustments. Ideally, sleep itself should also be monitored during treatment, but this is not yet feasible.

Positive airway pressure (PAP) is an often-used therapy for OSA. Typically, PAP devices also monitor the AHI by analysing the airflow signal, but do not assess sleep itself. Besides PAP therapy, alternative OSA treatments have been developed. An interesting alternative therapy targets specifically positional OSA, a phenotype where apnea and hypopnea events predominantly occur in supine posture. Positional therapy (PT) devices use a chest-worn accelerometer to assess posture and provide haptic feedback to avoid the supine position, reducing the burden of OSA. However, PT devices currently lack AHI measurement, as they typically are not able to measure airflow, and thus, miss a direct indication of decreased or interrupted breathing.

This thesis focuses on the development of a method to assess sleep and estimate the AHI using chest-wall accelerometry. The first part of the thesis explores the development of novel techniques to extract cardiorespiratory metrics from a chest-worn accelerometer, focusing on reliability during overnight use by patients with sleep disorders. Chapters 1 and 2 introduce novel methods to estimate respiratory effort. In chapter 1, respiratory effort is estimated using an adapted version of principal component analysis, which leads to results with substantially lower error than state-of-the-art methods for estimating respiratory effort in healthy individuals. To further improve performance and robustness in larger, sleep-disordered breathing populations, chapter 2 introduces a novel regression method based on a convolutional neural network. This method shows a further reduction of error in estimating respiratory effort, even in the presence of sleep disorders. Chapter 3 presents a novel method for heartbeat localization and inter-beat interval estimation. The method utilizes maximum a posteriori estimation, and an optimal solution is approximated by means of a Markov decision process.

The second part of the thesis investigates how methods, originally designed for estimating sleep stages and AHI from cardiorespiratory signals, obtained with electrocardiography (ECG) and respiratory belts, can also work with signals derived from a chest-worn accelerometer. In chapter 4, we investigate whether a neural network for cardiorespiratory sleep staging can be applied to the signals derived from a chest-worn accelerometer. We show that total sleep time, required for the computation of the AHI, can be computed accurately, and that estimated sleep stages show substantial agreement with those scored from polysomnography. In chapter 5, we adapt a neural network for AHI estimation from cardiorespiratory metrics to the specific characteristics of chest-wall accelerometry by means of transfer learning. The estimated AHI and corresponding severity classification reach a high correlation and show substantial agreement with that scored from polysomnography.

In summary, the work presented in this thesis significantly advances the performance and robustness of estimation of cardiorespiratory metrics from chest-wall accelerometry, which allow for the accurate estimation of sleep stages and AHI. This work has several applications, for example enabling the use of chest-wall accelerometry for monitoring of residual AHI during therapy for positional OSA. Furthermore, implementation in diagnostic devices such as those for extended Holter monitoring could expand the yield of those investigations to include sleep apnea screening.

Samenvatting

Slaap is belangrijk voor het dagelijks welzijn en de gezondheid op lange termijn. Obstructieve slaapapneu (OSA) is een veel voorkomende slaapstoornis die de slaapkwaliteit verslechtert en verschillende nadelige gezondheidseffecten heeft. OSA wordt veroorzaakt door herhaaldelijke gedeeltelijke of volledige obstructie van de bovenste luchtwegen tijdens de slaap, wat leidt tot respectievelijk hypopneus of obstructieve apneus, vaak gevolgd door zuurstof desaturatie en/of slaap fragmentatie.

Ondanks verschillende beperkingen wordt de ernst van OSA meestal gemeten met behulp van de apneu-hypopneu index (AHI), die het aantal obstructieve apneus en hypopneus per geslapen uur aangeeft. Het monitoren van de AHI is nuttig tijdens de behandeling van OSA omdat het een indicatie geeft of obstructies voldoende worden voorkomen en, als er een toename is, de therapie aangepast kan worden. Idealiter zou de slaap zelf ook gemonitord moeten worden tijdens de behandeling, maar dit is nog niet haalbaar.

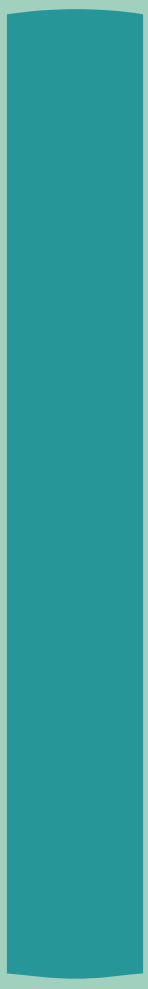
Positive airway pressure (PAP) is een veelgebruikte therapie voor OSA. PAP-apparaten meten meestal de AHI door de luchtstroom te observeren, maar beoordelen de slaap zelf niet. Naast PAP-therapie zijn er alternatieve OSA-behandelingen ontwikkeld. Een interessante alternatieve therapie is voor positieafhankelijke OSA, een fenotype waarbij apneus en hypopneus hoofdzakelijk voorkomen in rugligging. Apparaten voor positietherapie (PT) maken gebruik van een op de borst gedragen accelerometer om de houding te meten en geven voelbare feedback om de rugligging te vermijden, waardoor de last van OSA wordt verminderd. PT-apparaten hebben momenteel geen AHI-meting omdat ze geen toegang hebben tot de luchtstroom en daarom geen indicatie hebben van obstructies in de luchtweg.

Dit proefschrift richt zich op het ontwikkelen van een methode om de slaap te beoordelen en de AHI te schatten met behulp van accelerometrie op de borstwand. Het eerste deel van het proefschrift richt zich op het extraheren van cardiorespiratoire meetgegevens uit een op de borst gedragen accelerometer. Hoofdstukken 1 en 2 introduceren nieuwe methoden om de ademhalingsinspanning te schatten. In hoofdstuk 1 wordt een variatie op principale-componentenanalyse gebruikt om resultaten te verkrijgen met een aanzienlijk lagere fout dan de tot dan toe bekende methoden voor het schatten van de ademhalingsinspanning bij gezonde personen. Om de prestaties en robuustheid bij grotere populaties met slaapstoornissen verder te verbeteren, introduceert hoofdstuk 2 een nieuwe regressiemethode op basis van een convolutioneel neuraal netwerk. Deze methode laat een verdere reductie van de fout zien in het schatten van de ademhalingsinspanning, zelfs in de aanwezigheid van slaapstoornissen. In hoofdstuk 3 wordt een

nieuwe methode gepresenteerd voor het lokaliseren van hartslagen in tijd en het schatten van het interval tussen de hartslagen. De methode maakt gebruik van maximum a posteriori schatting en een optimale oplossing wordt benaderd door middel van een Markov beslissingsproces.

In het tweede deel van het proefschrift wordt onderzocht of methoden die oorspronkelijk zijn ontworpen voor het schatten van slaapstadia en AHI verkregen uit sensoren voor elektrocardiografie en ademhalingsplethysmografie, ook werken met meetgegevens die zijn afgeleid van een op de borst gedragen accelerometer. In hoofdstuk 4 onderzoeken we of een neuraal netwerk voor slaapstadia verkregen uit cardiorespiratoire signalen uit elektrocardiografie en ademhalingsplehysmografie, ook kan worden toegepast op de signalen van een op de borst gedragen accelerometer. We tonen aan dat de totale slaaptijd, die nodig is voor de berekening van de AHI, nauwkeurig kan worden berekend en dat de geschatte slaapstadia een aanzienlijke overeenkomst vertonen met de stadia afgeleid uit polysomnografie. In hoofdstuk 5 passen we een neuraal netwerk voor het schatten van de AHI uit cardiorespiratoire meetgegevens aan op de specifieke kenmerken van accelerometrie op de borstwand door middel van overdrachtsleren. Daarmee bereiken de geschatte AHI en bijbehorende ernstclassificatie een hoge overeenstemming met die afgeleid uit polysomnografie.

Het in dit proefschrift gepresenteerde werk verbetert de prestaties en robuustheid van cardiorespiratoire metingen via borstwandaccelerometrie aanzienlijk, waardoor de accurate schatting van slaapstadia en AHI mogelijk wordt. Dit werk heeft verschillende toepassingen, bijvoorbeeld het gebruik van accelerometrie op de borstwand voor het volgen van AHI tijdens therapie voor positieafhankelijke OSA. Bovendien kan toepassing in diagnostische apparaten zoals die voor uitgebreide Holter-monitoring de opbrengst van dergelijke onderzoeken uitbreiden met slaapapneu screening.



Introduction

Introduction

The patient

The work in this thesis should lead to technological developments that ultimately improve the lives of patients. As an example, we followed a hypothetical patient. Our patient was a 54-year-old male living in a small village in the province of Limburg, the Netherlands. He was an operator at the local power station, where he controlled power delivery to the grid. His physical activities comprised occasional cycling and gardening, he was slightly overweight and could be short of breath. At the time, he had been experiencing unexplained episodes of sleepiness. He occasionally dozed off, not only while watching TV, but also at work, which was concerning given his responsibilities. In addition, his wife had started complaining about his increased snoring at night.

Worried about these symptoms, they decided to visit their physician. During the consultation, the physician asked about his overall well-being and specific complaints. Given the daytime sleepiness and reported snoring, the physician suspected obstructive sleep apnea (OSA) and proceeded with relevant follow-up questions. The physician inquired whether our patient's wife had observed any breathing pauses and asked the patient about his sleep habits, awakenings, and whether he woke up with a headache or dry mouth.

The physician also reviewed his medical history and checked for previously reported cardiovascular conditions known to be associated with OSA, such as hypertension, heart failure, and atrial fibrillation. During the physical examination, she found no cardiac or respiratory abnormalities but noted an elevated BMI (28.7 kg/m^2) and high blood pressure (140/95 mmHg). Based on these findings, she assessed the potential presence of OSA.

Obstructive sleep apnea (OSA) is characterized by a repetitive reduction or absence of respiration (apnea), caused by obstruction in the upper airways, as indicated in the box "Obstructive Sleep Apnea". Common symptoms include excessive daytime sleepiness, sleep disturbances, and snoring. OSA can have adverse long-term health effects. Increased sympathetic activation may persist throughout the day and cause hypertension[1]. Through pathways such as intermittent hypoxia and sleep fragmentation, OSA can contribute to the development of type 2 diabetes[2]. Cardiac functioning is also impacted as OSA was found to be an independent risk factor for atrial fibrillation and heart failure[3]. OSA may be asymptomatic and remain undiagnosed. Excessive sleepiness is reported by

as few as 15% to 50% of OSA patients and there is currently no high-quality evidence of linking asymptomatic OSA to cardiovascular events [4].

OSA is a heterogeneous disease with multiple causes, including anatomical factors like a narrowed airway and non-anatomical factors such as muscle unresponsiveness or respiratory control instability. Although OSA phenotyping is emerging, the current diagnostic and reimbursement standards still rely on the frequency of the obstructions per slept hour, expressed in the **apnea-hypopnea index** (AHI). The diagnostic criteria for the establishment of OSA have been described in the International Classification of Sleep Disorders (ICSD)[5].

OSA is highly prevalent. A recent meta-study estimated the global prevalence among adults aged 65 years and older to be 35.9%. Correlated risk factors include high BMI, increasing age, and male gender[6]. With the aging global population and the obesity epidemic, the prevalence of OSA is expected to rise even further.

In the case of our patient, as the physician suspected OSA, she decided to refer him to the regional hospital for further diagnosis. Given our patient's work responsibilities and potential consequences of dozing off, she asked for a rapid follow-up. The next day, our patient received an invitation from the hospital to see a pulmonologist. At the visit one week later, the specialist reviewed the physician's notes, discussed our patient's symptoms, and performed a general examination of his lungs and upper airways. Like the physician, the pulmonologist also suspected OSA and promptly scheduled him for an overnight sleep study.

Obstructive Sleep Apnea

Normal respiration involves ventilation: air movement into and out of the lungs. A feedback system controls minute volume (respiration rate × tidal volume) to maintain stable CO_2 and O_2 blood concentrations. During inhalation, respiratory muscles contract, lowering lung pressure and increasing lung volume as air flows in. Exhalation is a passive process as muscles relax, and elastic recoil reverses the flow. Respiration is typically measured with multiple sensors. Thoracic and abdominal belts provide an estimate of the lung volume and are used as a surrogate for respiratory effort. An oronasal thermal flow sensor (thermistor) measures the sum of flow through mouth and nose A nasal pressure transducer measures the nasal airflow. An SpO_2 sensor measures the O_2 blood oxygen concentration. An example of the signals from these sensors during normal breathing is given in Figure 1.

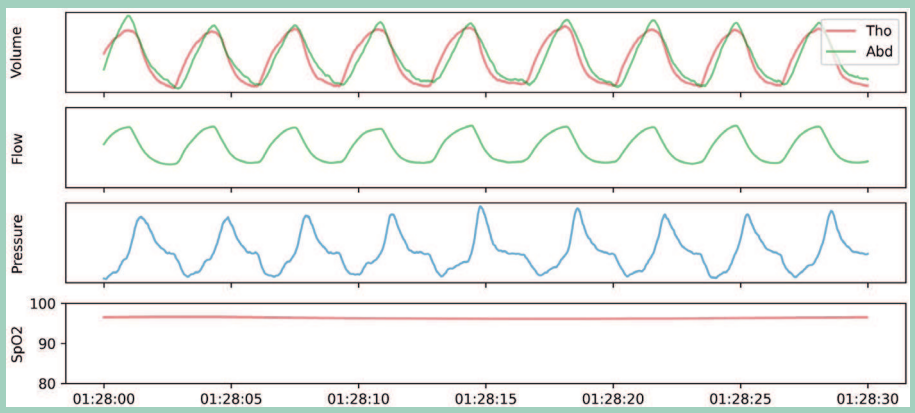


Figure 1 Signals during normal respiration (eupnea). Thoracic and abdominal volume measured by thoracic (Tho) and abdominal (Abd) respiratory belts; Airflow measured by an oronasal airflow sensor; Nasal pressure; blood oxygen saturation (SpO2).

Obstructive sleep apnea occurs when pharyngeal soft tissue collapses, reducing air flow during inhalation. An obstructive apnea is defined as a reduction in airflow of at least 90% (full collapse), while a hypopnea is defined as a reduction in airflow of ≥30% but < 90% (partial collapse) that is linked to either oxygen desaturation (SpO_2) of 3% or more, or a cortical arousal. Mechanical and chemical receptors sense the obstructive event and cause activation of the sympathetic nervous system, which initiates muscle contractions to remove the obstruction. Figure 2 displays a sequence of apnea events, causing intermittent reductions in pressure, flow, volume, and O_2 concentration (SpO_2).

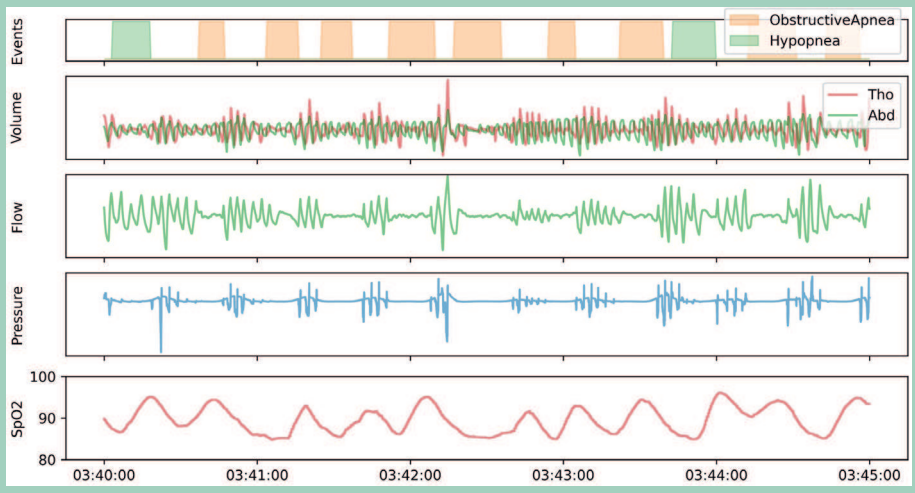


Figure 2 Signals during OSA respiration (apnea). Thoracic and abdominal volume measured by respiratory belts; Airflow measured by an oronasal airflow sensor; Nasal pressure; blood oxygen saturation (SpO_2). Airflow reduces by ≥ 90% and oxygen desaturates up to 10%.

The sleep study

The purpose of a sleep study is to measure sleep and detect sleep-related events, including apneas and hypopneas. The clinician uses the study results, along with other information, to diagnose sleep disorders. For the diagnosis of OSA, the AHI is used

$$AHI = \frac{\#apnea\ events + \#hypopnea\ events}{total\ sleep\ time}.$$

The AHI represents the frequency of **respiratory events** (here: apnea and hypopnea events) over the time slept in events per hour. To obtain the total sleep time, it is essential that sleep is measured as part of an OSA diagnosis. In addition, the measurement of sleep provides an overview of the sleep architecture (see the box "The measurement of sleep"). The sleep architecture offers crucial information for diagnosing other sleep disorders, such as those related to circadian rhythm or sleep continuity. Furthermore, understanding sleep architecture can aid in the phenotyping of OSA, such as REM-OSA, where respiratory events predominantly occur during REM sleep.

There are various types of sleep studies, depending on the physiological signals used (see the box "Types of sleep studies"). The gold standard is supervised in-laboratory polysomnography (PSG), designated as "Type I." Since a Type I sleep study is relatively expensive, alternative methods have been developed, ranging from Type II (ambulatory PSG, typically at home) to Type IV. Types III and IV sleep studies are referred to as home sleep apnea tests (HSAT), as they aim to measure OSA but typically do not (directly) measure sleep since they lack the required neurological signals. Some Type III and IV sleep tests estimate total sleep time and sleep architecture using surrogate sensors, but their agreement with a gold-standard measurement may be poor [7].

A Type III sleep test typically employs respiratory sensors, such as a thoracic belt and a nasal cannula, and cardiac sensors, like finger plethysmography and SpO2, while a type IV study does not. An example of a Type IV sleep study is peripheral arterial tonometry, which may include a sensor on the suprasternal notch to detect snoring, and usually includes a finger sensor to measure SpO2. In contrast, a Type II (or Type I) sleep study uses neurophysiological signals, including electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG), allowing for a gold-standard measurement of sleep (see the box "The measurement of sleep").

Types of sleep studies

The US Centers for Medicare & Medicaid Services (CMS) define four types of sleep study in their medicare National Coverage Determinations (NCD) manual:

- **Type I**: Supervised overnight polysomnography in a sleep laboratory. Considered the reference standard for the assessment of sleep, including the diagnosis of OSA.
- **Type II**: A portable monitor that includes sleep sensors and has a minimum of 7 channels. The sleep sensors comprise EEG, EOG, and Chin-EMG, and the remaining 4 channels are typically as in type III.
- **Type III**: A portable monitor with a minimum of 4 channels, including ventilation or airflow (2 respiratory belts or 1 respiratory belt and airflow), heart rate or ECG, and oxygen saturation. Sleep sensors are not required.
- **Type IV**: A portable monitor that measures 1, 2, 3, or more parameters but does not meet all the criteria of a higher category device.

For our patient, the pulmonologist prescribed a Type II sleep study. A few days later, he received an invitation for an overnight sleep study. The invitation included a folder instructing him to report to the hospital at 16:00, wearing loose night clothing on the upper body, with washed and dry hair and a shaved chin.

Our patient complied, arrived at the hospital, and went home equipped with sensors on the head, chest, face, and left hand. He had a quiet evening, went to bed around 22:00, and woke up at 06:30, after which he filled out a sleep diary: a log of activities, including sleep and wakefulness, as perceived by the patient.

Our patient had breakfast and returned to the hospital to have the sensors removed.

The measurement of sleep

The measurement of sleep has been standardized by the American Academy of Sleep Medicine (AASM). The measurement relies on **polysomnography** (PSG), either conducted in a clinic (type I sleep study), or at home (type II study). The patient is equipped with sensors to measure parameters including electro-encephalography (EEG: F4-M1 to O1-M2), electro-oculography (EOG: E1-M2 and E2-M2), electro-myography (EMG Chin), electro-cardiography (ECG), respiratory effort, sleeping posture, and leg movements.

All sensor signals are assembled into a polysomnogram and are subsequently divided into 30-second intervals called epochs. A sleep technician then scores each of the epochs into one of the sleep stages Wake, REM, and Non-REM, with Non-REM comprising stages N1, N2, and N3. Scorings are based on the visual analysis of patterns in the brain (EEG), eye (EOG), and chin muscle (EMG) signals. After the scoring of sleep stages for the complete night, they are visualised in a **hypnogram** showing the sleep architecture, see Figure 3. The height of a line indicates the stage, with REM sleep is usually drawn as a thicker red line. REM sleep is sometimes called paradoxical sleep, as the brain is highly active, while the skeletal muscles are paralysed. N3 sleep is sometimes referred to as slow-wave sleep, as the neurons in the cortex are all synchronized, firing in a slow rhythmic pattern.

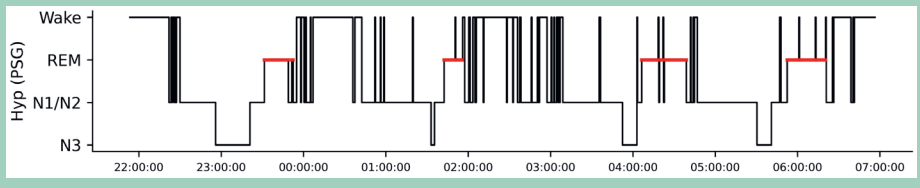


Figure 3 Example hypnogram. The study participant falls asleep around 22:30, first in light sleep (N1/N2) and shortly before 23:00 in deep sleep (N3). The first REM period starts around 23:30, while the participant wakes up around midnight. This pattern repeats three more times, with shorter periods of N3 and longer periods of REM sleep. The participant is awake for a few extended periods during the night. The total recording time (TRT) is 541 min, with a total sleep time (TST) of 394 min (73%).

The diagnosis

After removing the sensors from the patient, a pulmonary technician connected the sleep test device to a PC and uploaded the data. A quick inspection confirmed that the recording and data transfer were successful. Later in the week, the technician analysed the study data. Trained to score sleep stages and events according to the AASM rules for the scoring of sleep and associated events[8], the technician

first reviewed the initial scorings provided by the sleep test equipment's software and manually adjusted these where needed. Using the adjusted scorings, the software computed the hypnogram, sleep statistics, and event statistics (including the AHI), and generated a draft report. Finally, the technician added the sleep diary and analysis findings to the report and saved it for review by the pulmonologist.

The criteria for the diagnosis of OSA are described in the International Classification of Sleep Disorders (Third Edition, Text Revision, ICSD-3-TR)[5], and rely on OSA severity classes, which are based on the AHI: $5 \leq AHI < 15$ for Mild OSA, $15 \leq AHI < 30$ for Moderate OSA, and $AHI \geq 30$ for Severe OSA. However, the AHI alone is insufficient to diagnose OSA; the symptoms should not be better explained by other causes, and for the severity class Mild, there must be related complaints or symptoms. Current clinical guidelines indicate that diagnosed OSA should be managed, with the management including therapy[9].

One commonly used therapy, which directly targets the upper airway collapsibility, is positive airway pressure (PAP). By applying a positive pressure at the airway openings, airway collapse—the cause of the obstruction—can be effectively prevented. To facilitate the ongoing evaluation of therapy efficacy, and to allow pressure adjustments during therapy, modern PAP devices monitor the AHI on a nightly basis. Respiratory events (apneas and hypopneas) are detected through cessations and reductions of airflow and are then counted and divided by the total PAP usage time to obtain the so-called "residual AHI". Observing trends in the nightly residual AHI allow the caregiver to adjust the therapy when needed. A drawback of PAP therapy is that compliance can be problematic[10,11]. The positive pressure and the wearing of a mask can be uncomfortable and intrusive. This is one of the reasons why alternative therapies for OSA have been developed, which do not rely on PAP and therefore do not require the use of masks.

The pulmonologist reviewed the technician's report, and noted an AHI of 31 events per hour, placing it in the Severe OSA category. She concluded that the patient had OSA and required therapy, a judgment supported by the patient's complaints, his elevated BMI and blood pressure. She then considered which therapy to prescribe. Recently she had had a couple of OSA patients who had been struggling with PAP therapy. Their adherence was rather low; they only used the therapy during a part of the night or even skipped whole nights.

She then further reviewed the technician's report and noticed that the report highlighted a large difference in the AHI between the supine and non-supine sleeping positions, with a supine AHI of 45 events per hour (over 276 minutes of supine sleep) and a non-supine AHI of 5 events per hour (over 144 minutes non-supine sleep).

Therapy for positional OSA

In the context of alternative therapies, one phenotype of interest is positional OSA (POSA), where OSA is more severe in the supine sleeping position and less severe in other positions. Although various definitions have been developed for OSA being positional (see the box "Positional Sleep Apnea"), POSA was found to be highly prevalent among OSA patients. Prevalences of up to 75% have been reported, with up to 36% of OSA patients having exclusive POSA (AHI in non-supine sleep less than 5 events/h).

Positional Sleep Apnea

OSA was discovered not that long ago, in 1965, simultaneously in France by Gastaut et al.[12], and in Germany by Jung and Kuhlo[13]. Both were investigating the then so-called Pickwickian syndrome, named after an obese, daytime-napping character from the Pickwick papers by George Orwell. Later, in 1984, Cartwright et al. observed that, for some OSA patients, the AHI was much higher in the supine position than in the non-supine positions[14]. They attributed this effect to the back of the tongue being pulled down by gravity, thereby closing the upper airways. Over time, further studies have studied POSA and have used various definitions:

Cartwright, 1984[14]	$AHI_s \geq 2\, AHI_{ns}$
Marklund, 1998[15]	$AHI_s \geq 10$ and $AHI_{ns} < 10$
Mador, 2005[16]	$AHI_s > 2\, AHI_{ns}$ and $AHI_{ns} < 5$ and $AHI \geq 5$
Bignold, 2011[17]	$AHI \geq 15$ and $AHI_s \geq 2\, AHI_{ns}$ and $AHI_{ns} < 15$ and $\geq$ 20 minutes of sleep in both s and n^{o} positions
Levendowski, 2014[18]	$AHI \geq 5$ and $AHI_s \geq 1.5\, AHI_{ns}$ and $\geq$ 4 h Total Sleep Time
APOC, 2015[19]	(Establishment of OSA) and (Time slept in best and worst sleeping positions > 10% of Total Sleep Time) and (($AHI_{bsp} < 5$) or (AHI_{bsp} in lower category than overall AHI) or ($AHI \geq 40$ and $AHI_{bsp} \leq 0.75\, AHI$))

AHI_s = AHI in supine position, AHI_{ns} = AHI in non-supine position, AHI_{bsp} = AHI in best sleeping position

Therapy for POSA is in principle relatively straightforward and involves preventing the patient from sleeping in the supine position, such that periods of sleep with a higher AHI are avoided and the overall AHI is reduced. Early POSA therapies involved mounting a tennis ball on the patient's back[20]. Modern versions use an electronic device with haptic feedback (see box "Positional therapy"). These devices are worn on the chest and use an accelerometer to measure the sleeping position. Advantages of electronic versions include the ability to gradually "train" the patient to avoid sleeping in the supine posture. Additionally, these devices can monitor usage and sleeping position. Under the assumption that OSA remains positional, the time in (non-)supine position may be used as an indicator for therapy efficacy in addressing OSA, i.e. keeping the AHI low. However, this is a risky assumption.

The diagnosis of POSA is often based on only a single night of measurement. Furthermore, the percentage of time spent in non-supine positions during diagnostic PSG may be lower than during habitual sleep at home[21], partly due to physical constraints imposed by the actual monitoring equipment. Besides offering a skewed view of overall position-specific severity, it may mean that the diagnosis of POSA is solely based on the measurement of the (non-supine) AHI over a very short period. Additionally, OSA can worsen over time; aging and weight gain can increase OSA severity. It may well be that OSA (re-) occurs during non-supine positions even though it did not during the diagnostic measurement. As with PAP therapy, the ability to monitor the AHI during therapy could provide the caregiver with useful insights allowing for assessment of efficacy and potentially for the adaptation of the therapy.

Back to our patient. Given the clear positional character of OSA, the pulmonologist considered prescribing positional therapy. However, she was in doubt, especially in relation to our patient's job as a power plant operator. The overall AHI had been quite high with 31 events per hour. Although the non-supine AHI had been quite low with 5 events per hour, it had only been based on a 144-minute measurement. Sending the patient away with positional therapy without further monitoring could have left a potentially low therapy efficacy unnoticed. AHI monitoring during therapy could have mitigated the risk and provided reassurance. Despite these reservations, she decided to prescribe positional therapy. For certainty, she scheduled a follow-up appointment in three months.

Positional therapy

In electronic positional therapy[22], a device worn around the chest monitors the sleeping posture and provides vibrational stimuli when supine, see Figure 4. In this way, the wearer is conditioned to avoid the supine position. In a study among 117 participants with POSA, this therapy proved to be non-inferior to PAP and demonstrated a higher adherence[23]. The device comprises a battery, an accelerometer to measure the sleeping position, a vibrator to provide the stimuli, a memory to record the usage and sleeping posture, and a wireless connection to upload the usage and position to a phone and the cloud. In this way, adherence and effectiveness in avoiding the supine position can be reviewed.

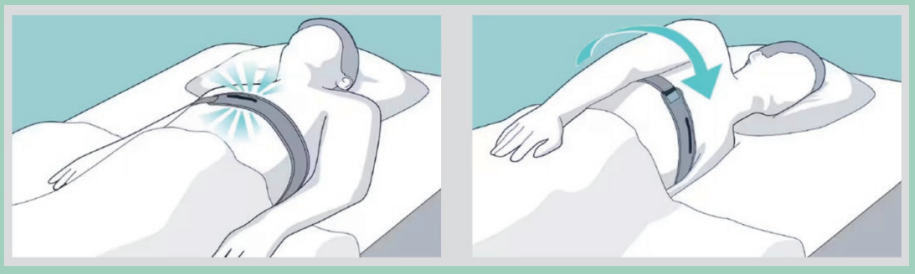


Figure 4 The working principle of the NightBalance sleep position trainer. When in supine position the device gently vibrates (left), and the wearer is conditioned to turn away into a non-supine position (right).

Cardiorespiratory measurements from chest-wall accelerometry

The work in this thesis relies on two important pillars: (1) cardiorespiratory measurements from chest-wall accelerometry and (2) estimation of sleep stages and AHI from cardiorespiratory measurements.

The choice for accelerometry is motivated by its ease of integration into positional therapy devices, which already require an accelerometer to monitor sleeping position. This allows respiratory and cardiac measurements without altering the device's enclosure or usage. In contrast, alternative sensing solutions such as RIP belts or ECG electrodes may offer higher accuracy for respiratory or cardiac signals individually, but they cannot capture both modalities with a single sensor. Moreover, their integration would require significant design changes—RIP belts involve electrical conductors in the belt, and ECG electrodes require direct skin contact, making the device unsuitable for use over clothing and sensitive to electrode wear out.

The chest-wall was selected as the sensing location primarily due to its immediate proximity to the lungs and heart. This anatomical position ensures high-quality mechanical coupling to both organs, allowing simultaneous respiratory and cardiac measurements to be obtained by a single sensor without the need for skin contact. Furthermore, placing the sensor on the chest leverages the standard mounting location of current positional therapy devices, reinforcing the goal of enabling dual-modality monitoring without altering the user experience.

Respiratory measurements

The measurement of respiration by means of chest-wall accelerometry relies on chest-wall movements due to inhalation and exhalation. Upon inhalation the chest wall expands and hinges outward, while it contracts inward upon exhalation.

In the past, studies have been performed into the use of accelerometers to measure these chest-wall movements (see box "How an accelerometer works"). In 2008, Hung et al. used a 2D Analog MEMS accelerometer mounted on the chest wall to measure respiratory waveforms, which they compared against those from respiratory impedance plethysmography[24]. Later, in 2009, Jin et al. used a 3D accelerometer in combination with principal component analysis (PCA) to estimate respiration rate[25]. More recently, in 2018, Cesareo et al. confirmed that PCA was thus far the best method for dimensionality reduction when estimating respiration rate[26]. While these studies have established the feasibility of measuring respiration with chest-wall accelerometry, we have found no larger, overnight studies, particularly for estimating the respiratory waveform in sleep-disordered patients. This is important as accurate reproduction of the waveform is necessary to identify and characterise respiratory events.

How an accelerometer works

An accelerometer is a device that measures acceleration. It typically does so by means of tiny proof masses suspended by microfabricated springs, with one mass for each of the axes x, y, and z. When the device accelerates, inertial forces cause the proof masses to shift slightly, with the movement inversely proportional to the acceleration. The movement is measured by the accelerometer and translated into an acceleration value. For example, the movement of the mass is often detected by measuring the change in electrical capacitance between the mass and nearby stationary electrodes, which is then translated into the final acceleration value. Beyond just measuring acceleration, the device also measures orientation because gravity also causes the masses to shift slightly, as is indicated in the figure. The masses and springs are typically manufactured by means of MEMS (Microelectro-

mechanical System) technology, with the proof masses and springs integrated into a die of only a few millimetres in size, see Figure 5. Almost all modern phones contain an accelerometer to measure their orientation.

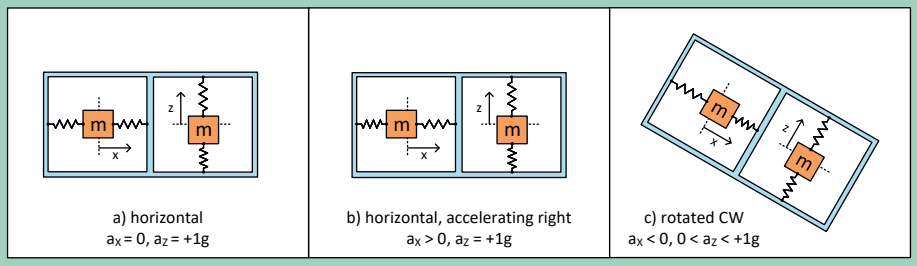


Figure 5 Proof mass displacement in response to acceleration and rotation. (a) Horizontal and stationary: gravity displaces the z-mass downward, leaving the x-mass centered. The measured acceleration is $+1g$ in the z-direction and $0g$ in the x-direction (b) Accelerating right: an inertial force displaces the x-mass to the left, causing an acceleration of $a_x > 0$ to be measured in the x-direction. (c) Clockwise rotation: gravity displaces the x-mass in the positive x-direction, causing a_x to be < 0. The z-acceleration becomes slighty less positive ($a_z < +1g$). The symbol g stands for the acceleration of gravity (≈9.8m/s^2).

Cardiac measurements

The detection of heartbeats by means of chest-wall accelerometry relies on vibrations originating from ventricular ejection. This phenomenon follows Newton's Third Law: "*To every action there is always opposed an equal reaction: or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts*." Here: the force exerted by the ventricles on the blood volume causes an equal and opposite force of the blood on the ventricles, thus inducing measurable vibrations.

This effect was investigated as early as 1959 by Agress and Fields, who examined "Low Frequency Vibrations" and concluded that "further research would be rewarding in the diagnosis and evaluation of cardiac function". The study of whole-body cardiac vibrations was dubbed **ballistocardiography** (BCG). In 1964, Baevski et al. narrowed the focus to heart-induced chest-wall vibrations, introducing the term **seismocardiography** (SCG)[27]. This research continued, such as the 1991 study by Salerno and Zanetti into the relationship between SCG and left ventricular functioning during ischemia and SCG[28]. While early work focused on **intra-beat** waveforms and timings, later studies expanded to **inter-beat** timings for the purpose of heart rate estimation. A key paper in this context was by Brüser et al. on robust inter-beat interval (IBI) estimation in cardiac vibration signals[29]. They successfully demonstrated that individual IBIs could be estimated from BCG signals with high accuracy, by using maximum-likelihood estimation. However, their method did

not yield IBI sequences that were temporally consistent, making it less suitable for robust, continuous tracking of heartbeats. For the research described in this thesis, obtaining consistent IBI sequences is vital for deriving accurate instantaneous heart rate (IHR) signals. This requirement is especially challenging in the context of SCG with overnight measurements in sleep-disordered patients.

An example of 15-second chest-wall accelerometry signal is given in Figure 6. The **cardiac component** is mostly present in the Z-axis (in blue), with an amplitude of approximately 0.030 g, and to a lesser extent in the X-axis (in red). The **respiratory component** is mostly visible in the X-axis, with an amplitude of approximately 0.015 g. Albeit small, both components can be measured well by a sensitive accelerometer.

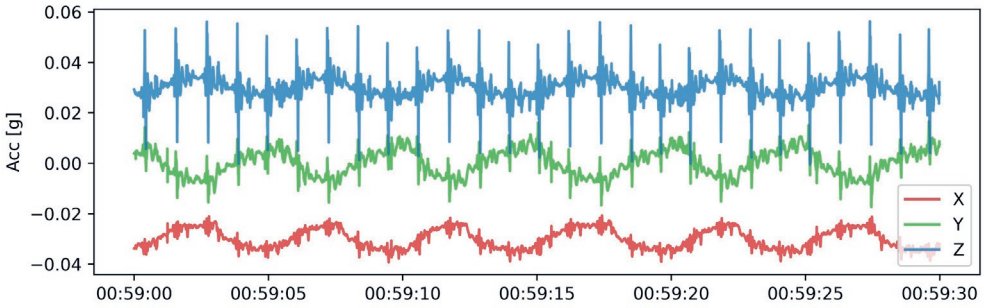


Figure 6 Example fragment of a chest-wall accelerometery signal. The peaks in the blue signal (Z) are due to heartbeats. The slower variations in the red signal (X) are due to respiration. The offsets of the signals have been adjusted for proper plotting.

Estimation of AHI and sleep stages from cardiorespiratory measurements

Apnea-Hypopnea Index

In estimating the AHI, the assessment of airflow is essential, as is indicated in the manual for the scoring of sleep and associated events[8]. However, the manual also states that if an airflow sensor is not available or not functioning, an **alternative apnea sensor** may be used. Among these alternative sensors are those for **respiratory inductance plethysmography** (RIP)[30]. RIP sensors measure the volumetric expansion of the chest and abdomen and are used as a surrogate for respiratory effort. This is also why a thoracic RIP belt is an accepted means for the assessment of respiration in a type III sleep study (see the box "Types of sleep studies").

In addition to changes in respiratory effort, events are often accompanied by a distinct cardiorespiratory response: a reduction in heartrate during the event, followed by an

increase afterwards (the latter is sometimes dubbed *postevent tachycardia*)[31,32]. However, there is variability in this effect, and the degree of increase (ΔHR) has been related to adverse cardiovascular outcomes of OSA[33]. To capture this abrupt change, the **instantaneous heartrate** (IHR) is of relevance, with the IHR defined as the reciprocal of the RR interval (in seconds), such that $IHR = 60/RR\,interval$ (in beats per minute). An average over time would smooth out changes (ΔHR) too much, due to the relatively short time scales, while the IHR reflects immediate changes and is thus a measure of heart rate variability. An example of cardiorespiratory responses to successive hypopnea events is given in Figure 7.

The need for comfortable, long-term home sleep apnea testing devices has driven research into estimating the AHI without a dedicated airflow sensor, focusing instead on cardiac signals (ECG) and respiratory effort (RIP) as surrogates. In 2019, Papini et al. estimated the AHI in a sleep-disordered population by automatically processing and selecting engineered features derived from electrocardiogram (ECG) data[34]. Soon after that, Iwasaki et al. investigated the estimation of the AHI based on heart rate variability using long short-term memory neural networks[35]. More recently, Fonseca et al. used ECG-derived IHR and respiratory effort (from RIP) in combination with an encoder-decoder neural network to estimate the AHI[36]. They reached an intraclass correlation coefficient of 0.91 when comparing the estimated AHI against that derived from polysomnography, over a cohort of 653 patients with varying OSA severities.

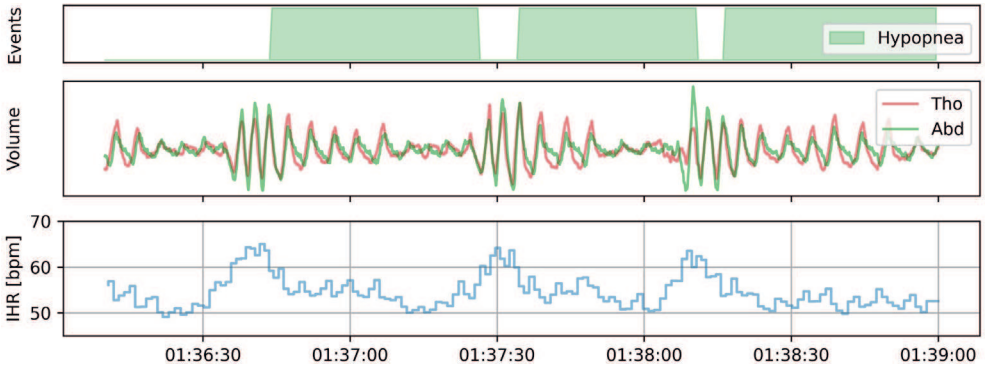


Figure 7 Example of cardiorespiratory responses to hypopnea events. Top: hypopnea events as scored from the PSG, middle: the thoracic and abdominal RIP signals, bottom: the IHR derived from ECG. After each of the hypopnea events, the respiratory rate and amplitude increases and the IHR increases.

Sleep stages

To obtain an accurate AHI estimate, determining the total sleep time (TST) is essential. TST is the cumulative time spent in all sleep stages but Wake (REM, N1, N2, N3). Traditionally, calculating TST requires neurological signals, as detailed in the box "The Measurement of

Sleep." Consequently, research has focused on alternative measurement methods. Since the sympathetic and parasympathetic branches of the autonomic nervous system reflect changes in sleep stages, and cardiac and respiratory activity mirror these shifts (e.g., heart rate and respiration become slower and more regular during N3), studies have explored cardiorespiratory sleep staging. This technique estimates sleep stages solely from cardiorespiratory signals.

A 2022 systematic review by Ebrahimi and Alizadeh investigated techniques for cardiorespiratory sleep staging. They categorized methods based on the signals used—ECG, respiratory signals, or a combination of both—and concluded that deep learning could significantly improve performance. One notable example is Cardiorespiratory Sleep Staging (CReSS). This method demonstrated good performance by utilizing IHR derived from ECG and respiratory effort from thoracic RIP. It achieved a Cohen's kappa coefficient of 0.64 for the classification of four classes: REM sleep, light sleep (N1+N2), deep sleep (N3), and Wake. The N1 and N2 stages were combined because the main differences between them—sleep spindles and K-complexes—are primarily visible in EEG signals and do not manifest distinctly in cardiorespiratory activity.

Research goals and thesis outline

The primary goal of the research in this thesis is to evaluate the feasibility of AHI estimation using chest-wall accelerometry. This comprises the detection of respiratory events (apneas and hypopneas) as well as the measurement of total sleep time, both required to compute the AHI. The choice of chest-wall accelerometry stems from its ease of integration into a chest-worn device, such as used in positional therapy for OSA. Feasibility is evaluated by comparing performance against the AHI derived from a type-I or type-II sleep test.

A secondary goal is to assess the feasibility of sleep staging. This can be used to obtain total sleep time, required for the calculation of the AHI. Furthermore, sleep stages themselves can provide valuable information for diagnosis; co-occurrence of respiratory events with sleep stages can provide information useful for OSA phenotyping and the sleep architecture (the progression of sleep stages over time) can provide information for the diagnosis of other conditions, such as sleep continuity and circadian issues.

The approach we follow in this thesis is to estimate respiratory effort and instantaneous heart rate from chest-wall accelerometry and to use these as input to the methods for sleep staging and AHI estimation. As the characteristics of the estimated cardiorespiratory signals may differ from those derived from ECG and respiratory impedance plethysmography, we include the tuning of the existing methods into the scope of this work.

The first three chapters focus on the robust and accurate estimation of instantaneous heart rate (IHR) and respiratory effort (RE) from chest-wall accelerometry in a sleep-disordered population.

The first three chapters are outlined in Figure 8 and comprise:

Chapter 1, Respiratory Effort Estimation with Principal Component Analysis, describes an algorithm for the estimation of RE from chest-wall accelerometry that uses an improved version of principal component analysis, in combination with a statistical method for the estimation or the inspiratory and expiratory direction. It was developed and evaluated on a small set of daytime measurements obtained among healthy participants.

Chapter 2, Respiratory Effort Estimation with Deep Learning, introduces an improved algorithm, using a convolutional neural network of which the hyperparameters were automatically optimized to reach a low estimation error. The network also estimated the inspiratory and expiratory directions. It was trained and evaluated on a larger overnight dataset that was obtained among a sleep-disordered population, with a wide variety of OSA severities.

Chapter 3, Maximum a posteriori detection of heartbeats, finds a temporally consistent set of beat localization times and interbeat intervals by means of *maximum a posteriori* (MAP) estimation. This approach combines two sources of information. First, we use the likelihood function (the data evidence), which describes how probable our observed acceleration signal is, given a specific set of beat times. Second, we introduce a prior probability, which represents the expectation about what a reasonable set of beat times should look like. MAP estimation then finds the specific set of beat times and intervals that maximizes this combined probability—the one that is both highly supported by the data (high likelihood) and matches the prior probability.

As finding the optimal solution is computationally intractable, the problem is reformulated into a Markov decision process, for which an optimal solution is approximated by means of depth-constrained search using the Bellman equation (the foundational equation for dynamic programming and reinforcement learning).

The third and fourth chapters focus on the use of IHR and RE derived from chest-wall accelerometery for sleep staging and AHI estimation. These are outlined in Figure 8 and comprise:

Chapter 4, Sleep Stage Estimation using Chest-Wall Accelerometry, leverages a previously developed method for cardiorespiratory sleeps staging from IHR derived from ECG and RE from thoracic respiratory impedance plethysmography. The method, using a neural network, was used as-is, without any retraining. However, the input signals obtained from the accelerometer did not match the characteristics of the signals used to train the neural network; in particular, the respiratory impedance plethysmography signals included motion artifacts from large body movements, and it was necessary to introduce similar motion artifacts into the RE signal derived from chest-wall accelerometry. Therefore, motion artifacts are estimated from chest-wall accelerometery and mixed into the RE signal.

Chapter 5, Apnea-Hypopnea Index Estimation using Chest-Wall Accelerometry, leverages a previously developed neural network to estimate OSA severity from IHR derived from ECG and RE from thoracic respiratory impedance plethysmography. The neural network is tuned to the specific characteristics of IHR and RE derived from chest-wall accelerometry, by means of transfer learning, where a small fraction of the parameters is retrained on part of the study data. Subsequently, the adapted network is used to estimate the probabilities of respiratory events, which are then compared to sleep stage-dependent thresholds to perform event detection. Finally, the events are counted and divided by the total sleep time to estimate an AHI.

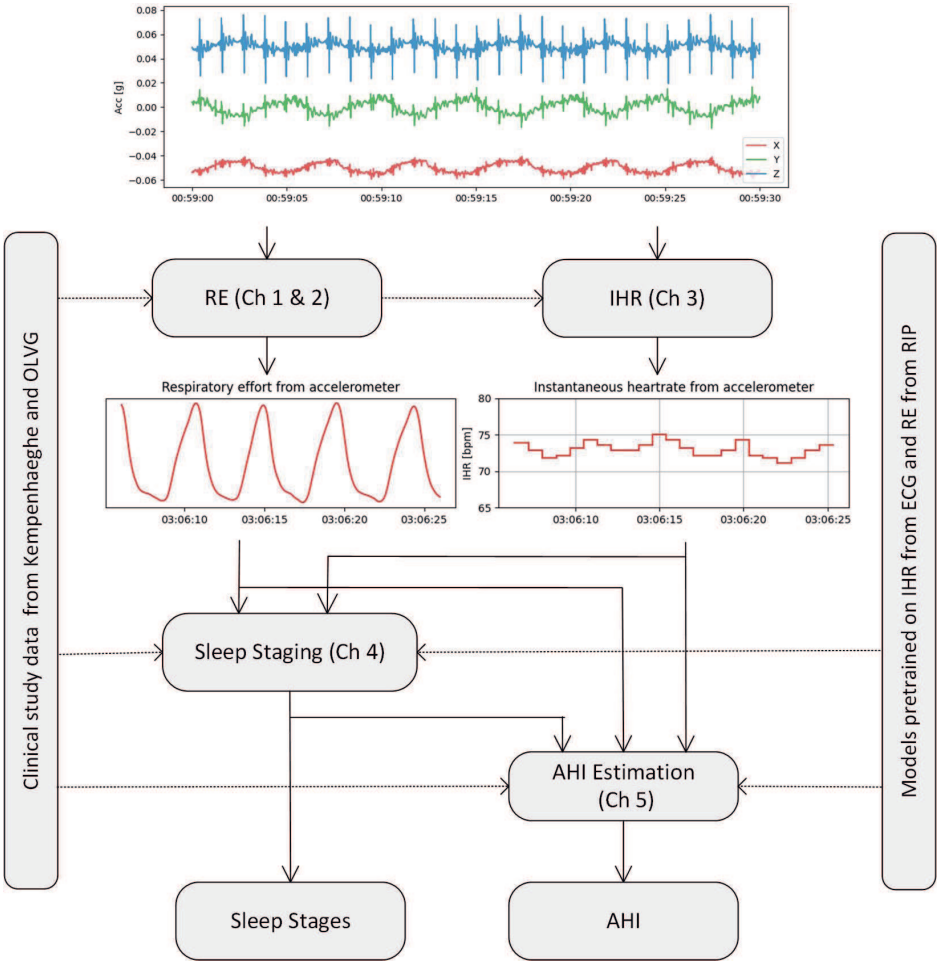


Figure 8 Graphical outline of the thesis. Chapters 1, 2, and 3 focus on cardiorespiratory measurements respiratory effort (RE) and instantaneous heartrate (IHR) from chest-wall accelerometry. Chapters 4 and 5 leverage previously developed methods to estimate sleep stages and the apnea-hypopnea index (AHI) from cardiorespiratory measurements derived from ECG (IHR) and respiratory effort (RE) from respiratory impedance plethysmography (RIP). Clinical study data from Kempenhaeghe and OLVG are used to develop the cardiorespiratory measurements (RE and IHR) and to tune the methods to estimate sleep stages and AHI to the specific characteristics of cardiorespiratory measurements from chest-wall accelerometry.



Respiratory Effort Estimation with Principal Component Analysis

This chapter corresponds to the manuscript:

F. Schipper, R. J. G. van Sloun, A. Grassi, R. Derkx, S. Overeem, and P. Fonseca, "Estimation of respiratory rate and effort from a chest-worn accelerometer using constrained and recursive principal component analysis," Physiol. Meas., vol. 42, no. 4, p. 045004, Apr. 2021, doi: 10.1088/1361-6579/abf01f.

Some minor editorial updates have been made.

CHAPTER 1

1.1 Abstract

Objective - Measurement of respiratory rate and effort is useful in various applications, such as the diagnosis of sleep apnea and early detection of patient deterioration in medical conditions, such as infections. A chest-worn accelerometer may be an easy and non-intrusive method, provided it is accurate and robust. We investigate the use of a novel method that can perform under realistic sleeping conditions such as variable sensor positions and body posture.

Approach - Twenty subjects (aged 46– 65 years) wore an accelerometer on the chest and a respiratory impedance plethysmography band as a reference. The subjects underwent an experimental protocol lasting approximately 90 min, under various postures and with different sensor positions. We used a novel, constrained, and recursive form of principal component analysis (PCA) to estimate the respiratory effort signal robustly. To obtain an estimate for the respiratory rate, first, multiple estimates were aggregated into a single frequency. Subsequently, a quality index was determined, such that unreliable estimates could be identified, and a trade-off could be made between coverage (percentage of time that the quality index is above a threshold) and limits of agreement.

Main results - Results were determined over all recorded data, including changes in sensor position and posture. For respiratory effort, it was found that recursive and constrained computation of PCA reduced the estimation error significantly. For respiratory rate, a relation between coverage and limits of agreement was determined. If a minimum coverage of 80% was required, the limits of agreement could be kept below 1.45 breaths per minute. If the limits of agreement were constrained to 0.2 breaths per minute, a mean coverage of 5% was still attainable.

Significance - We have shown that chest-worn accelerometery can be a robust and accurate method for measurement of respiratory features under realistic conditions.

1.2 Introduction

Measurement of respiratory features, like respiratory rate and respiratory effort, is essential in various clinical fields. Respiratory rate, one of the descriptive characteristics of respiration, is among the most important vital signs and is an essential indicator for patient deterioration in various medical conditions, such as infections[37]. The methods commonly used in clinical practice for measuring the respiratory rate are manual counting, capnometry, and impedance pneumography[38]. Respiratory effort is critical in the diagnosis of sleep-disordered breathing[39]. The respiratory effort signal varies during sleep and may be used for sleep staging[40]. The gold standard for measuring respiratory effort is esophageal manometry[41]. However, this method is highly invasive and is therefore seldom used. A more often used method is respiratory impedance plethysmography (RIP). RIP provides a surrogate of respiratory effort and is the de facto standard in, for example, nocturnal polysomnography[39]. Therefore, the RIP signal will be used throughout this paper as a reference.

As the methods commonly used in clinical practice are intrusive, alternatives have been investigated. These alternatives rely on various physiological phenomena and may require body contact or may be contactless, such as cameras[38,42–44]. Among these alternatives is chest-worn accelerometery, which may be quite suitable for non-intrusive measurement. It only requires a single point of mechanical contact with the chest. There is no need for electrical contact and, therefore, it may be worn in clothing, over clothing, or integrated in an adhesive patch. It can be very compact and due to low power consumption, allow for a long operating time, even when using a small battery.

The use of accelerometery for the measurement of respiratory features has been investigated in the past. In 2007 Morillo et al described how a single-axis accelerometer was used to obtain a respiratory signal at various chest positions[45]. Later, Hung et al used the sum of the two axes of a bi-axial accelerometer to yield a respiratory signal[24]. In 2009, Jin et al fused the axes of a tri-axial accelerometer with principal component analysis (PCA)[25]. As an alternative to PCA, Bates et al used a single-axis rotation model[46]. More recently, Cesareo et al used a combination of an accelerometer and a gyroscope[26]; various fusion techniques were investigated to obtain a one-dimensional respiratory signal and a derived respiratory rate. Agreement analysis showed that PCA was the best-performing fusion method.

Although literature suggests that a respiratory signal can be obtained using an accelerometer, the motion of such a device is not only caused by respiration, but also by other sources, like limb movements and posture changes. Such movements may fully mask the respiratory component, or distort it, such that measurement becomes impos-

sible or erroneous respectively. The impact of such movements has been addressed, for example, by detecting movement[46] or using peak-regularity metrics[47]. However, certain application scenarios would likely benefit from an indication of quality of the measurements produced, rather than (or in addition to) an attempt to limit the impact that these movements are likely to have.

In this paper, we propose a novel constrained and recursive form of PCA to estimate the respiratory effort (RIP) signal robustly. By constraining PCA both spatially (the horizontal plane, the direction along the principal axis) and temporally (consistency of principal axis), we demonstrate increased robustness over prior art. To obtain a robust estimate of the respiratory rate, a post-processing step, in which multiple estimates are aggregated into a single frequency, is introduced. Finally, by introducing a quality index, and by using a threshold to reject estimates with low quality, it is shown how a trade-off can be obtained between coverage (percentage of time that the quality index is above the threshold) and limits of agreement in the estimation of respiratory rate.

1.3 Methods

1.3.1 Background

The respiratory cycle manifests itself by the motion of the ribs and sternum into the ventral and cranial directions[48], which cause an attached accelerometer to move accordingly. Figure 1 shows an example of respiratory motion, by identification of the position of a chest-worn accelerometer at the start (yellow) and the end of inspiration (red). The displacement in both ventral and cranial direction is approximately 10 mm in this case. Although the displacement is clearly visible, it is very difficult to measure by an accelerometer, as the acceleration is very small, in the order of 1 mg (milli-g, with the g the acceleration of gravity, approximately 9.8 ms^{-2}). In comparison, changes in the orientation of the accelerometer (due to the respiratory cycle) are much better measurable.

An accelerometer measures orientation by the projection of gravity on its axes, as shown in Figure 2. The projection of gravity on the (vertical) z-axis changes by the derivative of the cosine, which is close to zero for small angles. The projection of gravity on the (horizontal) y-axis in changes by the derivative of the sine, which is close to one for small angles. Hence, orientation changes have the most effect on horizontal axes.

In general, the z-axis of the accelerometer will not be vertical, nor will the y-axis be horizontal. For this reason, the coordinate system is first transformed into one where two axes are always horizontal, and one is always vertical. For the first horizontal axis, we choose the craniocaudal axis of the human body (the axis from the feet to the head), and the second horizontal axis automatically follows from this, while the vertical axis follows from gravity. In this way, we obtain a right-handed orthogonal coordinate system that is independent of how the accelerometer is oriented in relation to the body.

The respiratory pattern illustrated in Figure 3 (a) shows the two-dimensional projection of the respiratory pattern into the horizontal plane as formed by the craniocaudal and horizontal axis. Figure 3 (b) illustrates the three components of the respiratory pattern for each of the axes (craniocaudal, horizontal, vertical) showing that the respiratory pattern manifests itself mainly in the horizontal plane. An additional, albeit more intuitive observation is that the respiratory pattern is often very repetitive, and the shape and orientation of the pattern only change slowly over time (except for disruptive events like posture changes). These two aspects will be exploited by introducing spatial and temporal constraining when processing the accelerometer signal.

1.3.2 Data collection

The data used in this study were collected from 20 healthy volunteers aged 46–65 years, with a BMI in the range 20–30 kg m^{-2}, without any previous diagnosis of cardiac or respiratory disorders. The size of 20 was chosen as previous studies used samples sizes between 8 and 20[26,47,49–54]. The study was reviewed and approved by the Philips Research Ethical board.

The subjects were equipped with a thoracic RIP band (Alice 6 PSG, Philips Respironics), on which an accelerometer device was placed (ADXL355, Analog Devices), as illustrated in Figure 1. The RIP band was placed successively at positions at the bottom of the sternum (xiphoid process) and the third rib. These positions were chosen as they were practical and likely to capture rib cage motion due to breathing. For both positions, the accelerometer was placed on the line mid-vertical between the sternum and the left nipple. The subjects were asked to breathe normally and to lay down in supine, left-lateral, and right-lateral postures. All changes of posture and sensor position were manually annotated by taking timestamps. Synchronization was performed by adapting the timestamps of the accelerometer signal such that they matched the annotations and the RIP band signal.



Figure 1 (a) Left-lateral view of a chest-worn accelerometer at the start of inspiration (yellow) and the end of inspiration (red). (b) Corresponding anterior view. The dimensions of the accelerometer device (white box) are 62×48×11 mm.

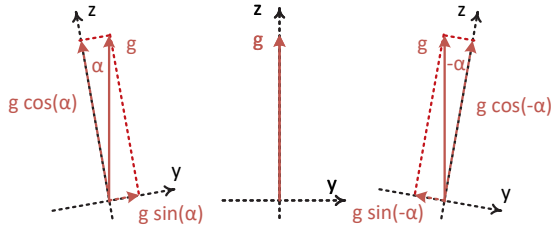


Figure 2 The change of the projection of gravity (g) on the vertical axis (z-axis) and a horizontal axis (y-axis) of an accelerometer due to a clockwise rotation by an angle 2α. Before the rotation (a), with positive y-component, halfway the rotation (b), with zero y-component, and after the rotation (c), with negative y-component. The z-component changes by the derivative of cos (α), which is close to zero for small α, while the y-component changes by the derivative of sin(α), which is close to one for small α.

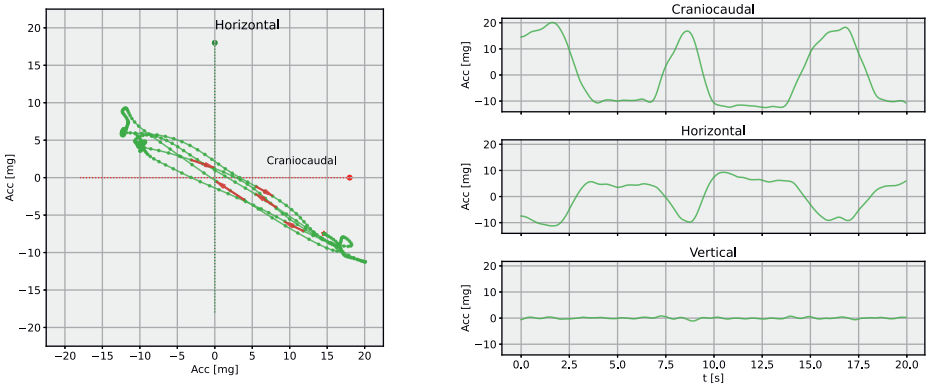


Figure 3 (a) Projection of a respiratory pattern into the horizontal plane. The red axis is the craniocaudal axis, of which the positive side points into the cranial direction, the green axis is the horizontal axis, of which the positive side points to the right side of the body. A respiratory cycle starts with inspiration from the top left to the bottom right, followed by expiration in the opposite direction. (b) The three components of the same respiratory pattern.

1.3.3 Data processing

Data from the 3D accelerometer is processed to estimate a respiratory effort, from which a respiratory rate is derived. The objective is to achieve a respiratory effort estimation as close as possible to the RIP band signal. The impact of confounding factors on the motion of the accelerometer, such as limb movements or posture changes, is minimized by constraining the respiratory effort signal and by determining a signal quality index, and by using it to exclude intervals for which accuracy is expected to be poor.

Overall signal processing chain

Figure 4 shows an overview of the complete signal processing chain. Notably, and unlike previous work[25,26,51], we predominantly leverage recursive processing to both reduce computational load and enhance the stability of the algorithm by adding temporal momentum to the states. We first pre-process the three-dimensional accelerometer signal and then project it on the horizontal plane. We then leverage a recursive form of PCA to obtain a (one-dimensional) respiratory estimate, which is further used to estimate the Inspiratory and Expiratory directions (IE directions). The IE direction estimation feeds back the detected direction to the PCA, where it is used in the recursive estimation. Finally, the (one-dimensional) estimate is used to obtain the respiratory rate. A movement signal indicates the presence of acceleration levels beyond those from respiration (for example due to posture changes) and it is used to adapt processing during and after such events.

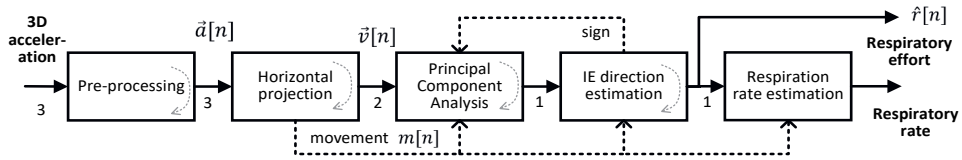


Figure 4 Overview of the processing steps. The numbers show the dimensionality of the signals. The dashed line marked 'sign' is a feedback signal, which causes the principal component analysis to adapt to the estimation of inspiratory and expiratory directions. The dashed line marked 'movement' causes later processing steps to adapt to the presence of movement. A dashed line inside a processing step denotes that it uses recursive computation.

The 3D acceleration signal is acquired with a sampling rate of 250 Hz. It is low-pass filtered by means of a Butterworth filter (order = 3, cut-off frequency $f_0 = 3.5$ Hz, see Table 1), and subsequently decimated to 16 Hz to reduce the processing load of the remaining steps. We denote this pre-processed acceleration signal as $\vec{a}[n]$, for sample index n. The parameters of the Butterworth filter, as well as most of the other parameters in the signal processing chain were determined by means of an optimization procedure, which is described in the section *Parameter optimization*. Note that, next to the low-pass filter in

this block, a second low-pass filter, with a lower cut-off frequency, is used further downstream, after the motion signal has been determined.

As described in the background section, the respiratory pattern of interest predominantly takes place in the horizontal plane, allowing us to project the full 3D pattern onto that 2D plane. The horizontal plane is derived from the gravity vector, to which it is perpendicular. The gravity vector is estimated recursively by means of exponential smoothing:

$$\vec{g}[n] = \alpha_h \vec{a}[n-1] + (1-\alpha_h)\vec{g}[n-1], \quad (1)$$

in which $\vec{g}[n-1]$ is the previous gravity estimate, $\vec{g}[n]$ is the updated gravity estimate, and the parameter α_h $(0 < \alpha_h < 1)$ determines how quickly the estimated gravity adapts to the incoming acceleration. A large value of α_h causes a quick adaptation but allows artifacts to pass through, while a small value of α_h causes the opposite.

In case of movement (like for example, during a posture change), the respiratory component in $\vec{a}[n]$ is fully obscured by the much larger movement component. In this case, the method is not able to extract the respiratory component anymore, and respiratory estimation is halted until the movement stops. For this reason, the downstream processing steps use a movement indicator $m[n]$, to stop estimation during movement, and to temporarily speed up adaptation after movement. The adaptation after movement is achieved by using different parameter values (see Table 1). Movement is detected if the new acceleration value differs from the gravity estimate by more than a threshold t_m:

$$m\left[n\right] = \left\| \vec{a}\left[n\right] - \vec{g}\left[n\right] \right\| > t_m \quad (2)$$

After having detected gravity, by means of (1), horizontal projection can be performed. Although the gravity vector determines the horizontal plane, it does not yet determine the (horizontal) axes that will be used to represent the (two-dimensional) respiratory pattern. To keep computational complexity low, arbitrary horizontal axes are used and not the craniocaudal axis and its perpendicular horizontal axis, as was done in the background section. Note that the choice of the craniocaudal axis was meaningful for analysis purposes, but for performance, the choice is irrelevant as PCA is applied, which is insensitive to the choice for the two axes. The horizontal projection comprises two successive rotations, a first, R_1, around the x-axis of the accelerometer and a second, R_2, around the y-axis of the accelerometer. The rotated acceleration vector then becomes (the sample index n has been left out for brevity):

$$\vec{b} = R_2 R_1 \vec{a}, \quad (3)$$

with:

$$R_1 = \frac{1}{\rho}\begin{bmatrix}\rho & 0 & 0\\ 0 & g_z & g_y\\ 0 & -g_y & g_z\end{bmatrix} \qquad R_2 = \begin{bmatrix}\rho & 0 & g_x\\ 0 & 1 & 0\\ -g_x & 0 & \rho\end{bmatrix} \qquad \rho = \sqrt{{g_y}^2 + {g_z}^2}. \tag{4}$$

The horizontal projection is performed by taking the first two elements of $\overrightarrow{b}$, as these are the coordinates in the horizontal plane:

$$\vec{h} = \begin{bmatrix}h_1\\ h_2\end{bmatrix} = \begin{bmatrix}b_1\\ b_2\end{bmatrix}. \tag{5}$$

After horizontal projection, the two-dimensional signal $\overrightarrow{h}$ is normalized and filtered. Normalization comprises division by the norm p:

$$\vec{u}\left[n\right] = \vec{h}\left[n\right]/p[n], \tag{6}$$

with:

$$p\left[n\right] = \alpha_n \left\|\vec{h}[n]\right\| + (1-\alpha_n)\, p\left[n-1\right] \tag{7}$$

The initial value of p is determined by taking the mean norm of ten initial samples from $\overrightarrow{h}$. After normalization, $\overrightarrow{u}$ is filtered with a Butterworth low-pass filter of order $n_1 = 3$ and cut-off frequency f_1= 0.8 Hz, see Table 1. The output of the filter, denoted $\overrightarrow{v}$, is input into the Principal Component Analysis.

$$\vec{v}\left[n\right] = Butterworth\left(\vec{u}[n \ldots n-3], \vec{v}\left[n-1 \ldots n-3\right]\right) \tag{8}$$

Principal Component Analysis

The horizontal projection already reduces the number of dimensions to two. In this step, we reduce the number of dimensions to one, by means of PCA. As with all steps so far, this is done recursively, such that computational complexity is kept low and rapid deviations, which are generally due to artifacts, are penalized. The recursive method implements Hebbian Learning [55], which models the learning of neurons; neural paths that cause high output activations are reinforced through the feedback of the activation. As, in this case only the first principal component is used, Oja's rule can be utilised instead of the generalized Hebbian rule.

In Oja's rule, in each iteration a new input sample $\vec{v}\left[n\right]$ (8) is projected onto a projection axis $\vec{w}\left[n\right]$, such that the respiratory estimate $\hat{r}\left[n\right]$ can be calculated ($\hat{r}\left[n\right]$ is an estimate for $r[n]$, being the respiratory reference waveform):

$$\hat{r}\left[n\right] = \vec{w}\left[n\right] \bullet \vec{v}\left[n\right], \tag{9}$$

with:

$$\vec{w}[n+1] = \eta\, \hat{r}[n]\vec{v}[n] + \left(1 - \eta\, \hat{r}[n]^2\right) \vec{w}[n]. \tag{10}$$

Equation (9) is the projection of $\vec{v}[n]$ on $\vec{w}[n]$, while (10) is the recursive update of $\vec{w}[n+1]$, which changes the projection axis $\vec{w}[n+1]$ in the direction of $\vec{v}[n]$, with learning rate η (see Table 1). The initial value of $\vec{w}[n]$ is chosen arbitrarily: $\vec{w}[1] = [1 \quad 0]^T$. Notably, the factor $\eta\,\hat{r}[n]^2$ causes the norm of the projection axis, $\|\vec{w}[n]\|$, to converge to one. This can be understood by analysing the steady-state behaviour (where $\vec{w}[n+1] = \vec{w}[n]$), for which the equality $\eta\,\hat{r}\,\vec{v}[n] = \eta\,\hat{r}^2\,\vec{w}[n]$ must hold. This equality holds if $\vec{w}[n]$ and $\vec{v}[n]$ align, $\hat{r}[n] = \|\vec{v}[n]\|$, and $\|\vec{w}[n]\| = 1$.

Estimation of the inspiratory and expiratory directions

In the respiratory estimate, inspirations should correspond to positive signal transitions, while expirations should correspond to negative signal transitions. There are multiple benefits for such a convention; in sleep-disordered breathing, snoring occurs during inspiration, and upper airway obstructions cause inspirations to take longer. Also, consistency is desired; if the sign would change on the fly, then extra transitions would be introduced, which do not correspond to inspirations or expirations and, therefore, would cause the respiratory rate to be over-estimated.

A feature that correlates with the proper choice of the direction is the skewness of the distribution of $\hat{r}$. For $\hat{r}$ in Figure 5(a), positive signal transitions correspond to inspirations, while negative signal transitions correspond to expirations. The distribution of the signal values, shown in Figure 5 (b), is peaked at the left side, which corresponds to the inspiratory pause. The skewness of the distribution is 0.7, which is larger than zero and, therefore, the distribution is right-skewed.

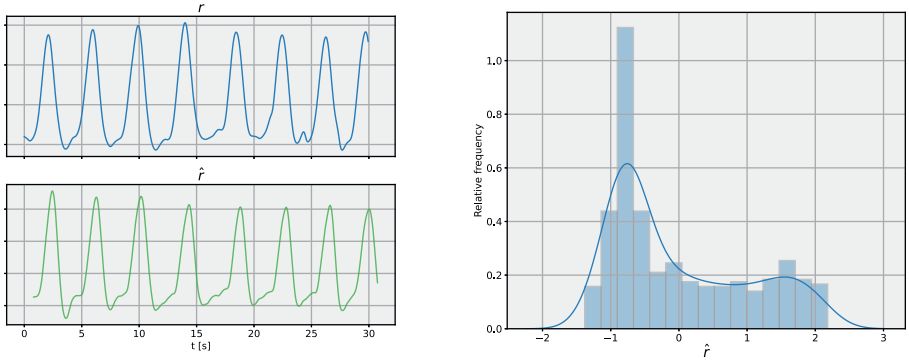


Figure 5 a) Respiratory reference (r) and respiratory estimate (r̂), with inspirations corresponding to positive transitions and expirations corresponding to negative transitions. b) Distribution of the respiratory estimate (r̂). The skewness is 0.7, meaning the distribution is right-skewed.

The skewness *s* is defined as:

$$s = \frac{\sum_{n=1}^{N} (\hat{r}[n]-\mu)^3 / N}{\sigma^3}, \tag{11}$$

with σ the standard deviation, μ the mean, and N the number of samples. Although the skewness is generally positive for a respiratory estimate with the correct sign, this may not be true for short analysis periods. To be robust against this, and to get consistency over time, exponential smoothing is performed; the skewness is not computed as in the definition, but in a recursive manner, just as all other computations so far, and rapid changes get penalized. As the respiratory estimate $\hat{r}$ has zero-mean ($\mu = 0$) and a standard deviation close to one ($\sigma \approx 1$), the following formula approximates the skewness:

$$s\left[n\right] = \alpha_s \, \hat{r}[n]^3 + (1 - \alpha_s) \, s\left[n-1\right]. \tag{12}$$

If the skewness is above a threshold ($s[n] \geq t_s$), then the data processing continues unchanged. However, if the skewness falls below a threshold ($s[n] < t_s$), then the respiratory estimate has the wrong sign, and inspirations correspond to negative transitions. In that case we invert the skewness ($s[n] = -s[n]$) and the projection axis ($\overrightarrow{w}[n] = -\overrightarrow{w}\left[n\right]$). The sign change of $\overrightarrow{w}\left[n\right]$ corresponds to the feedback arrow in Figure 4.

Respiratory rate estimation

The respiratory rate is derived by means of frequency analysis, as this provides robustness against amplitude artifacts. Because variations in the respiratory rate should be tracked, the Short-time Fourier Transform (STFT) is used. In the STFT, the signal is divided into windows of equal length (30 s) and frequency analysis is performed on each window. The windows are 5 s apart and therefore overlap by 25 s. Tracking of peaks in the spectrum over time provides a respiratory rate signal (with a sampling period of 5 s). The rate signal is fed into a statistical post-processing unit to obtain a final estimate, by combining twelve estimates, such that one estimate per minute is obtained.

Prior to application of the STFT, the respiratory effort signal is down sampled to 4 Hz, which suffices for the purpose of respiratory rate analysis. After down sampling, the signal is segmented into windows and for each window, the signal is multiplied by a Hanning function, such that artifacts due to segmentation are minimized. For each window, the 120 samples (30 s @ 4 Hz) are zero-padded to obtain 256 samples, after which an *N*-point Fourier transform is applied ($N = 256$). The zero padding is performed to increase the accuracy of the peak height estimate; more points reduce the bin-width and therefore peaks are less averaged-out. The bin width is:

$$\Delta f = \frac{f_s}{2\,N} 60 = \frac{240}{512} = 0.5 \text{ min}^{-1} \tag{13}$$

For each window *m*, the spectrum is normalized, and, for the highest spectral peak, its frequency and height $(f_m,\ h_m)$ are recorded. The peak height serves as a quality metric: the higher the value, the higher the quality. During statistical post-processing, the peaks of 12 successive windows (corresponding to one minute) are combined into one single estimate for the respiratory rate. Only peaks that are higher than a threshold ($h_m > t_p$) are selected. Over these peaks (f_m), a weighted mean is computed, using the peak heights h_n as weights:

$$f_B = \frac{\sum_{(m\ for\ which\ h_m > t_p)} h_m f_m}{\sum_{(m\ for\ which\ h_m > t_p)} h_m}. \tag{14}$$

If no peaks exist for which $h_m > t_p$ for a window, no respiratory rate is determined, and it is considered not to be available. The Coverage is computed as the number of intervals for which a respiratory rate is available, divided by the total number of intervals analysed.

1.3.4 Performance metrics

Respiratory effort

The Mean Squared Error (MSE) is used to assess how well $\hat{r}[n]$ resembles $r[n]$. The MSE is calculated over 60-second windows which are time-shifted by 15 seconds. The MSE is only calculated for windows that are free of movement, meaning that $m[n] = 0$ for all samples in the window (see (2) for $m[n]$). Prior to calculation of the MSE, $\hat{r}[n]$ and $r[n]$ are normalized on a window-by-window basis, by dividing through by their standard deviations and are time-aligned such that a minimum MSE is obtained.

Respiratory rate

Respiratory rate evaluation is performed over the same 60-second windows as used for the evaluation of respiratory effort. The rates are computed for both the estimate and the reference, using the respiratory rate estimation algorithm presented before. As a respiratory rate may not be available due to insufficient quality, agreement analysis is only performed for windows for which both a reference and an estimate are available. In addition, coverages are computed; for the reference, the Coverage is the fraction of windows for which a reference rate is available; for the estimate, the Coverage is the number of windows for which both a reference and an estimated rate are available, divided by the number of windows for which a reference rate is available.

1.3.5 Parameter optimization

All parameters used in the algorithms were automatically determined using an optimization procedure. This comprised Bayesian optimization using as objective function the 90th percentile of the distribution of the MSE over the time windows. The resulting parameter values are indicated in Table 1. If a parameter has a different value during and after

movement, then it is always a factor 10 higher than under non-movement conditions. This factor was not part of the optimization process, and it is introduced to speed up the adaptation to new conditions. Half of the dataset was used for optimization, while results were calculated over the full dataset.

Table 1 Parameter values determined by means of the optimization process.

Parameter	Equation / section	Normal value	Value during and after movement	Description
n_0, f_0	Pre-processing	3, 3.5 Hz	same	Order and cut-off frequency of Butterworth low-pass filter
α_h	(1)	0.03	0.3	Smoothing constant for gravity estimation
t_m	(2)	0.06 g	same	Movement threshold
n_1, f_1	Horizontal projection	3, 0.8 Hz	same	Order and cut-off frequency of Butterworth low-pass filter
α_n	(7)	0.01	0.1	Smoothing constant for normalization
η	(9)	0.8e-3	8e-3	Learning rate for recursive PCA
α_s	(11)	70e-6	0.7e-3	Smoothing constant for skewness
t_s	Estimation of IE dir.	5e-3	same	Threshold for skewness to change the direction of projection
t_p	(13)	0.8	same	Threshold for the height of the spectral peak

1.4 Results

1.4.1 Respiratory effort

For the (60-second) windows without movement, an MSE value was calculated: $MSE = \frac{1}{60f_s}\sum_{n=1}^{60f_s}\left(\hat{r}[n] - r[n]\right)^2$, with $\hat{r}[n]$ the estimated samples in the window. Attributes of each window are subject and posture. For each of the subjects, means over these MSE values were calculated, over each posture separately and over all postures, see Table 2. In addition, coverages were calculated, being the ratio between the number of movement-free windows and the total number of windows. Furthermore, means and standard deviations over the subjects means for each of the postures were calculated, while skipping subjects with less than ten movement-free windows for that posture. It was found that the mean error was lowest in the right posture (0.14), while it was higher for the left (0.26) and supine postures (0.33), see Table 2. The variation over the subjects was computed by means of the standard deviations (SD). The lowest variation was found for the right posture (0.11), with higher values for the left (0.17) and supine postures (0.21). Similarly, means and standard deviations were calculated over the subject

coverages. Figure 6 shows the distribution of the MSE-values over the subjects, over all postures (of which the first row in Table 2 is the mean value).

1.4.2 Comparison with block-wise, unconstrained PCA

To quantify the added value of recursive, constrained PCA, the data were also processed utilizing block-wise (non-recursive), unconstrained PCA. For each window, the three-dimensional accelerometer signal $\overrightarrow{a}$ was de-trended (by subtracting a least-squares linear fit for each axis), and low pass filtered (using a Butterworth low-pass filter of order n_1 and cut-off frequency f_1). Finally, the covariance matrix was calculated, and the projection axis $\overrightarrow{w}$ was determined by taking the eigenvector with the largest eigenvalue. The de-trended and filtered acceleration signal was projected upon $\overrightarrow{w}$, such that the (one-dimensional) respiratory estimate $\hat{r}$ was obtained. The comparison against the reference r was done in the same way as in the previous paragraph, such that results could be compared, see Table 3.

Table 3 Comparison of the mean MSE of respiratory effort over the subjects.

		Recursive, constrained PCA	Block-wise, unconstrained PCA
Metric	Posture	Mean (SD)	Mean (SD)
Mean MSE	All	0.32 (0.16)	0.53 (0.16)
	Supine	0.33 (0.21)	0.52 (0.20)
	Left	0.26 (0.17)	0.41 (0.17)
	Right	0.14 (0.11)	0.31 (0.14)

1.4.3 Respiratory rate

For each of the covered windows, a respiratory rate difference was calculated. Attributes of each window were subject and posture. For each of the subjects, agreement ranges were calculated, over each posture separately and over all postures, see Table 4. As the differences were not normally distributed (D'Agostino-Pearson test for normality, p=0.00), the 2.5th and 97.5th percentiles of the distribution of the differences were used for the agreement range. Also, coverages were calculated, being the ratio between the number windows for which a respiratory rate difference was available, and the total number of windows. Finally, means and standard deviations over the agreement ranges were calculated, while skipping subjects with less than ten covered windows. Similarly, means and standard deviations were calculated over the subject coverages. The coverages are combined coverages, meaning that respiratory rates must be available for both the reference and the estimate.

A graphical overview of the distribution of the respiratory rate difference over the subjects (over all postures) is shown in Figure 7. The mean bias is 0.007 min^{-1}. For the

postures, the mean agreement ranges for supine are significantly lower than those for left (Mann-Whitney U-test, p=0.001) and for right (Mann-Whitney U-test, p=0.006). A Bland-Altman plot is given in Figure 8 for the rate differences over all subjects and postures.

For some subjects, under certain postures, the number of covered windows is 0 (for example for subject s04 in supine posture). This is due to the selection of the operating point; the agreement analysis was performed at the default operating point ($t_p = 0.8$, see Figure 9), which results in one combination of an agreement interval and Coverage.

The same analysis may be performed for alternative operating points (other values of t_p), which results in different combinations of coverages and agreement ranges. One expects that a higher value of t_p leads to a lower coverage and a shorter agreement range, which is indeed the case, as can be found in Figure 9.

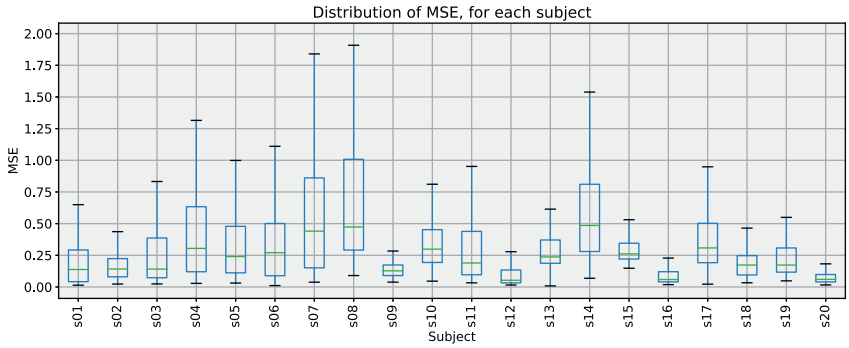


Figure 6 Box-and-whisker plot of the MSE of the respiratory effort, for each of the subjects, over all postures.

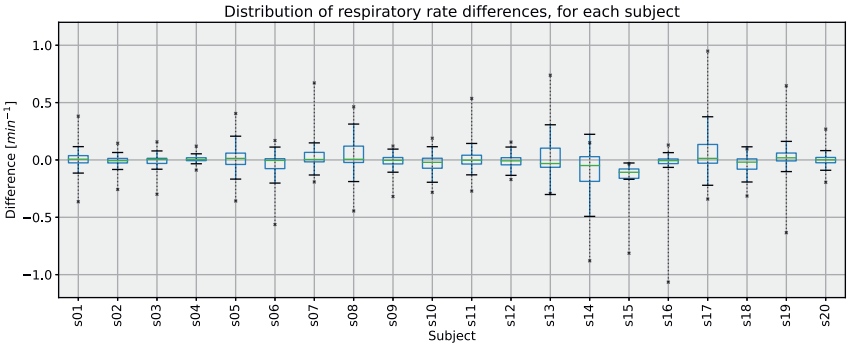


Figure 7 Distribution of the respiratory rate differences, for each subject, over all postures. Next to the box-and-whiskers, the dashed lines with 'x'-markers show the agreement range (from the 2.5% percentile to the 97.5% percentile).

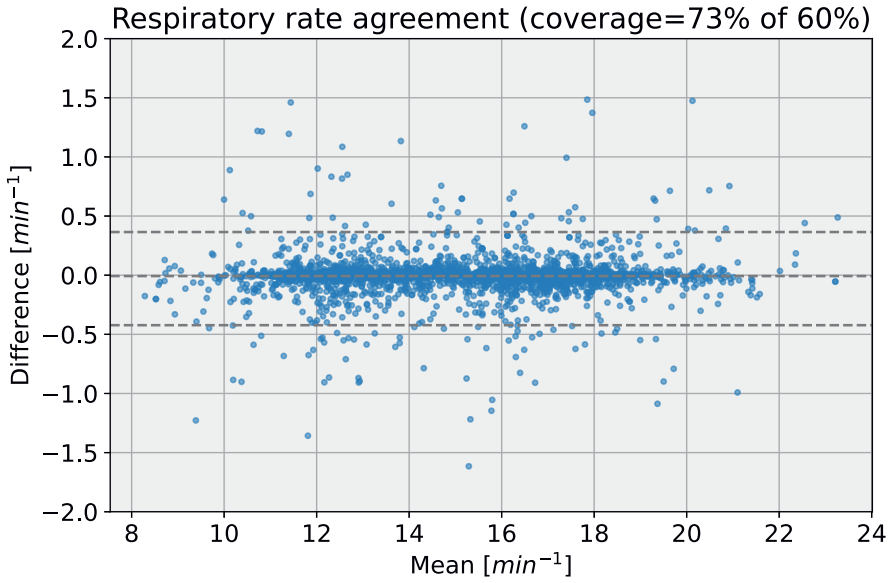


Figure 8 Bland-Altman plot for the respiratory rate. The grey dashed lines show, from bottom to top, the 2.5th percentile, the mean, and the 97.5th percentile of the distribution of differences. The bottom and top dashed lines form the agreement interval. The combined coverage is 73% x 60% = 44%.

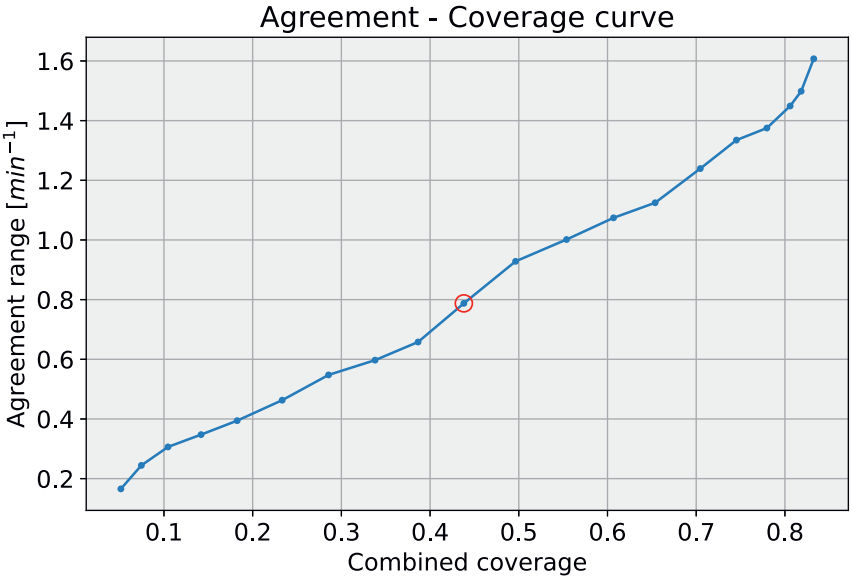


Figure 9 Mean agreement interval against mean Coverage, for various operating points (values of t_p). The values of t_p increase from right to left (higher values of t_p lead to smaller agreement intervals and lower coverages). A smaller agreement interval means better agreement. The default operating point is marked by the red circle ($t_p = 0.8$).

Table 2 Respiratory effort

Metric	Posture	s01	s02	s03	s04	s05	s06	s07	s08	s09	s10	s11	s12	s13	s14	s15	s16	s17	s18	s19	s20	Mean (SD)
Mean MSE	**All**	0.21	0.21	0.27	0.43	0.34	0.37	0.55	0.67	0.16	0.38	0.32	0.18	0.31	0.60	0.39	0.13	0.38	0.19	0.25	0.10	0.32 (0.16)
	Supine	0.31	0.21	0.30	0.53	0.31	0.52	0.73	0.83	0.19	0.36	0.32	0.08	0.28	0.56	0.33	0.11	0.34	0.18	0.13	0.07	0.33 (0.21)
	Left	0.07	0.14	0.41	0.29	0.48	0.09	0.10	0.34	0.09	0.28	0.21	0.59	0.04	0.00	1.57	0.04	0.05	0.00	0.49	0.00	0.26 (0.17)
	Right	0.07	0.16	0.12	0.07	0.10	0.02	0.26	0.38	0.10	0.06	0.16	0.00	0.22	0.23	0.00	0.02	0.40	0.14	0.43	0.08	0.14 (0.11)
Coverage	**All**	163/240	231/329	306/367	278/358	273/334	343/389	301/360	266/344	141/159	301/353	297/334	174/261	214/228	216/234	213/237	213/243	298/319	207/244	159/218	226/245	83% (8%)
	Supine	42/42	108/109	135/135	118/118	116/116	182/182	132/132	63/63	79/83	102/102	135/135	105/105	177/177	146/146	179/179	188/189	247/247	142/142	55/55	172/172	100% (1%)
	Left	27/32	62/62	52/52	55/55	32/32	49/49	60/60	49/49	35/35	32/32	44/44	14/14	7/7	0/0	1/1	9/9	9/9	0/0	6/6	0/0	99% (4%)
	Right	65/65	45/45	74/74	56/57	46/46	41/41	59/59	39/39	21/21	1/1	65/65	0/0	3/3	4/4	0/0	3/3	8/8	4/4	9/9	2/2	100% (1%)

Mean MSE and Coverage for all subjects, over individual postures and over all postures. For coverage, the number of covered windows and total windows are given, separated by a '/'. The total number of windows (under All) is less than the sum of the windows under Supine, Left, and Right, because the latter only covers windows where the posture is stable and in the respective posture.

Table 4 Respiratory rate

Metric	Posture	s01	s02	s03	s04	s05	s06	s07	s08	s09	s10	s11	s12	s13	s14	s15	s16	s17	s18	s19	s20	Mean (SD)
Agreement range	All	0.75	0.40	0.46	0.21	0.76	0.73	0.86	0.91	0.44	0.47	0.81	0.33	1.03	1.03	0.78	1.20	1.29	0.41	1.28	0.46	0.73 (0.33)
	supine	1.03	0.36	0.57	0.00	0.45	0.99	1.00	0.62	0.44	0.55	0.78	0.23	1.11	0.81	0.05	1.21	0.97	0.23	0.29	0.45	0.67 (0.33)
	left	0.41	0.14	0.10	0.19	0.72	0.70	0.34	0.09	0.23	0.19	0.21	1.13	0.24	0.00	0.00	0.02	0.30	0.00	0.05	0.00	0.30 (0.22)
	right	0.29	0.14	0.07	0.14	0.70	0.10	0.48	1.29	0.24	0.00	0.34	0.00	0.01	0.04	0.00	0.00	0.28	0.07	0.09	0.00	0.30 (0.20)
Coverage	All	136/240	245/329	48/367	121/358	143/334	208/389	103/360	62/344	85/159	270/353	142/334	177/261	58/228	106/234	6/237	116/243	132/319	170/244	107/218	104/245	46% (17%)
	supine	19/42	103/109	19/135	0/118	45/116	57/182	19/132	5/63	38/83	82/102	64/135	102/105	36/177	72/146	2/179	111/189	96/247	101/142	35/55	64/172	50% (24%)
	left	21/32	54/62	14/52	47/55	17/32	39/49	52/60	18/49	32/35	30/32	27/44	3/14	7/7	0/0	0/1	2/9	9/9	0/0	6/6	0/0	70% (22%)
	right	61/65	39/45	7/74	45/57	28/46	41/41	11/59	7/39	12/21	1/1	23/65	0/0	2/3	4/4	0/0	0/3	5/8	4/4	9/9	2/2	66% (27%)

Agreement range (in breaths per minute) and Coverage for all subjects, over all postures and for the individual postures. For a description. For coverage, the number of covered windows and total windows are given, separated by a '/'. The total number of windows (under All) is less than the sum of the windows under Supine, Left, and Right, because the latter only includes windows where the posture is stable and in the respective posture.

1.4.4 Computational complexity

As the data processing needs to run in an embedded processor of which the power consumption should be kept low, the computational complexity should be kept low as well. This holds for all the blocks in Figure 4, of which the four leftmost blocks implement recursive computations, while the rightmost block implements a block-based computation.

Each of the recursive processing blocks requires one computational step per sample and is therefore $O(n)$, with n the number of samples, and $O()$ the notation for computational complexity (big-O notation). The complexity $O(n)$ also holds for the recursive PCA. If, instead of a recursive PCA, non-recursive and non-constrained PCA would be applied, then the complexity would be $O(d^2nr)$, with *d* the number of dimensions ($d = 3$), and *r* the ratio of the window length and window shift ($r = 30/5 = 6$). Therefore, the complexity gain due to the recursive implementation becomes $d^2r = 54$.

The rightmost block in Figure 4 (Respiratory rate estimation) is not recursive, but block based. The computational complexity is that of an *N*-point FFT ($N = 256$), which is equal to $O(N \log N)$. As the FFT needs to be computed once every five seconds, this can easily be implemented on a small processor.

1.5 Discussion

In this paper, we propose a novel method for estimating respiratory features from a chest-worn accelerometer. We have evaluated the method on data recorded on twenty middle-aged volunteers, undergoing posture changes and sensor relocations, without manual selection of recording fragments. For estimating the respiratory effort, we have shown that that constraining and recursively computing PCA significantly improves the estimate. For the respiratory rate, we have shown that aggregating multiple estimates into a one, in combination with quality-based thresholding, allows for a trade-off between coverage and limits of agreement and yields accurate estimates.

1.5.1 Respiratory effort

Accuracy means that the estimate, if available, is close to the reference, and robustness means that accuracy is never compromised. In this work, several measures have been taken to achieve this. The PCA constraints in both space and time help suppress signal components that are not due to respiration, such as movement caused by posture changes and sensor relocations The effect of this becomes apparent from the comparison with unconstrained PCA; the mean MSE increases from 0.32 (0.16) to 0.52 (0.16) when not constraining (Table 3). However, the constraining does not prevent variations in MSE over

subjects and postures. To illustrate this difference, we further investigate subject s06, for which the mean MSE for the right posture is 0.02, while it is 0.52 for the supine posture. For both right and supine, we randomly selected three windows from the set of windows that have their mean MSE in the 45th to 55th percentile. Figure 10 shows these windows; each graph corresponds to one 60-second window and shows the raw acceleration waveforms a_x, a_y and a_z, the reference r, and the estimate $\hat{r}$. When looking at a waveform, respiratory and noise components can be identified. For example, in the bottom left graph, frequency analysis reveals a respiratory component at 0.23 Hz (13.8 min^{-1}) and a noise component at 1.3 Hz (78 min^{-1}). Visual inspection of the raw acceleration waveforms shows that the ratio between the respiratory and noise components is much higher in the top row (right posture) than in the bottom row (supine posture). The probable cause of the noise is the beating of the heart, of which a relation to posture is known [56]. In the right posture, gravity pulls the heart away from the chest wall, which causes less interference.

The acceptability of these MSE values depends on the application. There are various applications, such as the detection of respiratory events. Here, we elaborate further on the assessment of respiratory rate, as this is relevant for the assessment of patient deterioration.

1.5.2 Respiratory rate

To improve the estimates of respiratory rate, a statistical post-processing step was introduced, which relies on the signal quality index, allowing for both weighed averaging in determining estimates over one-minute periods, and for the exclusion of intervals with low signal quality. The mean agreement interval in the estimation of respiratory rate was 0.8 min^{-1}, while the mean coverage was 44%. This performance holds for the default operating point (see Figure 9). Depending on the application needs, also other operating points may be chosen. For example, for a mean coverage > 80%, a mean agreement interval of 1.5 min-1 is possible, while for a mean agreement interval < 0.2 min^{-1}, a mean coverage of 5% is achievable.

Earlier studies have described how to compute the respiratory rate using an accelerometer attached to the suprasternal notch, the waist, or the chest. It is difficult to compare the results of this study with those, as there are many differences, in, for example, protocols, postures, body positions, which parts of the protocol were considered when computing performance, and the quality of reference. Table 5 summarizes the performance of several of these studies. In comparison with literature, the performance obtained with our algorithm seems higher, even for the operating point with the poorest agreement interval (and highest Coverage). Other studies only looked at short intervals of controlled breathing under controlled conditions, see Table 6. By choosing a proper operating point for our method, the same or even better levels of agreement can be reached.

The results of this study suggest that it is possible to achieve an accurate and robust estimation of respiratory effort and respiratory rate only using an accelerometer, without requiring a gyroscope, as used in some studies [26]. In addition, the algorithm has a low computational complexity of only $O(n)$, as opposed to $O(d^2nr)$ for non-recursive, unconstrained PCA ($d^2r = 54$, see the section on computational complexity).

Some limitations can be mentioned for this study. First, the use of only a RIP band as a reference is not ideal; it is subject to movement just as the chest-worn accelerometer. As no data were discarded, the quality of the reference signal had to be taken into account in the data processing. Another limitation is that the method for deriving the respiratory rate from the accelerometer signal is the same as that for the RIP band signal. Furthermore, a high-accuracy accelerometer was used and the effect of reducing the accelerometer performance was not investigated. Finally, we chose for voluntary breathing, as this would allow us to test the performance on natural variations in the breathing pattern and rate, but we do recognize this does not necessarily cover all cases one could cover using controlled breathing.

We have shown that a chest-worn accelerometer can accurately and robustly measure respiratory effort and respiratory rate under realistic conditions. To further develop the method towards practical application, several steps need to be taken. The relation between posture and the interference of the beating heart is worth investigating further, even if only to improve the quality of the respiratory estimate. In addition, a larger cohort size, with more diversity in age, breathing patterns and movements will add value. Furthermore, improved signal processing techniques, like non-linear filtering and alternative rate estimation methods, such as the Discrete Wavelet Transform, may improve performance. To conclude, future work will focus on evaluating the robustness, validity, and diagnostic utility in clinical populations, for example, in patients with sleep-disordered breathing conditions such as obstructive sleep apnea.

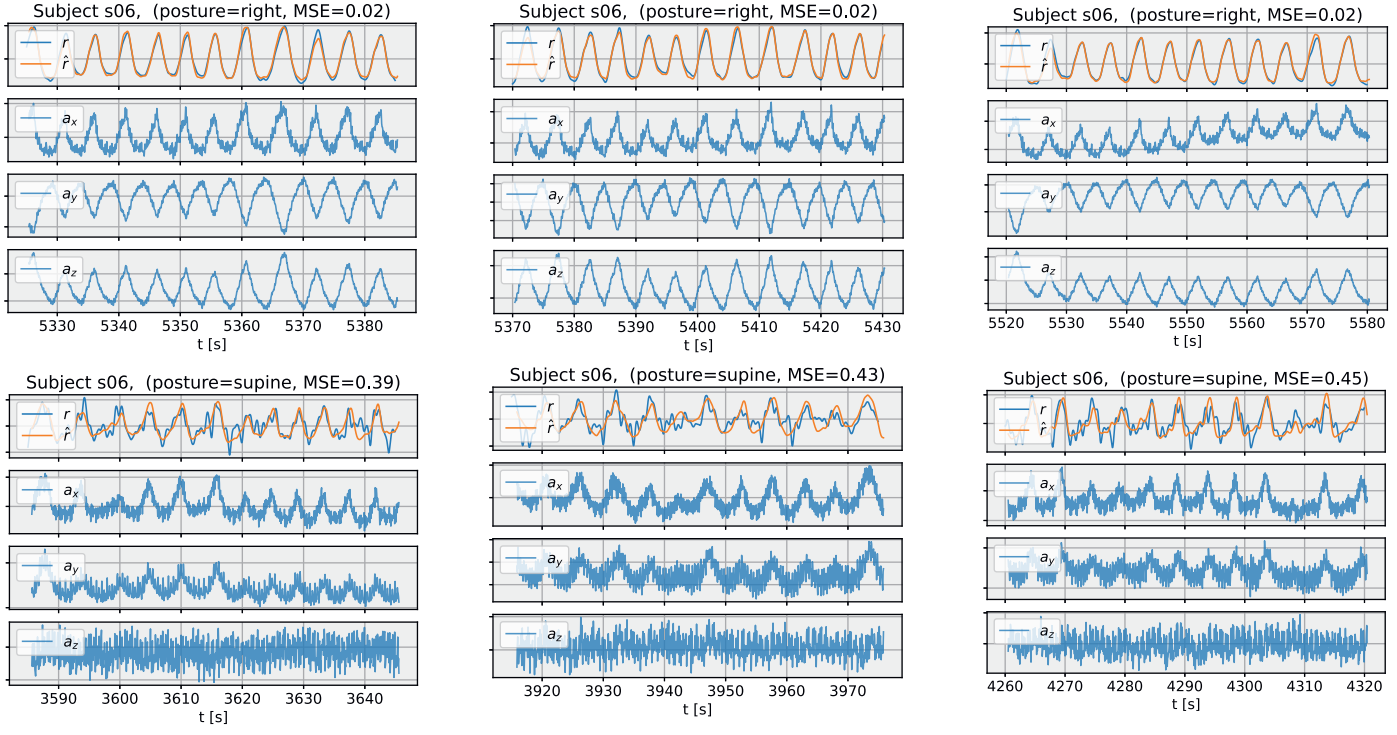


Figure 10 Waveforms for six windows of subject s06. The top graphs are for the right posture and have a low MSE (0.02), while the bottom graphs are for the supine posture and have a high MSE (0.39-0.45). For each window, the reference r, the estimate $\hat{r}$, and the raw acceleration a_x, a_y and a_z are plotted. When visually inspecting the raw acceleration signals, respiratory and noise components can be identified. The respiratory components have lower frequency (in the left bottom graph 0.23 Hz, or 13.8 min^{-1}), while the noise components have a higher frequency (in the left bottom graph 1.3 Hz, or 78 min^{-1}). The amplitude of the noise component is much larger for the supine posture than for the right posture.

Table 5 Comparison of the respiratory rate with other methods that cover ambulatory use-cases.

Dimensionality reduction method	Rate estimation method	Agreement interval [min^{-1}]	Comment	Reference
Constrained PCA	Frequency domain with statistical post-processing	0.8	The agreement interval is at the default operating point. The range is from 0.2 to 1.6 min^{-1}, over all operating points	The method in this paper
PCA	Frequency domain	5.3	The accelerometer and gyroscope signals are fused. The agreement interval is calculated at 10 min^{-1} as limits of agreement are heteroscedastic	[26]
Non-PCA	Time-domain	8.2	Results for guided breathing	[57]
PCA	Frequency domain	16.9	Based on one SD = 4.3 min^{-1}	[51]
Non-PCA	Frequency domain	5.0	At the suprasternal notch	[58]
Non-PCA	Time-domain	5.5	At the suprasternal notch	[54]

For each method, the dimensionality reduction method refers to how a one-dimensional signal is derived, while the rate estimation method refers to how the respiratory rate is derived from this one-dimensional signal. The agreement interval is the difference between the upper and lower limits of agreement.

Table 6 Comparison for the respiratory rate with other methods that cover short intervals of guided breathing.

Dimensionality reduction method	Rate estimation method	Agreement interval	Comment	Reference
Non-PCA	Time-domain	1.4 min^{-1}	Three ways of breathing, three postures, 60 s each (81 measurement points)	[50]
Non-PCA	Time-domain	0.2 %	Seven patterns of instructed breathing	[53]
Non-PCA	Time-domain	2.1 min^{-1}	Two intervals of voluntary breathing for 60 s, with 15 s breath-hold in between	[52]

For each method, the dimensionality reduction method refers to how a one-dimensional signal is derived, while the rate estimation method refers to how the respiratory rate is derived from this one-dimensional signal. The agreement interval is the difference between the upper and lower limits of agreement.

2

Respiratory Effort Estimation with Deep Learning

This chapter corresponds to the manuscript:

F. Schipper, R. JG van Sloun, A. Grassi, S. Overeem, and P. Fonseca, "A deep-learning approach to assess respiratory effort with a chest-worn accelerometer during sleep," Biomedical Signal Processing and Control, vol. 83, p. 104726, May 2023, doi: 10.1016/j.bspc.2023.104726.

Some minor editorial updates have been made.

2.1 Abstract

Objective - To develop a new deep learning method for the estimation of respiratory effort from a chest-worn accelerometer during sleep. We evaluate performance, compare it against a state-of-the art method, and assess whether it can differentiate between sleep stages.

Methods - In 146 participants undergoing overnight polysomnography data were collected from an accelerometer worn on the chest. The study data were partitioned into train, validation, and holdout (test) sets. We used the train and validation sets to generate and train a convolutional neural network and perform model selection respectively, while we used the holdout set (72 participants) to evaluate performance.

Results - A convolutional neural network with 9 layers and 207,855 parameters was automatically generated and trained. The neural network significantly outperformed the best performing conventional method, based on Principal Component Analysis; it reduced the Mean Squared Error from 0.26 to 0.11 and it also performed better in the detection of breaths (Sensitivity 98.4%, PPV 98.2%). In addition, the neural network exposed significant differences in characteristics of respiratory effort between sleep stages ($p < 0.001$).

Conclusion - The deep learning method predicts respiratory effort with low error and is sensitive and precise in the detection of breaths. In addition, it reproduces differences between sleep stages, which may enable automatic sleep staging, using just a chest-worn accelerometer.

2.2 Introduction

Respiratory effort refers to the force applied by the respiratory muscles during breathing and is a key mediator in the regulation of ventilation[59]. The measurement of respiratory effort is a central part of the diagnosis of sleep-disordered breathing, such as Obstructive Sleep Apnea (OSA). Therefore, the measurement of respiratory effort is a standard element in overnight diagnostic polysomnography[39]. Furthermore, respiratory effort may be used to determine respiratory rate[26,60,61]. Interestingly, analysis of respiratory effort may also be used for the determination of sleep stages themselves[40,62,63]. The gold standard for the measurement of respiratory effort is esophageal manometry[59]. Esophageal pressure is a measure of pleural pressure, which in turn is a measure of the force applied by the respiratory muscles. Given its intrusiveness, esophageal manometry is impractical for overnight polysomnography, which is why Respiratory Impedance Plethysmography (RIP) is used as the standard method[39]. RIP measures the volumetric expansion of the chest and abdomen, and is a surrogate measurement, meaning that it does not directly measure respiratory effort, but a derived effect. A typical RIP measurement uses two belts, one on the chest and one on the abdomen. While this is the standard for overnight polysomnography, home sleep testing typically often relies on the use of a single belt, usually worn around the thorax. However, RIP bands are not suitable for all applications, for example due to their high power consumption (23 mW)[64], or because the wearing of a band around the thorax is considered uncomfortable. Therefore, alternative methods have been or are being developed, which may be divided into contactless and contact-based methods. Contactless measurement of chest wall movements may be done by a camera, radar, or even WiFi[38,65]. Contact-based measurement of chest wall movements may utilize strain sensors, impedance sensors, or movement sensors, of which the accelerometer is one [44]. An accelerometer has a low power consumption (typically < 1 mW) and it only requires a single point of mechanical contact with the thorax, such that it may be worn in clothing or be part of a patch.

The use of an accelerometer for the measurement of chest-wall movements has long been studied[45,66]. Most of the studies focus on the measurement of respiratory rate[26,38,61], as this is an important parameter in the health deterioration of patients[37]. As respiratory rate is determined by peak counting or by frequency analysis, the shape of the respiratory waveform is less relevant. However, this is not the case for the purpose of sleep staging or apnea detection, where the shape of the respiratory waveform is important[39,62]. Towards this end, several studies focused on estimating the shape of the respiratory waveform; in one study, the time derivative of the orientation of the chest wall was used to estimate airflow[46], while in another study the angle of the chest-wall orientation is compared against RIP[24]. However, these studies only comprise a few exemplary cases (a 6.9-hour capture of one overnight patient in the first study, and two subjects immobilizing for 90 s

in two postures in the second study). A related study, among 15 participants undergoing overnight polysomnography, looked at the use of an accelerometer at the suprasternal notch instead of the chest[49]. To the best of our knowledge, no study beyond these few exemplary cases investigated the estimation of the respiratory waveform from a chest-worn accelerometer, neither in a general population nor in a population with sleep-disordered breathing (SDB).

All studies into the use of a chest-worn accelerometer for respiratory purposes contain a step in which the raw accelerometer signal is converted into a one-dimensional respiratory waveform. Usually, this comprises some form of pre-processing, like filtering and resampling, and, as most accelerometers measure acceleration in multiple dimensions, a form of dimensionality reduction. Various methods for dimensionality reduction have been used; one method uses the signal of a single axis[45], one uses a fixed linear combination of axes[24], one uses a rotational model[46], while another uses Principal Component Analysis (PCA)[25]. A comparison of methods showed that PCA performed best[26]. Recently, we showed that constrained and recursive PCA, in combination with the estimation of the inspiratory direction, outperformed conventional PCA[61]. However, this study was only performed during daytime, among a relatively small population of healthy participants, with a recording duration of only 90 minutes per participant. In addition, we found that the estimation of the inspiratory direction sometimes failed, causing the estimated waveform to have the wrong sign.

The primary objective of the current study is to assess the performance of a chest-worn accelerometer in estimating the respiratory waveform during sleep; during all sleep stages, in the presence of SDB and under realistic conditions such as body and leg movements. We aim to determine if PCA is the processing method that performs best, or if a different processing method, in particular deep learning, performs better. The reason to investigate deep learning, is that PCA can only model linear relationships, while deep learning can model much more complex and non-linear relationships.

In a set of 146 participants undergoing overnight polysomnography while wearing a chest-worn accelerometer, we evaluate the performance of our PCA method and introduce a new method based on Deep Learning (DL). We present techniques for automatic hyperparameter search and neural network generation, and for the automatic compensation of misalignments between the predicted signal and the reference signal, for which a new step to the Adam algorithm for stochastic optimization was introduced [67].

2.3 Methods

2.3.1 Chest-worn accelerometery and respiration

The main objective is to derive a signal from a chest-worn accelerometer, which resembles the reference RIP band signal (the reference) as closely as possible. As the accelerometer is in mechanical contact with the chest, it measures chest-wall movements, in the form of acceleration and orientation changes[61]. An example waveform from a chest-worn accelerometer is shown in Figure 1 (a), an overview of a whole night is shown. Abrupt, step-like changes in signal amplitude are related to changes in posture (such as at 02:30), while large but brief excursions represent other body movements (such as at 00:15 and 06:15). The magnitudes of these signals are in the range of 1 g (with g the acceleration of gravity). In Figure 1 (b), a detailed view is given, which has a duration of 15 s and starts at 23:18:05 (with horizontal offsets adapted for better viewing). Next to other physiological phenomena, such as heartbeats and snoring (which are outside the scope of this paper), we see respiration as relatively slow variations in the three channels X, Y, and Z. The magnitude of these signals is in the range of mg (milli-g). Finally, Figure 1 (c) shows the reference RIP band signal (y) and the prediction ($\hat{y}$), which is derived from the signals X, Y, and Z in Figure 1 (b).

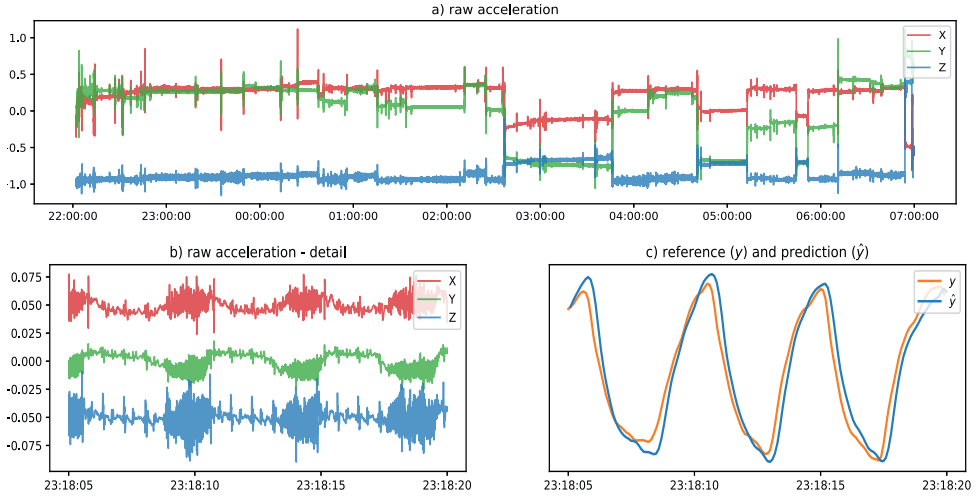


Figure 1 Example waveforms from chest-worn accelerometery: a) view over a whole night, showing body movements, posture and turns, b) detailed view (with horizontal offsets adapted for better viewing), showing respiration and other physiological signals, such as snoring and heart beats (which are outside the scope of this paper), and c) the RIP band reference (y) and the predicted signal $\hat{y}$.

2.3.2 Data collection

Data were collected during a study conducted at Sleep Medicine Center Kempenhaeghe, Heeze, the Netherlands, as part of the *Sleep and OSA Monitoring with Non-Invasive Applications* (SOMNIA) project[68]. Recruitment was done among individuals scheduled for overnight polysomnography (PSG) as part of the standard diagnostic process. Participants undergoing CPAP titration, with intellectual disabilities, or with an earlier diagnosis of Atrial Fibrillation were excluded. Next to standard PSG, the participants wore a 3D accelerometer (ADXL355, Analog Devices, Wilmington MA, USA) mounted on the thoracic RIP band, to measure the respiratory motion of the chest wall. The raw accelerometer signals (X, Y, and Z, sampled at 250 Hz) were recorded in the accelerometer device and were later downloaded into a computer for analysis. Sleep stages and SDB events were scored manually, according to the AASM guidelines[39].

The SOMNIA study was reviewed by the medical ethical committee of the Maxima Medical Center (Eindhoven, the Netherlands. File no: N16.074). All participants provided consent prior to study initiation. The protocol for data analysis was approved by the Medical Ethical Committee of the Kempenhaeghe hospital and by the Internal Committee of Biomedical Experiments of Philips Research.

A total of 146 participants were included (68 female, 78 male), with ages ranging from 17 to 80 years (mean 46.9), demographics are shown in Table 1. The mean recording duration was 8.3 h (SD 0.5 h). Table 2 provides information on the Apnea-Hypopnea Index (AHI).

Table 1 Demographics of the for the full dataset and subsets.

Subset	Count	Female	Male	Age
Full	146	68 (6.0%)	78 (53.4%)	46.9 [17.0-80.0] {33.0-58.8}
Training	59	29 (49.2%)	30 (50.8%)	44.7 [21.0-74.0] {30.0-56.5}
Validation	15	8 (53.3%)	7 (46.7%)	43.1 [18.0-61.0] {34.0-52.5}
Holdout	72	32 (43.1%)	41 (56.9%)	49.4 [17.0-80.0] {36.0-62.0}

The Age column shows the mean age, the range between square brackets and the first and third quartiles between curly brackets.

Before further processing, the RIP and the accelerometer signals were synchronized by visual inspection. Offsets in time, as well as drifts in sampling rates (due to different recording devices), were corrected. Subsequently, the data were randomly split into three subsets, on a participant basis (Table 1). The holdout set was used to evaluate the performance, while the training and validation sets were used to train and validate the DL method.

Table 2 Number of participants and AHI for the full dataset and its partitioning into training, validation, and holdout sets.

Dataset	Count	AHI Mean (SD)	Normal (AHI<5)	Mild (5≤AHI<15)	Moderate (15≤AHI<30)	Severe (30≤AHI)
Full	146	16.5 (18)	42	48	32	24
Training	59	16.8 (19)	19	19	10	11
Validation	15	13.9 (13)	4	6	3	2
Holdout	72	16.8 (16)	19	23	19	11

2.3.3 Performance evaluation

To evaluate performance, data from the holdout set were processed with our DL method. For comparison, the same data were also processed with the PCA method, see Figure 2. The two methods take as input the 3D acceleration signal and produce an estimate of the respiratory effort ($\hat{y}$), which is compared to the reference from the RIP band (y). Using performance metrics, statistics were determined, including factor analysis, in which the factors were the sleep stage (NREM, REM, Wake), the presence of SDB events (Apnea, Hypopnea, None), the presence of snoring, and leg movements.

In the remainder of the methods section, we will describe the PCA and DL methods and the performance metrics.

2.3.4 PCA method

As the PCA method has been extensively described in our earlier study[61], we only give a summary here. The raw acceleration signal is first low-pass filtered and decimated. Because the respiratory pattern manifests itself mainly in the horizontal plane, a projection is made on that plane, which reduces the dimensionality from three to two. The projected signal is used as input to a recursive PCA computation, from which the dominant component is used as an estimate for the respiratory signal. As PCA only estimates a component and not a sign, a classifier is used to estimate the sign, such that inspirations are represented in the positive direction and expirations in the negative direction.

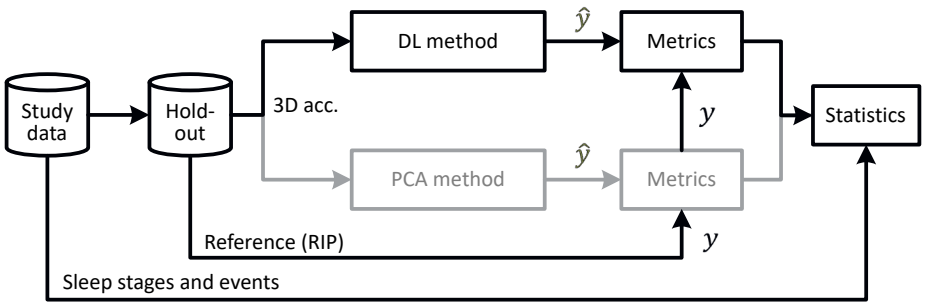


Figure 2 Processing block diagram.

2.3.5 Deep learning method

Architecture

Inspired by the recent successes of 1-D convolutional neural networks[69] and the much older universal approximation theory[70], in which it is shown that neural networks can approximate arbitrary functions, we investigate whether Convolutional Neural Networks (CNNs) can be a good alternative to PCA for this application.

An advantage of using a CNN is that it comes with an associated receptive field and processing equivariance[71]. A receptive field is the time interval that is required to predict one output sample; constraining it makes the system robust against disturbances (such as body movements) that occur outside the receptive field, as these do not influence the predicted sample. Processing equivariance for time shift means that if the same input is offered at a different moment in time then the same output is produced; this is desired as we don't want the mapping of input to output to change over time; if an input segment in the morning is the same as one in the evening before, then the predicted outputs should be the same as well. Furthermore, a CNN reduces the complexity of the training as well as the computational requirements during prediction.

A block diagram of the chosen CNN architecture can be found in Figure 3. It comprises L layers, of which each has kernel size $h^{[l]}$, and $c^{[l]}$ output channels. The kernel sizes $\{h^{[1]}, h^{[2]}, \ldots, h^{[L]}\}$ and output channels $\{c^{[0]}, c^{[1]}, \ldots, c^{[L]}\}$ are the hyperparameters. The number of input channels ($c^{[0]}$) is always 3 and the number of output channels ($c^{[L]}$) is always 1. The number of samples in each channel is kept constant throughout the network, which enables us to keep the sampling rate of input and output the same (10 Hz). Each layer starts with 1-D convolution, followed by batch normalization[72], such that convergence during training is improved. As the signals are normalized and have a mean of 0, we chose tanh-activation, which is symmetric around 0 (while a ReLU and a sigmoid are not). Finally, the last layer (L) has a trainable multiplication factor after its tanh-activation, to allow for amplitudes outside the -1 to +1 range.

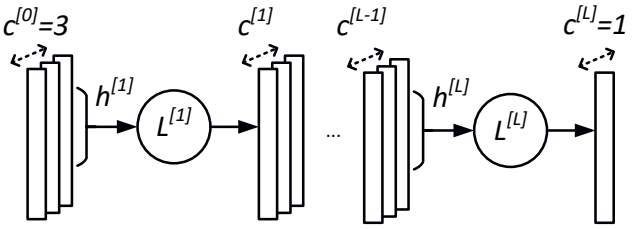


Figure 3 The CNN architecture, comprising L layers. Input is the pre-processed 3-D acceleration signal, output is the estimated respiratory effort signal ($\widehat{Y}$). Each layer L maps $C[L-1]$ input channels to $C[L]$ output channels. A layer comprises 1-D convolution, batch normalization and tanh-activation.

Preprocessing

The CNN is deterministic, so the relationship between its input and output is the same for all participants, for all postures, and all sensor orientations. Preprocessing the raw 3D acceleration signal can help reduce variability across these variables, which makes it easier for the CNN to generalize (to work well under variability), especially if the amount of training data is limited. In our method, the preprocessing helps reduce the variability due to changes in sensor orientation and amplitude, the former through horizontal transformation and gravity removal, the latter through normalization.

A block diagram can be found in Figure 4, which shows that preprocessing comprises three blocks that all operate recursively. The first block takes as input the raw acceleration (at 250 Hz), filters it, and decimates it to 10 Hz. The purpose of the decimation is twofold; it suppresses components that are not due to respiration (such as heartbeats and snoring), and it reduces computational load, size of intermediate storage, and bitrate during further processing. The frequency of 10 Hz was chosen empircally, such that the non-respiratory components were sufficiently suppressed, while the predicted signal ($\hat{y}$) still adequately resembles the reference (y). The second block performs a coordinate transformation, such that the first two dimensions lie in the horizontal plane (and consequently, the third aligns with the acceleration of gravity). The third bock performs normalization.

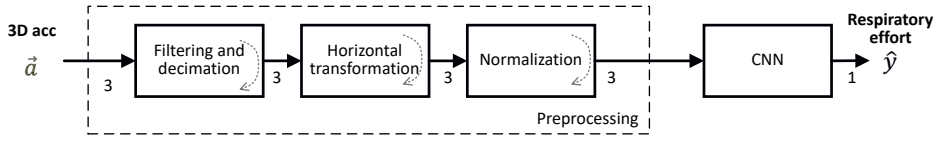


Figure 4 Block diagram of the DL method, comprising preprocessing and a CNN. Input is the raw 3D acceleration signal ($\vec{a}$) and output is the respiratory effort ($\hat{y}$). Preprocessing comprises filtering and decimation, horizontal transformation, and normalization. The dashed arrows in the processing blocks denote recursive operation. The numbers below the arrows represent the dimensionality of the signal.

Example waveforms are shown in Figure 5. The top graph shows the raw signal, including noise and other components, such as heart beats and snoring; the middle graph shows the filtered and decimated signal, in which components other than respiration have been suppressed; the bottom graph shows the signal after horizontal transformation. Note that, although not shown in the figure, the gravity component is still present in the first two signals, and is only removed during the horizontal transformation, such that the three components (H1, H2 and V) are zero-mean. It may be observed that the respiratory component predominantly takes place in the horizontal plane (components H1 and H2). Further details on recursive processing can be found in our earlier work[61].

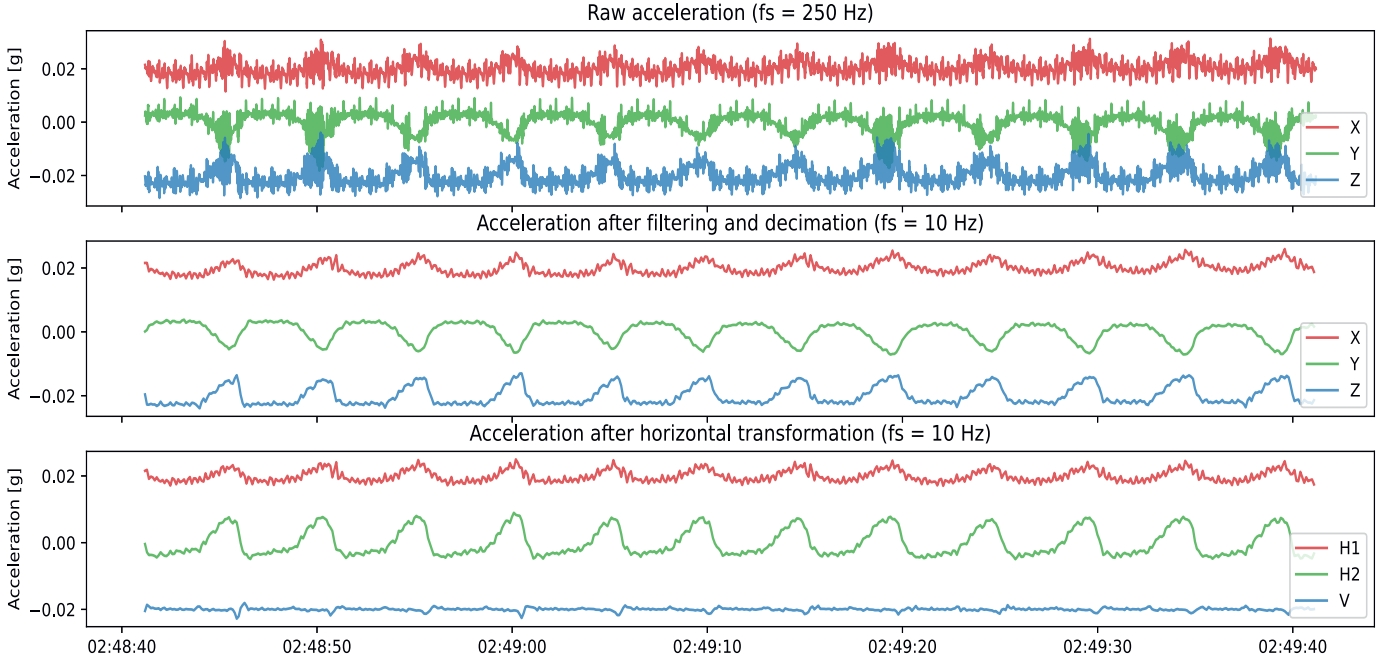


Figure 5 Example waveforms during preprocessing, with adjusted vertical offsets to facilitate good visibility of the respiratory component. Top: raw accelerometer signal. Middle: acceleration after decimation and filtering. Bottom: acceleration after horizontal projection. X, Y and Z are the axes of the accelerometer, H1, H2 are the horizontal components, V is the vertical component, and fs stands for sampling frequency.

One reason to use recursive processing (as opposed to block-based processing) is that it allows for seamless prediction of signals that last a whole night, while only requiring access to a time interval of the acceleration signal with the size of the receptive field.

Automatic hyperparameter search

Many of the CNN parameters are application specific, and architectural specifications do therefore not directly map one sequence-to-sequence application to the other. For example, the required temporal receptive field depends on the frequency spectrum of important concepts in the input, in our case breaths. To optimize the architectural structure of our CNN to predict respiratory effort, we thus perform an automatic hyperparameter search across the number of layers, number of channels, and kernel sizes. We use a combined method for selecting the hyperparameters and for training the network, which is loosely based on One-State NAS [73], in which architecture optimization (hyperparameter search) and network parameter training are performed simultaneously. A block diagram of the method can be found in Figure 6.

In the initial step, P hyperparameter sets are randomly sampled from a constrained hyperparameter space ($P = 600$). For each number of layers L, $L \in \{5,\ 7,\ 9,\ 11\}$, an equal number of samples ($P/4$) is taken. For a specific sample (with given L), the hyperparameters are determined through rejection sampling. First, the kernel sizes $\{h^{[1]}, h^{[2]}, \ldots,\ h^{[L]}\}$ are sampled independently. For each layer l, an auxiliary variable $u^{[l]}$ is sampled: $u^{[l]} \sim U(1,\ 63)$, after which the kernel size is computed as $h^{[l]} = 2u^{[l]} + 1$, such that kernel sizes are odd and in the range from 3 to 127. Secondly, the number of output channels $\{c^{[1]}, c^{[2]}, \ldots,\ c^{[L-1]}\}$ are sampled independently: $c^{[l]} \sim U(8,\ 64)$. Finally, as the total number of parameters of the network may become very large, we reject a sample if the total number of network parameters exceeds 250,000. We continue this process until we have $P/4$ instances for each L.

After the initial sampling, we generate network instances from the samples and perform simultaneous selection and training over successive rounds. Within each round, the instances are trained, and, in between the rounds, instances are either rejected or selected to proceed to the next round. In each round r, C_r instances are trained for a fixed number of epochs, or iterations (20), after which their losses over the validation set are computed. The process is stopped if the decrease in validation loss of the best performing instance in the current round (r), compared to that of the best performing instance in the previous round ($r-1$), is less than 1%. Upon stopping, the best performing instance is selected as the final network, including the network parameters as they were in the final round. The process continues if the decrease in validation loss is larger than 1%, in which case C_{r+1}instances are selected for continued training in round $r+1$. The selection of the C_{r+1} instances for the next round is done by taking the instances with the lowest training loss,

the lowest validation loss, and the largest decrease in training loss, such that C_{r+1} is as small as possible, but $C_{r+1} \geq max(0.5 \bullet C_r,\ 1)$.

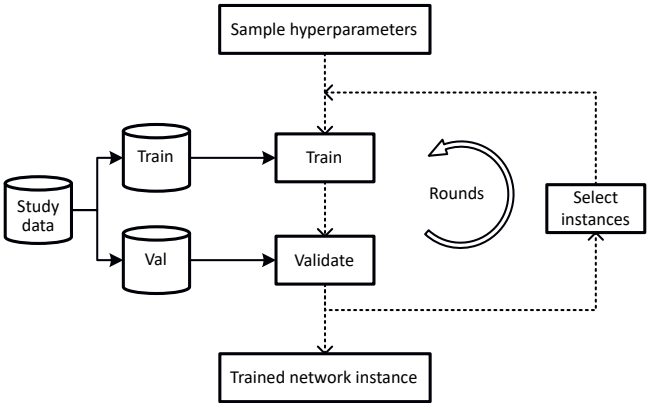


Figure 6 Block diagram of the hyperparameter search and training. First, 600 sets of hyperparameters are sampled, after which, for each set, a CNN instance is generated. In each round, CNN instances are trained (using the training set), after which they are evaluated (using the validation set), and CNN instances are selected for further training in the next round. Finally, a single, trained, CNN instance is obtained.

Network training

The training of a network instance during a round constitutes the (further) adaptation of its parameter values, such that its loss over the training set decreases. If the network instance generalizes well (which depends on the hyperparameters), then also its loss over the validation set will decrease. For computing the loss over a dataset (either the training or the validation set), the data are partitioned into segments of equal length, and the loss over the set is computed as the average loss over the segments. The segmentation concerns the input data (the pre-processed 3D acceleration $\vec{a}$) as well as the reference (y). The prediction for a segment s ($\hat{y}_s$) is computed by applying the network to the input data $\vec{a}_s$ for that segment: $\hat{y}_s = CNN_\theta\left(\vec{a}_s\right)$, with CNN_θ being the network instance with (partially trained) parameters θ.

Even though the reference (y) and inputs $\left(\vec{a}\right)$ were synchronized, small misalignments may remain, which vary over time, for example due to variations in the sampling rate of the recording devices. Although small, these misalignments may significantly influence the loss (for example a misalignment of 1 s may change the loss by 50%, when using a sinusoidal waveform at 15 min^{-1}), which is why we compensate for them by means of automatic reference alignment. Automatic reference alignment is performed by adapting the timing of the reference, using a delay parameter d_s being the number of samples that y_s should be delayed to best match $\hat{y}_s$. As a result, the loss function is computed over $y_s[n - d_s]$ and $\hat{y}_s[n]$, with $n = m + 1 \ldots M - m$, with M the number of samples in a segment,

and $-m \leq d_s \leq m$. We (empirically) chose $m = 15$, and $M = 900$; the choice for m corresponds to 1.5 s, which is sufficient to compensate for the delays as they occurred in our study; the choice for M corresponds to 90s, yielding good convergence of the stochastic optimization. As we do not know what the optimal delays should be ($\hat{y}_s$ must be first predicted for that), we optimize these during the training process. Due to variability in the optimal alignment, each segment s gets its own delay d_s, such that the set of delays becomes $D = \{d_1, \ldots, d_N\}$, with N the number of segments in the data set.

The loss function should direct the hyperparameter search and training process towards good performance according to the metrics as defined in the section on performance metrics. Our loss function quantifies the difference between the predicted waveform ($\hat{y}_s$) and the delay-compensated reference waveform (y_s):

$$L_s\left(\theta, D\right) = \frac{1}{M-2m} \sum_{n=m+1}^{M-m} \left(y_s[n - d_s] - \hat{y}_s[n]\right)^2, \quad (1)$$

with L_s the loss of segment s and θ the (partially trained) parameters of the network instance (note that the overall loss L is the mean over the segment losses L_s, with $s = 1 \ldots N$). The objective of training a network instance is to minimize the loss over the parameter space (θ, D):

$$(\theta_o, D_o) = argmin_{\theta,D}\, L(\theta, D), \quad (2)$$

with θ_o the optimal network parameters and D_o the optimal delays. Clearly, $\theta_o$and $D_o$are interdependent, such that we must optimize them simultaneously, which is why we apply coordinate descent. We update θ and D in turn, each by a small amount at a time. We use the Adam method for stochastic optimization [67], to which we add an additional step. The step comprises the update of the delay, and is inserted at the start of an iteration, prior to computing the gradient with respect to θ (which is normally the first step in Adam), see below. As the loss function (1) is not differentiable in respect to d_s, we compute it for both $d_s - 1$ and $d_s + 1$ and use the sign of the gradient $\{-1,\ 0,\ 1\}$ instead of the gradient itself, such that d_s is only changed in steps of 1 sample.

1. $t \leftarrow t + 1$,
2. $D[t] \leftarrow D[t-1] - \frac{\partial L_s(\theta, D)}{\partial D}$, update the delay parameters (automatic reference alignment)
3. $g[t] \leftarrow \nabla_\theta L_s(\theta[t-1],\ D[t])$, compute the gradient with respect to θ, as is normally done in Adam
4. compute $\hat{m}$[t] and $\hat{v}[t]$ from $g[t]$, estimate moments, as is normally done in Adam
5. $\theta[t] \leftarrow \theta[t-1] - \alpha \bullet \frac{\hat{m}[t]}{\sqrt{\hat{v}[t]} + \epsilon}$. update the network parameters, as is normally done in Adam

2.3.6 Performance metrics

To quantify performance, we compute the mean squared error (MSE), the correlation coefficient, the positive predictive value (PPV), and the sensitivity:

- *MSE* – Once an estimate (ŷ) has been produced then its performance can be determined by comparing it against the reference (y). To quantify how well the predicted waveform resembles the reference, we use the MSE.

$$MSE = \frac{1}{N} \sum_{n=1}^{N} \left(y[n] - \hat{y}[n] \right)^2 \quad (3)$$

To allow the study the influence of factors such as the sleep stages, we split a recording into separate observations. Observations correspond to 60-second segments of a recording, distanced 15 s apart such that they overlap by 45 s. Before computing the MSE, the reference is normalized and automatically aligned to the prediction.

- *Correlation coefficient* – To further quantify the similarity between the estimate ($\hat{y}$) and the reference (y), we compute the Pearson's correlation coefficient for each observation. As the correlation coefficient can only be computed for observations of which the standard deviation of both y and $\hat{y}$ are not equal to zero, observations with zero standard deviation are skipped. The standard deviation of an observation may (temporarily) be zero during non-wearing or the occurrence of a technical problem.
- *PPV and sensitivity* – For the detection of breaths, from which the PPV ($TP/(TP+FP)$) and sensitivity ($TP/(TP+FN)$) are computed, peaks are detected in the waveforms, see Fig. 7. A peak is a local maximum, with a height > 0, a distance > 1.2 s to neighboring peaks (1.2 s corresponds to one period at 50 breaths per minute), and a prominence > 0.5 (the prominence is the vertical distance between the peak and its lowest contour line). PPV and sensitivity are computed by first matching the peaks from y and $\hat{y}$; two peaks match if they are closer in time than 1 s. Matched peaks are true positives; peaks in y that could not be matched are false negatives; peaks in $\hat{y}$ that could not be matched are false positives, see Table 3. We do not compute the PPV and sensitivity for individual observations, but only for complete recordings as the duration of an observation (60 s) is too short to get meaningful results.

Table 3 Classification of peaks into TP, FN and FP categories and the definitions of PPV and Sensitivity.

		Peak is in prediction $(\hat{Y})$		
		Y	N	
Peak is in reference (Y)	Y	TP (The peaks in y and $\hat{y}$ match)	FN (The peak in y does not match with any peak in $\hat{y}$)	**Sensitivity** = TP / (TP + FN)
	N	FP (The peak in $\hat{y}$ does not match with any peak in y)	-	
		PPV = TP / (TP + FP)		

We investigate how various factors influence performance. As a performance metric we use the MSE, as that is computed for each observation. First, we label observations with levels of factors (for example an observation may get the level REM for the factor Sleep stage and get the level Apnea for the factor SDB event), after which we compute the mean and median MSE for the different levels and perform a Mann-Whitney U test between the observations for the different levels to see if differences between levels are significant. We investigate the following factors:

- *Sleep stage* – As levels we have the sleep stages NREM (comprising N1, N2 and N3), REM, and Wake.
- *SDB event* – As levels we have Apnea (obstructive, central and mixed), Hypopnea, and None (neither apnea, nor hypopnea).
- *Leg movement* – Since leg movements may be a source of motion of the chest wall (next to respiration) and as such may be a source of noise, we investigate its influence. As levels we have Movement and None.
- *Snoring* – Snoring may be another source of motion of the chest wall, which is why we investigate its influence. As levels we have Snore and None.

Finally, we investigate if differences between sleep stages can be reproduced by the respiratory effort estimation. From earlier work on cardiorespiratory sleep staging [62], we know that the regularity of breathing varies over the sleep stages and therefore may be used for automatic sleep staging. To quantify regularity, we use two metrics: periodicity and dynamic range. Periodicity is computed through spectral analysis and it indicates the relative height of the spectral peak compared to that of a purely periodic signal[61]. Dynamic range is the difference between the highest and lowest peak in the estimated signal (see Fig. 7). Accordingly, a signal can be considered regular if it has a low dynamic range and a high periodicity, and irregular if it has a high dynamic range and a low periodicity. We investigate if there are significant differences of the regularity between the sleep stages in the reference signal and if these differences are reproduced in the predicted signal. We also check if regularity is correlated to performance, using Spearman rank correlation.

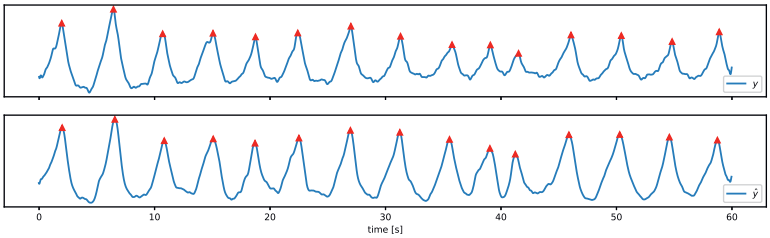


Figure 7 The waveforms of an observation. The top graph shows the reference, while the bottom graph

shows the prediction. Triangles indicate automatically detected peaks in the waveforms, considered landmarks of individual breaths and further used to compute PPV and sensitivity.

2.4 Results

2.4.1 Hyperparameter search and training of the CNN

First, 600 hyperparameter sets were sampled from a constrained space, by means of rejection sampling, meaning that samples not satisfying the constraints were rejected, such that samples that were selected did satisfy the constraints (for further details, see section 2.5.3). For each of the parameter sets, a network instance was generated, such that the number of layers ranged from 5 to 11, the number of parameters ranged from 23,612 to 250,422, and the receptive field ranged from 121 to 865 samples. The training and search process ran for about a week (on an Intel Xeon E-2146G CPU @ 3.50 GHz, and an Nvidia Quadro RTX 4000 GPU card), and terminated after 6 rounds, because the stop criterion was met. From the 10 remaining network instances, the one with the lowest mean validation loss was selected (mean training loss = 0.258, mean validation loss = 0.330, mean test loss = 0.324). The selected instance (see Table 4) had 9 layers, 207,855 parameters, and a receptive field of 297 (29.7 s), which means that to compute a sample of the predicted signal at $t = t_0$, input samples from $t_0 - 14.85$s to $t_0 + 14.85$s are used. The number of input channels is 3 (being the 3 output channels of the preprocessing, H1, H2 and V), while the number of output channels is 1 (being the predicted respiratory signal, $\hat{y}$). The sampling frequency is 10 Hz for inputs, outputs, and all signals throughout the network. To be able to use stochastic optimization, we trained the network using inputs with a length of 900 samples. Because the network is convolutional, it works by sliding kernels over the input sequence. As this allows for seamless processing of input sequences of arbitrary length, we process a recording in its entirety during prediction, alleviating us from the need of assembling an output sequence from individual segments.

Table 4 Overview of the selected model, with the number of output channels, kernel size and number of parameters for each layer (comprising convolution and batch normalization).

Layer	Output channels	Kernel size	Number of parameters
1	51	13	2040 + 102
2	20	11	11,240 + 40
3	50	17	17,050 + 100
4	15	75	56,265 + 30
5	26	101	39,416 + 52
6	20	9	4,700 + 40
7	28	5	2,828 + 56
8	49	53	72,765 + 98
9	1	21	1,030 + 2 + 1
			207,855

The extra parameter of layer 9 is a trainable multiplication factor after the last tanh-activation, such that predicted outputs can become less than -1 or greater than 1.

2.4.2 Performance results

Mean squared error and correlation

Both the DL and PCA methods were evaluated over a total of 146,283 (60-second) observations from the holdout set, corresponding to 610 h of data collected from 72 participants. As no observations were excluded (the coverage was 100 %), the results comprise the full PSG session, including body movements such as posture changes. The summary statistics for the MSE are shown in Table 5, which shows the median, mean and standard deviation (SD) for all observations, and for the factors Sleep stage, SDB event, Leg movement and Snore. The median MSE over all observations is 0.11 for the DL method, while it is 0.26 for the PCA method. The DL method significantly outperforms the PCA method (Wilcoxon signed-rank test, $p < 0.001$) when computed over all observations and when computed over the different factors. For 70 out of the 72 subjects the median MSE obtained with the DL method is less than or equal to the median MSE obtained with the PCA method (Fig. 8). The summary statistics for the correlation can be found in Table 6. Like the MSE, the DL method outperforms the PCA method, with the same significance (with the difference that, for the MSE, outperforming means having a lower value, while for the correlation, it means having a higher value). Note that the number of observations in the correlation table (Table 5) can be lower than that in the MSE table (Table 6) because not for all observations a correlation coefficient could be computed (see section "Performance metrics").

Table 5 Summary statistics for the MSE, for the two processing methods (DL and PCA), for all observations and for the factors sleep stage, SDB event, leg movement and snore.

			DL Method		PCA Method	
		Observations	Median	Mean (SD)	Median	Mean (SD)
All	All	146283	0.11	0.33 (0.5)	0.26	0.53 (0.6)
Sleep stage	NREM	89153	0.06	0.19 (0.3)	0.14	0.34 (0.4)
	REM	15032	0.14	0.26 (0.3)	0.28	0.46 (0.4)
	Wake	26973	0.58	0.69 (0.6)	1.03	1.00 (0.6)
SDB event	None	137213	0.10	0.33 (0.5)	0.24	0.52 (0.6)
	Apnea	1362	0.25	0.43 (0.5)	0.71	0.86 (0.4)
	Hypopnea	7708	0.25	0.38 (0.4)	0.57	0.68 (0.5)
Leg movement	Movement	32416	0.32	0.53 (0.5)	0.70	0.83 (0.6)
	None	113867	0.08	0.27 (0.4)	0.19	0.44 (0.5)
Snore	Snore	4948	0.07	0.25 (0.4)	0.19	0.39 (0.4)
	None	141335	0.11	0.33 (0.5)	0.27	0.54 (0.6)

Table 6 Summary statistics for the correlation coefficient, for the two processing methods (DL and PCA), for all observations and for the factors sleep stage, SDB event, leg movement and snore.

			DL Method		PCA Method	
		Observations	Median	Mean (SD)	Median	Mean (SD)
All	All	145508	0.94	0.83 (0.2)	0.87	0.73 (0.3)
Sleep stage	NREM	89017	0.97	0.90 (0.2)	0.93	0.83 (0.2)
	REM	14945	0.93	0.87 (0.2)	0.86	0.77 (0.2)
	Wake	26479	0.71	0.64 (0.3)	0.43	0.47 (0.3)
SDB Event	None	136438	0.95	0.83 (0.2)	0.88	0.74 (0.3)
	Apnea	1362	0.87	0.78 (0.2)	0.64	0.56 (0.2)
	Hypopnea	7708	0.87	0.81 (0.2)	0.71	0.65 (0.2)
Leg movement	Movement	32375	0.83	0.73 (0.3)	0.64	0.58 (0.3)
	None	113133	0.96	0.86 (0.2)	0.91	0.77 (0.3)
Snore	Snore	4942	0.96	0.88 (0.2)	0.9	0.80 (0.2)
	None	140566	0.94	0.83 (0.2)	0.87	0.73 (0.3)

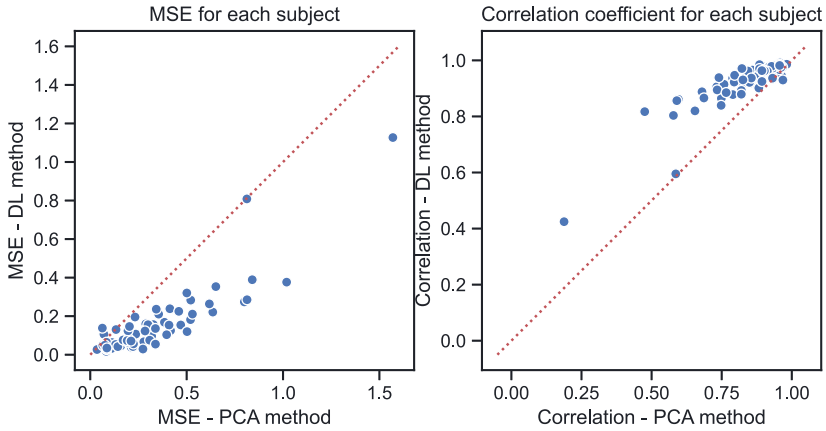


Figure 8 Scatter plots of the median MSE and correlation values for each subject. Left: The MSE for the PCA and DL methods; a point below the dashed red line means that the MSE of the DL method is lower than that of the PCA method. Right: The correlation coefficient for the PCA and DL methods; a point above the dashed red line means that the correlation coefficient for the DL method is higher than that of the PCA method.

Influence of factors on the performance of the DL method

In this section we investigate how factors influence the performance of the DL method. The performance figures for the various levels of each factor may be found in Table 5 for the MSE, and in Table 6 for the correlation. All mentioned differences are significant (Mann-Whitney U test, $p < 0.001$). We have the following factors:

- Sleep stage – Performance is better during non-REM (NREM) sleep than during REM sleep and better during REM sleep than during Wake.
- Leg movement – Performance is better in absence of leg movements than in their presence. Leg movements have a high impact on performance, as they occur quite frequently (in 32416 out of the 146283 observations, corresponding to 22%) and because they increase the MSE from 0.08 to 0.32 and decrease the correlation from 0.96 to 0.83.
- SDB event – Performance is better in absence of SDB events (apnea, hypopnea) than in their presence. Even though the presence of SDB events increases the MSE (from 0.10 to 0.25) and decreases the correlation (from 0.95 to 0.87), their impact on the overall performance is limited, as they occur relatively infrequently (7708 observations for hypopnea and 1362 for apnea).
- Snoring – The presence of snoring decreases the MSE (from 0.11 to 0.07) and somewhat increases the correlation (from 0.94 to 0.96). This is unexpected as snoring is a disturbance in the acceleration signal (see Figure 1); a decrease in performance is expected. The possible explanation is that the presence of snoring excludes other factors that decrease performance, such as being awake and the presence of obstructive apnea events.

Breath detection (PPV and sensitivity)

For breath detection, PPV and Sensitivity were computed for the PCA and DL methods, of which the results can be found in Table 7. DL significantly outperforms PCA (Wilcoxon signed-rank test, $p < 0.001$). Note that once more the metrics are computed over all observations, without discarding any, such that the coverage is 100%.

Table 7 PPV and sensitivity for breath detection, for the DL and PCA methods.

	DL Method		PCA Method	
Metric	Median	Mean (SD)	Median	Mean (SD)
PPV	94.5	93.5 (4.9)	91.1	90.0 (6.4)
Sensitivity	95.5	93.7 (5.3)	91.6	90.8 (6.1)

2.4.3 Relationship between sleep stages, performance, and the regularity of breathing

Figure 9 represents the relationship between the regularity of breathing and the sleep stages. The left pane shows the regularity of the reference signal for each of the sleep stages, while the right pane shows the same for the predicted signal (using the DL method). We have found that breathing is more regular during NREM sleep than during REM sleep, and more regular during REM sleep than during Wake (Mann-Whitney U test,

with $p < 0.001$). This holds for the reference as well as for the predication, and holds when comparing all observations, as well as when comparing subject medians.

Regarding the relation with performance, we have found a correlation of 0.73 between MSE and dynamic range (Spearman rank correlation, $p < 0.001$) and a correlation of -0.67 between MSE and periodicity (Spearman rank correlation, $p < 0.001$), suggesting a relation between regularity and performance (better performance when breathing is more regular). These correlations were for the DL method, for the PCA method they were 0.69 and -0.63, respectively.

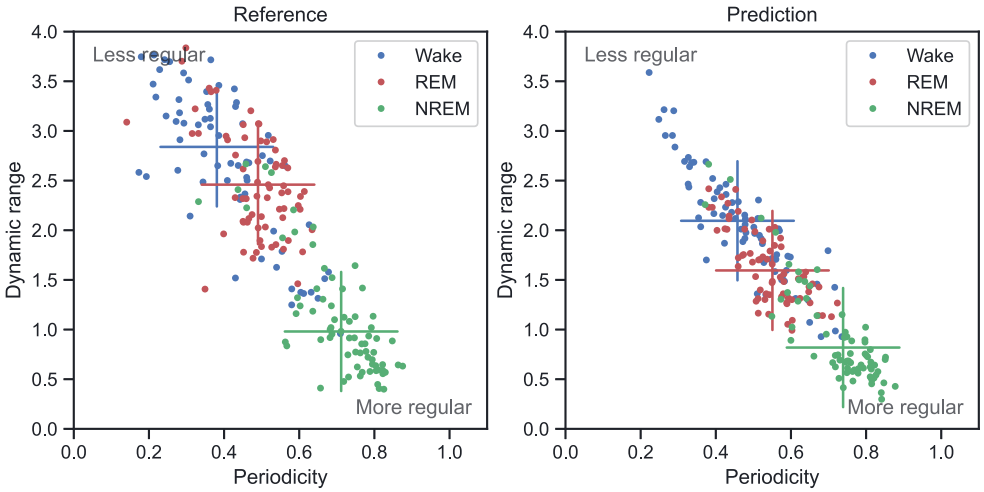


Figure 9 Regularity for the sleep stages NREM, REM and Wake. The left pane shows the regularity computed from the reference signal, while the right pane shows it for the predicted signal. Each dot represents the median value for Periodicity and Dynamic range for a subject, while the cross shows the median over all observations.

2.4.4 Waveform examples

To illustrate the results, Figure 10 presents the waveforms from three different observations (60-second segments), during NREM, REM and Wake. The observations were randomly selected such that their MSE was close to the median for the corresponding sleep stage (as in Table 5). The waveform during NREM is most regular and has the lowest MSE, while the waveform during wake is the least regular and has the highest MSE, which corresponds to our findings.

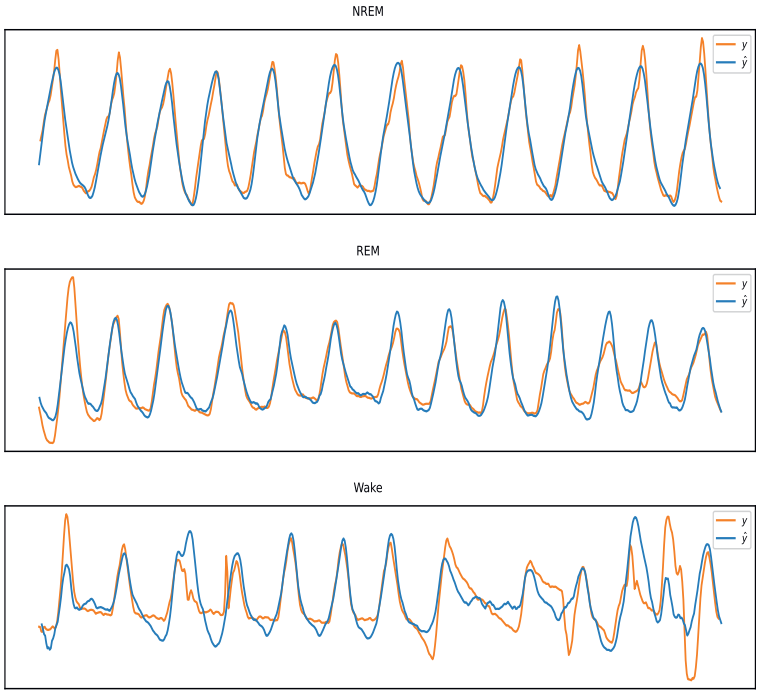


Figure 10 Three examples of waveforms during NREM, REM and Wake. The MSE increases (0.05, 0.14, 0.57), the dynamic range increases (0.41, 0.96, 2.08), and the periodicity decreases (0.81, 0.79, 0.49).

2.5 Discussion

2.5.1 Principal findings and relation to other methods

We evaluated the estimation of respiratory effort from a chest-worn accelerometer in a holdout set of 72 participants undergoing overnight PSG. The study covered the various sleep stages, included body and leg movements, and contained intervals with SDB. The new DL method was trained and validated on a set of 74 subjects. We found that the performance depends on the sleep stage (NREM better than REM and REM better than Wake) and on the presence of leg movements (better in absence than in presence). A relationship between the regularity of breathing and performance was shown; performance was better when breathing was more regular. This matches with the performance difference between the sleep stages; performance was better in sleep stages where breathing was more regular.

In a previous study it was shown that PCA was the best performing method so far[26]. However, we found that DL significantly outperforms PCA; the median MSE decreased with more than a factor 2 from 0.26 to 0.11. The median PPV for breath detection increased from 91.1 to 94.5%, and the median sensitivity increased from 91.6 to 95.5%. These in-

creases correspond to a decrease of the false discovery rate (100 - PPV) from 8.9 to 5.5% and to a decrease of the false negative rate (100 - sensitivity) from 8.4 to 4.5%. It is important to highlight that these numbers were attained without excluding any recorded intervals, such as those during body movements and leg movements.

A reason that DL outperforms PCA may be that PCA is a linear coordinate transformation on a single sample, and taking the dimension with the highest variability is merely an approximation that discards part of the information in the acceleration signal. In contrast, DL uses all the information (all three axes), uses all samples in the receptive field, and allows for non-linear relationships. As per the universal approximation theory[70], it can approximate any function (as long as it is deterministic, which is inherent to a function). Another factor affecting performance is that DL also provides an estimate of the inspiratory and expiratory directions (the sign of the respiratory effort signal), while for PCA an additional estimator is required, as PCA only estimates the axis with maximal variability, but not its sign.

The relationship between the regularity of breathing and the estimation performance may be principal for the use of an accelerometer for this application. When breathing becomes more regular, a better estimate becomes possible; this is supported by the positive correlation between performance and regularity, and by performance being better in sleep stages where breathing is more regular. The relationship not only holds for the DL method but also for the PCA method, which is an indication that estimation performance increasing with regularity may be inherent to the use of a chest-worn accelerometer and does not depend on the processing method. The root cause may be that an accelerometer only measures the orientation of the chest wall at a single point, whereas a RIP belt measures the total circumference of the body. It may be that when breathing becomes less regular, the relationship between the total circumference and the chest-wall orientation at a single point becomes more complex and less deterministic.

2.5.2 The DL method

As for the DL method, techniques beyond neural network training were required to obtain the results as presented, being automatic hyperparameter search, automatic reference alignment, and recursive preprocessing. With the first technique, automatic hyperparameter search, we did not only determine the number of layers, kernel sizes, and the number of trainable parameters to what was optimal for this architecture, but we also determined the receptive field. A receptive field of 29.7 s (from -14.85 to +14.85 s) gave optimal results for the CNN architecture. This matches with the intuition that acceleration values further away in time should have little to no influence on the estimate, but also that a considerable number of samples (297 @ 10 Hz) is required to produce a reliable estimate. The equivariance over time, which comes with a CNN, is a limitation of the

model, as the relationship between the acceleration signals and the respiratory effort is not equivariant; it varies with posture and between participants. However, imposing this limitation (by the choice of the CNN architecture and hyperparameters), does not overly limit the model and allows it to significantly outperform PCA.

The second technique, automatic reference alignment, allowed for the training of the CNN to converge well to an optimal solution. If no automatic reference alignment would be done, the CNN would have needed to compensate for this misalignment by for exampling implementing a phase delay, for which it would require information that is not present in the input signal and to which it therefore has no access. However, the addition of automatic reference alignment increased the complexity of the loss function $L(\theta, D)$, as it became a function of not only the network parameters θ (as is normally the case), but also of the delays D. This was solved by adding a new step to the Adam algorithm, with which we have extended it to perform coordinate descent. Our technique for automatic reference alignment is not restricted to the estimation of respiratory effort and may be used for other applications where alignment with a reference is important, for example in the image domain.

The third technique, recursive preprocessing, reduces variability in amplitude and orientation, such that the CNN generalizes better than without (we found: mean validation loss = 0.330, mean test loss = 0.324). In addition, due to its recursive nature, recursive preprocessing does not extend the receptive field, which allows for processing long recordings while only a part of the signal with the duration of the receptive field (29.7 s) needs to be stored in memory at one time.

The difference in input length during training (900 samples) and prediction (all samples in a recording), causes differences between the prediction from segmented and non-segmented samples, because of padding (affecting the output samples at the boundaries of a segment for a length of half the receptive field). We accept this, because otherwise we would not be able to benefit from stochastic gradient descent (lower computational load and better convergence).

Finally, it seems that the CNN has a regularizing effect on the signal, meaning it tends to decrease dynamic range and increase periodicity, in case the reference is irregular. This is supported by Figure 9, which shows a decrease in dynamic range and an increase in periodicity from reference to prediction and by the positive correlation between performance and regularity that we found.

2.5.3 Applications and future research

Certain applications may become feasible when using a chest-worn accelerometer in combination with the DL method to estimate respiratory effort. A first application is automatic sleep staging. The DL method can reproduce differences in regularity between the sleep stages. As methods exist that use cardiorespiratory signals for sleep staging[63], it may be possible to automatically detect sleep stages with respiratory effort estimated from a chest-worn accelerometer. The required cardiac signals may originate from ECG (in case of a patch), or, if possible, from the accelerometer signal itself. A second application may be the detection of SDB events; even though performance degrades a bit during such events, detection may be feasible. Another application is the estimation of respiratory rate. Although out of scope for this paper, we have already shown good results in combination with PCA [61] and we may expect that these only get better by using DL, as DL increases sensitivity and PPV and reduces MSE.

Some limitations remain and may yield opportunities for future research. Performance varied over the subjects (see Figure 8). Although the impact of factors such as sleep stages, regularity of breathing, and leg movements were studied, there may be other factors that affect performance, such as whole-body movements, accelerometer signal-to-noise ratio, and the quality of the mechanical contact with the chest wall. Related to this, for applications that do not need to monitor respiration continuously, it may be beneficial to know when estimation performance is temporarily deteriorated, such that these time intervals can be discarded, or weighed less. For such applications, it would be beneficial if the signal quality (or estimation uncertainty) could be assessed automatically, using only the accelerometer signals.

In conclusion, we presented a Deep Learning method for estimating respiratory effort from a chest-worn accelerometer that significantly outperforms PCA. As it reproduces differences between sleep stages, it may enable automatic sleep staging from just an accelerometer worn on the chest.

3

Maximum A Posteriori Detection of Heartbeats

This chapter corresponds to the manuscript:

Schipper F, van Sloun RJG, Grassi A, et al. Maximum a posteriori detection of heartbeats from a chest-worn accelerometer. Physiol Meas. 2024;45(3). doi:10.1088/1361-6579/ad2f5e

Some minor editorial updates have been made.

3.1 Abstract

Objective - Unobtrusive long-term monitoring of cardiac parameters is important in a wide variety of clinical applications, such as the assessment of acute-illness severity and unobtrusive sleep monitoring. Here we determined the accuracy and robustness of heartbeat detection by an accelerometer worn on the chest.

Approach - We performed overnight recordings in 147 individuals (69 female, 78 male) referred to two sleep centres. Two methods for heartbeat detection in the acceleration signal were compared: one previously described approach, based on local periodicity, and a novel extended method incorporating maximum a posteriori estimation and a Markov decision process to approach an optimal solution.

Main Results - The maximum a-posterior estimation significantly improved performance, with a mean absolute error for the estimation of inter-beat intervals of only 3.5 ms, and 95% limits of agreement of -1.7 to +1.0 beats per minute for heartrate measurement. Performance held during posture changes and was only weakly affected by the presence of sleep disorders and demographic factors.

Significance - The new method may enable the use of a chest-worn accelerometer in a variety of applications such as ambulatory sleep staging and in-patient monitoring.

3.2 Introduction

The monitoring of cardiac parameters has a variety of applications in medicine. One example is the assessment of acute-illness severity, where heartrate (HR) is measured to compute an early warning score for patient deterioration[37]. Another is the monitoring of sleep using cardiac parameters, as an alternative to polysomnography[63]. For these applications, precise measurement of instantaneous heart rate (IHR) is paramount. IHR is derived from heartbeats, specifically the inter-beat intervals (IBIs) and beat localization times; it is computed as the inverse of the IBIs and can change at every heartbeat. A common characteristic of these applications is the need for prolonged monitoring of patients in bed, either continuously or night-after-night, which is why the measurement must be unobtrusive, be robust against differences in wearing conditions and postures and allow for easy re-attachment of the sensor.

The gold standard for cardiac measurements is electrocardiography, or ECG, in which the QRS complexes are the basis of measurement. The R-peaks are used to localize heartbeats and the intervals between R-peaks (R-R intervals) are used to measure IBIs [74] . Alternative detection methods have been developed, which may or may not require body contact. Among the contactless methods are millimetre wave radar [75] and remote photoplethysmography (PPG)[76]. An often used method that requires body contact is PPG, where the blood volume pulse is measured by means of a sensor on the finger or on the wrist[77].

Another approach is ballistocardiography (BCG), which measures cardiac vibrations through the structure upon which the body resides. The vibrations can be measured using sensors installed on the bed, either on top of or below the mattress. In a comparable approach, seismocardiography (SCG), cardiac vibrations are measured on the chest wall. SCG has long been studied[78] and has regained attention since micro-electromechanical systems such as accelerometers, have become widely available at low cost[79,80]. An SCG measurement may performed by means of an accelerometer inside a device worn on the chest, as an adhesive patch or through a band. The advantages of using a chest-worn accelerometer are multiple; modern accelerometers have a small footprint, are relatively low in cost, have low power consumption, and only require mechanical contact with the chest at a single point. There is no need for skin contact, neither electrical nor optical, which means that biocompatibility issues leading to allergy, or skin irritations can be avoided, and skin preparation is not required. These properties make it suitable for prolonged wearing, or nightly re-application.

Many studies use SCG to investigate cardiac functioning, specifically cardiac time intervals, such as the pre-ejection period and the left ventricular ejection time[81,82]. Some studies also look at the IBIs. In one study, 66 participants equipped with an array of 32 IMUs,

were measured in supine posture for one minute, and IBIs were computed by means of a Hidden Markov Model[83]. Other studies used templates. In one study, 13 participants wore an accelerometer and rested in a chair for four minutes and templates were automatically calibrated for each participant[84]. In another study, 100 participants wore an accelerometer and a gyroscope and were measured in supine position for 6 minutes and 48 seconds in average, after which participant-specific templates were automatically matched[85,86]. In yet another study, 21 subjects equipped with an IMU were measured for two minutes in supine posture, and heart beats were detected by envelope detection using a Hilbert transformation[87].

A common characteristic of all these studies is that the recording times were relatively short and that subjects followed a protocol in which they were asked to not move and remain lying in the same posture. However, during protocol-free conditions, such as sleep, subjects are free to move their body and limbs and take any posture they prefer. A method for estimating the heart rhythm should therefore be robust against these conditions and be accurate at the same time.

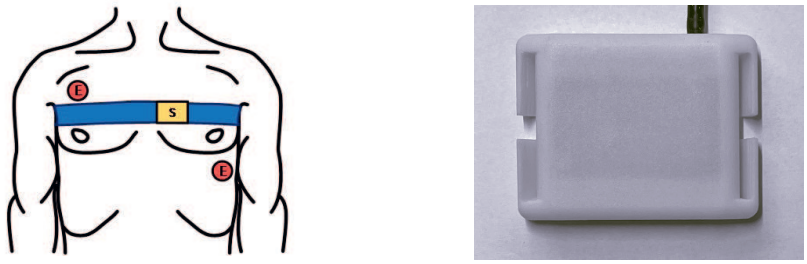
Freedom of movement brings additional challenges to the processing of the data. Global template approaches, for example, cannot be used, as the shape of the waveform may change with posture. This problem has been addressed in the past by exploiting local periodicity, rather than global templates[29]. This method used maximum-likelihood estimation (MLE) for IBIs and was evaluated on BCG signals. It could estimate individual IBIs very accurately. However, IBI sequences could be temporally inconsistent, meaning that sums of successive IBIs could be unequal to the elapsed time, and that time differences between successive beats could be unequal to estimated IBIs. Furthermore, performance was not evaluated on SCG signals.

Here, we evaluate the detection of heartbeats using SCG signals originating from a chest-worn accelerometer. We evaluate the previous approach based on local periodicity[29] and we introduce a new method, which enforces temporal consistency and relies on maximum a-posteriori estimation (MAP) rather than MLE. We compare the two methods on beat localization, IBI estimation, and on the derived metrics IHR and HR. We use overnight sleep recordings from a population that is diverse in gender, body-mass index, age, and includes participants with sleep disorders.

3.3 Methods

3.3.1 Data collection

Data were collected at Sleep Medicine Centre Kempenhaeghe, Heeze, the Netherlands, as part of the SOMNIA project[68]. Recruitment was done among individuals scheduled for overnight polysomnography as part of the standard diagnostic process. Participants undergoing CPAP titration, with intellectual disabilities, or with an earlier diagnosis of atrial fibrillation or other rhythm disorders were excluded. Next to the polysomnographic montage, the participants wore a prototype chest-worn device including a 3D accelerometer (ADXL355, Analog Devices, Wilmington MA, USA), see Figure 1. The device had a cable that was connected to a remote recorder, and staff were instructed to mount the sensor device on the thoracic respiratory belt, such that the cable was at the top right. The raw accelerometer signals (X, Y, and Z, each sampled at 250 Hz) were recorded in the recorder and were later downloaded into a computer for analysis. ECG was recorded by the PSG system, using a modified lead II montage sampled at 512 Hz.



a) b)

Figure 1 a) Location of the sensor device (S) and ECG electrodes (E) on the chest wall of the female/male participant. b) The 3D printed device including the accelerometer, measuring 62 x 48 x 14 mm and weighing 30 grams, with the cable at the top-right.

The SOMNIA study was reviewed by the medical ethical committee of the Maxima Medical Centre (Eindhoven, the Netherlands. File no: N16.074). All participants provided consent prior to study initiation. The protocol for data analysis was approved by the Medical Ethical Committee of the Kempenhaeghe hospital and by the Internal Committee of Biomedical Experiments of Philips Research. A total of 147 participants was included (69 female, 78 male), with ages ranging from 17 to 80 years (mean 46.8, SD 15.9). BMI ranged from 18 to 42 kg/m^2 (mean 26, SD 4). The three sleep disorders with the highest prevalence among the included individuals were Sleep-Disordered Breathing (71), Insomnia (58), and Movement Disorders (31). Recording durations ranged from 6.7 to 9.8 h (mean 8.3, SD 0.5).

3.3.2 Data preprocessing and transformation

Prior to detecting heartbeats, we preprocess the data and transform it into spectral and temporal representations D_S and D_T, as is shown in Figure 2. First, the three components of the acceleration signal a are summed and bandpass filtered with cutoff points of 3 Hz and 25 Hz, such that a one-dimensional SCG signal s is obtained. The 3 Hz cutoff point is chosen to suppress lower-frequency components such as respiration. The summation reduces the sensitivity for changes in the orientation of the sensor with respect to the heart. Even though the instructions in the study prescribe a fixed orientation of the sensor in relation to the chest, in applications in which the sensor does not have a cable, users may mount the sensor in a different orientation. Especially in free-living applications where users re-apply the sensor nightly, there can be quite some variation. Furthermore, the orientation with respect to the heart may vary within a night due to posture changes and arm movements. Finally, the one-dimensional SCG signal s is used for the spectral and temporal transformations.

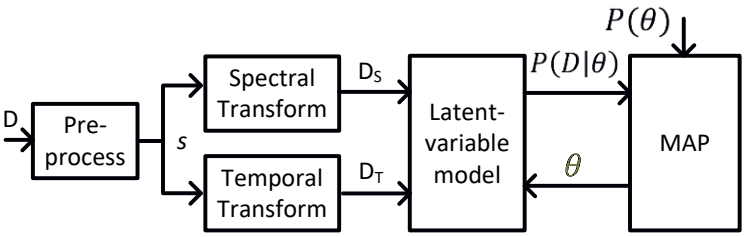


Figure 2 Block diagram of the data processing, with the 3D acceleration signal a, the one-dimensional SCG signal s, the spectrally transformed data D_S, the temporally transformed data D_T, the likelihood $P(D\,|\theta)$, the prior $P(\theta)$, and the MAP estimation θ_{MAP}.

Next to the transformations, the SCG signal s is partitioned into segments. A segment is an interval defined by consecutive samples for which $|s| \leq s_{th}$, with s_{th} defining a threshold (empirically set to 0.06 g), such that large motion artifacts and posture changes are excluded from segments. Heartbeat detection is only performed within segments. In time intervals external to segments, no heartbeat detection is performed, as motion artifacts mask any heartbeats. An example of partitioning into segments may be found in Figure 5a.

The spectral representation D_S comprises local periodicity spectra:

$$D_S = \{S_1(\tau'), S_2(\tau'), \ldots\}, \tag{1}$$

with the spectrum $S_j(\tau')$ representing local periodicity as a function of τ', around a center time $t_m^j = (j-1)\,T$, with T = 200 ms. The local periodicity expresses the similarity of two intervals from s as a function of the period time τ', with one interval ranging

from $t_m^j - \tau'$ to t, and one interval ranging from t_m^j to $t_m^j + \tau'$, see Figure 3. Due to the prominence of the heartbeats in s, high periodicity typically occurs when τ' is equal to the time distance between the last heartbeat before t_m^j and the first heartbeat after t_m^j (which together constitute an IBI). As a typical IBI is much longer than T, multiple spectra will have their peak at the same τ'. When the periodicity over time is plotted on a gray scale, as is shown in Figure 4c, IBIs will become visible as horizontal bars, with height and duration equal the IBI.

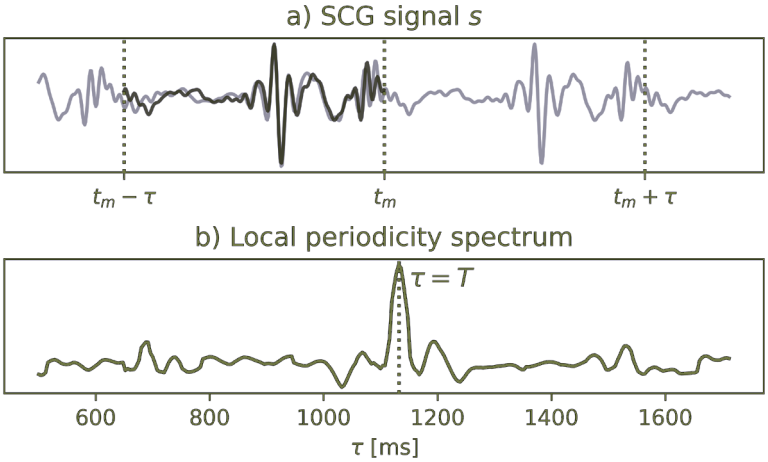


Figure 3 The local periodicity spectrum and how it relates to the SCG signal: a) the SCG signal s, with the interval from t_m^j to $t_m^j + \tau'$ overlayed on the interval from $t_m^j - \tau'$ to t_m^j (in dark black), and b) the local periodicity spectrum, with a peak at $\tau' = 1130$ ms, being the time distance between the high-amplitude vibrations (heartbeats) in the first and second intervals.

The temporal representation D_T expresses the locally normalized magnitude of s, with separate components for the positive and negative parts, see Figure 4d. The normalization for each sample n is relative to the maximum of the product of the absolute value of s and a Gaussian window function centered at the sample n, with a standard deviation of 1.7 s (corresponding to an IBI at a heartrate of 35 beats per minute). If we denominate the normalized positive component f_p and the normalized negative component f_n, then we get:

$$D_T = (f_p(t), \ f_n(t)) \tag{2}$$

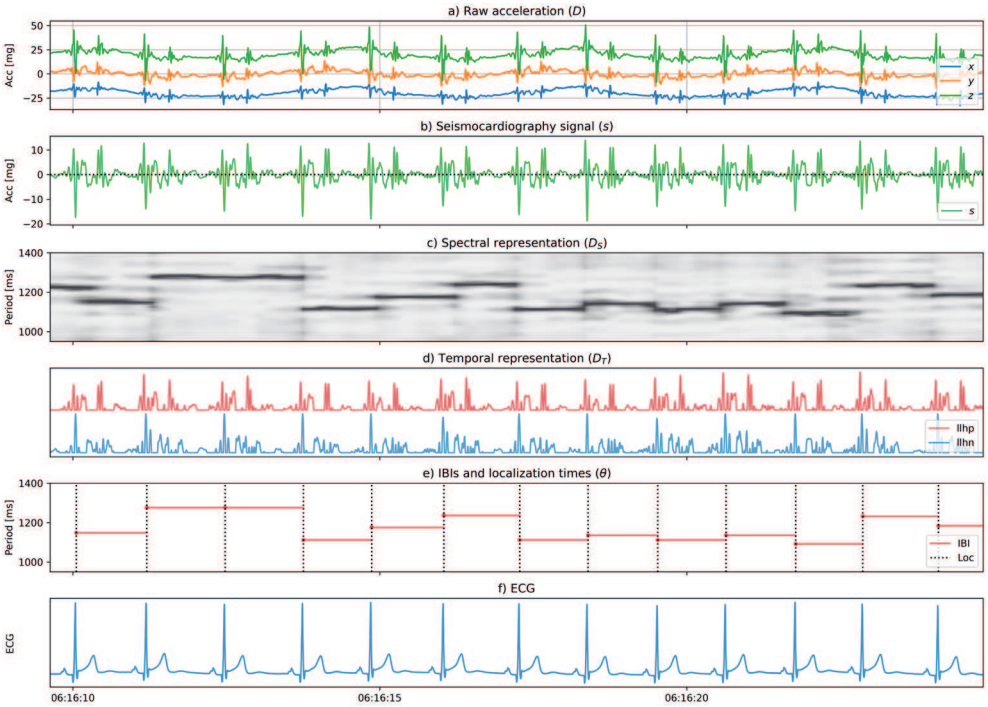


Figure 4 Waveforms for a full example, from top to bottom: a) the raw acceleration data D (with arbitrary offsets for x, y and z), b) the SCG signal s, c) the spectral representation D_S with higher periodicity in darker shades of grey, d) the temporal representation D_T with the positive and negative components ($f_p(t)$ and $f_n(t)$), e) the latent variable θ with the IBIs τ_i, horizontal, at a height equal to the IBI, and the localization times as dashed lines, and f) the ECG signal.

3.3.3 MAP estimation

Core to our method is a latent-variable model. This describes the stochastic relation between observed data D_S and D_T, and a latent random variable $\theta = (t_1, \tau)$, describing the heartbeats with the starting time t_1 and the IBIs $\tau = (\tau_1, \tau_2, \ldots)$. The starting time t_1 is the localization time of the first beat and the IBIs are equal to the difference between the localization times of beats i and $i+1$: $\tau_i = t_{i+1} - t_i$. Given knowledge of t_1, the most probable value of τ can be determined by means of MAP estimation:

$$\hat{\tau} = \ argmax_\tau \ P(\tau \mid D_S, D_T, \ t_1)\Big, \tag{3}$$

which can be rewritten by means of Bayes' rule into:

$$\hat{\tau} = argmax_\tau \ P(D_S, D_T \mid t_1, \ \tau) \ P(\tau \mid t_1)\Big, \tag{4}$$

with likelihood $P(D_S, D_T \mid t_1, \ \tau)$ and prior $P(\tau \mid t_1)$.

Computing the likelihood

First, we factorize the likelihood into individual factors for each IBI i:

$$P(D_S, D_T \mid \tau) \cong \prod_i P(D_S^i,\ D_T^i \mid t_i, \tau_i), \tag{5}$$

with D_S^i and D_T^i the parts of the data in a time interval that is spanned by τ_i (from t_i to $t_i + \tau_i$). The approximation uses the assumption that the likelihood of data from t_i to $t_i + \tau_i$ only depends on the beats i and $i+1$, which is reasonable. Next, we model the likelihood for each IBI as the product of the spectral likelihood $P(D_S^i \mid t_i, \tau_i)$ and the temporal likelihood $P(D_T^i \mid t_i, \tau_i)$:

$$P(D_S^i,\ D_T^i \mid t_i, \tau_i) \cong P(D_S^i \mid t_i, \tau_i)\ P(D_T^i \mid t_i, \tau_i). \tag{6}$$

The temporal likelihood is high for data of which the acceleration has a high magnitude at the occurrence of a heartbeat. Therefore, we interpret the locally normalized magnitude (see equation 2) as the (unnormalized) temporal likelihood:

$$P(D_T^i \mid t_i, \tau_i) = c_T f_a(t_i), \tag{7}$$

with c_T a normalizing constant, t_i the localization time of heartbeat i, and f_a either f_p or f_n, depending on whether we choose to localize beats at the positive or negative component of s, as we will describe later. Figure 4d shows temporally transformed data D_T of high likelihood, as the peaks in $f_n(t)$ coincide with the heartbeats, that are shown as dashed lines in Figure 4e.

The spectral likelihood is high for data of which the spectral peaks match with the IBIs, as is the case in Figure 4c, where the dark horizontal bars (the spectral peaks) match in height and duration with the IBIs in Figure 4e. To compute the spectral likelihood, we define D_S^i as a subset of spectra with their center times within the interval spanned by IBI i:

$$D_S^i = \{S_{n(i)}(\tau'), \ldots, S_{n(i)+N(i)}(\tau')\}, \tag{8}$$

with $n(i)$ the index of the first spectrum with $t_m^{n(i)} \geq t_i$, and $n(i) + N(i)$ the index of the last spectrum with $t_m^{n(i)+N(i)} < t_i + \tau_i$. We interpret the geometric mean of the spectral values in D_S^i as the (unnormalized) spectral likelihood:

$$P(D_S^i \mid t_i, \tau_i) = c_S \sqrt[N(i)]{\prod_{j=n(i)}^{n(i)+N(i)} S_j(\tau_i}, \tag{9}$$

with c_S a normalizing constant. $P(D_S^i \mid t_i,\ \tau_i)$ will be high if all $S_j(\tau')$ in D_T^i have a high peak at $\tau' = \tau_i$.

Computing the prior

The prior $P(\tau \mid t_1)$ in equation (4) is computed for each recording individually, by pre-inspecting D_S. First, for each of the spectra $S_j(\tau')$, the values v_j of the highest peaks (

$v_j = max_{\tau'}\{S_j(\tau')\}$) and the period times p_j of the highest peaks ($p_j = argmax_{\tau'}\{S_j(\tau')\}$) are computed. Next, the period times p_j of high peaks ($v_j \geq v_{th}$, with v_{th} experimentally set to 0.70), are put in a sequence, which is then median filtered (kernel size = 151) to remove outliers. The median-filtered period times, $p_j^{'}$, are put in a new sequence (at their original positions j), and the period times for the positions k for which no sufficiently strong peak was found ($v_k < v_{th}$), are linearly interpolated. The resulting sequence $(p_1^{'}, p_2^{'}, \ldots)$ gives a period time for each time point t_m^j, which are used for the computation of the prior. The prior is then computed as follows:

$$P(\tau \mid t_1) = \prod_i P(t_i, \tau_i), \tag{10}$$

in which each factor $P(\tau_i)$ follows a Gaussian distribution with mean μ_i and standard deviation σ:

$$P(t_i, \tau_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\tau_i - \mu_i}{\sigma}\right)^2}. \tag{11}$$

The mean μ_i is the geometric mean over the periods $p_j^{'}$ with j from $n(i)$ to $n(i) + N(i)$, as in equation (9), and σ is the standard deviation of the period times p_j.

Finding the maximum

With (5), (6), and (10), the expression for the MAP estimate $\widehat{\tau}$ (5) can be rewritten as follows:

$$\widehat{\tau} = argmax_{\tau} \prod_i P\left(D_T^i \mid t_i, \tau_i\right) P\left(D_S^i \mid t_i, \tau_i\right) P(t_i, \tau_i), \tag{12}$$

in which the expressions for the factors in the right-hand side are given by equations (7), (9), and (11).

As there may be many heartbeats in one recording (typically ~30.000), finding the maximum becomes intractable. Fortunately, the problem can be reformulated into a Markov Decision Process, which breaks down the problem into steps. In each step, a current localization time t_i is advanced to a next localization time t_{i+1} (with $t_{i+1} = t_i + \tau_i$) by deciding for a value of τ_i. In each step, a reward $R(t_i, \tau_i)$ is obtained, equal to the log of the corresponding factor in (12):

$$R(t_i, \tau_i) = log\left\{P\left(D_T^i \mid t_i, \tau_i\right) P\left(D_S^i \mid t_i, \tau_i\right) P(t_i, \tau_i)\right\}. \tag{13}$$

The value $V(t_i)$ of a localization time t_i is the maximum of the sum of rewards that may be obtained from t_i onwards:

$$V(t_i) = max_{\tau_i}\{ R(t_i, \tau_i) + V(t_i + \tau_i)\}, \tag{14}$$

with $V(t_i + \tau_i) = V(t_{i+1})$, such that (14) is recursive. Equation (14) is known as the Bellman Equation[88] and provides guidance on decisions for τ_i such that an optimal solution

for equation (12) is found. During each step, the process aims to decide for a value of τ_i that maximizes $V(t_i)$, for which it inspects C IBI candidates, being the τ_i for the C highest peaks in $R(t_i, \tau_i)$. For each IBI candidate, equation (14) is evaluated recursively, up to a finite horizon, or recursion depth, H. The IBI candidate τ_i for which $R(t_i, \tau_i) + V(t_i + \tau_i)$ is maximal is then selected, after which the localization point is advanced to t_{i+1}, until all data have been processed. We chose $C = 3$ and $H = 8$.

The decision process needs a starting point t_1. To this end, the data are partitioned into segments, in which a segment is an interval in which $|s| \leq s_{th}$, with s the SCG signal and s_{th} a threshold (empirically set to 0.06 g). We only process segments with a duration $\geq$ 15 seconds as in shorter segments the impact of body and sensor settling may be too prominent. For each segment, a starting point t_1 is searched such that $R(t_1, \tau_1)$ is maximal. This is done for both polarities of the SCG signal, so with $f_a = f_p$ and $f_a = f_n$, see (7). The polarity (p or n) with the highest reward is then chosen and the decision process commences from t_1 in forward direction as described above, and in reverse direction, in a similar manner, but then with $t_{i+1} = t_i - \tau_i$.

3.3.4 Evaluation

Operating point selection

After estimating the IBIs and localization times, the rewards $R(t_i, \tau_i)$ for each of the beats (t_i, τ_i) are compared against a threshold R_{th} and beats that have their reward below the threshold are discarded. The thresholding enables the selection of an operating point; decreasing R_{th} decreases the number of detected beats and increases the mean posterior probability of the remaining beats.

Synchronization and pairing with ECG beats

We compute the localization times for the ECG signal, through the detection of R-peaks, using a nonlinear transformation in combination with first-order Gaussian differentiation[89]. The IBIs are then computed as the time-differences between the localization times and IBIs that exceed a threshold T_{th} are discarded (we chose T_{th}=1714 ms, corresponding to 35 beats per minute).

Using the localization times of the SCG and ECG beats, we synchronize the two acquisitions, globally and locally. The global synchronization aims at reducing overall clock offset and drift and was performed by minimizing the localization error between all matched ECG and SCG beats within each recording. An ECG beat and SCG beat become a matched pair if they are each other's nearest neighbours in time, and their distance is less than 250 ms. The localization error, with an example shown in Figure 5, may be positive or negative, and may exhibit jumps (caused by changes in the SCG signal morphology) and

trends (caused mainly by clock drift). Using a manual procedure, the offset and drift of the SCG signal are adapted such that the localization error trends disappeared, and the mean localization error becomes zero (globally).

In addition to the global synchronization, we also performed a local optimization aimed at estimating a lower limit for the localization error. For each segment within a recording, the timing offset of the SCG signal is adjusted starting from the global clock offset, until the mean localization error within the segment becomes zero. No drift adjustment is performed for the local synchronization.

Performance metrics and statistics

For performance on beat detection, localization, and IBI estimation, we synchronize over segments without body motion as well as over full recordings. For both cases, we compute the same performance metrics. First, we label all beats; ECG and SCG beats that are in a pair get the label True Positive (TP), ECG beats that are not in a pair get the label False Negative (FN), and SCG beats that are not in a pair get the label False Positive (FP). From these labels we compute the metrics Sensitivity = TP / (TP + FN) and Positive Predictive Value (PPV) = TP / (TP + FP). To get meaningful Sensitivity and PPV metrics, in case of synchronization over segments, we only consider segments with a duration of at least 5 minutes. In case of synchronization over full recordings, all processed segments are included and the Sensitivity and PPV are computed over the full recording.

For the TP beats, we further define the IBI error as the difference between the SCG and the ECG IBIs and the localization error as the difference between the SCG and ECG localization times. Finally, we perform factor analysis for the RMSE IBI error. For the numeric factors (BMI and Age), we perform Spearman rank correlation, while we perform a Mann-Whitney U test for the categorical factors (Sex and presence of the three most prevalent disorders in the study).

For performance of HR and IHR, we synchronize over full recordings. For the IHR, for both SCG and ECG, a stepwise constant function is computed, with levels equal to the inverse of the IBIs and level changes at the localization times. The two functions are then resampled to 10 Hz, after which the error is computed as the difference between the two, thereby discarding gaps. For the HR, for both SCG and ECG, we take 15-second intervals over which we compute the HR as the inverse of the average of the IBIs, thereby discarding intervals with less than 5 IBIs. The error is then computed as the difference between the SCG and ECG HR.

As our method comprises the extension of an existing method [29] with MAP, we also compute the results for the existing method (without MAP). Furthermore, we tune the

operating point of the method with MAP, such that its Sensitivity is close to, but not lower than that of the method without MAP. We compute the statistics for both methods and compare them by means of a Wilcoxon signed-rank test. As we have more than one statistic on the same data, we use a Bonferroni-corrected significance level of α/m ($\alpha = 0.001$, $m = 9$) to prevent false discoveries.

3.4 Results

The overnight recordings of 147 participants were preprocessed and segmented, and the median coverage was 79.9% (Q1 = 65.7%, Q3 = 86.2%). This coverage is the combined result of segmentation, with median coverage of 96.6% (Q1 = 94.7%, Q3 = 97.6%), and thresholding within each segment, as described in the section on operating point selection. We obtained performance metrics for the methods with and without MAP, see Table 1. All errors were significantly lower with MAP than without. The median RMS IBI error decreased by a factor 2.9 from 25.2 to 8.8 ms (the median IBI was 938 ms). The median RMS localization error decreased by a factor 1.3 from 73.2 to 58.1 ms. Although significant, this reduction factor (1.3) is rather limited when compared to the reduction factor for the RMS IBI error (2.9).

Table 1 IBI, Localization, and Detection results when synchronizing over full recordings, computed over subjects (N=147).

	Without MAP			With MAP			
	Med	IQR	95% intv.	Med	IQR	95% intv.	p
IBI							
RMSE [ms]	25.2	18.9 - 38.5	11.1 - 270.2	8.8	6.5 - 13.1	4.3 - 68.3	< 0.001
MAE [ms]	7.4	5.9 - 12.4	4.8 - 116.4	4.1	3.2 - 5.5	2.3 - 15.6	< 0.001
Localization							
RMSE [ms]	73.1	58.7 - 89.0	31.1 - 139.1	58.1	39.5 - 80.7	18.5 - 129.0	< 0.001
MAE [ms]	59.4	44.3 - 74.2	21.1 - 125.5	45.9	29.0 - 65.9	13.5 - 115.6	< 0.001
Detection							
PPV [%]	92.9	85.6 - 97.1	69.4 - 99.3	98.0	93.5 - 99.3	75.3 - 99.9	< 0.001
Sensitivity [%]	76.9	67.5 - 85.1	53.7 - 94.2	79.3	67.5 - 86.0	34.2 - 93.2	-

Figure 5 shows how the offset in the localization error can change within one recording and explains the limited decrease of the localization error in case of synchronization over full recordings. Figure 6 shows two short intervals, one from a segment before a posture change and another from a segment after it. In the first segment, the method localized on an earlier positive peak in the SCG signal, while it localized to a later negative peak in

the second. The difference in localization caused a change in the offset of the localization error, which could not be compensated for when synchronizing over full recordings.

Table 2 Localization and Detection results when synchronizing over segments, pooled over segments (N=3339).

	Without MAP			With MAP			
	Med	IQR	95% intv.	Med	IQR	95% intv.	p
IBI							
RMSE [ms]	15.8	8.1 - 31.0	5.0 - 114.4	5.6	3.8 - 9.7	2.4 - 65.6	< 0.001
MAE [ms]	6.1	4.9 - 9.6	4.1 - 51.1	3.5	2.7 - 5.1	1.9 - 21.8	< 0.001
Localization							
RMSE [ms]	44.6	24.8 - 70.8	8.6 - 132.7	7.9	4.4 - 24.6	2.7 - 109.6	< 0.001
MAE [ms]	30.4	11.7 - 55.5	5.7 - 120.7	4.7	3.2 - 11.5	2.1 - 97.4	< 0.001
Detection							
PPV [%]	98.4	91.9 - 99.7	40.3 - 100.0	100.0	99.7 - 100.0	41.7 - 100.0	< 0.001
Sensitivity [%]	87.4	72.7 - 95.0	30.0 - 99.1	88.9	76.3 - 95.2	7.4 - 98.9	-

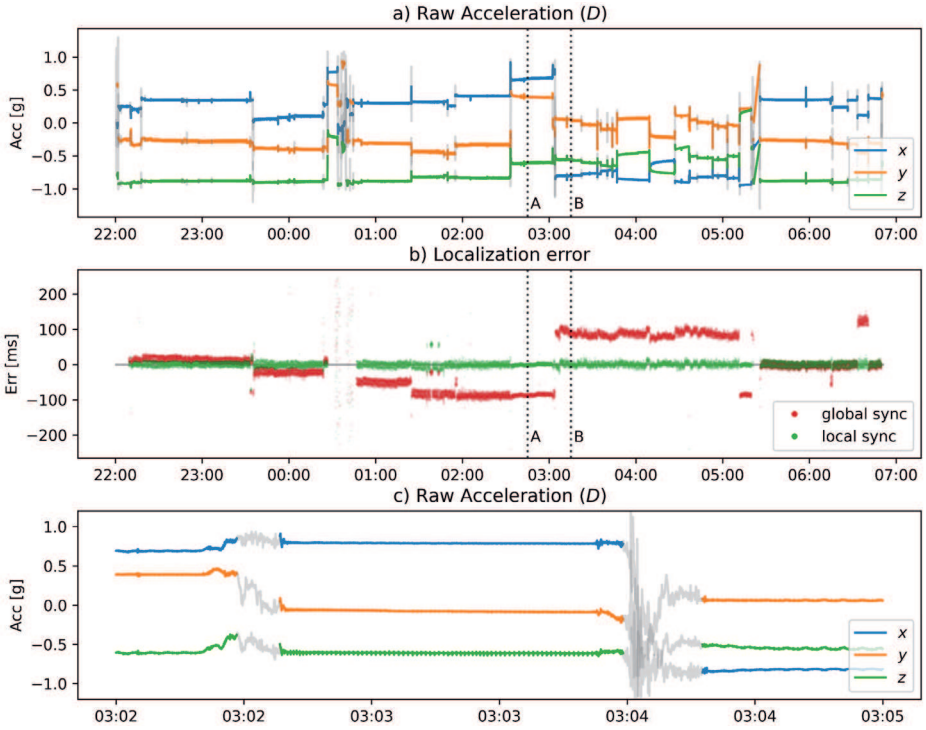


Figure 5 Segments and localization errors. a) The raw acceleration signals for a full recording. The signals are coloured within segments and greyed outside of segments. A total of 70 segments is present with a mean duration of 443 seconds, covering 97.4% of the recording time. b) The localization errors for synchronization of SCG and ECG over the whole recording (global sync) and over segments (local sync). The

markers A and B are timepoints before and after a change in sleeping posture from left to right, and they are detailed further in Figure 6. c) The raw acceleration zoomed in around the posture change between A and B (at t= 03:04:00).

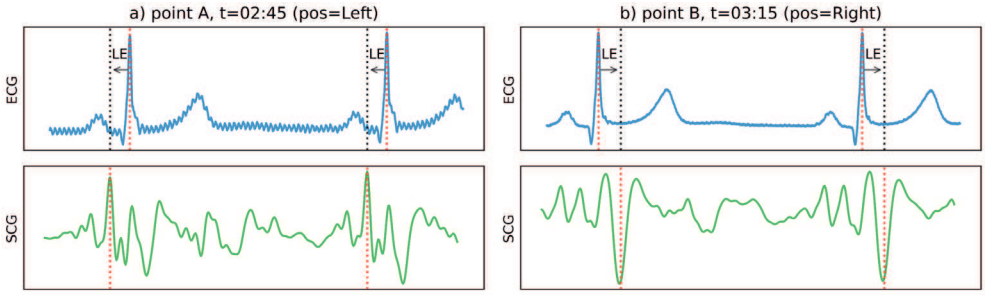


Figure 6 ECG, SCG and beat localizations for timepoints A and B from Figure 5, after global synchronization (LE = localization error, pos = sleeping posture). The top graphs show the ECG signal with localized R-peaks. The bottom graphs show the SCG signals with localized heartbeats. The dashed arrows show the localization error. As the morphology of the SCG signal changes, the beat detection algorithm localizes at two different points in the cardiac cycle:at t= 02:45 it localizes at an early positive peak of the SCG signal, while at t= 03:15 it localizes at a later negative peak. As a result, the localization error jumps from negative to positive.

When synchronizing over segments, on the other hand, the offset in the localization error could be compensated for, and the median RMS localization error decreased from 58.1 ms (as in Table 1, with MAP) to 7.90 ms (as in Table 2, with MAP). This decrease shows that the higher localization error in table I was only due to changes in offset and that the MAP method consistently localizes at the same point (phase) in the cardiac cycle. Again, all errors were significantly lower with MAP than without. The IBI localization error decreased by a factor 2.8 from 15.8 to 5.6 ms, while the RMS localization error decreased by a factor 5.6 from 44.6 ms to 7.9 ms. As only segments longer than 5 minutes were included in Table II, the coverage was 77.9% (Q1 = 67.0%, Q3 = 84.9%).

The IBI error for the method with MAP, when aggregated over all recordings ($N=3.5{\cdot}10^6$ beats), had 95% limits of agreement of -10.6 and 11.1 ms, while the quantiles Q1, Q2, and Q3 were -2.8, 0.0, and 2.8 ms respectively. Mean coverage was 74.6%, over 1236.8 recorded hours.

We investigated the influence of various factors on the performance. The RMS IBI error increased ($p < 0.001$) due to the presence of sleep-disordered breathing (median from 5.2 to 6.2 ms, +19%), movement disorders (median from 5.4 to 6.4 ms, +19%), and sex (median in males 5.4 ms, median in females 5.9 ms, +9%). Furthermore, positive rank correlations for the RMS IBI error ($p < 0.001$) were found for BMI (0.12) and age (0.22).

These numbers relate to the method with MAP in combination with synchronization over segments.

We computed the derived metrics IHR and HR, of which Bland-Altman density plots are shown in Figure 7. For the method with MAP, the median error and 95% limits of agreement were 0.00, -1.20 and 1.25 bpm for the IHR and 0.00, -1.65 and 1.03 bpm for the HR. If we define a measurement as being accurate if it is within a range around the reference of either ±5 bpm or ±10% (whichever is larger), then 98.8% of the IHR measurements and 100.0% of the HR measurements were accurate. These percentages were obtained at median coverages of 80.2% and 79.7% respectively. The Spearman correlation coefficient was 0.992 for the IHR ($p < 0.001$) and 0.997 for the HR ($p < 0.001$).

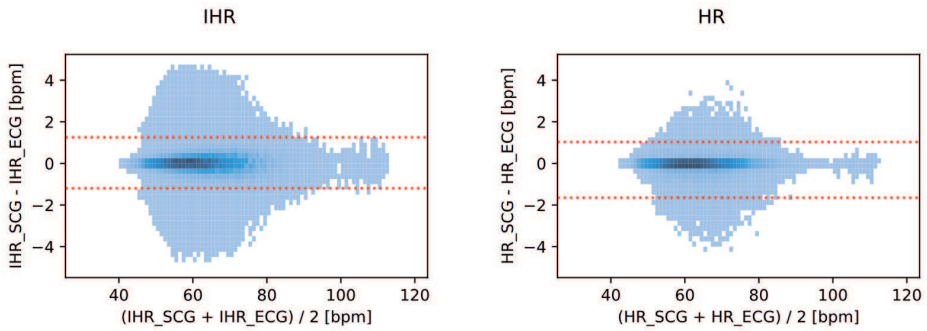


Figure 7 Bland-Altman density plots for IHR and HR (method with MAP). The red lines show the 95% limits of agreement. Density cutoff levels are 300 (IHR, N=4.5·106 beats) and 8 (HR, N=3·105 intervals).

3.5 Discussion

We evaluated heartbeat detection from a chest-worn accelerometer in a set of 147 overnight recordings among a diverse population with a variety of sleep disorders. We determined performance for the primary metrics beat detection, beat localization, and IBI estimation, and for the derived metrics HR and IHR. IBI estimation was accurate with an MAE of 3.5 ms, which is less than the sampling period of the SCG signal (4.0 ms) and less than 0.5% of the median IBI time (938 ms). For the derived metrics IHR and HR, we found that 98.8% and 100.0% of all measurements were accurate according to the American National Standard "Cardiac monitors, heart rate meters, and alarms" [90]. It should be noted that these percentages were obtained over the covered parts of the recordings (median around 80%).

We have found that adding MAP to the detection method significantly improved the performance. Without reducing sensitivity, the RMS IBI error decreased by a factor 2.8,

while the RMS localization error decreased by a factor of 5.6. A possible reason for the improvement is the larger scope of the algorithm; when estimating one IBI (in the Markov Decision Process), all data up to 8 IBIs ahead contribute to the choice for that single IBI.

Table 3 compares our results with other studies on seismocardiography, which reported IBI performance. It should be emphasized that a direct benchmark is not possible, since studies used different protocols, and ultimately, different measuring setups, accelerometers and even performance metrics, so this serves merely as a high-level comparison. Our study shows that it is possible to achieve a high level of accuracy in IBI, HR and IHR estimation also on longer recordings, including changes of posture and body movements, achieving errors at least comparable with the best studies reported even if those were obtained in shorter recordings with subjects laying still in fixed postures (mostly supine).

Table 3 Comparison of results against other studies.

	Year	Study	Sensor	N	Dur	Posture	Results	Method
A	2017	[83]	Acc	66	1 min	Supine	IBI MAE = 6.10 +/− 31.89 ms	Hidden Markov model
B	2019	[91]	IMU	25	7 min	Supine	HR LoA = -5.06 to 5.66 bpm	Autocorrelated differential algorithm
C	2020	[84]	Acc	20	50 min	Supine	IBI RMSE = 4.2 ms	Automatic template calibration and matching
				13	4 min	Resting in chair	IBI RMSE = 6.2 ms	
D	2021	[92]	Gyroscope at head	19	5 min	Supine	HR MAE ≈ 2.5 bpm	Envelope based (Hilbert transform)
E	2021	[93]	Acc	30	40 min	All	HR LoA = -19.3 to 17.8 bpm	Envelope based
F	2023	[87]	IMU	21	2 min	Supine	LoA = -26.43 to 26.43 ms	Envelope based (Hilbert transform)
G	2023	This study	Acc	147	8.3 hrs	All	IBI RMSE = 5.6 ms IBI MAE = 3.5 ms LoA: - IBI: -10.6 to 11.1 ms - HR: -1.6 to 1.0 bpm - IHR: -1.2 to 1.3 bpm	Maximum a posteriori estimation

As for the localization error, we found that it was substantially higher when the synchronization was performed over whole recordings (MAE = 45.9 ms, Table 1), than over segments without body movements (MAE = 4.7 ms, Table 2). The body movements, and perhaps more importantly, posture changes, cause the MAP search algorithm to use dif-

ferent localization points in the cardiac cycle. These cause a change in the offset in the localization error, which is removed by synchronization within a stable posture period. In these intervals, the algorithm consistently assigns localization points to the same point in the cardiac cycle, achieving a relatively small localization error. Note that these intervals were detected by the algorithm itself from the acceleration data, without manual annotations of any kind.

The influence of demographics and sleep disorder prevalence on the performance was limited. Spearman correlation coefficients for age and BMI were only 0.22 and 0.12, while the largest increase in RMS IBI error was due to the occurrence of sleep- disordered breathing (from 5.2 to 6.2 ms, +19%), such that still a high level could be maintained. Only a small increase (+9%) in IBI error was found for female over male participants. This shows that the method is robust against body movements and posture changes but also against the presence of sleep disorders and different demographic factors, such as gender.

As stated, the MAP estimation was beneficial for the robustness and accuracy of the method. The probabilistic approach with the latent variable model, Bayes' rule, and the MAP estimation, provided a structured framework for estimation. Separate models could be created for the likelihood and the prior, and the product of the two automatically yielded an objective function. The chain rule of probability allowed for the reformulation of the MAP problem into a Markov Decision Process problem, of which an optimum solution is described by the Bellman Equation. The optimum can be approximated by a tree-search algorithm, of which the width and depth can be adjusted to trade accuracy against run time.

A limitation of this study is that subjects with prior diagnosis of cardiac rhythm disorders, most notably atrial fibrillation, were excluded from the study. Especially during sleep monitoring, atrial fibrillation is highly prevalent among patients with sleep-disordered breathing[94]. During cardiac arrhythmias, local periodicity, upon which the method relies, may be absent and therefore results may be worse. Future work should include patients with cardiac arrhythmias and annotation of cardiac rhythms should be made, such that performance under such conditions can be investigated. Another limitation is that the computation of the prior was done by inspection of the full recording; we did not investigate the effect of a constrained look-ahead, which is required for low-latency applications. Furthermore, the Markov Decision Process may be improved; instead of exhaustive search, dynamic programming or reinforcement learning may be applied. Finally, further research may be conducted in the mapping of the three accelerometer axes into the one-dimensional SCG signal. The summation that we used in the present study, reduces sensitivity to changes in the orientation of the sensor with respect to the heart and maintains periodicity, as any linear combination of signals with period T is also periodic with

T. However, noise may not be uniformly distributed over the axes, so alternative methods may be evaluated in future work, such as a vector sum which may increase SNR while preserving amplitude periodicity.

The possibility to accurately and robustly estimate HR and IHR, as presented in this paper, together with our earlier published work on respiration[61,95], may enable cardio-respiratory applications to work with a single sensor. The sensor can be small and only needs mechanical attachment at a single point on the chest, without the need for skin contact. One example of such an application is the detection of patient deterioration using respiration rate and HR[93]. Another application may be in positional therapy devices for obstructive sleep apnea, which use a chest worn tactile feedback system to entrain the patient to avoid supine sleeping, thereby reducing the apnea burden[23]. The addition of our method may enable the estimation of residual apnea burden, quantified in the form of the apnea-hypopnea index by using IBIs and respiratory effort from an accelerometer in the therapy device itself [96].

In conclusion, we propose a method that detects heartbeats from a chest-worn accelerometer, demonstrating robustness and high accuracy during overnight recordings with unrestricted postures and body movements, in a diverse population and in the presence of disorders. This enables applications, that directly measure respiration, to monitor cardiac activity robustly and accurately without the need for an additional sensor.



Sleep Stage Estimation using Chest-Wall Accelerometry

This chapter corresponds to the manuscript:

Schipper F, Grassi A, Ross M, Cerny A, Anderer P, Hermans L, van Meulen F, Leentjens M, Schoustra E, Bosschieter P, van Sloun R, Overeem S, Fonseca P. Overnight Sleep Staging Using Chest-Worn Accelerometry. Sensors. 2024;24(17). doi:10.3390/s24175717.

Some minor editorial updates have been made.

4.1 Abstract

Overnight sleep staging is an important part of the diagnosis of various sleep disorders. Polysomnography is the gold standard for sleep staging, but less-obtrusive sensing modalities are of emerging interest. Here, we developed and validated an algorithm to perform "proxy" sleep staging using cardiac and respiratory signals derived from a chest-worn accelerometer. We collected data in two sleep centres, using a chest-worn accelerometer in combination with full PSG. A total of 323 participants were analysed, aged 13–83 years, with BMI 18–47 kg/m2. We derived cardiac and respiratory features from the accelerometer and then applied a previously developed method for automatic cardio-respiratory sleep staging. We compared the estimated sleep stages against those derived from PSG and determined performance. Epoch-by-epoch agreement with four-class scoring (Wake, REM, N1+N2, N3) reached a Cohen's kappa coefficient of agreement of 0.68 and an accuracy of 80.8%. For Wake vs. Sleep classification, an accuracy of 93.3% was obtained, with a sensitivity of 78.7% and a specificity of 96.6%. We showed that cardiorespiratory signals obtained from a chest-worn accelerometer can be used to estimate sleep stages among a population that is diverse in age, BMI, and prevalence of sleep disorders. This opens the path towards various clinical applications in sleep medicine.

4.2 Introduction

Assessing overnight sleep structure is a cornerstone of the somnological diagnostic process and is an important step in the diagnosis of various sleep disorders. The gold standard for the scoring of sleep stages uses the rules of the American Academy of Sleep Medicine (AASM) and is based on neurological signals measured during polysomnography (PSG), specifically electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG)[8]. Sleep stages encompass "light sleep" (stage N1 and N2), "deep sleep" (slow wave sleep, stage N3) and REM sleep. The occurrence of sleep stages overnight is typically plotted in a hypnogram to visualize sleep structure. Furthermore, derived metrics such as total sleep time (TST) or sleep efficiency are often some of the diagnostic indicators describing the burden or effect of a sleep disorder. One important example is the Apnea-Hypopnea Index (AHI) as a severity estimator for obstructive sleep apnea (OSA). The AHI is defined as the number of apnea and hypopnea events per hour of sleep. As such, estimation of total sleep time is crucial for a proper determination of the AHI. The absence of sleep staging (e.g., in reduced montage polygraphy) forces the use of total recording time instead of TST and leads to an associated underestimation of the AHI[97].

PSG may be conducted in a clinic or at home but is often restricted to a single night because of its obtrusiveness and associated handling costs. PSG is therefore impractical for the long-term monitoring of sleep at home. Furthermore, to determine the efficacy of a sleep-related therapy, prolonged measurements are required as well. To enable such measurements, alternatives to PSG have been studied over the past decade. One approach uses EEG measurements with a reduced number of electrodes, which can be self-attached[98,99]. In other, increasingly popular approaches, sleep staging is performed without EEG, but with the measurement of cardiac and/or respiratory activity instead. These exploit the increasingly well-understood link between autonomic nervous system activity and sleep stages[100–102]. For example, heart rate decreases during deeper non-REM sleep stages (N2-N3), while REM is a period of autonomic instability, with heart rate variability reflecting the constant change in balance between sympathetic and parasympathetic tone. Comparably, respiration becomes more regular in non-REM stages and less regular during REM sleep [40,62,103].

With advances in machine learning it has become possible to exploit these variations in cardiac and respiratory signals for the estimation of sleep stages. Cardiorespiratory signals thus act as a "surrogate", as opposed to the neurological signals that are the formal foundation for sleep staging[104]. In a recent review, researchers show that approaches based on the simultaneous use of cardiac and respiratory signals tend to achieve the most accurate results[105]. Note that, as differences between sleep stages may be less apparent

in cardiorespiratory signals, classification is often performed with a reduced number of stages, for example Wake/REM/Non-REM (3-class), or Wake/REM/N1+N2/N3 (4-class). The CardioRespiratory Sleep Staging (CReSS) algorithm uses both cardiac and respiratory signals to perform 4-class staging. CReSS was validated using 296 overnight PSG recordings and a substantial agreement was found between sleep stages derived from the AASM-recommended signals for sleep staging (EEG, EOG, EMG) and those derived from only the cardiorespiratory signals in the PSG[63].

Recent advances in wearable cardiorespiratory sensors, for example based on photoplethysmography (PPG), but also low-power and low-noise accelerometers, enable new approaches to longitudinal and minimally obtrusive measurement of sleep in general, and sleep staging in particular[106]. These developments have been the driver behind a plethora of sleep trackers, that to some degree, allow for the assessment of sleep-wake patterns in healthy individuals[107,108]. Further developments in artificial intelligence, in combination with the availability of clinical datasets that combine PSG with new sensors, have led to substantial performance improvements, also in simplified sleep staging tasks. In one study for example, the use of heart rate variability derived from wrist PPG, in combination with actigraphy derived from wrist accelerometry, enabled sleep staging in a clinical population[109].

Another interesting sensing modality in this setting is chest-worn accelerometry. These sensors are typically very small, consume little power, and can be integrated in devices that are worn on the chest anyway, such as monitoring patches and therapy devices for positional obstructive sleep apnea[23,110]. In addition, such a device only requires a mechanical coupling with the chest at a single point, without the need for direct skin contact, so that it can be worn in or over clothing and does not depend on galvanic or optical contact properties.

Recent studies have shown that cardio-respiratory metrics can be accurately and robustly derived from a chest-worn accelerometer. By measuring small changes in the orientation of a chest-worn accelerometer, and processing these with a deep neural network, respiratory effort can be estimated[61]. Cardiac activity can be measured by a method that uses tiny vibrations in the chest wall, in combination with maximum a-posteriori estimation to localize heartbeats and to estimate inter-beat intervals (IBIs)[111].

The objective of this study is to investigate whether sleep stages and associated sleep metrics can be accurately estimated from a chest-worn accelerometer, leveraging the recently developed methods for estimation of cardiac and respiratory activity from this sensor, in combination with established methods for cardiorespiratory sleep staging[63].

4.3 Methods

4.3.1 Data collection

Data were collected in two sleep centres, from participants undergoing full overnight PSG, as part of a routine investigation because of a suspected sleep disorder. Simultaneous with the PSG, a triaxial accelerometer device (ADXL355, Analog Devices, Wilmington MA, USA) was mounted in a housing measuring (62 x 48 x 11 mm) on the thoracic respiratory band, approximately 50 mm left of the sternum. The three channels of the accelerometer (x, y, and z) were sampled and recorded at 250 Hz.

The first centre (KH) was the Sleep Medicine Centre Kempenhaeghe, Heeze, the Netherlands. Data collection was part of the SOMNIA project, reviewed by the Medical Ethical Committee of the Maxima Medical Centre (Eindhoven, the Netherlands. File no: N16.074)[68]. The second centre (OLVG) was the Sleep Center of the OLVG Hospital in Amsterdam, the Netherlands. The study protocol was reviewed and approved by the local scientific advice committee of OLVG (Amsterdam, The Netherlands. File no: WO 20.134). Both protocols were approved by the Internal Committee of Biomedical Experiments of Philips Research. Recruitment was among participants scheduled for overnight PSG as part of the standard diagnostic process. Participants undergoing CPAP titration, with intellectual disabilities, or with an earlier diagnosis of atrial fibrillation were excluded. All participants provided consent prior to study initiation. In OLVG, 10% of the recordings were made in the clinic, with in-lab PSG, while the remaining 90% were made at home, with ambulatory PSG. All sensors, including the accelerometer device, were mounted by the clinical staff in the hospital, also for the ambulatory recordings.

4.3.2 Data preprocessing

After each nightly recording, data from the accelerometer device and from the PSG were downloaded. As the accelerometer and the PSG signals were recorded with two different devices, the clocks of the two signals needed to be synchronized before sleep stages could be compared. To synchronize, time series with IBIs were derived from the PSG and from the accelerometer signal. For the PSG, R-peaks were localized in the ECG signal, by means of a non-linear transformation in combination with first-order Gaussian differentiation[89]. IBIs were derived from the R-peaks by computing the time differences. For the accelerometer signal, the recently developed method for IBI estimation was used[111]. Clock synchronization was finally achieved by using the PSG clock as reference, and by adjusting the rate and offset of the clock of the acceleration signal such that the cross correlation between the two IBI time series was maximized. To this end, we recomputed the IBI sequence from the accelerometer signal$(t_i,\ IBI_i)$, by retiming it with offset o and rate r, such that the adjusted IBI sequence became: $(o + r\, t_i,\ IBI_i)$, with $i = 1 \ldots N$, t_i the localization time of beat i, and IBI_i the time difference between beat $i+1$ and beat i.

We then plotted the IBI sequence derived from the ECG signal of the PSG, together with the adjusted IBI sequence derived from the accelerometer, using the times given by the PSG clock on the horizontal axis and the two IBI time series on the vertical axis. In an iterative way, we manually updated o and r until visually an optimal alignment between the time series was achieved, corresponding to maximum cross correlation. As both the PSG system and the accelerometer device used crystal clocks, the rate r was always very close to 1.0.

For the scoring of reference sleep stages, we employed Somnolyzer, an automatic sleep staging method that takes PSG signals as input and combines deep Convolutional Neural Networks with Long-Short Term Memory units to predict sleep stages for each 30-second epoch[112,113]. Somnolyzer was validated in multiple studies, against up to 12 different scorers from up to 6 different centres and reached a Cohen's kappa coefficient of agreement (kappa) versus consensus scoring of 0.79, i.e. levels of agreement comparable to human interrater reliability[114].

4.3.3 Sleep stage estimation

Sleep stages were estimated using the 3D accelerometer as outlined in the block diagram in Figure 1. In a first stage, respiratory effort, heartbeats, and activity counts were derived, which were subsequently processed and offered as inputs to the sleep stage classifier. Examples of the waveforms during the various stages of processing are shown in Figure 2.

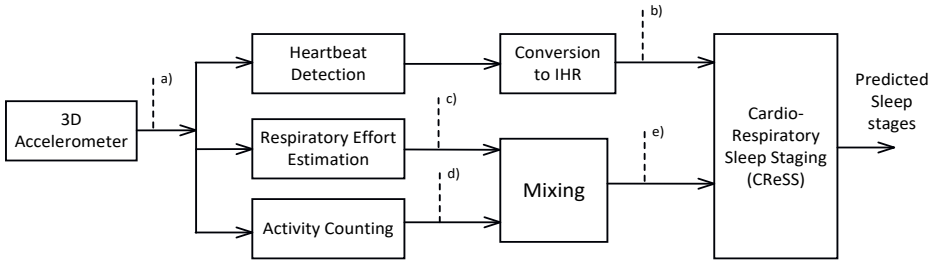


Figure 1 Block diagram of the sleep staging algorithm. Blocks depict algorithmic steps, while arrows carry the signals. The dashed lines with labels refer to the signals in Figure 2. Heartbeats, respiratory effort, and activity counts are derived from the 3D accelerometer signal (a). The instantaneous heart rate (IHR, b) is derived from the heartbeats. The respiratory effort (c) and activity counts (d) are mixed to obtain a signal (e), which combines respiratory effort with body motion. Finally, sleep stages are derived by means of cardio-respiratory sleep staging.

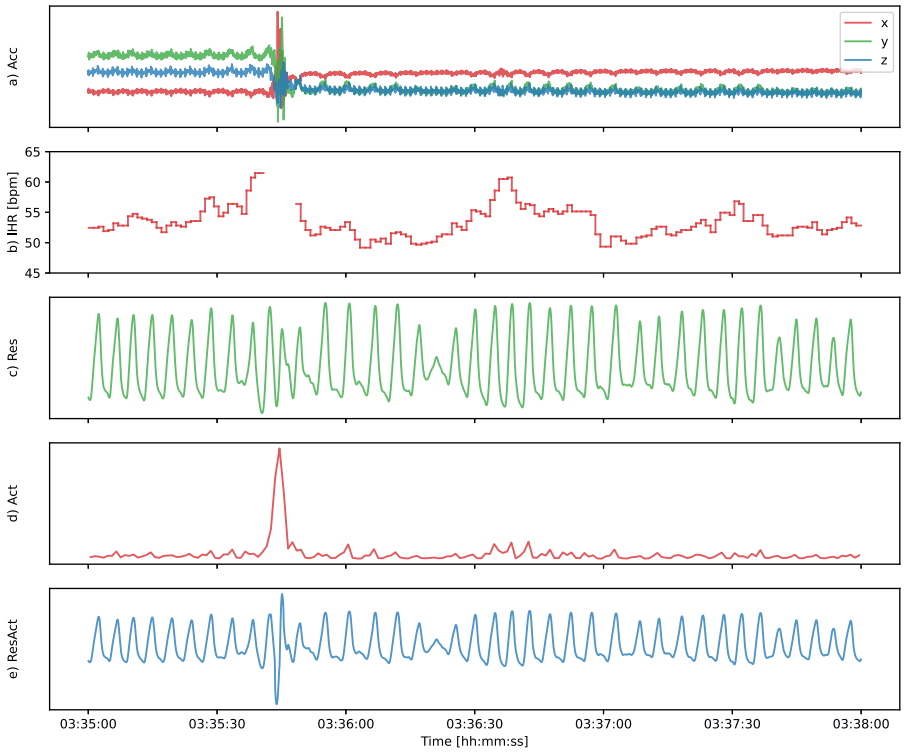


Figure 2 Waveforms when estimating sleep stages from the accelerometer. a) The acceleration components x, y and z (with arbitrary offsets for better visibility), with respiration as slow variations, heartbeats as small peaks, and body movement at t= 03:35:40. b) The instantaneous heart rate (IHR) in beats per minute, with a gap at the time of motion. c) The estimated respiratory effort (continuing during motion). d) The activity counts, with a high amplitude at the time of motion. e) The respiratory signal, with body movements mixed in.

For heartbeat detection, we used a recently published method[111], which first reduced the dimension of the acceleration signal from 3 to 1 by performing a simple arithmetic addition of the channels. Next, using maximum a posteriori (MAP) estimation, it localized heartbeats and estimated IBIs, such that heartbeat localization times coincided with points of maximum acceleration (the small peaks in Figure 2a) and the IBIs matched local periodicity (the time difference between the small peaks). Instantaneous heart rate (IHR) was then computed based on the IBI to the next beat $IHR_i = 60/IBI_i$. Subsequently, the IHR signal was upsampled, with sample-and-hold, to 10 Hz. As a result, we got a signal that was constant in between heartbeat localization times and that potentially changed at heartbeats, see Figure 2b. Using the signal quality self-assessment, which is part of the algorithm[111], estimations during periods of low signal quality were removed, as was the case at the motion artifact in Figure 1a, which resulted in a gap in the IHR in Figure 1b, at t = 03:35:45.

To obtain the respiratory effort signal, we used a previously published method[95] which estimated respiration using the slower variations in the acceleration signal (Figure 2a). It first filtered and decimated the 3D acceleration components x, y and z to 10 Hz and then transformed these into an orientation signal, comprising two horizontal components (h_1, h_2), and a vertical component (v). The components (h_1, h_2) may be interpreted as the coordinates of the bubble in a 2D (planar) spirit level and represent the orientation of the sensor, which follows a pattern that represents the respiratory cycle. A convolutional neural network, which was trained on a separate data set (that was not used in the present study), estimated a respiratory signal that resembles the signal that, in the absence of body movements, resembles the signal that is otherwise acquired by a thoracic band (Figure 2c).

In respiratory effort signals, measured with recommended sensors such as thoracic belts, body movement artifacts are inevitably present. CReSS was trained on these signals and therefore relies on such artifacts to be present as well. The respiratory effort neural network is very accurate in the estimation of respiration in the absence of movement artifacts, and during periods of movement it tends to produce a relatively clean signal, unlike traditional sensors. An example is the body movement artifact at t=03:35:45 in Figure 2a, where the respiratory effort signal (in Figure 2c) continues without much disturbance. To avoid having to retrain CReSS, we aimed at mimicking body movement artifacts, such that the respiratory effort signal better resembles traditional sensors. Towards this end we estimate gross body movements, quantified with so-called "activity counts". The body movements were estimated each second, using the acceleration signal over the last second, by removing the mean for each of the components (x, y and z), by taking the absolute value, and by summing over all samples and components: $Act = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \overline{x}\right| + \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \overline{y}\right| + \frac{1}{N}\sum_{i=1}^{N}\left|z_i - \overline{z}\right|$, with Act the activity counts, and N the number of samples per second. The resulting signal (Figure 2d) was scaled and upsampled to 10 Hz, after which it was mixed with the respiratory signal to obtain an actigraphy-modified respiratory effort signal $ResAct$. Mixing was performed by multiplying the respiratory effort signal Res by the activity counts Act, after offsetting the latter by 1.0 to avoid attenuation: $ResAct = (1.0 + Act) * Res$, with Res, Act, and $ResAct$, as in Figure 2 c, d and e.

Sleep stage estimation was finally achieved with CReSS, presented in earlier work [63]. CReSS was designed to be device agnostic and work with respiratory signals (airflow, respiratory effort, or a combination of the two) and instantaneous heart rate inputs. CReSS uses different branches of deep convolutional networks to extract high-level features from both cardiac and respiratory signals. The feature vectors from the individual branches are merged to create a 192-element high-level cardiorespiratory feature vector for each 30-second epoch. Finally, 3 layers of bidirectional Long-Short Term Memory units

introduce global context from the entire recording and a fully connected dense layer assigns probabilities for the stages Wake, REM, Light Sleep (N1+N2), and Deep Sleep (N3), from which the stage with the highest probability is then selected.

4.3.4 Performance evaluation

Sleep stage estimation was evaluated in terms of epoch-per-epoch agreement with the reference sleep stages, using accuracy and Cohen's kappa coefficient of agreement. We computed agreement for classification of four classes (Wake/N1+N2/N3/REM), three classes (Wake/NREM/REM; obtained by merging N1+N2 and N3 in a single non-REM or NREM class) and for a binary classification of each of the four classes against the rest (Wake vs. combined N1+N2, N3 and REM, etc.). In addition, we computed the confusion matrix from which we derived sensitivity, specificity, and positive predictive value (PPV) for the binary classification tasks.

Furthermore, we evaluated the estimations of sleep metrics by calculating the total sleep time, sleep efficiency, sleep latency, and time in each stage from the sleep stages estimated from the accelerometer, and from the PSG.

Finally, we evaluated the impact of different factors on sleep staging performance, by assessing sleep stage agreement, using a Mann-Whitney U test across gender, presence of each sleep disorder. For age and BMI, we performed a Spearman's rank correlation on kappa. Finally, we used a Benjamini-Hochberg correction with an acceptable false discovery rate of 5%, to prevent false discoveries.

4.4 Results

PSG and accelerometer data were collected for a total of 323 participants (Table 1). Sleep disorders were diagnosed following standard clinical practice in each study, according to the criteria of the international classification of sleep disorders (ICSD-3) [5]. The most prevalent sleep disorders were sleep-disordered breathing (SDB, N=175, 54%) and insomnia (N=76, 24%).

Table 1 Study demographics. Entries are given with standard deviation (±) and ranges ([low-high]).

Dataset	N	Female	Age[1] (years)	BMI[1] (kg/m^2)
KH	195	77 (39%)	52 ± 17 [13-83]	26 ± 4 (19-37)
OLVG	128	41 (32%)	46 ± 13 [21-78]	28 ± 5 [19-47]
KH+OLVG	323	118 (37%)	50 ± 15 [13-83]	28 ± 4 [19-47]

Sleep stages were derived from PSG and from the accelerometer. Table 2 gives the performance of sleep stage classification for each task, with the inter-quartile ranges (if not normally distributed) or the standard deviations (if normally distributed). The sample statistics are calculated across the performance per participant (N=323). We found a median four-class (Wake/N1+N2/N3/REM) kappa of 0.678 and a median three-class (Wake/NREM/REM) κ of 0.745. As expected, the three-class kappa was higher because confusion between N1+N2 and N3 is removed when merging these two into the (single) NREM class. For the binary classification task Wake versus Sleep (Wake detection), we found a median κ of 0.723, a median sensitivity of 78.7%, and a median specificity of 96.6%, which is in line with the last row of the confusion matrix (Table 3), which shows that 20.9% of the Wake epochs are classified as N1+N2, 0.4% as N3, and 2.8% as REM. Confusion between REM, Wake and N3 is all < 5%, confusion > 5% is always in relation to N1+N2.

Table 2 Results for sleep stage agreement.

Task	Cohen's kappa[3]	Accuracy[3] (%)	Sensitivity[3] (%)	Specificity[3] (%)	PPV[1,3] (%)
Wake/N1+N2/N3/REM	0.678 {0.577, 0.749}	80.8 {75.6, 84.6}			
Wake/NREM/REM	0.745 {0.650, 0.808}	87.5 {83.1, 90.6}			
N1+N2[2]	0.617 {0.510, 0.699}	81.8 {77.2, 85.3}	85.4 {80.2, 90.4}	79.1 {70.0, 85.1}	82.6 {76.8, 88.7}
N3[2]	0.645 {0.363, 0.769}	93.6 {91.1, 95.7}	66.3 ± 27.4	97.6 {94.5, 99.4}	65.3 ± 32.3
REM[2]	0.779 {0.641, 0.861}	94.8 {92.1, 96.7}	76.8 ± 21.1	97.9 {95.7, 99.1}	78.0 ± 23.9
Wake[2] (vs. Sleep)	0.723 {0.609, 0.804}	93.3 {89.7, 95.4}	78.7 {66.2, 88.3}	96.6 {93.9, 98.1}	81.1 {68.4, 90.0}

[1] PPV = Positive Predictive Value. [2] Binary classification tasks were evaluated in a one vs. rest strategy, where one single class (e.g., Wake, REM, etc.) was considered the "positive" class, and the remaining were aggregated in a single "negative" class. [3] Entries are given as mean with standard deviation (±) if normally distributed else as median with first quartile and third quartile ({Q1, Q3}).

Table 3 Confusion matrix for four-class sleep staging.

Prediction → Reference ↓	N1+N2	N3	REM	Wake
N1+N2	144578 (84.2%)	10334 (6.0%)	6558 (3.8%)	10150 (5.9%)
N3	10943 (33.6%)	21238 (65.2%)	271 (0.8%)	126 (0.4%)
REM	9744 (20.7%)	45 (0.1%)	36395 (77.1%)	1001 (2.1%)
Wake	14009 (20.9%)	293 (0.4%)	1865 (2.8%)	50898 (75.9%)

Confusion is over all epochs and all participants (N=318,448). Percentages are the number of epochs of the predicted class relative to those of the reference class.

Note that Table 3 gives numbers over all epochs of all participants, while Table 2 gives the means and quartiles over performance statistics computed for each participant individually. Therefore, the sensitivities, specificities and PPV values in Table 2 are not the exactly equal to the ones that may be derived from Table 3.

Figure 3 illustrates three examples of sleep stage classifications, chosen as the recordings with the four-class κ closest to the first quartile (κ = 0.579), to the median (κ = 0.678), and to the third quartile (κ = 0.748). Furthermore, and beyond the κ metric, intervals of occurrence often match; if we define a REM interval as the longest possible interval in which REM occurs, without it being interrupted by more than M epochs, then we find four REM intervals in the reference in Figure 3a (M=20). We also find four REM intervals in the estimation in Figure 3a, and we find that these intervals overlap with the reference. Similarly, we find four N3 intervals in the reference of Figure 3c, with four overlapping N3 intervals in the estimation. However, occurrences do not always match, as may be found in Figure 3b, where four REM intervals match, but where the first REM interval in the reference (t ≈ 22:30) has no matching REM interval in the estimation.

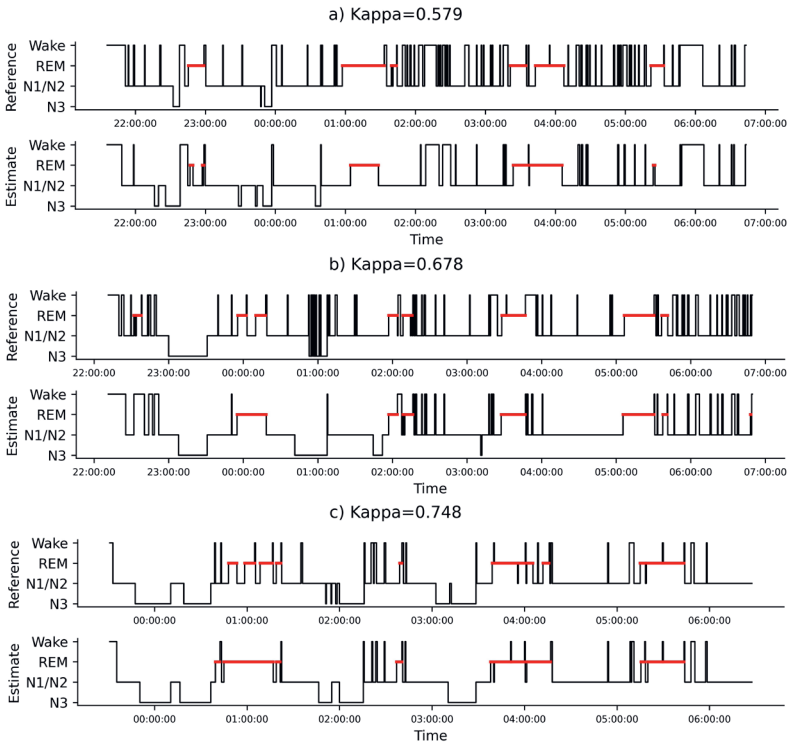


Figure 3 Reference and estimated hypnograms for three recordings. a) near the first quantile (κ = 0.579), b) near the median (κ = 0.678), and c) near the third quantile (κ = 0.748).

Table 4 gives the performance of the estimation of sleep metrics, with the interquartile ranges for the reference value and for the estimation error in between curly brackets. Furthermore, the 95% Limits of agreement are given in the rightmost column. The total sleep time was overestimated with a median error of 2.5 minutes, with first and third quartiles at -9.5 and +17.0 minutes, and 95% limits of agreement at -51.0 and +107.7 minutes. The third quartile and upper limit of agreement having higher magnitudes than the first quartile and lower limit of agreement, is in line with the confusion matrix in Table 3. Relatively many wake epochs were scored as sleep (20.9% N1+N2, 0.4% N3, 2.8% REM, bottom row), while relatively few sleep epochs were scored as wake (5.9% N1+N2, 0.4% N3, 2.1% REM, rightmost column). For the sleep efficiency, we found a median error of 0.5%, with first and third quartiles at -2.0 and +3.0%, and 95% limits of agreement at -10.9 and +21.1%.

Regarding the influence of factors on the epoch-by-epoch agreement (κ), and after application of the Benjamini-Hochberg correction for false discoveries (FDR < 0.05), we found a significant influence of age. Spearman's rank correlation was -0.195, hence yielding better performance for younger participants. We did not find significant influences of the presence of SDB and insomnia, which were the two most prevalent disorders in our study (54.2% and 23.5%).

Table 4 Sleep statistics.

Statistic	PSG[3]	Error[1,3]	95% LoA[2]
Total sleep time (min)	412 {338, 454}	2.5 {-9.5, 17.0}	-51.0, 107.7
Sleep latency (min)	16 {6, 37}	1.5 {-2.5, 5.5}	-42.8, 27.9
Wake after sleep onset (min)	60 {32, 106}	-3.0 {-16.8, 8.0}	-101.5, 45.0
Stage R latency (min)	120 {86, 170}	0.5 {-1.5, 3.0}	-140.8, 131.2
Sleep efficiency (%)	83 {72, 90}	0.5 {-2.0, 3.5}	-10.9, 21.1
Time in Wake (min)	84 {48, 138}	-2.5 {-17.0, 9.5}	-107.8, 50.9
Time in REM (min)	76 {48, 97}	-3.5 {-15.5, 7.5}	-53.0, 41.3
Time in N1 + N2 (min)	274 {229, 308}	10.0 {-16.8, 36.0}	-72.5, 114.0
Time in N3 (min)	50 {25, 71}	-1.5 {-19.5, 16.8}	-64.5, 61.9
Percentage of TST in Wake (%)	20 {11, 40}	-0.6 {-5.9, 3.2}	-55.7, 21.9
Percentage of TST in REM (%)	19 {15, 24}	-1.2 {-4.2, 1.9}	-13.3, 10.0
Percentage of TST in N1 + N2 (%)	67 {61, 74}	2.1 {-4.3, 7.3}	-18.1, 23.2
Percentage of TST in N3 (%)	13 {7, 18}	-0.6 {-5.6, 4.5}	-17.1, 17.2

The error is computed as the value derived from the accelerometer minus the value derived from the scoring of PSG. [2] The 95% limits of agreement (LoA) refer to the error. [3] Each entry contains the median followed by the first and third quantile in between curly brackets.

4.5 Discussion

In this study, we evaluated the performance of sleep staging from a chest-worn accelerometer among a group of 323 participants with various sleep disorders. Making use of existing methods for cardio-respiratory sleep staging in combination with estimation methods for cardiac and respiratory characteristics, we evaluated the performance over studies in two sleep centers and compared against reference scoring from PSG. For sleep stage classification we found a kappa for four-class scoring of 0.678, and a kappa of 0.723 for wake (versus sleep) detection, both indicative of "substantial agreement" with PSG [115].

To put the performance obtained in our study in perspective, in a systematic review on methods for cardiorespiratory sleep staging, one can find kappa values ranging from 0.55 to 0.76 (median = 0.55), and accuracies ranging from 58% to 90% (median = 78%) [105]. It should be noted, however, that there are many differences between the reviewed studies and our work, such as the study populations, the number of sleep stages (the number of classes), and the way the cardiorespiratory signals were obtained (PSG versus "surrogate" signals). The algorithm that we used in the present study, CReSS, was validated earlier among a population of 296 participants with a wide range of SDB severity. Using IHR derived from ECG or finger-worn PPG, respiratory effort from thoracic belts, and airflow, it achieved a kappa of 0.643, without significant differences between different SDB severities [63]. These results are like those obtained in the present study, confirming the suitability of deriving cardiorespiratory inputs from an accelerometer instead.

All confusions > 5% occurred between N1+N2 on the one side, and Wake, REM, and N3 on the other (see Table 3). The CReSS validation study[63] provides confusion matrices for classification using airflow as input, but not specifically with respiratory effort in the absence of airflow (as is the case here). However, the study reveals that the highest confusion occurs between N1+N2 and the remaining sleep stages Wake, REM, and N3, which aligns with our findings. A recent review of cardiorespiratory sleep staging methods [105] unfortunately does not describe possible sources of confusion. It remains a topic for further research whether this is inherent to the intrinsic limitations of cardiorespiratory sleep staging, and the information is only present in the neurological signals; or if the confusion is a specific limitation of the chosen algorithm (CReSS).

Relatively many Wake epochs (20.9%, Table 3) were scored as light sleep (N1+N2). This elevated the upper bound on the 95% limits of agreement for Sleep Efficiency (21,1%, Table 4). Retraining the cardiorespiratory sleep staging on the accelerometer data, possibly targeting the accuracy of the total sleep time, may improve this aspect.

When developing a new sleep staging method, an important choice is how to obtain the reference ground-truth sleep stages. Most often, these are obtained by a single technician scoring PSG, according to the rules of the AASM[8]. To increase scoring reliability, the AASM has described the rules in detail and regularly issues updates. Despite these efforts, human interrater reliability remains imperfect, with kappa values reaching about 0.76[116], and with the percentage of epochs where all scorers fully agree progressively declining with an increasing number of scorers, down to 25% with 15 scorers or more[113]. This effect can be worse if the data are collected over different centers with staff that have different levels of experience.

One way to counteract this effect is to obtain a ground truth by having the data scored by multiple technicians, from accredited sleep centers, all trained to the same version of the rules. The consensus of these scorers can then be considered a "gold-standard" reference. However, the associated cost and organizational effort make this approach often not feasible, especially for larger datasets. In the present study, we addressed this issue by employing an automatic scoring method trained on the consensus scorings of multiple experts and widely validated in different external cohorts. Somnolyzer is such a method, representing the consensus opinion of multiple expert scorers and thereby reducing the epistemic uncertainty associated with inter-institutional, and scorer-dependent variability [113]. Somnolyzer has been cleared by the FDA as a tool for clinical and pharmacological studies and has been certified by the AASM for sleep staging through their Autoscoring Certification Program [117], a process during which a panel of 9 sleep technicians and 2 physicians assess performance by conducting a detailed review of 100 PSG recordings.

A possible limitation of the approach is that it may overestimate performance of the accelerometer-based sleep staging method, as there is partial overlap between the datasets on which Somnolyzer and CReSS have been trained. However, we believe this bias is limited, as shown by the hold-out validation of both algorithms in completely independent, external validation datasets[114].

Although we use an accelerometer, our method differs from wrist actigraphy. Wrist actigraphy solely relies on the measurement of body movements, which are used for the estimation of 'rough' parameters like the sleep-wake rhythm or the total time in bed. Wrist actigraphy detects wake with high sensitivity, but with low specificity, as wakefulness without motion is easily confused with sleep. A study into the validation of actigraphy showed a sensitivity of 93% to 99% and a specificity of 37% to 62% in 25 healthy young adults[118]. Our method also uses body movements but additionally uses small- amplitude accelerations to estimate cardio-respiratory activity. For comparison, our method detected sleep with a sensitivity of 96.6% and a specificity of 78.7% among a large SDB population with a wide range of age. For the related sensing modality of ballistocardi-

ography (where cardiac activity, and possibly respiration, are measured through the structure on which the body resides), several studies were published. For a small numbers of healthy participants, by using bed sensors, three-class kappa values were reported of 0.44 (18 participants)[119], 0.55 (17 participants)[120], 0.63 (17 participants)[121], and 0.74 (5 participants)[122]. The present method reached a three-class kappa of 0.745 among a diverse and disordered group of 323 participants.

We found that sleep staging performance was slightly better in younger individuals (Spearman's rank correlation coefficient = -0.19). This might be explained by younger participants having more pronounced heart rate variability, as is known from literature[123]. Notably, we found no significant influences of the presence of SDB and insomnia, which were the most prevalent sleep disorders in our study. The consistency of performance due to the presence of SDB is in line with an evaluation of CReSS among a cohort with a large range of SDB severities[63].

There are multiple advantages of performing sleep staging using a chest-worn accelerometer. The technology can be integrated into devices that are worn anyway on the chest during the night. Examples are devices for positional therapy for OSA (also known as sleep position trainers), and patches for extended Holter monitoring. Unlike most other technologies for sleep staging our method only needs a single point of mechanical contact with the human body, without the need for direct skin contact, which avoids the need for skin preparation (like shaving), avoids skin irritation, and makes the technology robust against contact deterioration due to for example the drying out of electrodes.

A limitation of this study was that participants with a known history of heart rhythm disorders were excluded. As atrial fibrillation is often comorbid with obstructive sleep apnea[124], the use of our method among populations with suspected obstructive sleep apnea may be problematic. Correlations between sleep stages and heart rate variability may be masked by the rhythm disorder. Further research may reveal whether this is the case. Another limitation was that the accelerometer was mounted by clinical staff, at a defined location on the thoracic band. The influence of self-attachment, as well as the influence of different sensor locations on the body as a result, could be investigated in a follow-up study. Finally, CReSS, which was trained on cardio-respiratory signals derived from ECG and respiratory impedance plethysmography, was taken as-is. As the characteristics of cardiorespiratory signals derived from chest-wall accelerometry are different, as was shown in our earlier work on the estimation of respiratory effort from an accelerometer[95], further research may be conducted in adapting the sleep staging to these signals, thereby possibly improving performance further.

5

Apnea-Hypopnea Index Estimation using Chest-Wall Accelerometry

This chapter corresponds to the manuscript:

F. Schipper, P. Fonseca, A. Grassi, M. Ross, F. van Meulen, M. v. Gilst, P. Bosschieter, E. Schoustra, R.J.G. van Sloun, F. Abdenbi, N. de Vries, R. Heinzer, J.L. Pépin, and S. Overeem, Apnea-Hypopnea Index estimation using overnight chest-wall accelerometry.

CHAPTER 5

5.1 Abstract

Rationale and Study Objectives - To develop a method for apnea-hypopnea index (AHI) estimation using a chest-worn accelerometer, as an approach to obstructive sleep apnea diagnosis and long-term monitoring of treatment, for example through positional therapy.

Methods - We developed a method for AHI estimation by combining a cardiorespiratory sleep staging algorithm with an adapted neural network for detecting respiratory events. Originally based on electrocardiography and respiratory impedance plethysmography, the network was retrained using chest-wall accelerometry-based instantaneous heart rate and respiratory effort. Training and validation were conducted using accelerometer data from 413 participants who underwent concurrent overnight polysomnography (PSG).

Results - Sleep staging reached substantial agreement with polysomnography, achieving a Cohen's kappa coefficient of agreement of 0.67 for four-class sleep staging (Wake/REM/N1-N2/N3). AHI was estimated with high reliability, showing an intraclass correlation coefficient of 0.89 (95% CI: 0.85 to 0.91) when compared to polysomnography-derived values. Performances were consistent in both supine and non-supine sleeping positions. Positive likelihood ratios were high (7.5, 12.8, and 153.4 for the mild, moderate, and severe OSA severity classes, using near-boundary double labelling). Negative likelihood ratios were low (0.076, 0.244, and 0.139 for the same cases).

Conclusion - AHI can be reliably estimated using chest-wall accelerometry. This approach may be used in diagnostic tests for OSA but also to assess residual AHI during positional therapy.

5.2 Introduction

Untreated obstructive sleep apnea (OSA) and its associated poor sleep quality can adversely impact short-term well-being and long-term health[125,126]. The management of OSA can benefit from continuous monitoring during sleep, focusing on the apnea-hypopnea index (AHI) which can be used as an indicator of disease severity and as a measurement of therapy efficacy.[5,127]

A commonly used OSA therapy is positive airway pressure (PAP). Modern PAP devices track the AHI on a nightly basis by measuring apnea and hypopnea events derived from airflow[128]. Given the challenges with PAP therapy compliance, alternative therapies such as positional therapy (PT) have been developed. PT treats positional obstructive sleep apnea (pOSA), a condition where patients have a significantly higher OSA severity in the supine position, and much lower severity in other sleeping positions[129]. High prevalence rates of pOSA have been reported, with over 50% of patients referred for OSA evaluation presenting this phenotype[130]. PT devices continuously monitor body position and provide feedback (usually with vibrations) when in supine position, triggering body position changes[131]. Although having demonstrated compliance and efficacy[23], PT devices do not measure airflow, and consequently cannot monitor AHI and provide a comprehensive assessment of therapy efficacy.

PT devices are typically mounted on the chest wall and already incorporate an accelerometer to monitor body position. Historically, accelerometers were used in sleep medicine to measure active-rest periods, and to aid in the diagnosis of, for example, circadian rhythm disorders[132]. However, the capability of these devices was limited to the measurement of gross body movements, and they were unable to measure sleep stages or to accurately detect motionless periods of wakefulness[118]. Crucially, accelerometers can also be used to measure respiration and cardiac activity. Morillo et al.[45] used an accelerometer mounted on the suprasternal notch to measure cardiac, respiratory and snoring signals. Hung and colleagues[24] assessed respiration with an accelerometer mounted on the chest wall. More recently, Ryser et al.[133] estimated respiratory rate and cessations of breathing in signals obtained from a chest-worn accelerometer.

In earlier work, we demonstrated that a chest-worn accelerometer could yield accurate measures of thoracic respiratory effort (RE) and instantaneous heart rate (IHR) throughout the night[111].

Using an algorithm originally developed with RE measured by a respiratory inductance plethysmography (RIP) belt and IHR measured by photoplethysmography or electrocardiography (ECG)[63], we showed that four-class sleep staging (Wake, REM, N1-N2, and N3)

could be accurately estimated using RE and IHR derived solely from chest-wall accelerometry[134]. Validation on a diverse group of participants with a variety of sleep disorders, including OSA, yielded a Cohen's kappa coefficient of agreement of 0.68 in comparison with reference sleep stages scored from PSG. Furthermore, we recently demonstrated that AHI could be estimated by means of a neural network that utilised RE from a RIP belt and IHR from ECG, achieving an intraclass correlation coefficient (ICC) of 0.91 when compared to reference AHI scored from PSG[36].

In the present study, we adapt this established neural network model to estimate AHI entirely from overnight chest-wall accelerometry recordings, utilizing transfer learning, and evaluate its performance against gold-standard PSG. A key aspect of our validation is to demonstrate the method's suitability for positional therapy (PT) by specifically evaluating its accuracy during non-supine body positions, which reflect effective therapy usage.

5.3 Methods

We adapted a previously developed neural network designed for respiratory event detection, which utilised ECG and Respiratory Inductance Plethysmography (RIP) belt data, to accept inputs derived entirely from chest-wall accelerometry. The adaptation was performed using transfer learning due to the limited availability of simultaneous recordings featuring both Polysomnography (PSG) and chest-wall accelerometry. Because the Apnea-Hypopnea Index (AHI) calculation requires Total Sleep Time (TST), we first evaluate the performance of the previously established accelerometer-based sleep staging method on the present dataset. Finally, we assessed the method's suitability for direct implementation in Positional Therapy (PT) devices by evaluating its performance separately for supine and non-supine body positions.

5.3.1 Dataset

The dataset comprised 413 subjects referred for overnight PSG as part of their standard clinical procedure. Data were collected from two distinct clinical sites. In the first site, the OLVG hospital (Amsterdam, the Netherlands) data were collected between June 3, 2021, and December 12, 2022, among 124 adults scheduled for ambulatory PSG. The study was reviewed and approved by the local scientific advice committee (Adviescommissie Wetenschappelijk Onderzoek, ACWO) of the OLVG hospital (File no: WO 20.134). In the second site (Sleep Medicine Center Kempenhaeghe Heeze, the Netherlands), data were collected between January 27, 2020, and November 11, 2022, and comprised 281 adults referred for overnight in-lab diagnostic studies, as part of the SOMNIA study[68].The SOMNIA study was reviewed by the medical ethical committee of the Maxima Medical Center (Veldhoven, the Netherlands; file no: N16.074 and W17.128).

In addition to the standard PSG setup at both sites, participants wore a single 3D accelerometer (ADXL355, Analog Devices, Wilmington MA, USA) mounted on the thoracic Respiratory Inductance Plethysmography (RIP) belt. This sensor was configured to acquire triaxial raw accelerometer data at 250 Hz.

All participants gave informed consent before participating. The studies met the ethical principles of the Declaration of Helsinki, the guidelines of Good Clinical Practice and the current legal requirements. The protocols for the data analysis described in the present study were approved by the local scientific advice committee of the OLVG hospital, by the review boards of the Medical Ethical Committee of the Kempenhaeghe hospital and by the Internal Committee of Biomedical Experiments of Philips Research.

To ensure standardized scoring of sleep stages and sleep-disordered breathing (SDB) events (specifically, apneas and hypopneas) across the recordings of both sites, the data were automatically scored using Somnolyzer 4.1 (Philips, Monroeville, PA, USA)[114]. Somnolyzer has been validated against human scoring following the American Academy of Sleep Medicine (AASM) rules for scoring of sleep stages, SDB events and cortical arousals and was certified by the AASM through their Autoscoring Certification Program[117]. Somnolyzer was configured to score events using the current recommended rules[8]: apneas were defined as a (near-)complete cessation of airflow (at least 90% reduction from baseline), for at least 10 seconds; hypopneas were scored based on a reduction of airflow between 30% and 90% from baseline, lasting at least 10 seconds, and associated with an oxygen desaturation of at least 3% or a cortical arousal.

5.3.2 Data preparation

Precise synchronization between the Polysomnography (PSG) and accelerometer recordings was essential for comparing the derived sleep stages and events, but the internal clocks of the independent acquisition devices were not aligned. Therefore, an intrinsic synchronization method was required.

The accelerometer data were first processed to obtain the precise interbeat intervals (IBIs) and temporal locations for each heartbeat. IBIs were detected by exploiting their local periodicity and were combined with peaks in the acceleration signal to accurately localize the heartbeats in time[111]. IBIs and temporal locations were derived from the PSG as well. R-peaks were localized in the ECG signal by means of a non-linear transformation in combination with first-order Gaussian differentiation[89]. IBIs were subsequently derived by computing the time differences between the localization times of the R-peaks. Synchronization was finally achieved by using the PSG clock as reference, and by adjusting the rate and offset of the clock of the acceleration signal such that the cross-correlation between the two IBI time series was maximized.

5.3.3 Data processing

All steps required to estimate sleep stages and AHI from the accelerometer signal are illustrated in Figure 1 and further described in detail in the following sections. The labels (a) to (f) refer to Figure 1 (block diagram) and Figure 2 (example waveforms).

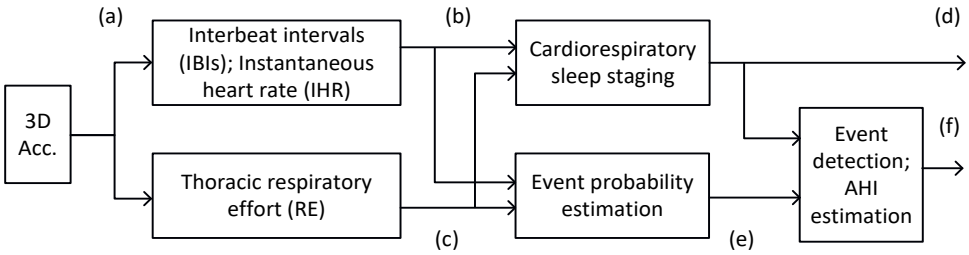


Figure 1 Data processing block diagram. The labels (a) to (f) refer to the example waveforms in Figure 2. Using the 3D acceleration (a), the IHR is estimated from IBIs (b). Estimated RE (c) is used for cardiorespiratory sleep staging (d) and estimation of probabilities of apnea and hypopnea events (e). Finally, using sleep stage-dependent thresholds, events are detected, counted, and divided by the total sleep time to obtain an estimate for the AHI (f).

5.3.4 Cardiorespiratory measurements

The methods to measure RE and IHR from chest-wall accelerometry have been described in detail in previous publications[95,111]; a summary is provided here. The RE signal (c) was estimated using a pre-trained convolutional neural network. This network modelled the complex relationship between the triaxial accelerometer signal (representing chest-wall movements) and the reference RIP signal used for training. The output of the network was a 10 Hz time series intended to represent the RE signal, mimicking the morphology of a reference RIP signal. The IHR signal (b) was derived from the accelerometer data exploiting the principles of Seismocardiography. IBIs were estimated by exploiting the local periodicity in the accelerometry signal and heartbeats were localized in time. The sequence of localization times and IBIs was converted into IHR by taking the reciprocal of the IBIs. Finally, the IHR time series was uniformly sampled at 10 Hz using a zero-order hold interpolation.

5.3.5 Cardiorespiratory sleep staging

The estimation of AHI requires an estimation of TST, which in turn requires a classification of the epochs during which the participant is awake or asleep. Furthermore, as the salience of respiratory events in terms of their cardiorespiratory expression depended on the sleep stage during which they occurred, the detection of respiratory events also required sleep stages (see Figure 1).

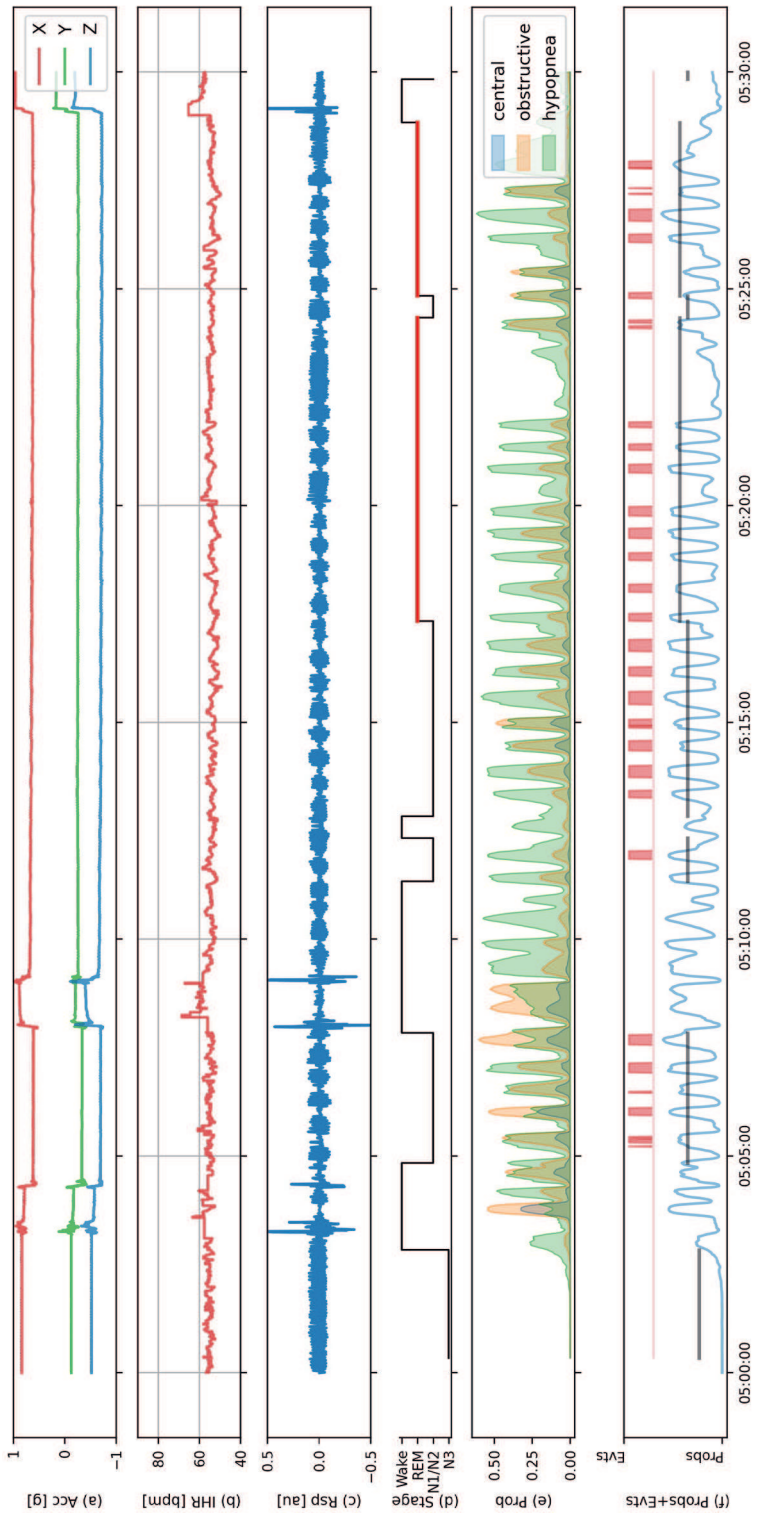


Figure 2 Waveforms. The labels (a) to (f) refer to outputs of processing blocks in Figure 1: (a) 3D acceleration from the accelerometer, (b) instantaneous heart rate, (c) respiratory effort, (d) sleep stages estimated from the instantaneous heart rate and respiratory effort, (e) SDB event probabilities estimated from instantaneous heart rate and respiratory effort, (f) combined SDB event probabilities with sleep stage-dependent thresholds; when the combined probability exceeds the threshold (bottom subgraph), an event is detected (top subgraph). During wakefulness, no events are detected.

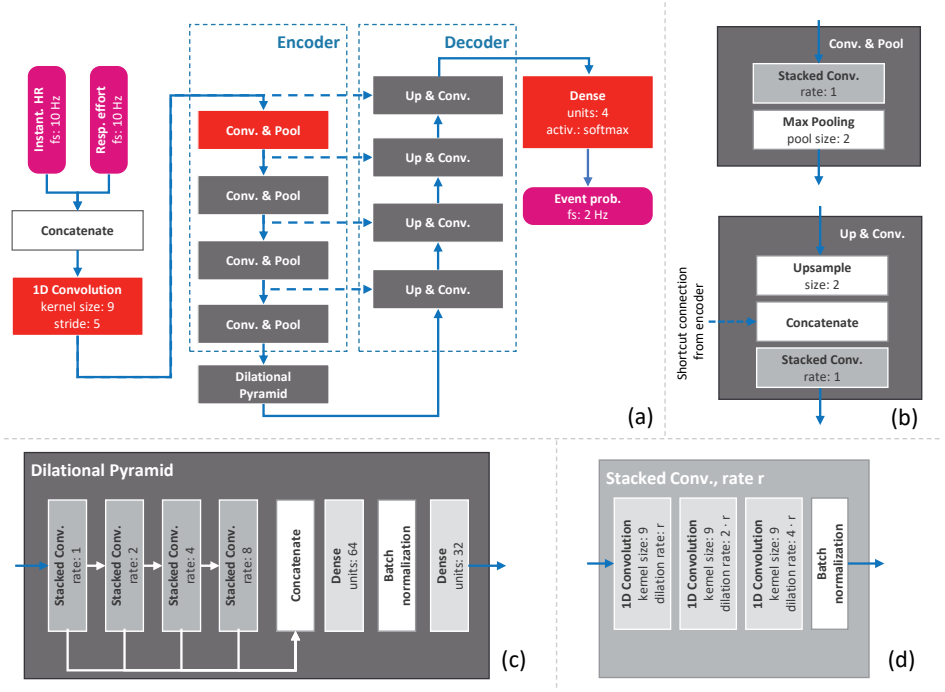


Figure 3 – (a) Architecture of the neural network, including the encoder-decoder, (b) encoder and decoder layers, (c) dilational pyramid, and (d) each stacked convolutional block. The blocks drawn in red in (a) were retrained using accelerometer data from the training part of the present study, while the blocks in grey were kept as pre-trained on ECG and RIP data from a previous study.

We used a previously developed algorithm form cardiorespiratory sleep staging (CReSS)[135]. The CReSS algorithm was originally developed to perform four-class sleep staging (Wake, N1-N2, N3, REM) using cardiorespiratory signals. We previously demonstrated that CReSS could be successfully used to perform the same four-class staging using only RE and IHR derived from chest-wall accelerometry[134]. Using that algorithm, TST was estimated simply by counting the time spent in any of the detected sleep stages (except Wake).

5.3.6 Event detection and AHI estimation

Apnea and hypopnea events (henceforth referred to simply as SDB events) were detected using a neural network previously developed to detect these from IHR derived from ECG and RE derived from a RIP band[36]. The original neural network was trained with over 2000 recordings and further validated on more than 600 additional recordings. Such amounts of data are not available for PSG recordings with concurrent chest-wall accelerometry; hence, we leverage transfer learning, whereby a neural network trained on a larger, more

readily available dataset, is fine-tuned using a smaller target-specific dataset. This technique has been applied successfully in the domains of ECG[136], PPG[137], and sleep[138].

The SDB event detection network uses an encoder-decoder architecture with a pyramid pooling block (Figure 3)[36]. The encoder represents cardiorespiratory information in feature maps with a progressively lower temporal resolution, while the decoder module upscales the abstract, low-resolution maps to match the desired temporal output resolution for event detection (2 Hz in this case). A dilational pyramid pooling module combines information at different scales to improve the precision in detecting event boundaries and to help separate events that are temporally close to each other. Inspired by pyramid scene parsing architectures, it exploits not only local but also neighbouring clues to improve the reliability of the final prediction.

The network outputs SDB event probabilities, (label (e) in Figures 1 and 2). Based on the observation that the salience of the expression of SDB events in cardiorespiratory activity depends on the sleep stage in which they occur, we opted to use sleep-stage dependent probability thresholds, optimized by grid search maximizing the ICC between the predicted and reference AHI values. Using these thresholds, individual events are detected as the group of consecutive output samples where the probability is higher than the threshold defined for the accelerometer-based sleep stage in which they occur (label (f) in Figures 1 and 2). These events are subsequently counted and divided by the TST to obtain an estimate of AHI based on accelerometer data (AHI_{ACC}).

5.3.7 Training and validation

Sleep staging

The CReSS sleep staging algorithm was used as is, without retraining or in any way tuning to the characteristics of the dataset used in the present study. To evaluate sleep staging performance, we first selected for each recording the interval for which both accelerometer and PSG data were available. Next, TST obtained from chest-wall accelerometry (TST_{ACC}) was compared with TST obtained from PSG scoring (TST_{PSG}). In addition, for each recording, epoch-per-epoch CReSS sleep stages were compared with the reference sleep stages scored from PSG. Performance was assessed using Cohen's κ of agreement (κ for short) and accuracy for four-class classification (Wake/N1-N2/N3/REM), three-class classification Wake/NREM/REM; obtained by merging N1-N2 and N3 in a single non-REM or NREM class) and for the binary classification of each of the four classes against the rest (Wake vs combined N1-N2, N3 and REM, etc.). For each of the binary classification tasks, sensitivity, specificity, and positive predictive value (PPV) were computed.

Event detection and AHI Estimation

Transfer learning was performed by freezing all weights of the pre-trained network from the previous study, except for the blocks at the input and at the output of the neural network, highlighted in red in Figure 3. To perform transfer learning, we divided the dataset (N=413) into two parts at subject-level, balancing the number of recordings of each OSA severity category between the two. The first part, the training set (N=206), was used for transfer learning. The second part, the hold-out set (N=207), was reserved for evaluating the performance of the adapted AHI estimation algorithm against gold-standard PSG and was completely unseen by any part of the complete system prior to evaluation.

The performance of the AHI estimation algorithm based on accelerometer inputs (AHIACC) was compared against the reference AHI determined from the PSG scorings (AHIPSG) using the ICC (two-way random-effects model for absolute agreement) with the corresponding 95% confidence intervals, and a Bland-Altman analysis of bias and 95% limits of agreement (LoA). Furthermore, we calculated the root mean square error and the mean error (bias) and standard deviation of the error for the complete dataset, as well as for the recordings from each of the two sleep centres.

The performance in establishing OSA severity was assessed with a four-way confusion matrix from which measures of agreement for the common severity classes 'None' (AHI < 5), 'Mild' (5 ≤ AHI < 15), 'Moderate' (15 ≤ AHI < 30), and 'Severe' (AHI ≥ 30) were derived. Furthermore, the algorithm was evaluated in terms of its diagnostic capacity at distinguishing OSA severity at different thresholds, i.e., AHI < x versus AHI ≥ x, with x corresponding to the severity thresholds 5, 15 and 30. Classification performance for each of these binary classification tasks was evaluated using measures of accuracy, sensitivity, specificity, positive and negative predictive values (PPV, NPV), and positive and negative likelihood ratios (LR+, LR-). To account for uncertainty around the canonical thresholds (5, 15, and 30 events/h), and to avoid giving an overly pessimistic view of performance due to small errors around these boundaries, which might not be clinically relevant, we used the near-boundary double-labelling procedure as proposed by Van Pee et al.[139] and Martinot et al.[140] Using the boundaries defined by Van Pee and colleagues, we considered the following boundaries: 'Normal' (AHIPSG < 2.4), 'Normal OR Mild' (2.4 ≤ AHIPSG < 7.0), 'Mild' (7.0 ≤ AHIPSG < 12.4), 'Mild OR Moderate' (12.4 ≤ AHIPSG < 17.4), 'Moderate' (17.4 ≤ AHIPSG < 26.6), 'Moderate OR Severe' (26.6 ≤ AHIPSG < 35.2), and 'Severe' (AHIPSG ≥ 26.6), in which the zones with 'OR' in the label are near-boundary zones. All reference AHI values were classified into one of these zones. The final confusion matrix was obtained after reducing the near-boundary double-labelled references to a single class out of the standard four severity classes ('None', 'Mild', 'Moderate', 'Severe'), using the following rules: if the predicted severity matched one of the double-labelled references, the reference was assigned to the row corresponding to the predicted class; otherwise, the reference

was assigned to the severity class according to the canonical AHI thresholds (i.e., 5, 15, 30).

5.3.8 Relation with body position

The performance of the algorithm was evaluated separately for supine and non-supine body positions. We used the body position sensor, which was part of the PSG, to determine intervals where the participant was lying in supine position and intervals where the participant was lying non-supine (i.e., any other body position except supine). An SDB event that overlapped at least partially with a supine interval during sleep was considered a supine-specific event, otherwise it was considered a non-supine event. Using this rule, we calculated the number of supine and non-supine SDB events detected by the accelerometer, as well as using the scoring from PSG. We also calculated the TST spent in supine and non-supine positions, based on CReSS using accelerometer inputs, and based on the reference scorings from PSG. For all recordings with a minimum of twenty minutes of TST (according to PSG) in both supine and in non-supine positions, the supine-AHI and non-supine-AHI were calculated for both the PSG and the accelerometer algorithms and evaluated in terms of ICC (two-way random-effects model for absolute agreement).

5.3.9 Statistics

The Shapiro-Wilk normality test was used to determine whether sample data was normally distributed (at a p-value of 0.05). Since nearly none of the data analysed was normally distributed, all sample statistics throughout the manuscript indicate the median, and between curly brackets, the 25th and 75th percentiles: "median {Q1, Q3}".

5.4 Results

5.4.1 Demographics

Table 1 presents participant demographics, stratified by participating site and dataset: training (N=206) and hold-out (N=207). SDB was the most common disorder in both datasets (training: 56%; hold-out: 51%), followed by insomnia (25%, 28%), movement disorders (15%, 14%), and REM parasomnia (14%, 13%).

Table 1 – Demographic information and AHI of participants in the training and hold-out sets.

Site	N	N Female (%)	Age (year)	BMI (kg/m^2)	AHI* (events/h)	TST (h)
Training set (N=206)						
OLVG	68	21 (30.9)	46 {36, 58} [21, 78]	27.0 {24.5,30.0} [19.0, 39.0]	16.1 {6.3, 27.5} [0.8, 69.4]	7.2 {5.9, 7.8} [1.8, 10.0]
Kempenhaeghe	138	61 (44.2)	52 {34, 62} [13, 83]	26.0 {23.2, 29.3} [18.7, 39.4]	7.0 {3.2, 19.3} [0.0, 78.8]	6.7 {5.5, 7.4} [0.0, 9.0]
Hold-out set (N=207)						
OLVG	60	20 (33.3)	46 {36, 55} [21, 71]	28.0 {25.0, 31.0} [19.0, 47.0]	16.7 {7.3, 27.8} [1.8, 72.8]	6.7 {5.3, 7.5} [2.9, 8.2]
Kempenhaeghe	147	58 (39.5)	53 {38, 64} [17, 80]	25.4 {23.4, 28.7} [17.8, 42.2]	7.5 {2.8, 17.9} [0.2, 72.2]	6.8 {5.5, 7.5} [0.1, 8.5]

*BMI: body mass index, AHI: apnea-hypopnea index, TST: total sleep time. For age, BMI, AHI and TST the table indicates the median, between curly brackets, the 25th and 75th percentiles, and between square brackets, the range. *AHI according to PSG.*

5.4.2 Sleep staging

Comparison of sleep stages derived from chest-wall accelerometry with those derived from PSG yielded a κ-coefficient of agreement of 0.665 for four-class sleep staging and of 0.713 for Wake versus Sleep classification. Using the agreement thresholds from Landis and Koch for interpreting κ, agreement with PSG was substantial for all classification tasks[141]. The best performing class was REM, followed by Wake, with a sensitivity of 76.7%, a specificity of 96.7% and a positive predictive value of 81.4%.

The median, first, and third quartiles of the error in TST were +2.5, -9.24, and +14.6 minutes for the hold-out set, and -1.43, -12.5, and +12.2 minutes for the training set. The algorithm achieved an ICC of 0.942, with a 95% CI of 0.920 to 0.960 ($p<0.001$) for TST on the hold-out set.

5.4.3 AHI estimation

Table 2 indicates the performance in AHI estimation over the hold-out set. The algorithm achieved an overall ICC of 0.89, with a bias of 0.62 events/h. The ICC for female (N=78) and male (N=128) participants was 0.90 and 0.87, respectively, not significantly different (Mann-Whitney U test, $p = 0.28$). Pearson's correlation coefficient reached a value of 0.81 on the AHI against PSG. Figure 4 illustrates the estimation error (AHI_{ACC} minus AHI_{PSG}) for each recording in the hold-out set, separately coloured per OSA severity according to AHI_{PSG}. The algorithm achieved a negligible bias of -0.62 events/h, with 95% LoA of -15.0 and +13.8 events/h. The error distribution is narrowest at $AHI_{PSG} < 5$ and becomes wider with increasing severity.

Table 2 - AHI estimation performance for hold-out set, overall and per center.

Centre	N	r	p-value (r)	ICC	95%CI	p-value (ICC)	RMSE	Bias	SD error
All	207	0.81	<0.001	0.89	[0.85, 0.91]	<0.001	7.34	0.62	*7.34*
Performance per centre									
OLVG	60	0.83	<0.001	0.85	[0.76, 0.91]	<0.001	9.03	0.40	9.10
Kempenhaeghe	147	0.78	<0.001	0.90	[0.86, 0.93]	<0.001	6.53	0.72	6.51

N: number of recordings for each performance evaluation ('overall' corresponds to the complete hold-out set), r: Spearman's r, ICC: intraclass correlation coefficient for two-way random-effects model for absolute agreement, 95%CI: 95% confidence interval for the ICC, RMSE: root mean square error, SD error: standard deviation of the estimation error.

Table 3 indicates the confusion matrix for OSA severity classification on the hold-out set (N=207), using near-boundary double labelling. The predicted severities matched the reference for 164 recordings, with the predicted severity class off by one class for 40 recordings, by more than one class for only 3 recordings. The four-way κ-coefficient of agreement was 0.71 and the accuracy was 79.2%.

Table 4 indicates the overall classification performance for each severity threshold, using near-boundary double labelling. Negative likelihood ratios were 0.076, 0.244, and 0.139 for the Mild, Moderate, and Severe thresholds.

Table 3 – Confusion matrix for OSA severity classification on the hold-out set using near-boundary double labeling and after reduction of the reference labels to standard categories.

Ref.↓ Pred.→	No	Mild	Moderate	Severe
No	**49 (83.1%)**	4 (4.8%)	2 (5.1%)	1 (3.8%)
Mild	10 (16.9%)	**62 (74.7%)**	5 (12.8%)	0 (0.0%)
Moderate	0 (0.0%)	17 (20.5%)	**28 (71.8%)**	0 (0.0%)
Severe	0 (0.0%)	0 (0.0%)	4 (10.3%)	**25 (96.2%)**

Each cell indicates the count, and between parenthesis, the distribution of predictions (per column) across the reference severity classes (the diagonal indicates the positive predictive value for the class).

Table 4 - Performance for classification at different severity thresholds on the hold-out set, using near-boundary double-labeling.

Min. severity	Prev.≥Sev.(%)	Acc.	Sens.	Spec.	PPV	NPV	LR+	LR-
Mild	72.9	0.918	0.934	0.875	0.831	0.953	7.5	0.076
Moderate	35.7	0.879	0.770	0.940	0.880	0.877	12.8	0.244
Severe	14.0	0.976	0.862	0.994	0.978	0.962	153.4	0.139

Min. severity: minimum severity defined for positive class (i.e., positive class includes this severity and higher) for each binary classification task; Prev.≥Sev.: Prevalence, or percentage of participants with severity equal or greater than the minimum severity; Acc.: accuracy; Sens.: sensitivity; Spec.: specificity;

PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR-: negative likelihood ratio.

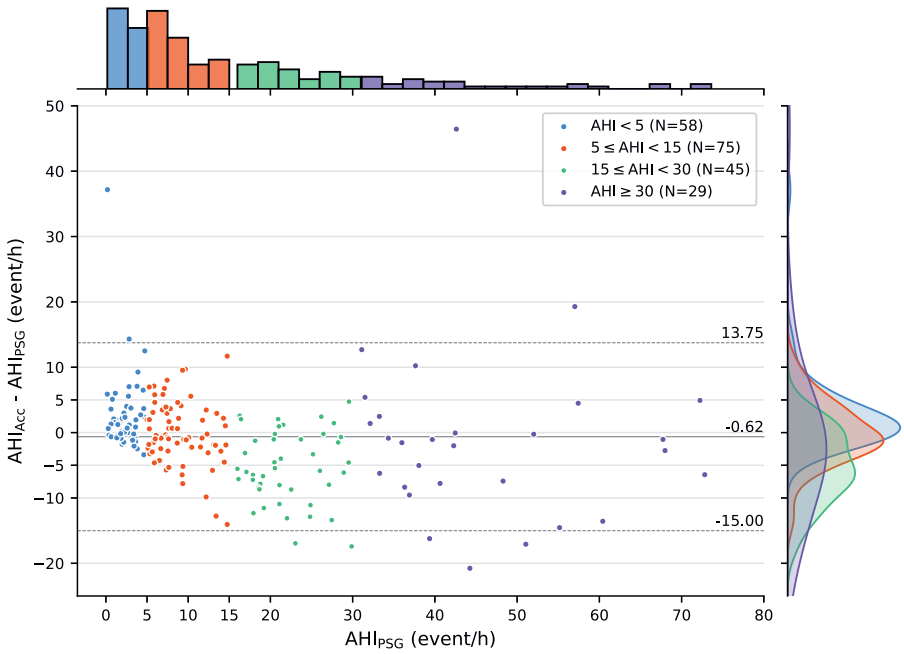


Figure 4 - Scatter plot of error per severity class (according to PSG). The histogram above the scatterplot shows the distribution of recordings per severity and the density plot at the right of the scatterplot shows the errors per severity. The solid horizontal line indicates the overall AHI estimation bias, and the dashed lines show the 95% limits of agreement.

5.4.4 Relation with body position

Table 5 compares the TST and AHI obtained from PSG and from the accelerometer overall, and per body position (supine, and combination of all non-supine positions), together with the ICC between the PSG and accelerometer-based estimations, for all recordings with a minimum of 20 minutes in supine and in non-supine positions. For the AHI, an overall ICC of 0.89 was reached, with an ICC of 0.87 in the supine position and of 0.86 in the non-supine position. The overall TST reached an ICC of 0.95, which is slightly different than the 0.942 which was reported earlier in this section for the complete hold-out set (N=207), because here we only consider recordings with at least 20 minutes in supine and non-supine positions (N=147). ICC values for separate body positions can be higher than the overall ICC value, because when calculating the ICC, data is first centred and scaled based on its mean and standard deviation. This eliminates the different biases obtained in TST estimation for each separate body position which led to lower agreement when computing the statistic for both positions combined.

Table 5 - Comparison between PSG- and accelerometer-derived TST and AHI for the supine and non-supine body positions, on the subset of 147 recordings with a minimum of 20 min of total sleep time (PSG) for each body position (supine and non-supine).

Body Position	TST_{PSG} (h)	TST_{ACC} (h)	ICC_{TST} [95% CI] (-)	Bias [LoA] (h)	AHI_{PSG} (events/h)	AHI_{ACC} (events/h)	ICC_{AHI} [95% CI] (-)	Bias [LoA] (events/h)
All	6.93 {5.90, 7.55}	6.92 {6.06, 7.58}	0.95 [0.93 0.96]	0.06 [-0.79, 0.90]	9.77 {4.89, 21.2}	10.1 {4.18, 17.9}	0.89 [0.85, 0.92]	-0.93 [-15.10, 13.24]
Supine	2.41 {1.35, 3.45}	2.32 {1.52, 3.47}	0.99 [0.99 0.99]	0.04 [-0.36, 0.44]	16.9 {7.33, 37.6}	15.2 {4.39, 34.1}	0.87 [0.82, 0.91]	-2.46 [-24.99, 20.07]
Non-supine	3.84 {2.77, 5.29}	4.01 {2.89, 5.42}	0.98 [0.98 0.99]	0.01 [-0.57, 0.60]	4.70 {2.66, 14.4}	6.95 {2.55, 13.2}	0.86 [0.81, 0.90]	-0.02 [-14.67, 14.62]

TSTPSG: total sleep time (TST), according to scoring from PSG; TSTAcc: TST derived from CReSS sleep staging based on accelerometer inputs; ICCTST: intraclass correlation coefficient for two-way random-effects model for absolute agreement in total sleep time estimation; 95%CI: 95% confidence interval for the (corresponding) ICC; LoA: 95% limits of agreement, after a Bland-Altman analysis of the estimation error; AHIPSG: apnea-hypopnea index (AHI) according to scoring from PSG; AHIACC: AHI estimated from the accelerometer-based algorithm; ICCAHI: intraclass correlation coefficient for two-way random-effects model for absolute agreement in AHI estimation, curly brackets: the 25th and 75th percentiles.

5.5 Discussion

We developed a method for AHI estimation using an accelerometer on the chest, by combining a cardiorespiratory sleep staging algorithm with an adapted neural network for detecting respiratory events. Performance was evaluated on a hold-out set of 207 participants. Total sleep time was estimated accurately with first and third error quartiles of −9.2 and +15 minutes, as a basis for accurate AHI estimates. Compared with the PSG-based AHI standard, chest-wall accelerometry achieved an ICC of 0.89 (95% CI: 0.85 to 0.91), a negligible bias of -0.62 events/h, and 95% LoA of -15.0 and +13.8 events/h. Using near-boundary double labelling, OSA severity classification reached a Cohen's κ coefficient of agreement of 0.71, indicating substantial agreement with severity classes determined from PSG[30]. The performance of the AHI estimation changed very little if the evaluation intervals were restricted to either supine or non-supine positions. In non-supine position, the algorithm achieved an ICC of 0.86 (95% CI: 0.81 to 0.90). This is important in the application of PT, as the goal is to have the patient sleep as long as possible in a non-supine position, and thus, the estimation of residual AHI should provide reliable results in this situation. The findings indicate that the method may be suitable for AHI estimation during PT.

In comparison with currently available methods for AHI estimation during OSA therapy, a recent meta-study of the accuracy of the AHI during PAP therapy by commercially available machines showed 95% LoAs ranging from -15.8 and -4.9, to 0.1 and 18.8 events/h, when evaluated over 12 studies[128]. Four of the studies reported ICC values, being: 0.69[142], 0.74[143], 0.79[144], and 0.96[145]. In a recent study of 300 OSA patients on PAP therapy[146], it was found that the negative likelihood ratio for AHI estimation was higher than 0.5 for all boundaries and considered by the authors as inconclusive (LR– was 0.58, 0.58, and 0.59 for the boundaries 5, 10, and 15 events/h, respectively). In the present study, we found LR– values of 0.20, 0.31, and 0.17 for the (classical) boundaries 5, 15, and 30 events/h, respectively.

The ICC, the 95% LoA, and the negative likelihood ratio of our method are well within the range of performances reported for PAP machines. This suggests that our method could be equally suitable for AHI monitoring during OSA therapy. However, it is relevant to remark that our participants were measured during diagnostic PSG, whereas the PAP studies evaluated participants during therapy. Therefore, range of the AHI values may vary between the two populations, with lower AHI values observed during PAP therapy and higher values in our study.

Regarding the potential use during PT, it should be noted that of course performance in different body positions in this study was evaluated without any form of therapeutic intervention. During actual positional therapy, the stimulation used to trigger the patient to change from a supine to a non-supine position might influence sleep architecture and perhaps introduce artifacts that could negatively impact AHI estimation performance. This effect may be limited however, given the relatively small number of positional changes during a night, compared to for example the number of naturally occurring arousals, or even residual apnea and hypopnea events. AHI estimation performance may in practice also be impacted by the accelerometer placement and body contact; here the devices were mounted on a respiratory thoracic RIP belt by professional staff, while during therapy the devices are mounted by the patients themselves.

In the context of a diagnostic application, the method demonstrated relatively high positive likelihood ratios for AHI severity classification (2.64, 7.94, and 73.7 for the Mild, Moderate, and Severe classes, respectively, using classical boundaries). For comparison, in a recent meta-study on peripheral arterial tonometry (PAT)[147], positive likelihood ratios of 1.66, 3.34, and 5.74 were found for the same boundaries. A few studies on PAT reported an ICC for the AHI, over cohorts with varying characteristics. In one study, among 30 participants suspected of having OSA, an ICC of 0.88 was found when comparing against PSG[148]. In an early PAT meta study, a Pearson correlation of 0.89 was found[149] when com-

paring the AHI against that derived from PSG, while the overall value found in the present study was 0.81.

Although not fully comparable with previous studies because of variations in study cohorts and in scoring parameters (for example 3% versus 4% desaturation in case of hypopneas), the relatively high positive likelihood ratios found in this study indicate that the method might be usable for diagnostic purposes. An interesting practical use case may be to integrate the method in wearable patches for extended Holter ECG monitoring. In this case, the IHR may be derived from the ECG signal, while respiration can be derived from an added accelerometer[150]. Since OSA is common in patients with cardiac disorders including arrythmias population, with prevalences reported up to 74%[94] and the burden of OSA may worsen the cardiac condition[151], there is clinical value in having an estimation of possible OSA severity for further follow up[152].

Our method might also be useful as a backup sensing modality during sleep recordings, for example in type II sleep tests (unsupervised polysomnography). These tests require a high number of sensors that need proper galvanic or optical skin contact, and poor recording quality is not uncommon. Using alternative signals from an accelerometer device may save otherwise unusable recordings. Body position sensors in current PSG systems may have insufficient sensitivity to detect cardiac vibrations or chest wall movements but could be easily replaced by commercially available accelerometers with adequate sensitivity.

A limitation of this study was that patients with known heart rhythm disorders were not included. In the presence of arrhythmias, the cardiac measurements themselves may be less reliable and the relationship between these measurements and sleep apnea may have been insufficiently captured in the present dataset. Furthermore, the proposed method is not capable of distinguishing obstructive from central apnea events. In future research, it could be investigated if full absence of respiratory movements could be used to discern central from obstructive events. Furthermore, the method was evaluated among patients undergoing polysomnography and the sensors were mounted by clinical staff. In future studies, the method should be evaluated in application scenarios, possibly with self-attachment of the sensor and during actual OSA therapy, in particular PT, which currently lacks any measurement of AHI or sleep.

In summary, this study demonstrates the potential of accurately measuring AHI using chest-worn accelerometry, achieving substantial agreement with PSG-derived AHI. Importantly, accelerometer-based AHI estimation performance remained consistent across all body positions, and comparable to the performance of residual AHI measurement during PAP therapy. This method may be of value in various clinical applications, in particular

in the treatment of OSA with PT. By offering the possibility of measuring sleep stages and AHI on a nightly basis for patients undergoing this therapy, it enriches the currently available treatment information which consists simply of usage time, and a sleep position summary. As the severity of the disorder may vary or worsen over time, and eventually even manifest itself in non-supine positions, continuous monitoring of sleep and AHI during treatment may substantially enhance the assessment of therapy efficacy and the management of the disorder.

6

General Discussion and Perspectives

CHAPTER 6

6.1 Research goals

The goal of this thesis was to develop methods for the monitoring of sleep and sleep apnea using chest-wall accelerometry and to evaluate their effectiveness by comparing sleep stages and AHI against those derived from PSG.

To enable sleep staging and AHI estimation, we developed robust and accurate methods for cardiorespiratory measurements using chest-wall accelerometry. A neural network was developed that that could estimate (surrogate) respiratory effort with very low error when compared to state-of-the art methods. Additionally, a statistical estimation method was developed that could robustly measure instantaneous heartrate with very high accuracy, such that results were compliant with the American national standard on cardiac monitoring equipment.

With these cardiorespiratory measurements, we leveraged a previously developed method for cardiorespiratory sleep staging. To mimic the respiratory impedance plethysmography signal used for training the sleep staging method, we injected typical gross body movement artifacts, usually found in respiratory belt signals, into our estimated respiratory effort signal. Four-class sleep staging with chest-wall accelerometry as input showed substantial agreement with sleep stages derived from PSG when evaluated on a diverse cohort including sleep-disordered participants.

We then leveraged a neural network for AHI estimation from ECG and RIP signals that was previously trained on a large PSG dataset. By means of transfer learning, we tuned this network to the characteristics of the cardiorespiratory signals derived from chest-wall accelerometry. The AHI derived from chest-wall accelerometry achieved a high correlation with that from PSG. Low negative likelihood ratios indicated the suitability of chest-wall accelerometry for use during OSA therapy, where it is important to determine that the (residual) AHI is below a certain severity level, indicating that the patient is adequately treated. We showed that performances were comparable to those obtained for the estimation of residual AHI during PAP therapy, although the performances for PAP therapy were obtained during therapy, while the results here were not. Furthermore, positive likelihood ratios were high and comparable to those obtained for peripheral arterial tonometry (PAT), an often-used home sleep apnea test, indicating suitability of our work for diagnostic use.

6.2 Interpretation of results

Respiratory Effort

For the estimation of respiratory effort, we first developed an enhanced version of principal component analysis (PCA), which yielded a 40% lower estimation error compared to the state of the art. However, the method was developed on a relatively small group of 20 healthy individuals, and when applying it on sleep-disordered participants, its performance deteriorated. We then leveraged the suitability of convolutional neural networks (CNNs) to represent complex, non-linear relationships. To find the optimum CNN architecture, we automatically searched for the best number of layers, channels, and kernel sizes. This yielded a 9-layer CNN with 207,855 parameters that was able to reduce the estimation error by an additional 60% compared to the enhanced PCA method.

Besides yielding a signal closely resembling the respiratory impedance plethysmography (RIP) signals during normal breathing, this method allowed for an accurate representation of breathing during and around respiratory events, as illustrated in Figure 1, where the reference RIP signals are shown in the middle and the accelerometry-derived signal at the bottom. Respiratory events can lead to changes in signal amplitude and frequency, which are present and visible in the RIP signals as well as in the accelerometry-derived signal.

While Esophageal Manometry is considered the gold standard for measuring respiratory effort[41], RIP was selected as the reference for this study. The rationale is that RIP is the standard sensor used in clinical PSG and was the specific modality used to train the neural networks for sleep staging and AHI estimation in our target application.

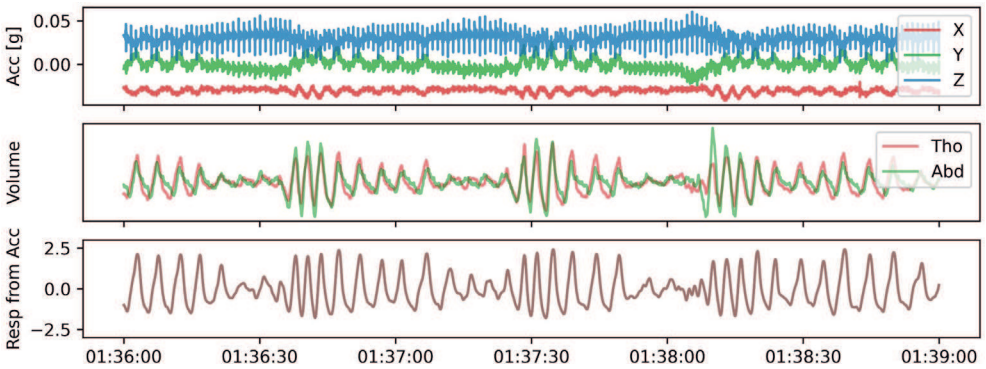


Figure 1 Example of estimated respiratory effort (bottom) derived from the 3D accelerometer signal (top, with arbitrary offsets for x, y, and z), with the reference thoracic and abdominal respiratory impedance plethysmography signals (middle). Note how the variations in amplitude and frequency in the plethysmography signals, potentially indicative of respiratory events, are also present in the accelerometry-derived signal.

Instantaneous heartrate

For the estimation of instantaneous heartrate (IHR), we first needed to localize heartbeats in time (t_n) and estimate the associated interbeat intervals (IBI_n). The instantaneous heartrate can then be computed by taking the reciprocal of the IBIs (i.e. $IHR_n = 60/IBI_n$, in beats per minute when IBI_n in s) and by resampling these at an equidistant time grid.

For IBI estimation, we started from a method that was shown to be able to estimate these in seismocardiography (the measurement of whole-body cardiac vibrations)[29]; this method, in particular, relies on **maximum-likelihood estimation** (MLE) for IBI estimation. Although this method could estimate **individual** IBIs very accurately, the resulting IBI sequences could be temporally inconsistent (i.e. $t_{n+1} \neq t_n + IBI_n$ and $\sum_{n=1}^{M-1} IBI_n \neq t_M - t_1$, with t_n the localization time of beat n). Consequently, the IHR sequence could become inaccurate.

Our contribution was that we enforced **temporally consistent sequences** of IBIs and localization times ($t_{n+1} = t_n + IBI_n$, and $\sum_{n=1}^{M-1} IBI_n = t_M - t_1$). In this way, the localization of one heartbeat could support the localization of other heartbeats and vice versa. Similarly, estimations of individual IBIs could support each other. As a result, the estimations became more robust and accurate. In addition, we used **maximum a posteriori estimation** (MAP) rather than MLE.

Here, MAP aims at finding a temporally consistent beat sequence for which the posterior probability is maximized. The posterior probability is the product of the likelihood and the prior. The likelihood expresses how well the observed chest-wall accelerometry data fits the estimated beat sequence. The likelihood is high if the data are locally periodic at time distances equal to the IBIs (with the likelihood based on the previous method, that used MLE for individual IBIs[29]). The prior represents the prior belief in the intrinsic plausibility of the beat sequence per se, preventing the fitting of unplausible, erratic beat sequences to noisy acceleration data.

On the downside, constraining the search space to temporally consistent IBIs sequences rather than individual IBIs caused a substantial increase in computational complexity. A combinatorial optimization problem was created that was of non-polynomial (exponential) complexity. To efficiently approximate a maximally probable beat sequence, we transformed the MAP problem into a Markov decision process. We used the Bellman equation (a foundational concept in reinforcement learning), to approximate the Bayes optimal solution. With depth-constrained search we achieved accurate results, while keeping computation time limited.

The result is a robust and highly accurate estimation of IBIs and localization times. The mean absolute IBI error was only 3.5 ms, less than the sampling time of the acceleration signal (4 ms), when evaluated over a diverse cohort, including sleep disorders. The 95% limits of agreement for IHR ranged from -1.20 to +1.25 beats per minute and met the accuracy requirements of the American national standard on cardiac monitoring equipment[90].

The example in Figure 2 shows the IHR for the same interval as in Figure 1. It shows the IHR from ECG in blue and the one from chest-wall accelerometery in red. In both signals increases in the IHR can be seen that occur simultaneously with increases in respiratory rate.

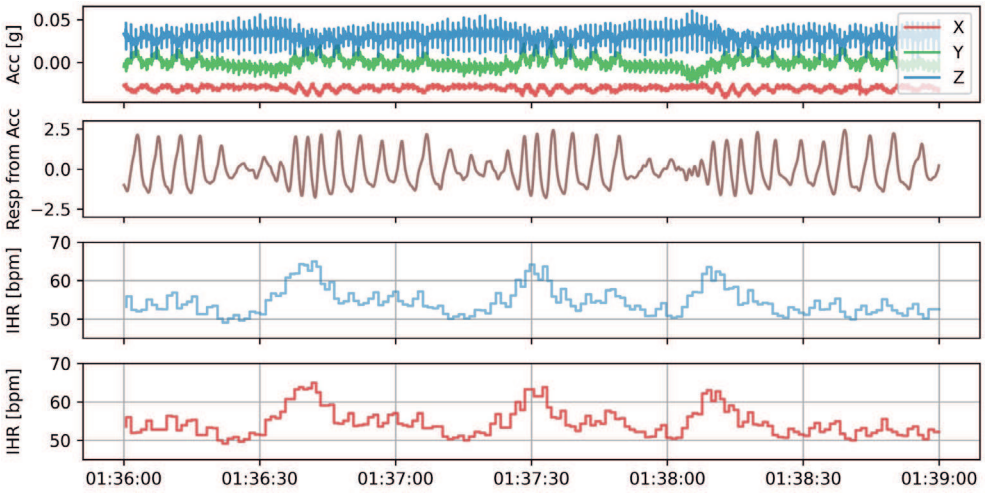


Figure 2 Example of estimated respiratory effort and instantaneous heart rate (IHR). The respiratory effort signal in and the IHR signal in red (bottom) were derived from the accelerometer signal, while the IHR signal in blue (third graph) was derived from the ECG signal. The increases in IHR coincide with the temporary increases in respiratory rate, and both follow decreases in amplitude of the (surrogate) respiratory effort signal.

AHI estimation

We leveraged a previously developed method that could estimate the AHI from IHR derived from ECG, and (surrogate) respiratory effort from RIP. The method uses a neural network to detect respiratory events, which it then counts and divides by an estimated total sleep time to obtain the AHI. Its performance was evaluated on a cohort of 653 participants with a large range of OSA severities and reached an intraclass correlation coefficient (ICC) of 0.91[36].

Our contribution was that we used this network in combination with respiratory effort and IHR derived from chest-wall accelerometry. To tune the network to the specific properties of inputs from chest-wall accelerometery, we applied transfer learning. A small fraction of the network parameters at the inputs and outputs was retrained while keeping the remaining parameters frozen. The retrained network then was used to estimate probabilities of respiratory events, see Figure 3. Using (sleep-stage dependent) thresholds, event probabilities could be converted into event occurrences, which were counted. Dividing the number of events by the total sleep time (derived from sleep staging) then yielded an estimate for the AHI.

A high ICC of 0.89 was reached for the AHI when compared to that derived from PSG. Using near-boundary double labelling, Cohen's kappa coefficient of agreement on OSA severity classes (None, Mild, Moderate, and Severe) reached 0.71, indicating substantial agreement and showing that the method may be usable for classification of AHI severities.

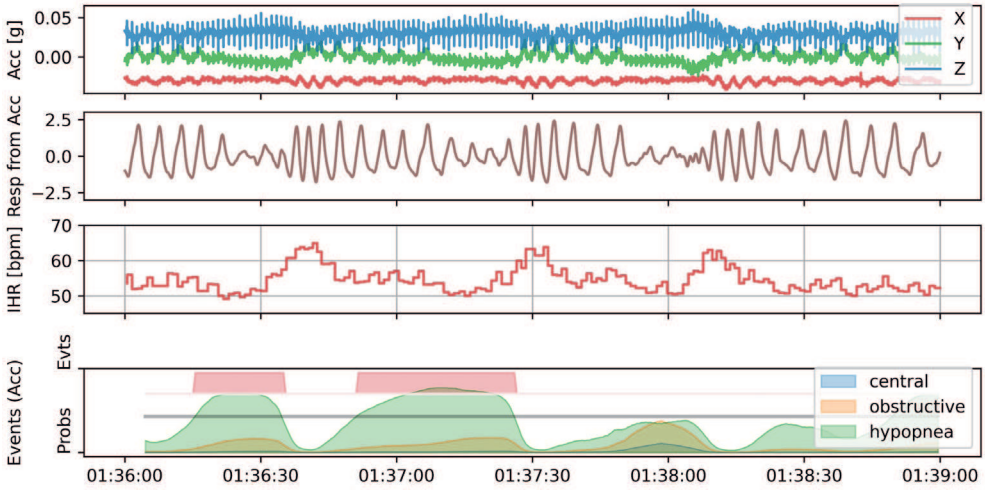


Figure 3 Example of event detection. When event probabilities (filled areas in bottom graph) exceed a (sleep-stage dependent) threshold, an event is detected (red areas).

Application in OSA diagnostics

The positive likelihood ratios were 2.64, 7.94, and 73.7 for the Mild (AHI between 5 and 15 events/hour), Moderate (AHI between 15 and 30 events/hour), and Severe (AHI above 30 events/hour) OSA classes, respectively, using classical boundaries. These relatively high values indicate that the method's ability to establish if a patient's AHI is at least as severe as the respective severity class. This is important in the diagnosis of OSA, which relies on the AHI boundaries for Mild and Moderate OSA, as specified in the international

classification of sleep disorders[5]. Here, a comparison with the widely applied peripheral arterial tonometry (PAT) is relevant. PAT has been accepted for home sleep studies by the US Centers for Medicare and Medicaid Services (CMS, setting health insurance standards) and the AASM[153]. A recent meta-study of peripheral arterial tonometry (PAT)[147] found positive likelihood ratios of 1.66, 3.34, and 5.74 for the same boundaries as indicated above. One study among 30 patients suspected of having OSA reported an ICC of 0.88[148]. Although not fully comparable, the results obtained in this thesis indicate the method may be usable for diagnostic purposes at a level similar to PAT.

Application in therapy monitoring

The negative likelihood ratios were 0.20, 0.31, and 0.17 for the Mild, Moderate, and Severe classes, respectively, using classical boundaries. These relatively low values indicate that the method's ability to determine whether a patient's AHI falls below a respective severity class. This is important during OSA therapy to assess its efficacy, where the confirmation that the AHI during treatment remains below a threshold means that the therapy is successfully working at reducing the AHI.

For hypothetical use during therapy, a comparison with PAP therapy may be insightful: the AHI estimation performance of PAP devices is accepted for general clinical use. In a recent study among 300 OSA patients on PAP therapy[146], it was found that the negative likelihood ratios were all higher than 0.5 for all examined boundaries (0.58, 0.58, and 0.59 for 5, 10, and 15 events/h, respectively) and were therefore considered inconclusive. Furthermore, four studies reported ICC values between the estimated and the PSG reference AHI of 0.69[142], 0.74[143], 0.79[144], and 0.96[145]. These results show that our method can achieve performance comparable to that of PAP machines. However, a direct comparison should be made with caution, as our performances were obtained in absence of therapy while that of PAP devices was determined during therapy.

An important aspect of the present study is that performance was consistent across sleeping positions, with an ICC of 0.87 in the supine position and of 0.86 in non-supine positions. This consistency is relevant in positional therapy for OSA, where the supine sleeping position is avoided.

Sleep staging

For sleep staging, we employed a previously developed method for cardiorespiratory sleep staging (CReSS)[63]. CReSS utilizes a deep neural network with convolutional and long short-term memory layers. The CReSS algorithm is device agnostic and requires airflow and heart rate inputs with an option to rely on a respiratory effort signal during periods when airflow is absent or invalid. CReSS combines the stages N1 and N2 into a single class as the difference between N1 and N2 can mainly observed in cortical activity (spindles,

k-complexes) and less in cardiorespiratory characteristics. This is a known limitation of cardiorespiratory sleep staging[105].

In earlier work, CReSS was evaluated on a set with 296 PSGs with a wide range of OSA severities, and epoch-by-epoch agreement with scorings from PSG reached a Cohen's kappa coefficient of agreement of 0.64 when using and flow and heart rate, and 0.68 with respiratory effort as additional input.

Our contribution was that we used CReSS in combination with heart rate and respiratory effort derived by chest-wall accelerometery. The chest-wall accelerometry input signals were dissimilar to the neural network's training data. Because the original training data (RIP signals) contained motion artifacts from large body movements, these same artifacts had to be introduced into the RE signal derived from chest-wall accelerometry.

As a result, we obtained sleep stages that showed substantial agreement with those derived from PSG, with a Cohen's kappa coefficient of agreement of 0.68. This result matches the highest reported performance (kappa = 0.68) of the CReSS algorithm when using a traditional respiratory effort sensor, thereby indicating the use of chest-wall accelerometry as a reliable input.

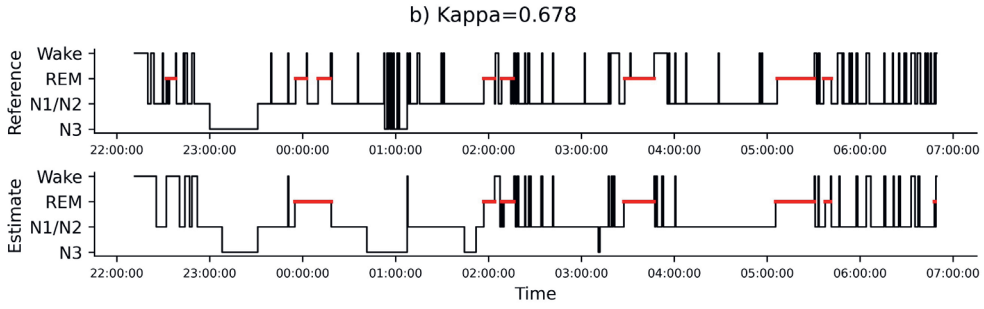


Figure 4 Example of median agreement (kappa = 0.68) between reference (top) and estimated (bottom) sleep stages.

The relatively high kappa value of 0.68 indicates that sleep stage estimation agrees substantially with sleep stages derived from PSG, see Figure 4 as an example. For comparison, reported kappa values for four-class sleep staging with PAT were 0.42[7] (N=16) and 0.48[154] (N=227, multi-centre), both indicating moderate agreement.

Sleep-wake detection reached a Cohen's kappa of 0.723, also indicating substantial agreement, which is important for AHI estimation. In addition to the accurate estimation of total sleep time, required to calculate the AHI, events should only be included in the numerator of the AHI when they occur during sleep. If sleep-wake detection if wrong, then events may erroneously be counted in or out. An accurate estimation of total sleep time was obtained, with a negligible bias and an inter-quartile error range from -9.5 to +17 minutes.

Performance was evaluated over a diverse cohort, including participants with sleep disorders. The presence of sleep-disordered breathing as well as insomnia had no significant effect on performance, and there was only a slight dependency on age, with better performance for younger participants. This indicates that sleep staging and the estimation of total sleep time may be used in cohorts with suspicion, or diagnosis of sleep disorders.

6.3 Strengths and Limitations

Sleep stages and respiratory events

The combined provision of sleep stages and respiratory events is a strength of this method. Not only are the sleep stages required to estimate a total sleep time, thereby enabling accurate AHI estimation, but also the sleep stages themselves can provide useful clinical information. In the example illustrated in Figure 5, respiratory events occur predominantly during REM sleep, which could be indicative of REM-OSA. The relationship between the occurrence of events and sleeps stages may help the clinician in refining the OSA diagnosis and prescribing a personalized treatment. To establish the reliability of such analyses, future research could go beyond evaluating the agreement in AHI estimation, and into evaluating the agreement between individual event detection from chest-wall accelerometry and PSG.

Furthermore, the sleep stages themselves may provide additional insights to a clinician. For example, insomnia may be or become comorbid with OSA. The night-to-night detection of sleep stages may enable the clinician to assess extended wake periods which might pertain to this condition and possibly adapt the therapy.

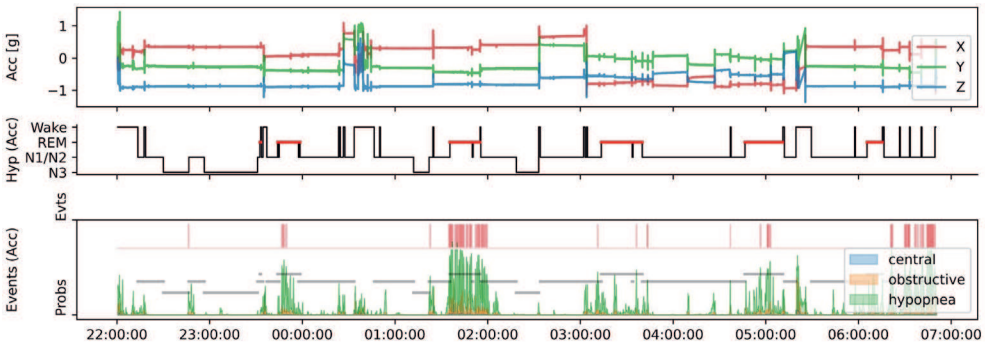


Figure 5 Example of sleep stages and events derived from chest-wall accelerometry. The bottom graph shows event probabilities with sleep stage-dependent thresholds. When probabilities exceed thresholds, events are detected. Events predominantly occur during REM sleep, which may be indicative of REM-OSA.

A diverse dataset collected over two centres

Another strength of this research was that it was conducted over a relatively large and diverse cohort (N = 413) recorded at two different sleep centres, with a large range of OSA severities. The simultaneous use of PSG (type I or type II home sleep test) and chest-wall accelerometry allowed for the development of methods for accurate and robust estimation of instantaneous heart rate and respiratory effort, that worked well across the target population. Furthermore, it allowed us to evaluate the performance on a hold-out set of 207 participants, acquired across two different sleep centres, with part of the recordings made at home.

Sleep staging and AHI estimation from cardiorespiratory metrics

The availability of methods for sleep staging and AHI estimation from instantaneous heart rate and respiratory effort, that were previously trained on large external datasets with a wide range of OSA severities, allowed our method to generalize well towards the hold-out set in our cohort. Training these methods only on the training set of the present cohort, would have led to a much higher risk of overfitting. Furthermore, the required adaptations for tuning the method to the acceleration signals were limited. The network for sleep staging could be left unmodified, only requiring a modification to one of the inputs: movement artifacts as present in the respiratory belt signals, on which the network was trained, had to be injected in the accelerometry-derived respiratory effort. The neural network for AHI estimation was tuned by means of transfer learning, with only 0.1% of the weights (at the first and last layers) being retrained.

Reference scorings

The gold standard references for sleep and AHI, against which performance was determined, were based on PSG. The rules for scoring sleep stages and AHI are extensively documented by the American Academy of Sleep Medicine in their guidelines for the scor-

ing of sleep and associated events[8]. Despite these efforts, a high inter scorer variability is inherently present in the scoring process. In a previous study of multi-scorer agreement it was found that with three scorers, only 60% of the scored sleep stages showed full agreement; with more scorers this proportion decreased to only 25%[113]. To have a reliable reference in this work, multiple scorers would have been required, which was impractical. Using only a single scorer for each recording, especially in the case of different centres with different levels of scoring experience, could have adversely affected the reproducibility of the results.

To alleviate this problem, we resorted to automatic scoring using the commercial software Philips Somnolyzer system[114]. Somnolyzer was trained on different datasets, in which recordings were scored by multiple accredited scorers. To deal with the inter scorer variability, Somnolyzer was trained against the consensus across the scorers. In this way, Somnolyzer carried the knowledge of multiple experts into its scoring logic. In our study, it allowed for an expert-level, consistent standard of scoring regardless of the origin of the data.

Use during therapy for positional OSA

A primary limitation of this study is that all data collection occurred during PSG diagnosis, thus in the absence of therapy for positional OSA. We compared our device's performance, in part, to the estimated residual AHI observed during PAP therapy. A direct comparison, however, warrants caution. During successful positional therapy, the AHI is expected to be reduced, significantly narrowing the range of values. Furthermore, PAP likely alters the event distribution, potentially reducing the ratio of apneas to hypopneas as positive pressure may convert full obstructions into partial ones even if not completely resolved. Although we also evaluated performance in the absence of the supine position (which mimics successful positional OSA therapy), this may not be fully representative for the use during actual therapy in the target population. Therefore, future research should integrate our device with positional OSA therapy in an ambulatory PSG setting to validate performance under therapeutic conditions.

Furthermore, in the present studies, the devices were carefully mounted by clinical staff on the thoracic respiratory belt, ensuring optimal placement. In a real-world application, however, the device's mounting, location (which may vary), and the fact that it would be applied by the patient rather than clinical staff, could significantly differ. This variability in placement and user application may influence the results and warrants investigation into robustness across various real-world mounting scenarios.

Potential influence of arrhythmia

A limitation of the study was that patients with known cardiac disorders such as atrial fibrillation, were excluded. Especially heartbeat detection, which relies on local periodicity may fail here, as a characteristic of atrial fibrillation is that local periodicity is often not (fully) present. This topic requires further research; a different beat detection method may be required.

Furthermore, it not yet known how well cardiorespiratory sleep staging works in the presence of arrhythmia. Publications on sleep staging using cardiorespiratory signals generally do not elaborate on the influence of heart rhythm disorders [105,135]. However, it is plausible that the erratic and chaotic nature of atrial fibrillation masks the autonomic nervous system's influence on the observable heart rhythm. This, in turn, could obfuscate the relationship between the heart rhythm and both sleep stages and OSA. However, there is evidence that cardiorespiratory sleep staging works while using beta blockers[155], a common medication used to treat known comorbidities of OSA.

6.4 Applications

A system architecture

Figure 6 outlines a prototype system architecture for the measurement of sleep and respiration using chest-wall accelerometry. The system comprises a wearable device and a cloud component. The wearable device can be designed as a patch or integrated into a band. It operates without skin contact or a user interface, is low in power consumption, and compact, making it unobtrusive and suitable for continuous or nightly use. The lack of skin contact eliminates the need for skin preparation (shaving, abrasion), prevents skin irritations, and avoids issues related to contact deterioration over time. Additionally, as the use of accelerometery only requires mechanical contact, it allows the sensor to be fully embedded within the device, rather than integrated into the device's enclosure, thereby reducing cost and enhancing reliability.

The device comprises a 3D accelerometer, data compression, memory, and a connectivity unit (either wired or wireless). The data compression concerns dimensionality reduction, filtering, downsampling, scaling, quantization, delta coding, and finally exponential Golomb-Rice coding. The data compression is lossy, and parameters were tuned to obtain a good compression ratio without compromising end-to-end performance. For IHR, IBI detection was executed on the embedded device, IBIs were transmitted from the device and IHR was computed in the cloud.

The memory may store the data for one or more nights, depending on the application requirements. Connectivity may be through a pod (in case the device is returned to the provider), Narrowband Internet of Things (in case of direct communication to a host), or Bluetooth Low Energy (in case of a phone).

All further processing, including all neural network inferencing, may run in the cloud where suitable processing power is available. As the neural networks used are relatively small compared to modern language models (with a few hundred thousand parameters versus billions), they can also run on phones, which have ample processing power and memory.

The cost of the device may be low, provided a cost-down of the accelerometer is possible. The ADXL355 accelerometer used in the study has an excellent signal-to-noise ratio but is relatively expensive. The ADXL380 may be a suitable successor. Furthermore, sensitivity analysis may indicate that performance holds for lower-cost accelerometers, something that was not done in the scope of this thesis.

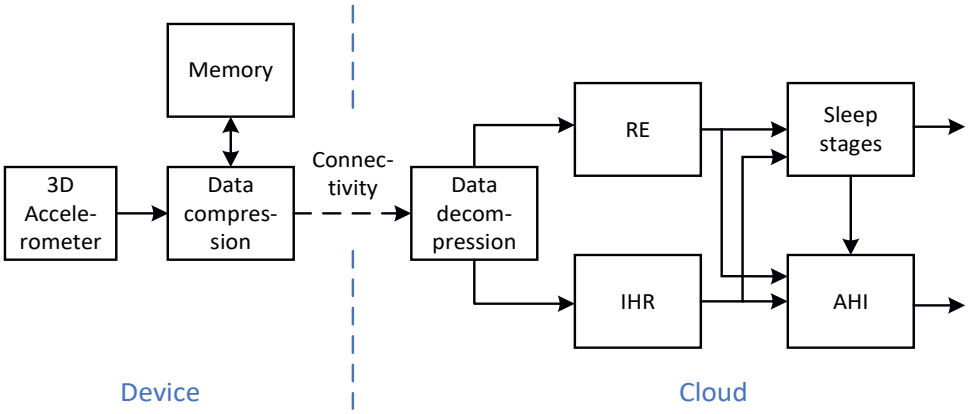


Figure 6 Prototype system architecture, comprising a wearable module that sends compressed data to a processing module in the cloud. RE = Respiratory Effort, IHR = Instantaneous Heart Rate, and AHI = Apnea-Hypopnea Index.

Implementations

To collect the initial raw acceleration data, a first prototype was created. This system comprised a portable recorder connected by cable to a device housing an accelerometer (see Figure 7 left), which allowed data download to a PC via USB. This initial setup served as a proof of concept for the data acquisition phase. Subsequently, to demonstrate technical feasibility and realize the architecture proposed in Figure 6, an improved prototype was developed. This device (see Figure 7 right) was entirely cable-free, incorporating internal

memory and Bluetooth Low Energy (BLE) connectivity to a smartphone. The smartphone could then upload the compressed data to a cloud service for decompression and further processing, specifically to derive sleep stages and the Apnea-Hypopnea Index (AHI). This improved device was built by modifying a commercial Philips NightBalance Sleep Position Trainer, internally integrating a sensitive ADXL355 accelerometer. Critically, this modification maintained the original enclosure and operation, requiring no external changes to the device.

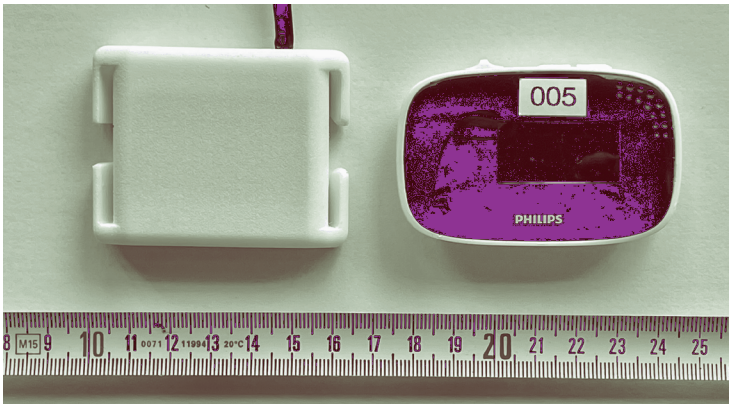


Figure 7 Prototype Evolution. The Left device is the initial wired prototype, comprising an accelerometer connected to an external recorder for raw data collection. The Right device implements the architecture in Figure 6, featuring integrated memory, data compression, and Bluetooth Low Energy (BLE) connectivity. Scale is in centimetres.

Therapy for Positional OSA

As indicated in the introduction, the target application for this work was therapy for positional OSA. The use of chest-wall accelerometry stemmed from positional therapy, where an accelerometer is already used to measure the sleeping position. Therefore, the integration of AHI estimation through chest-wall accelerometry can be relatively straightforward. The accelerometer may need to be replaced by one with a sufficiently high signal-to-noise ratio and data needs to be collected and uploaded towards a phone or the cloud. The AHI and sleep statistics may then be computed and inspected off-line. The relative cost increase may be limited, and the therapy and the use of the device do not need to change.

Extended Holter Monitoring

OSA is often comorbid with cardiovascular disease and atrial fibrillation. OSA can worsen cardiac outcomes, with one contributing factor the hypoxic burden resulting from apnea and hypopnea events [124,151]. Consequently, cardiac patients may benefit from OSA treatment. Although subject of debate, evidence suggests that cardiac patients may benefit

from CPAP treatment[156]. However, the screening for OSA is often not conducted as part of cardiac screening and diagnostics [157]. An opportunity for OSA diagnosis may be during extended Holter monitoring by means of patches, which are typically worn at the same body position as used the study data of this thesis. Often, an accelerometer is already present in an extended Holter monitor, such as for example in mobile cardiac outpatient telemetry (MCOT)[158]. If it is sufficiently sensitive, then respiration can be measured by means of chest-wall accelerometry. In addition, cardiac measurements can be done with the already present ECG sensor, thereby enabling the application of the present work for OSA diagnosis in cardiac patients without needing to change the externals or the usage of the Holter patch.

The device impact can be limited. The internal accelerometer should be sensitive enough and the raw acceleration data should be captured in memory (next to the raw ECG data). The full data analysis can be done off-line. The raw acceleration data may be downsampled as only respiration is required (IBIs can come from ECG), such that the data storage and transmission load only increase marginally.

Redundancy in Home Sleep Testing

The present method may also be of value in type II or type III ambulatory sleep tests. Here, a patient wears a device on the chest wall, which connects to various sensors on the body (including EEG, EMG, and EOG in case of a type II sleep test). As these methods are unsupervised and conducted at the home, there is a risk that one or more of the sensors detach or suffer from contact deterioration. Having an accelerometer in the device and implementing an architecture as in Figure 6 may serve as a backup measurement that can save otherwise unusable tests. The extra cost for the device may be very limited as most of the architecture, if not all, is already present in such a device. The accelerometer needs to be sufficiently sensitive, and, as home sleep test devices typically already have a relatively large internal memory, and upload data over an electrical connection, data compression may not be needed.

Non-sleep applications, patient deterioration

When patients who had an intervention are dismissed from care departments to a general ward or home, there is often a need to monitor them for a period as there is a risk of deterioration or readmission. Monitoring is typically based on physiological parameters such as pulse (heartrate), respiration rate, blood pressure, and SpO_2. Each of the parameters gets a score that increases with deviation from normal. Based on an aggregate score, an early warning score is computed[37].

In one approach[93], patients are monitored by means of an accelerometer device placed on the lower-left rib. This device monitors pulse and respiration rate and sends regular

updates to a monitoring centre. The contribution of our work to this application can be twofold.

First, there may be additional diagnostic value in the monitoring of sleep and OSA. OSA may to negatively affect cardiovascular outcomes and treatment of OSA may improve outcomes. In one study the negative effect of surgery on sleep architecture is demonstrated[159], while in another study sleep disruptions are associated with postoperative pain and impaired later functioning[160].

Second, the work in this thesis may improve the estimation of heart rate and respiratory rate[93], as these are the inputs to the computation of the early warning score. The narrow 95% limits of agreement for breathing rate and the compliance with the American national standard for cardiac monitoring equipment, may further enhance the accuracy of the vital parameter assessment.

6.5 Future Perspectives

Snoring

Snoring is a relevant factor in the diagnosis and management of OSA, and the ability to detect snoring would be a valuable addition to the technology in this thesis.

Snoring is the production of sound by vibrations in the upper airways, for which a partial closure or narrowing is required, and as such is an indicator for the potential presence of OSA. During full obstructions, no sound is produced and therefore no snoring sounds are generated. Because of this, the time pattern of snoring may be used for the detection of OSA [161]. As the upper airways are connected through the trachea to the lungs, the chest wall can serve as a sounding board for snoring. An illustration is given in Figure 7, which shows snoring signals from the contact microphone (mounted near the throat) and the chest-wall accelerometer. The green annotations in the spectrograms represent sound bursts, which correspond to snoring events. The annotations indicate the duration as well as the frequency of the snore.

As may be seen Figure 8, the accelerometer spectrogram closely matches that of the contact microphone, up to a frequency of 125 Hz. This is because the sampling frequency of the accelerometer is 250 Hz and the spectrogram only shows content up to the corresponding Nyquist frequency (250 Hz / 2 = 125 Hz). Visual inspections have revealed that some snoring sounds with fundamental frequencies higher than 125 Hz may be observed as aliased components (folding back into the spectrum from above), indicating that the accelerometer can potentially also detect snoring sounds with higher frequencies.

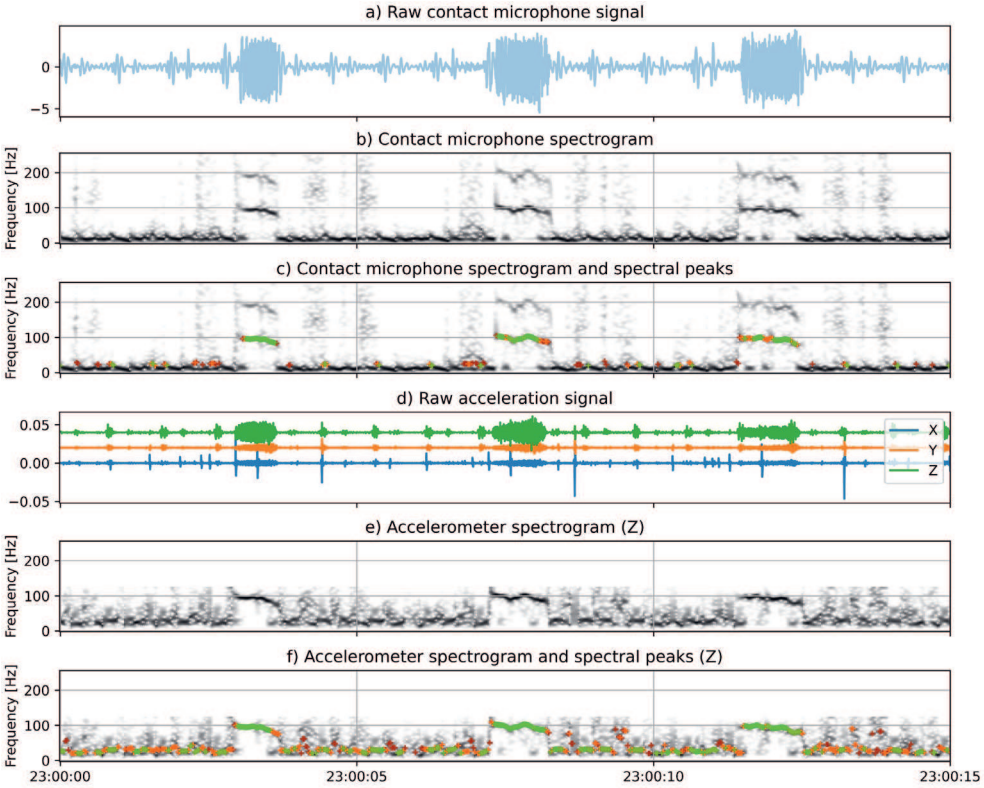


Figure 8 Signals from contact microphone and accelerometer: a) raw contact microphone signal, b) spectrogram of the contact microphone signal, c) spectrogram of the contact microphone signal with annotated peaks, d) raw acceleration signal, e) spectrogram of the Z-axis of the raw acceleration signal, f) spectrogram of the Z-axis of the raw acceleration signal with annotated peaks. The peak prominences are color-coded, with red, orange, dark green and light green representing increasing prominences. The traces at approximately 100 Hz indicate snores.

For a proper evaluation of the potential of snoring detection by means of chest-wall accelerometry acquisition of new data with a higher sampling frequency is likely required. The benefits of snoring detection may be in improved AHI estimation, e.g. through observation of the distribution of the snoring events over time[161], besides providing additional clinical feedback regarding (positional OSA) therapy efficacy in alleviating this symptom. The advantages of using the accelerometer for snoring are that no (extra) sensor, such as a microphone, is required. It remains to be investigated if a sufficient percentage of the snoring sounds are indeed observable on the chest wall and if a sufficiently sensitive and precise classification can be achieved.

Hypnodensity analysis

It is quite common that when multiple scorers score the same PSG, that they score a percentage of the epochs with different sleep stages[113]. Instead of representing each epoch with a single, categorical sleep stage, it could instead be represented with a distribution that represents the relative scoring frequencies. These relative frequencies may be interpreted as probabilities and visualized in a hypnodensity graph, with stacked bar plots representing the probabilities[162]. An example of such a hypnodensity graph is illustrated in Figure 9, with different colours indicating the probabilities of different sleep stages.

Hypnodensities can provide a richer, more detailed, and more accurate view on sleep and especially sleep stage transitions, thereby offering increased diagnostic power [162].

Hypnodensities can also be useful during evaluation of new methods for sleep staging. Instead of comparing predicted sleep stages against consensus scorings (the sleep stages scored by most expert scorers), one might use the distribution of the export scorers as a reference, through the evaluation against hypnodensities. As neural networks typically predict class probabilities (here: sleep stage probabilities) and use the class with the highest probability as the predicted label, one could also directly compare the predicted probabilities against the hypnodensities. Agreement could for example be quantified by means of the cosine similarity, or intraclass correlation[113]. However, care should be taken as the probabilities generated by the neural network are "uncalibrated", meaning they do not necessarily reflect the true probabilities, i.e. they may be over-confident in some classes.

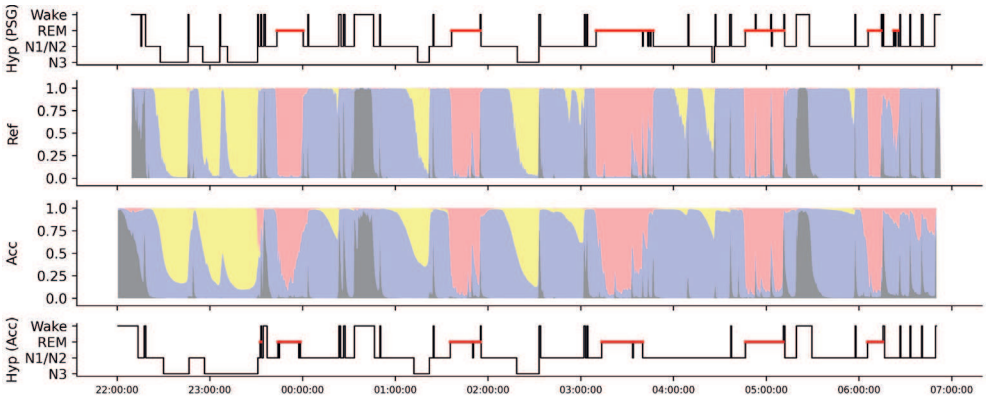


Figure 9 Hypnogram and hypnodensities from PSG, and hypnodensities and hypnogram from chest-wall accelerometry, Red = REM, Gray = Wake, Purple = light sleep (N1+N2), Yellow = deep sleep (N3). Hypnodensities from PSG were derived by means of Somnolyzer, while hypnodensities from chest-wall accelerometry were obtained with CReSS.

Such an evaluation could give a more nuanced view on the performance of a sleep staging method, as illustrated in Figure 9, which shows reference and predicted hypnodensities. As can be seen in the hypnograms, some short transitions to wakefulness that are present in the PSG-derived hypnogram are missed in the predicted hypnogram. However, one can see increased probabilities of wakefulness at the corresponding moments in the predicted probabilities. These probabilities just did not cross the probability boundary to become the most likely among the predicted classes.

Beyond evaluation, PSG-derived hypnodensities may be used for neural network training as well. In this way, the to-be-trained method could better represent the knowledge of the expert scorers, including their (inherent) differences in opinion. In the case of new sensing modalities, for which the number of simultaneous recordings with gold-standard PSG is often limited, this may alleviate the requirements for larger datasets, potentially through techniques like transfer learning which utilize knowledge encoded in the hypnodensities.

Future research on this topic may include evaluation of agreement on hypnodensities and the tuning of the sleep staging network used in this thesis to better represent hypnodensities in case of cardiorespiratory features derived from chest-wall accelerometry. The use of learning from soft targets (the hypnodensities) is known from *knowledge distillation*, a process where a smaller network (the student) learns from a larger network (the teacher). In this process, the student is trained on a combined objective function that minimizes the divergence between the student's and teacher's soft outputs, alongside the standard loss on the original hard labels (the sleep stages).

Detection of atrial fibrillation and other hearth rhythm disorders

Future research may be aimed at the detection of heart rhythm disorders by means of chest-wall accelerometry, particularly atrial fibrillation (AFib), the most prevalent heart rhythm disorder. AFib and other arrhythmias often go undiagnosed, and screening could prevent adverse outcomes, such as stroke[163]. Furthermore, AFib and OSA are often comorbid, with prevalences reported up to 74%[94] and (untreated) OSA being able of worsening the cardiac outcome[151]. An AFib detection method based on chest-wall accelerometry could bring value in various applications of chest-wall accelerometry, such as positional OSA therapy.

Detection methods for AFib frequently rely on the irregularity of IBIs, as it is one of the most characteristic features of AFib (next to absence of P-waves in the ECG). In ECG-based detection methods, IBIs may be derived from the timings of R-peaks in the form of RR intervals. The variability of RR intervals may then be quantified by statistical measures,

such as standard deviation of the RR intervals, or sample entropy. In other approaches, machine learning algorithms are used, such as for example Hidden Markov Models[164].

In methods based on PPG, this approach may be used as well. A hidden Markov model, trained on a large ECG dataset may be tuned to the characteristics of IBIs derived from PPG. In one study, among a population undergoing 24h Holter monitoring, this approach was used in combination with IBIs detected from wrist-PPG[165]. However, IBI detection during AFib is difficult. Not only the beat timings are variable, but the ejected blood volume is also variable. Consequently, the blood volume pulses measured by PPG are also variable, which makes beat detection difficult. Not only are the timings and amplitudes irregular, but the amplitudes can become very low and become hard to distinguish from noise. This difficulty in beat detection required substantial adaptations of the hidden Markov model, thereby potentially limiting the generalization capabilities.

It is hypothesized that this problem also occurs with IBI detection from chest-wall accelerometry during AFib, as from Newton's laws it follows that the magnitude of the acceleration signal is proportional to the ejected blood volume. This problem is illustrated in Figure 10, where two examples are given, one with a relatively regular heart rhythm (a) and one with an irregular heart rhythm (b). Differences in the quality of IBI estimation are clearly visible, with a relatively good agreement between IBIs derived from ECG and chest-wall accelerometry for the more regular heart rhythm and a relatively poor agreement for the irregular heart rhythm.

To detect AFib from chest-wall accelerometry, and to avoid the difficulties with IBI detection during AFib, an alternative approach is of interest. Instead of relying on IBIs, an alternative method could be based directly on the local periodicity representations (LPR, see chapter 3), which directly represent regularity of beats, in timing (IBIs) as well as in amplitude (magnitude). The LPR for regular heart rhythms significantly differ from those for irregular heart rhythms.

As can be seen in Figure 10, the LPR for a regular heart rhythm, shows a clear trace, while the LPR for an irregular heart rhythm (suspected AFib) does not have this clear trace but shows a scattered pattern. A classifier could take the LPR as input (similar to a 2D image) and could classify it as arrhythmic or non-arrhythmic. Future research could include patients with arrhythmias and detection performance could be evaluated. Note that in Figure 10, as hypothesized, the detected IBIs in case of the irregular heart rhythm (in red) are in poor agreement with the ECG IBIs (in blue).

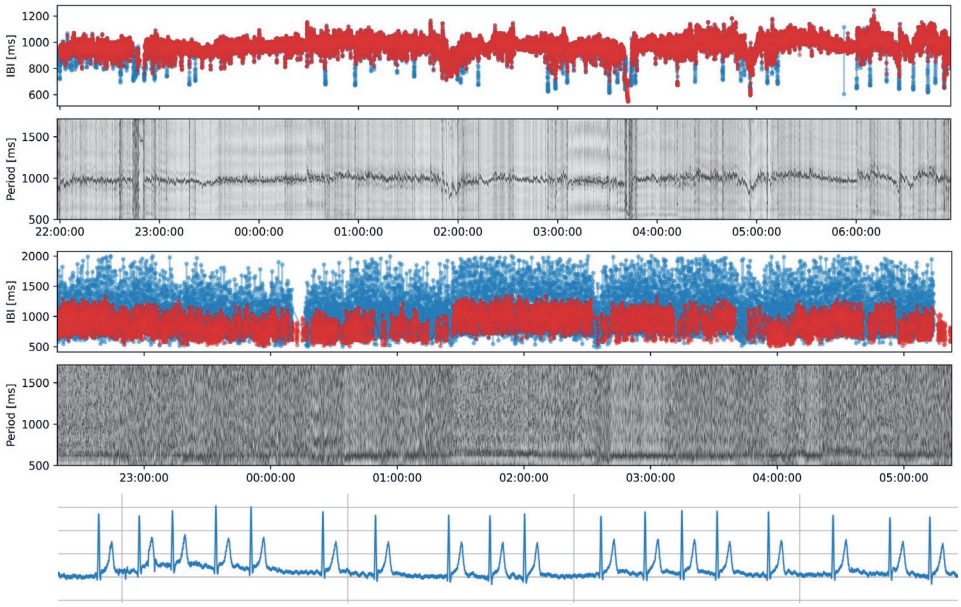


Figure 10 Examples of recordings with a regular (a) and an irregular heart rhythm (b). The top graphs show the IBIs from ECG (blue) and chest-wall accelerometry (red), while the bottom graphs show the local periodicity representation (LPR). For the regular heart rhythm (a), the ECG and chest-wall accelerometry IBIs match relatively well, and the spectrogram of periodicity shows a clear trace, while for the irregular heart rhythm (b), the IBIs don't match very well, and the spectrogram of periodicity is scattered. The ECG signal for the irregular heart rhythm recording shows an irregular sinus rhythm and has no visible P-wave, characteristic for atrial fibrillation (c).

6.6 Conclusions

This thesis presents a novel method that can successfully measure sleep and sleep apnea using chest-wall accelerometry. Its ability to confirm the absence of OSA suggests that it may be used for the monitoring of OSA therapy efficacy. Its ability to detect the presence of OSA shows that it may be used for the detection of OSA as well. In addition, it can provide measurement of sleep architecture in substantial agreement with that derived from PSG.

The technology can be integrated relatively easily in existing devices such as those for positional OSA therapy and extended Holter monitoring. Patients could benefit in multiple ways: otherwise undetected OSA could be detected and treated; otherwise ineffective therapy could be adapted or even replaced to properly treat OSA, thereby improving short term well-being and long-term outcomes.

Returning to the patient, who should ultimately benefit from this work, the clinical implications are multi-faceted. When prescribing positional therapy, the pulmonologist faced a dilemma: the overall high AHI severity led to hesitancy in recommending this less-intensive treatment. Our monitoring approach could resolve this uncertainty, either by reassuring the clinician that positional therapy is effective, or by promptly signaling inefficacy and allowing for timely therapeutic adaptation or even switching to a different therapy. Furthermore, continuous monitoring would enable clinicians to assess the sleep structure and detect issues like sleep continuity problems. Ultimately, the future development of algorithms for cardiac rhythm assessment could leverage this platform to detect conditions such as atrial fibrillation, allowing for early treatment and mitigating the risk of serious adverse outcomes.

Dankwoord

Eindelijk is het dan zover: ik kan het dankwoord van mijn proefschrift schrijven. Talloze uren heb ik doorgebracht op zaterdag- en zondagochtenden, werkend vanuit een stil huis terwijl iedereen nog sliep. Wat een geweldige reis was het, wat een inspirerende mensen heb ik ontmoet en wat heb ik veel nieuwe dingen geleerd. Het samenwerken met collega's van Philips, TU/e, Kempenhaeghe, OLVG en zelfs ver daarbuiten was enorm waardevol. Het werk dat ik hier heb mogen presenteren is tot stand gekomen dankzij velen. Ik ben dankbaar dat ik dit werk van hen en mij in deze vorm mag afronden.

Nu denk ik aan mijn ouders, **Theo** en **Fia**, die een voorbeeld voor me waren en me altijd gestimuleerd hebben. Die vertrouwen in me hielden, ook als de resultaten er nog niet waren. Wat zouden zij nu trots zijn geweest! Dankbaar ben ik voor de steun van **Barbara**, mijn vrouw, die me de afgelopen 6 jaar heel veel ruimte heeft gegeven, ondanks haar drukke baan en vele andere activiteiten. Trots ben ik op **Meike**, **Emiel**, en **Lotte**, hoe ze zich op hun eigen manier ontwikkelen, in de geneeskunde, in de rechten, en in het leven daaromheen. Het is fijn om ze erbij te hebben en dat geldt natuurlijk ook voor **Lars**, **Finn**, **Bryan**, en **Eva**.

Mijn reis naar dit werk begon vele jaren geleden met het meten van de hartslag op de pols, in een klein kantoortje met **Ad** en **Koen**. We deden allerlei dingen zelf, solderen, meten op atleten in Papendal en het in productie brengen in China. Wat een prachtige tijd was dat. In dit project ging ik met **David** meewerken in de signaalverwerking, zoals het meten van wandel- en fietscadans met een accelerometer; een ervaring waar ik veel van heb geleerd. Van **John** kreeg ik de ruimte om met **Alberto**, **Jenny** en **Linda** te werken aan het detecteren van atrium fibrillatie. We behaalden mooie resultaten en ik leerde hoe waardevol het was om samen te werken met een universiteit en een kliniek: het zaadje voor deze thesis was geplant.

Toen Philips de activiteiten in optische hartslagmeting afbouwde, ging ik op zoek naar een nieuwe uitdaging. **Paul** had al eens gezegd dat een PhD wel iets voor mij zou zijn, en mede daardoor werd mijn doel niet alleen het vinden een nieuwe positie, maar óók het doen van een PhD. Met **Pedro** vond ik een goede match. Naast z'n werk bij Philips had hij een aanstelling bij de TU/e en begeleidde al een aantal promovendi. Kort daarop maakte ik kennis met **Sebastiaan,** die mijn promotor zou worden. **Henning** regelde voor mij een plaats bij Philips Research, waar **Angela** en Pedro net waren gestart met een project om de AHI te schatten op basis van de signalen van een borst-gedragen accelerometer in de Philips *NightBalance*. Dat leek me een fantastisch PhD-project waar ik m'n eerdere ervaringen goed kon gebruiken. Sebastiaan was hier snel mee akkoord en we hadden een plan.

Omdat er de nodige klinische data verzameld moesten worden, ontwierpen en bouwden Ad, **Henry**, **Jan**, en **Henry** klinische prototypes voor datacollectie, de ISA en de NB. Angela ging voortvarend aan de slag en organiseerde diverse studies waaronder die in Kempenhaeghe en het OLVG. **Fokke**, **Bertram** en **Stan** zorgden vanuit Kempenhaeghe voor een vloeiende datacollectie en overdracht naar Philips Research. Vanuit het OLVG lieten **Pien**, **Meerie**, **Mickey** en **Emily** de zaak ook op rolletjes lopen. Met de prototypes en dit team konden we een fantastische dataset opbouwen. Zonder de instemming en hulp van meer dan **400 patiënten** was dit werk niet mogelijk geweest. Ik hoop van harte dat de resultaten op een bepaald moment vele andere patiënten ten goede mogen komen.

Het starten van de PhD betekende ook het werken op de TU/e. Allereerst kwam **Ruud** het team van begeleiders versterken en bracht heel veel kennis met zich mee. Daarnaast kwam ik in het warme bed van de *Sleep Group* terecht, met **Bernice**, **Iris**, **Marco**, **Hans**, **Jaap**, **Merel**, **Jiali**, **Marco**, **Xi**, **Martine**, **Luca**, **Jasmin**, **Laura**, **Guido**, **Abhirup** en **David**. De vele interacties waren heel waardevol.

The medical advisory board, instantiated by **Fayçal**, with members **prof. Pépin**, **prof. Heinzer**, and **prof. de Vries**, supported this work and gave me the opportunity to present to them, which was very motivating and inspiring. Furthermore, **Philips Sleep and Respiratory Care** (currently represented by **Hari** and **Tamara**), **Kempenhaeghe**, the **TU/e** (through **e/MTIC**), and **OLVG** have contributed significantly to this work. **Yavuz** and **Michiel** gave their support regarding IP matters.

Ten slotte gaat mijn dank uit naar de leden van de commissie, **prof. Verbraecken**, **prof. Aarts**, en **dr. Peri**. Zij hebben in hun ongetwijfeld drukke schema de tijd gevonden om mijn proefschrift te lezen en te beoordelen.

About the author



Fons Schipper was born in Purmerend on August 11, 1962, as the oldest son of builder Theo Schipper and schoolteacher Sofia van Baar. Fons grew up in the small village of Ilpendam, in Waterland, just north of Amsterdam. Among his early memories are the first moon landings in 1969, which his parents were watching live on a Philips black-and-white TV. These memories boosted his interest in science and technology and set Fons' path, which led to a study electrical engineering at the *Hogere Technische School Amsterdam*. Eager to learn more and gain deeper understandings, he continued his studies in electrical engineering at the *Technische Hogeschool Twente*, where he graduated *cum laude*.

After his studies, in 1987, he was fortunate to be hired by the *Philips Natuurkundig Laboratorium*, where he started in the new *IC design centre*, to work on optimisation algorithms for the design of large-scale integrated circuits. Wanting to see more of the world and Philips, Fons then moved to the *IC Lab* to work on ICs for digital audio and video. His first product to enter the market was the *SAA2502 ISO/MPEG Audio Source Decoder*. Moving from hardware to software, he joined Digital Systems Labs; as a software architect, he helped bring the first DVD recorder to market.

Later, as a systems architect, Fons developed a clinical prototype of an optical mammography system, reconstructing 3D breast images using laser scanning. Getting involved with measurement data and its interpretation, he changed his path to signal processing and algorithms. With that, he was part of the team that brought the world's first optical heart rate watch to the market at the *Consumer Electronics Show* in 2012. Soon after that, Fons started exploring the possibility to detect Atrial Fibrillation with this new technology. The cooperation with the *Eindhoven University of Technology* and the *Catharina Hospital* on that work boosted his interest in research and science, which finally led to this thesis.

Publications

Peer reviewed journal papers on the topic of this thesis

- Schipper F, Sloun RJG van, Grassi A, Derkx R, Overeem S, Fonseca P. Estimation of respiratory rate and effort from a chest-worn accelerometer using constrained and recursive principal component analysis. *Physiol Meas*. 2021;42(4):045004. doi:10.1088/1361-6579/abf01f
- Schipper F, JG van Sloun R, Grassi A, Overeem S, Fonseca P. A Deep-Learning Approach to Assess Respiratory Effort with a Chest-Worn Accelerometer During Sleep. *Biomedical Signal Processing and Control*. 2023;83:104726. doi:10.1016/j.bspc.2023.104726
- Schipper F, van Sloun RJG, Grassi A, Brouwer J, van Meulen F, Overeem S, Fonseca P. Maximum a Posteriori Detection of Heartbeats from a Chest-Worn Accelerometer. *Physiol Meas*. 2024;45(3). doi:10.1088/1361-6579/ad2f5e
- Schipper F, Grassi A, Ross M, Cerny A, Anderer P, Hermans L, van Meulen F, Leentjens M, Schoustra E, Bosschieter P, Sloun RJG, Overeem S, Fonseca P. Overnight Sleep Staging Using Chest-Worn Accelerometry. *Sensors*. 2024;24(17). doi:10.3390/s24175717
- Schipper F, Fonseca P, Grassi A, Ross M, van Meulen F, van Gilst M, Bosschieter P, Schoustra M, van Sloun RJG, Abdenbi F, de Vries N, Heinzer R, Pépin JL, Overeem S, Apnea-Hypopnea Index Estimation Using Overnight Chest-Wall Accelerometry, in preparation for resubmission
- Fonseca P, Ross M, Cerny A, Anderer P, Schipper F, Grassi A, van Gilst M, Overeem S. Estimating the Severity of Obstructive Sleep Apnea Using ECG, Respiratory Effort and Neural Networks. *IEEE J Biomed Health Inform*. 2024;PP. doi:10.1109/JBHI.2024.3383240

Peer reviewed journal papers on related topics

- Bonomi AG, Schipper F, Eerikäinen LM, et al. Atrial Fibrillation Detection Using a Novel Cardiac Ambulatory Monitor Based on Photo-Plethysmography at the Wrist. *J Am Heart Assoc*. 2018;7(15) 009351. doi:10.1161/JAHA.118.009351
- Eerikäinen LM, Bonomi AG, Schipper F, et al. Comparison between electrocardiogram- and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions. *Physiol Meas*. 2018;39(8):084001. doi:10.1088/1361-6579/aad2c0
- Eerikainen LM, Bonomi AG, Schipper F, et al. Detecting Atrial Fibrillation and Atrial Flutter in Daily Life Using Photoplethysmography Data. *IEEE J Biomed Health Inform*. 2020;24(6):1610-1618. doi:10.1109/JBHI.2019.2950574

Patent applications and patents on the topics in this thesis

- Schipper F, Fonseca P, van Sloun RJG, Overeem S, Method to enhance the detection of periodical physiological phenomena, EP4493049, granted 23-Apr-2025
- Schipper F, Derkx R, van Sloun RJG, Estimation of respiratory effort and respiratory rate from a chest-worn accelerometer, EP4216812A1, granted 2025-Oct-24

- Schipper F, Fonseca P, Brouwer J, System and method for determining a sleeping posture, EP4258974A1, published 16-Jun-2022
- Schipper F, Fonseca P, Overeem S, Detection of arrhythmic episodes in otherwise periodic signals-

Patent applications and patents on related topics

- Schipper F, Dillen P, Roovers D, Dynamic adaptation of the signal to noise ratio and power consumption of a sensor to the requirements of a processing algorithm, US11,986,275, granted 21-May-2024
- Bonomi A, Schipper F, Margarito J, de Morree, H, Method to improve the detection of atrial fibrillation using wearable photoplethysmography and motion sensors, US2021267471A1, published 10-Mar-2023
- Paulussen E, Coppola G, de Lange J, Schipper F, Animal vital sign detection system, EP3160346A1, published 31-Mar-2017
- Roovers D, Huijbregts L, Schipper F, Masculo F, Device for measuring a cycling cadence, US9,936,912, granted 10-Apr-2018
- Schipper F, Improved ambient light cancellation, US9,668,698, granted 6-Jun-2017
- Presura C, van Engen P, Schipper F, Geenen K, Lijten G, Optical sensor with frame protection for heart rate monitoring, US 108,567,138, granted 15-Feb-2022
- Schipper F, van Engen P, Geenen K, Light intensity control in optical heart rate monitor, EP2861131, granted 21-Feb-2018

The TU/e sleep series

1. Maaike Goelema (2018). Perceived sleep quality in a personal health monitoring context.
2. Pedro Fonseca (2018). Home sleep monitoring.
3. Mark van Gastel (2018). Remote photoplethysmography in infrared- towards contactless sleep monitoring.
4. Lieke Hermans (2020). Sleep structure & sleep perception in people with insomnia.
5. Gabriele Papini (2022). Taking the pulse: unobtrusive sleep apnea monitoring using cardiovascular features.
6. Laury Quaedackers (2024). Capturing the narcolepsy symptom spectrum: an mHealth approach.
7. Bernice Wulterkens (2024). Shining light on a dark duo: advancing the analysis of sleep in the assessment of COMISA.
8. Iris Huijben (2024). Uncovering sleep structure through discrete representation learning.
9. Jiali Xie (2025). Assessment of obstructive sleep disordered breathing using cardiorespiratory and audio signals.
10. Jaap van der Aar (2025). Optimizing automated sleep staging.
11. Hans van Gorp (2026). Deep generative modelling in sleep diagnostics.
12. Fons Schipper (2026). Measuring sleep and respiration with chest-wall accelerometry.

References

1. Narkiewicz K, Somers VK. The sympathetic nervous system and obstructive sleep apnea: implications for hypertension. *J Hypertens*. 1997;15(12):1613.
2. Muraki I, Wada H, Tanigawa T. Sleep apnea and type 2 diabetes. *J Diabetes Investig*. 2018;9(5):991-997. doi:10.1111/jdi.12823
3. Yeghiazarians Y, Jneid H, Tietjens JR, et al. Obstructive Sleep Apnea and Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation*. 2021;144(3). doi:10.1161/CIR.0000000000000988
4. Gottlieb DJ, Punjabi NM. Diagnosis and Management of Obstructive Sleep Apnea: A Review. *JAMA*. 2020;323(14):1389-1400. doi:10.1001/jama.2020.3514
5. International Classification of Sleep Disorders, third edition, text revision (ICSD-3-TR). Published online 2023.
6. de Araujo Dantas AB, Gonçalves FM, Martins AA, et al. Worldwide prevalence and associated risk factors of obstructive sleep apnea: a meta-analysis and meta-regression. *Sleep Breath*. 2023;27(6):2083-2109. doi:10.1007/s11325-023-02810-7
7. Holmedahl NH, Fjeldstad OM, Engan H, Saxvig IW, Grønli J. Validation of Peripheral Arterial Tonometry as Tool for Sleep Assessment in Chronic Obstructive Pulmonary Disease. *Sci Rep*. 2019;9(1):19392. doi:10.1038/s41598-019-55958-2
8. Troester MM, Quan SF, Berry RB, et al. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications version 3. Published online 2023.
9. Epstein LJ, Kristo D, Strollo PJ, et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2009;5(3):263-276.
10. Holley AB, Walter RJ. Does a Low Arousal Threshold Equal Low PAP Adherence? *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2018;14(5):713-714. doi:10.5664/jcsm.7088
11. Rotenberg BW, Murariu D, Pang KP. Trends in CPAP adherence over twenty years of data collection: a flattened curve. *J Otolaryngol - Head Neck Surg*. 2016;45(1):43. doi:10.1186/s40463-016-0156-0
12. Gastaut H, Tassinari C, Duron B. Etude polygraphique des manifestations épisodiques (hypniques et respiratoires), diurnes et nocturnes, du syndrome de Pickwick. *Rev Neurol Paris*. Published online 1965.
13. Jung R, Kuhlo W. NEUROPHYSIOLOGICAL STUDIES OF ABNORMAL NIGHT SLEEP AND THE PICKWICKIAN SYNDROME. *Prog Brain Res*. 1965;18:140-159. doi:10.1016/s0079-6123(08)63590-6
14. Cartwright RD. Effect of Sleep Position on Sleep Apnea Severity. *Sleep*. 1984;7(2):110-114. doi:10.1093/sleep/7.2.110
15. Marklund M, Persson M, Franklin KA. Treatment success with a mandibular advancement device is related to supine-dependent sleep apnea. *Chest*. 1998;114(6):1630-1635. doi:10.1378/chest.114.6.1630
16. Mador MJ, Kufel TJ, Magalang UJ, Rajesh SK, Watwe V, Grant BJB. Prevalence of positional sleep apnea in patients undergoing polysomnography. *Chest*. 2005;128(4):2130-2137. doi:10.1378/chest.128.4.2130
17. Bignold JJ, Mercer JD, Antic NA, McEvoy RD, Catcheside PG. Accurate Position Monitoring and Improved Supine-Dependent Obstructive Sleep Apnea with a New Position Recording and

Supine Avoidance Device. *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2011;7(4):376-383. doi:10.5664/JCSM.1194

18. Levendowski DJ, Seagraves S, Popovic D, Westbrook PR. Assessment of a neck-based treatment and monitoring device for positional obstructive sleep apnea. *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2014;10(8):863-871. doi:10.5664/jcsm.3956
19. Frank MH, Ravesloot MJL, van Maanen JP, Verhagen E, de Lange J, de Vries N. Positional OSA part 1: Towards a clinical classification system for position-dependent obstructive sleep apnoea. *Sleep Breath Schlaf Atm*. 2015;19(2):473-480. doi:10.1007/s11325-014-1022-9
20. Oksenberg A, Gadoth N. Are we missing a simple treatment for most adult sleep apnea patients? The avoidance of the supine sleep position. *J Sleep Res*. 2014;23(2):204-210. doi:10.1111/jsr.12097
21. Joosten SA, O'Driscoll DM, Berger PJ, Hamilton GS. Supine position related obstructive sleep apnea in adults: pathogenesis and treatment. *Sleep Med Rev*. 2014;18(1):7-17. doi:10.1016/j.smrv.2013.01.005
22. Eijsvogel Michiel M., Ubbink Rinse, Dekker Janita, et al. Sleep Position Trainer versus Tennis Ball Technique in Positional Obstructive Sleep Apnea Syndrome. *J Clin Sleep Med*. 2015;11(02):139-147. doi:10.5664/jcsm.4460
23. Berry RB, Uhles ML, Abaluck BK, et al. NightBalance Sleep Position Treatment Device Versus Auto-Adjusting Positive Airway Pressure for Treatment of Positional Obstructive Sleep Apnea. *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2019;15(7):947-956. doi:10.5664/jcsm.7868
24. Hung PD, Bonnet S, Guillemaud R, Castelli E, Yen PTN. Estimation of respiratory waveform using an accelerometer. In: *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE; 2008:1493-1496. doi:10.1109/ISBI.2008.4541291
25. Jin A, Yin B, Morren G, Duric H, Aarts RM. Performance evaluation of a tri-axial accelerometry-based respiration monitoring for ambient assisted living. In: *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2009:5677-5680. doi:10.1109/IEMBS.2009.5333116
26. Cesareo A, Previtali Y, Biffi E, Aliverti A. Assessment of Breathing Parameters Using an Inertial Measurement Unit (IMU)-Based System. *Sensors*. 2018;19(1):88. doi:10.3390/s19010088
27. Baevskii RM, Egorov AD, Kazarian LA. [SEISMOCARDIOGRAPHY]. *Kardiologiia*. 1964;4:87-89.
28. Salerno DM, Zanetti J. Seismocardiography for Monitoring Changes in Left Ventricular Function during Ischemia. *CHEST*. 1991;100(4):991-993. doi:10.1378/chest.100.4.991
29. Brüser C, Winter S, Leonhardt S. Robust inter-beat interval estimation in cardiac vibration signals. *Physiol Meas*. 2013;34(2):123-138. doi:10.1088/0967-3334/34/2/123
30. Cohn MA, Rao AS, Broudy M, et al. The respiratory inductive plethysmograph: a new non-invasive monitor of respiration. *Bull Eur Physiopathol Respir*. 1982;18(4):643-658.
31. Guilleminault C, Connolly S, Winkle R, Melvin K, Tilkian A. Cyclical variation of the heart rate in sleep apnoea syndrome. Mechanisms, and usefulness of 24 h electrocardiography as a screening technique. *Lancet Lond Engl*. 1984;1(8369):126-131. doi:10.1016/s0140-6736(84)90062-x
32. Azarbarzin A, Ostrowski M, Moussavi Z, Hanly P, Younes M. Contribution of arousal from sleep to postevent tachycardia in patients with obstructive sleep apnea. *Sleep*. 2013;36(6):881-889. doi:10.5665/sleep.2716
33. Azarbarzin A, Sands SA, Younes M, et al. The Sleep Apnea-Specific Pulse-Rate Response Predicts Cardiovascular Morbidity and Mortality. *Am J Respir Crit Care Med*. 2021;203(12):1546-1555. doi:10.1164/rccm.202010-3900OC

34. Papini GB, Fonseca P, van Gilst MM, et al. Estimation of the apnea-hypopnea index in a heterogeneous sleep-disordered population using optimised cardiovascular features. *Sci Rep*. 2019;9(1):1-16. doi:10.1038/s41598-019-53403-y
35. Iwasaki A, Nakayama C, Fujiwara K, et al. Screening of sleep apnea based on heart rate variability and long short-term memory. *Sleep Breath Schlaf Atm*. 2021;25(4):1821-1829. doi:10.1007/s11325-020-02249-0
36. Fonseca P, Ross M, Cerny A, et al. Estimating the Severity of Obstructive Sleep Apnea Using ECG, Respiratory Effort and Neural Networks. *IEEE J Biomed Health Inform*. 2024;PP. doi:10.1109/JBHI.2024.3383240
37. Royal College of Physicians. National Early Warning Score (NEWS) 2. Published online December 19, 2017. Accessed January 14, 2020. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2
38. Liu H, Allen J, Zheng D, Chen F. Recent development of respiratory rate measurement technologies. *Physiol Meas*. 2019;40(7):07TR01. doi:10.1088/1361-6579/ab299e
39. Berry R, Albertario C, Harding S, et al. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Published online 2018.
40. Fonseca P, Long X, Radha M, Haakma R, Aarts RM, Rolink J. Sleep stage classification with ECG and respiratory effort. *Physiol Meas*. 2015;36(10):2027-2040. doi:10.1088/0967-3334/36/10/2027
41. Vandenbussche NL, Overeem S, van Dijk JP, Simons PJ, Pevernagie DA. Assessment of respiratory effort during sleep: Esophageal pressure versus noninvasive monitoring techniques. *Sleep Med Rev*. 2015;24:28-36. doi:10.1016/j.smrv.2014.12.006
42. Kanna S, Rosenberg W von, Goverdovsky V, Constantinides AG, Mandic DP. Bringing Wearable Sensors into the Classroom: A Participatory Approach [SP Education]. *IEEE Signal Process Mag*. 2018;35(3):110-130. doi:10.1109/MSP.2018.2806418
43. Procházka A, Charvátová H, Vaseghi S, Vyšata O. Machine Learning in Rehabilitation Assessment for Thermal and Heart Rate Data Processing. *IEEE Trans Neural Syst Rehabil Eng*. 2018;26(6):1209-1214. doi:10.1109/TNSRE.2018.2831444
44. Massaroni C, Nicolo A, Lo Presti D, Sacchetti M, Silvestri S, Schena E. Contact-Based Methods for Measuring Respiratory Rate. *Sensors*. 2019;19(4):908. doi:10.3390/s19040908
45. Morillo DS, Ojeda JLR, Foix LFC, Rendon DB, Leon A. Monitoring and Analysis of Cardio Respiratory and Snoring Signals by using an Accelerometer. In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2007:3942-3945. doi:10.1109/IEMBS.2007.4353196
46. Bates A, Ling MJ, Mann J, Arvind DK. Respiratory Rate and Flow Waveform Estimation from Tri-axial Accelerometer Data. In: *2010 International Conference on Body Sensor Networks*. IEEE; 2010:144-150. doi:10.1109/BSN.2010.50
47. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE; 2013:4058-4061. doi:10.1109/EMBC.2013.6610436
48. Ward JPT, Ward J, Leach RM. *The Respiratory System at a Glance*. John Wiley & Sons; 2011.
49. Morillo DS, Ojeda JLR, Foix LFC, Jiménez AL. An Accelerometer-Based Device for Sleep Apnea Screening. *IEEE Trans Inf Technol Biomed*. 2010;14(2):491-499. doi:10.1109/TITB.2009.2027231
50. Dehkordi PKh, Marzencki M, Tavakolian K, Kaminska M, Kaminska B. Validation of respiratory signal derived from suprasternal notch acceleration for sleep apnea detection. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011:3824-3827. doi:10.1109/IEMBS.2011.6090950

51. Liu GZ, Guo YW, Zhu QS, Huang BY, Wang L. Estimation of Respiration Rate from Three-Dimensional Acceleration Data Based on Body Sensor Network. *Telemed E-Health*. 2011;17(9):705-711. doi:10.1089/tmj.2011.0022
52. Pandia K, Inan OT, Kovacs GTA, Giovangrandi L. Extracting respiratory information from seismocardiogram signals acquired on the chest using a miniature accelerometer. *Physiol Meas*. 2012;33(10):1643-1660. doi:10.1088/0967-3334/33/10/1643
53. Fekr AR, Radecka K, Zilic Z. Tidal volume variability and respiration rate estimation using a wearable accelerometer sensor. In: *2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. 2014:1-6. doi:10.1109/MOBIHEALTH.2014.7015894
54. Preejith SP, Jeelani A, Maniyar P, Joseph J, Sivaprakasam M. Accelerometer based system for continuous respiratory rate monitoring. In: *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2017:171-176. doi:10.1109/MeMeA.2017.7985870
55. Karhunen J, Joutsensalo J. Generalizations of principal component analysis, optimization problems, and neural networks. *Neural Netw*. 1995;8(4):549-562. doi:10.1016/0893-6080(94)00098-7
56. Leung RST, Bowman ME, Parker JD, Newton GE, Bradley TD. Avoidance of the left lateral decubitus position during sleep in patients with heart failure: relationship to cardiac size and function. *J Am Coll Cardiol*. 2003;41(2):227-230. doi:10.1016/S0735-1097(02)02717-1
57. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013:4058-4061. doi:10.1109/EMBC.2013.6610436
58. Morillo DS, Ojeda JLR, Foix LFC, Jiménez AL. An Accelerometer-Based Device for Sleep Apnea Screening. *IEEE Trans Inf Technol Biomed*. 2010;14(2):491-499. doi:10.1109/TITB.2009.2027231
59. de Vries H, Jonkman A, Shi ZH, Spoelstra-de Man A, Heunks L. Assessing breathing effort in mechanical ventilation: physiology and clinical implications. *Ann Transl Med*. 2018;6(19):387. doi:10.21037/atm.2018.05.53
60. Drummond GB, Fischer D, Lees M, Bates A, Mann J, Arvind DK. Classifying signals from a wearable accelerometer device to measure respiratory rate. *ERJ Open Res*. 2021;7(2). doi:10.1183/23120541.00681-2020
61. Schipper F, Sloun RJG van, Grassi A, Derkx R, Overeem S, Fonseca P. Estimation of respiratory rate and effort from a chest-worn accelerometer using constrained and recursive principal component analysis. *Physiol Meas*. 2021;42(4):045004. doi:10.1088/1361-6579/abf01f
62. Long X, Foussier J, Fonseca P, Haakma R, Aarts RM. Analyzing respiratory effort amplitude for automated sleep stage classification. *Biomed Signal Process Control*. 2014;14:197-205. doi:10.1016/j.bspc.2014.08.001
63. Bakker JP, Ross M, Vasko R, et al. Estimating sleep stages using cardiorespiratory signals: validation of a novel algorithm across a wide range of sleep-disordered breathing severity. *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2021;17(7):1343-1354. doi:10.5664/jcsm.9192
64. Zhang Z, Zheng J, Wu H, Wang W, Wang B, Liu H. Development of a Respiratory Inductive Plethysmography Module Supporting Multiple Sensors for Wearable Systems. *Sensors*. 2012;12(10):13167-13184. doi:10.3390/s121013167
65. Coronel C, Wiesmeyr C, Garn H, et al. Measurement of respiratory effort in sleep by 3D camera and respiratory inductance plethysmography. *Somnologie*. 2019;23(2):86-92. doi:10.1007/s11818-019-0203-y

66. Reinvuo T, Hannula M, Sorvoja H, Alasaarela E, Myllyla R. Measurement of respiratory rate with high-resolution accelerometer and emfit pressure sensor. In: *Proceedings of the 2006 IEEE Sensors Applications Symposium, 2006.* 2006:192-195. doi:10.1109/SAS.2006.1634270
67. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs*. Published online January 29, 2017. Accessed June 19, 2021. http://arxiv.org/abs/1412.6980
68. Gilst MM van, Dijk JP van, Krijn R, et al. Protocol of the SOMNIA project: an observational study to create a neurophysiological database for advanced clinical sleep monitoring. *BMJ Open*. 2019;9(11):e030996. doi:10.1136/bmjopen-2019-030996
69. Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1D convolutional neural networks and applications: A survey. *Mech Syst Signal Process*. 2021;151:107398. doi:10.1016/j.ymssp.2020.107398
70. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2(5):359-366. doi:10.1016/0893-6080(89)90020-8
71. Kondor R, Trivedi S. On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups. *ArXiv180203690 Cs Stat*. Published online November 10, 2018. Accessed December 16, 2021. http://arxiv.org/abs/1802.03690
72. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv150203167 Cs*. Published online March 2, 2015. Accessed June 12, 2021. http://arxiv.org/abs/1502.03167
73. He X, Zhao K, Chu X. AutoML: A Survey of the State-of-the-Art. *Knowl-Based Syst*. 2021;212:106622. doi:10.1016/j.knosys.2020.106622
74. Clifford GD, Azuaje F, McSharry PE. Advanced Methods and Tools for ECG Data Analysis. Published online 2006.
75. Chen J, Zhang D, Wu Z, Zhou F, Sun Q, Chen Y. Contactless Electrocardiogram Monitoring with Millimeter Wave Radar. *IEEE Trans Mob Comput*. Published online 2022:1-17. doi:10.1109/TMC.2022.3214721
76. de Haan G, Jeanne V. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Trans Biomed Eng*. 2013;60(10):2878-2886. doi:10.1109/TBME.2013.2266196
77. Hoog Antink C, Mai Y, Peltokangas M, Leonhardt S, Oksala N, Vehkaoja A. Accuracy of heart rate variability estimated with reflective wrist-PPG in elderly vascular patients. *Sci Rep*. 2021;11(1):8123. doi:10.1038/s41598-021-87489-0
78. Agress CM, Fields LG. New method for analyzing heart vibrations. *Am J Cardiol*. 1959;4(2):184-190. doi:10.1016/0002-9149(59)90246-2
79. Taebi A, Solar BE, Bomar AJ, Sandler RH, Mansy HA. Recent Advances in Seismocardiography. *Vibration*. 2019;2(1):64-86. doi:10.3390/vibration2010005
80. Rai D, Thakkar HK, Rajput SS, Santamaria J, Bhatt C, Roca F. A Comprehensive Review on Seismocardiogram: Current Advancements on Acquisition, Annotation, and Applications. *Mathematics*. 2021;9(18):2243. doi:10.3390/math9182243
81. Khosrow-Khavar F, Tavakolian K, Blaber A, Menon C. Automatic and Robust Delineation of the Fiducial Points of the Seismocardiogram Signal for Noninvasive Estimation of Cardiac Time Intervals. *IEEE Trans Biomed Eng*. 2017;64(8):1701-1710. doi:10.1109/TBME.2016.2616382
82. Dehkordi P, Khosrow-Khavar F, Di Rienzo M, et al. Comparison of Different Methods for Estimating Cardiac Timings: A Comprehensive Multimodal Echocardiography Investigation. *Front Physiol*. 2019;10:1057. doi:10.3389/fphys.2019.01057
83. Wahlström J, Skog I, Händel P, et al. A Hidden Markov Model for Seismocardiography. *IEEE Trans Biomed Eng*. 2017;64(10):2361-2372. doi:10.1109/TBME.2017.2648741

84. Cocconcelli F, Mora N, Matrella G, Ciampolini P. High-Accuracy, Unsupervised Annotation of Seismocardiogram Traces for Heart Rate Monitoring. *IEEE Trans Instrum Meas*. 2020;PP:1-1. doi:10.1109/TIM.2020.2967135
85. Centracchio J, Parlato S, Esposito D, Bifulco P, Andreozzi E. ECG-Free Heartbeat Detection in Seismocardiography Signals via Template Matching. *Sensors*. 2023;23(10):4684. doi:10.3390/s23104684
86. Parlato S, Centracchio J, Esposito D, Bifulco P, Andreozzi E. Heartbeat Detection in Gyrocardiography Signals without Concurrent ECG Tracings. *Sensors*. 2023;23(13):6200. doi:10.3390/s23136200
87. Milena Č, Romano C, De Tommasi F, et al. Linear and Non-Linear Heart Rate Variability Indexes from Heart-Induced Mechanical Signals Recorded with a Skin-Interfaced IMU. *Sensors*. 2023;23(3):1615. doi:10.3390/s23031615
88. Bellman R. On the Theory of Dynamic Programming. *Proc Natl Acad Sci U S A*. 1952;38(8):716-719.
89. Kathirvel P, Sabarimalai Manikandan M, Prasanna SRM, Soman KP. An Efficient R-peak Detection Based on New Nonlinear Transformation and First-Order Gaussian Differentiator. *Cardiovasc Eng Technol*. 2011;2(4):408-425. doi:10.1007/s13239-011-0065-3
90. Association for the Advancement of Medical Instrumentation. Cardiac monitors, heart rate meters, and alarms. *Am Natl Stand ANSIAAMI EC13 2002 Arlingt VA*. Published online 2002:1-87.
91. D'Mello Y, Skoric J, Xu S, et al. Autocorrelated Differential Algorithm for Real-Time Seismocardiography Analysis. *IEEE Sens J*. 2019;19(13):5127-5140. doi:10.1109/JSEN.2019.2903449
92. Gardner M, Randhawa S, Malouf G, Reynolds K. A Wearable Ballistocardiography Device for Estimating Heart Rate During Positive Airway Pressure Therapy: Investigational Study Among the General Population. *JMIR Cardio*. 2021;5(1):e26259. doi:10.2196/26259
93. Jacobs F, Scheerhoorn J, Mestrom E, Stam J van der, Bouwman RA, Nienhuijs S. Reliability of heart rate and respiration rate measurements with a wireless accelerometer in postbariatric recovery. *PLOS ONE*. 2021;16(4):e0247903. doi:10.1371/journal.pone.0247903
94. Linz B, Norup Hertel J, Hendriks J, et al. Sleep apnea and atrial fibrillation: challenges in clinical and translational research. *Expert Rev Cardiovasc Ther*. 2022;20(2):101-109. doi:10.1080/14779072.2022.2036606
95. Schipper F, JG van Sloun R, Grassi A, Overeem S, Fonseca P. A Deep-Learning Approach to Assess Respiratory Effort with a Chest-Worn Accelerometer During Sleep. *Biomed Signal Process Control*. 2023;83:104726. doi:10.1016/j.bspc.2023.104726
96. Xie J, Fonseca P, van Dijk JP, Long X, Overeem S. The Use of Respiratory Effort Improves an ECG-Based Deep Learning Algorithm to Assess Sleep-Disordered Breathing. *Diagnostics*. 2023;13(13):2146. doi:10.3390/diagnostics13132146
97. Bianchi MT, Goparaju B. Potential Underestimation of Sleep Apnea Severity by At-Home Kits: Rescoring In-Laboratory Polysomnography Without Sleep Staging. *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2017;13(4):551-555. doi:10.5664/jcsm.6540
98. Hofmann E. Insomnie-Screening mit dem Home Sleep Test. *Nervenheilkunde*. 2022;41(10):666-672. doi:10.1055/a-1826-7603
99. Tăutan AM, Rossi AC, de Francisco R, Ionescu B. Automatic Sleep Stage Detection: A Study on the Influence of Various PSG Input Signals. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2020:5330-5334. doi:10.1109/EMBC44109.2020.9175628

100. Stein PK, Pu Y. Heart rate variability, sleep and sleep disorders. *Sleep Med Rev*. 2012;16(1):47-66. doi:10.1016/j.smrv.2011.02.005
101. Penzel T, Kantelhardt JW, Lo CC, Voigt K, Vogelmeier C. Dynamics of Heart Rate and Sleep Stages in Normals and Patients with Sleep Apnea. *Neuropsychopharmacology*. 2003;28(1):S48-S53. doi:10.1038/sj.npp.1300146
102. Bonnet M, Arand D. Heart rate variability: Sleep stage, time of night, and arousal influences. *Electroencephalogr Clin Neurophysiol*. 1997;102:390-396. doi:10.1016/S0921-884X(96)96070-1
103. Gutierrez G, Williams J, Alrehaili GA, et al. Respiratory rate variability in sleeping adults without obstructive sleep apnea. *Physiol Rep*. 2016;4(17):e12949. doi:10.14814/phy2.12949
104. Hermans LWA, Huijben IAM, van Gorp H, et al. Representations of temporal sleep dynamics: Review and synthesis of the literature. *Sleep Med Rev*. 2022;63:101611. doi:10.1016/j.smrv.2022.101611
105. Ebrahimi F, Alizadeh I. Automatic sleep staging by cardiorespiratory signals: a systematic review. *Sleep Breath*. 2022;26(2):965-981. doi:10.1007/s11325-021-02435-8
106. Imtiaz SA. A Systematic Review of Sensing Technologies for Wearable Sleep Staging. *Sensors*. 2021;21(5):1562. doi:10.3390/s21051562
107. De Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable Sleep Technology in Clinical and Research Settings. *Med Sci Sports Exerc*. 2019;51(7):1538. doi:10.1249/MSS.0000000000001947
108. Depner CM, Cheng PC, Devine JK, et al. Wearable Technologies for Developing Sleep and Circadian Biomarkers: A Summary of Workshop Discussions. *Sleep*. Published online October 23, 2019:zsz254. doi:10.1093/sleep/zsz254
109. Wulterkens BM, Fonseca P, Hermans LWA, et al. It is All in the Wrist: Wearable Sleep Staging in a Clinical Population versus Reference Polysomnography. *Nat Sci Sleep*. 2021;13:885. doi:10.2147/NSS.S306808
110. Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armoundas AA. Wearable Devices for Ambulatory Cardiac Monitoring. *J Am Coll Cardiol*. 2020;75(13):1582-1592. doi:10.1016/j.jacc.2020.01.046
111. Schipper F, van Sloun RJG, Grassi A, et al. Maximum a Posteriori Detection of Heartbeats from a Chest-Worn Accelerometer. *Physiol Meas*. 2024;45(3). doi:10.1088/1361-6579/ad2f5e
112. Punjabi NM, Shifa N, Dorffner G, Patil S, Pien G, Aurora RN. Computer-Assisted Automated Scoring of Polysomnograms Using the Somnolyzer System. *Sleep*. 2015;38(10):1555-1566. doi:10.5665/sleep.5046
113. Bakker JP, Ross M, Cerny A, et al. Scoring Sleep with Artificial Intelligence Enables Quantification of Sleep Stage Ambiguity: Hypnodensity Based on Multiple Expert Scorers and Auto-Scoring. *Sleep*. 2023;46(2):zsac154. doi:10.1093/sleep/zsac154
114. Anderer P, Ross M, Cerny A, Shaw E. Automated Scoring of Sleep and Associated Events. In: Penzel T, Hornero R, eds. *Advances in the Diagnosis and Treatment of Sleep Apnea*. Vol 1384. Advances in Experimental Medicine and Biology. Springer International Publishing; 2022:107-130. doi:10.1007/978-3-031-06413-5_7
115. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica*. 2012;22(3):276-282.
116. Lee YJ, Lee JY, Cho JH, Choi JH. Interrater reliability of sleep stage scoring: a meta-analysis. *J Clin Sleep Med*. 18(1):193-202. doi:10.5664/jcsm.9538
117. Celmer L. AASM announces first recipient of autoscoring software certification. American Academy of Sleep Medicine – Association for Sleep Clinicians and Researchers. March 8, 2024. Accessed May 31, 2024. https://aasm.org/american-academy-of-sleep-medicine-announces-first-recipient-of-autoscoring-software-certification/

118. Gao C, Li P, Morris CJ, et al. Actigraphy-Based Sleep Detection: Validation with Polysomnography and Comparison of Performance for Nighttime and Daytime Sleep During Simulated Shift Work. *Nat Sci Sleep*. 2022;14:1801-1816. doi:10.2147/NSS.S373107
119. Kortelainen JM, Mendez MO, Bianchi AM, Matteucci M, Cerutti S. Sleep Staging Based on Signals Acquired Through Bed Sensor. *IEEE Trans Inf Technol Biomed*. 2010;14(3):776-785. doi:10.1109/TITB.2010.2044797
120. Migliorini M, Bianchi AM, Nisticò D, et al. Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors. *Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf*. 2010;2010:3273-3276. doi:10.1109/IEMBS.2010.5627217
121. Wang Z, Zha S, Yu B, Chen P, Pang Z, Zhang H. Sleep Staging Using Noncontact-Measured Vital Signs. *J Healthc Eng*. 2022;2022:e2016598. doi:10.1155/2022/2016598
122. Yi R, Enayati M, Keller JM, Popescu M, Skubic M. Non-Invasive In-Home Sleep Stage Classification Using a Ballistocardiography Bed Sensor. In: *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. 2019:1-4. doi:10.1109/BHI.2019.8834535
123. Voss A, Schroeder R, Heitmann A, Peters A, Perz S. Short-term heart rate variability--influence of gender and age in healthy subjects. *PloS One*. 2015;10(3):e0118308. doi:10.1371/journal.pone.0118308
124. Moula AI, Parrini I, Tetta C, et al. Obstructive Sleep Apnea and Atrial Fibrillation. *J Clin Med*. 2022;11(5):1242. doi:10.3390/jcm11051242
125. Kendzerska T, Mollayeva T, Gershon AS, Leung RS, Hawker G, Tomlinson G. Untreated Obstructive Sleep Apnea and the Risk for Serious Long-Term Adverse Outcomes: A Systematic Review. *Sleep Med Rev*. 2014;18(1):49-59. doi:10.1016/j.smrv.2013.01.003
126. Morsy NE, Farrag NS, Zaki NFW, et al. Obstructive Sleep Apnea: Personal, Societal, Public Health, and Legal Implications. *Rev Environ Health*. 2019;34(2):153-169. doi:10.1515/reveh-2018-0068
127. Pevernagie DA, Gnidovec-Strazisar B, Grote L, et al. On the Rise and Fall of the Apnea-Hypopnea Index: A Historical Review and Critical Appraisal. *J Sleep Res*. 2020;29(4):e13066. doi:10.1111/jsr.13066
128. Iftikhar IH, BaHammam A, Jahrami H, Ioachimescu O. Accuracy of Residual Respiratory Event Detection by CPAPs: A Meta-Analysis. *Sleep Breath*. Published online January 30, 2023. doi:10.1007/s11325-023-02780-w
129. Oksenberg A, Gadoth N, Töyräs J, Leppänen T. Prevalence and Characteristics of Positional Obstructive Sleep Apnea (POSA) in Patients with Severe OSA. *Sleep Breath*. 2020;24(2):551-559. doi:10.1007/s11325-019-01897-1
130. Heinzer R, Petitpierre NJ, Marti-Soler H, Haba-Rubio J. Prevalence and Characteristics of Positional Sleep Apnea in the Hypnolaus Population-Based Cohort. *Sleep Med*. 2018;48:157-162. doi:10.1016/j.sleep.2018.02.011
131. Ravesloot MJL. Positional Treatment of Obstructive Sleep Apnea. *Otolaryngol Clin North Am*. 2024;57(3):481-490. doi:10.1016/j.otc.2024.01.002
132. Smith MT, McCrae CS, Cheung J, et al. Use of Actigraphy for the Evaluation of Sleep Disorders and Circadian Rhythm Sleep-Wake Disorders: An American Academy of Sleep Medicine Clinical Practice Guideline. *J Clin Sleep Med*. 2018;14(07):1231-1237. doi:10.5664/jcsm.7230
133. Ryser F, Hanassab S, Lambercy O, Werth E, Gassert R. Respiratory Analysis During Sleep Using a Chest-Worn Accelerometer: A Machine Learning Approach. *Biomed Signal Process Control*. 2022;78:104014. doi:10.1016/j.bspc.2022.104014
134. Schipper F, Grassi A, Ross M, et al. Overnight Sleep Staging Using Chest-Worn Accelerometry. *Sensors*. 2024;24(17). doi:10.3390/s24175717

135. Bakker JP, Ross M, Vasko R, et al. Estimating Sleep Stages Using Cardiorespiratory Signals: Validation of a Novel Algorithm Across a Wide Range of Sleep-Disordered Breathing Severity. *J Clin Sleep Med*. 2021;17(7):1343-1354. doi:10.5664/jcsm.9192
136. Gliner V, Makarov V, Avetisyan AI, Schuster A, Yaniv Y. Using domain adaptation for classification of healthy and disease conditions from mobile-captured images of standard 12-lead electrocardiograms. *Sci Rep*. 2023;13(1):14023. doi:10.1038/s41598-023-40693-6
137. Davies HJ, Zylinski M, Bermond M, Liu Z, Khaleghimeybodi M, Mandic DP. Feasibility of Transfer Learning from Finger PPG to In-Ear PPG. *Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf*. 2023;2023:1-4. doi:10.1109/EMBC40787.2023.10340172
138. Waters SH, Clifford GD. Comparison of deep transfer learning algorithms and transferability measures for wearable sleep staging. *Biomed Eng Online*. 2022;21(1):66. doi:10.1186/s12938-022-01033-3
139. Van Pee B, Massie F, Vits S, et al. A Multicentric Validation Study of a Novel Home Sleep Apnea Test Based on Peripheral Arterial Tonometry. *Sleep*. 2022;45(5):zsac028. doi:10.1093/sleep/zsac028
140. Martinot JB, Pépin JL, Malhotra A, Le-Dong NN. Near-Boundary Double-Labeling-Based Classification: The New Standard When Evaluating Performances of New Sleep Apnea Diagnostic Solutions Against Polysomnography? *Sleep*. 2022;45(10):zsac188. doi:10.1093/sleep/zsac188
141. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159-174.
142. Nigro CA, González S, Arce A, Aragone MR, Nigro L. Accuracy of a Novel Auto-CPAP Device to Evaluate the Residual Apnea-Hypopnea Index in Patients with Obstructive Sleep Apnea. *Sleep Breath*. 2015;19(2):569-578. doi:10.1007/s11325-014-1048-z
143. Baek JH, Jeon JY, Lee SA. Accuracy of the Auto Scoring by the S9 CPAP in Patients with Obstructive Sleep Apnea. *Sleep Med Res*. 2016;7(1):26-32. doi:10.17241/smr.2016.00059
144. Berry RB, Kushida CA, Kryger MH, Soto-Calderon H, Staley B, Kuna ST. Respiratory Event Detection by a Positive Airway Pressure Device. *Sleep*. 2012;35(3):361-367. doi:10.5665/sleep.1696
145. Gagnadoux F, Pevernagie D, Jennum P, et al. Validation of the System One RemStar Auto A-Flex for Obstructive Sleep Apnea Treatment and Detection of Residual Apnea-Hypopnea Index: A European Randomized Trial. *J Clin Sleep Med*. 2017;13(02):283-290. doi:10.5664/jcsm.6464
146. Fanfulla F, D'Artavilla Lupo N, Malovini A, et al. Reliability of Automatic Detection of AHI During Positive Airway Pressure Treatment in Obstructive Sleep Apnea Patients: A "Real-Life Study." *Respir Med*. 2021;177:106303. doi:10.1016/j.rmed.2021.106303
147. Iftikhar IH, Finch CE, Shah AS, Augunstein CA, Ioachimescu OC. A Meta-Analysis of Diagnostic Test Performance of Peripheral Arterial Tonometry Studies. *J Clin Sleep Med*. 2022;18(4):1093-1102. doi:10.5664/jcsm.9808
148. Pittman SD, Ayas NT, MacDonald MM, Malhotra A, Fogel RB, White DP. Using a Wrist-Worn Device Based on Peripheral Arterial Tonometry to Diagnose Obstructive Sleep Apnea: In-Laboratory and Ambulatory Validation. *Sleep*. 2004;27(5):923-933. doi:10.1093/sleep/27.5.923
149. Yalamanchali S, Farajian V, Hamilton C, Pott TR, Samuelson CG, Friedman M. Diagnosis of Obstructive Sleep Apnea by Peripheral Arterial Tonometry: Meta-Analysis. *JAMA Otolaryngol-- Head Neck Surg*. 2013;139(12):1343-1350. doi:10.1001/jamaoto.2013.5338
150. Janbakhshi P, Shamsollahi MB. Sleep Apnea Detection from Single-Lead ECG Using Features Based on ECG-Derived Respiration (EDR) Signals. *IRBM*. 2018;39(3):206-218. doi:10.1016/j.irbm.2018.03.002

151. Azarbarzin A, Sands SA, Taranto-Montemurro L, et al. The Sleep Apnea-Specific Hypoxic Burden Predicts Incident Heart Failure. *Chest*. 2020;158(2):739-750. doi:10.1016/j.chest.2020.03.053
152. Linz D, McEvoy RD, Cowie MR, et al. Associations of Obstructive Sleep Apnea with Atrial Fibrillation and Continuous Positive Airway Pressure Treatment: A Review. *JAMA Cardiol*. 2018;3(6):532-540. doi:10.1001/jamacardio.2018.0095
153. Rimler R. AASM OK with PAT Technology for Home Sleep Studies. Sleep Review. April 27, 2017. Accessed May 4, 2025. https://sleepreviewmag.com/sleep-health/prevailing-attitudes/academies-associations/aasm-ok-pat-technology-home-sleep-studies/
154. Hedner J, White DP, Malhotra A, et al. Sleep Staging Based on Autonomic Signals: A Multi-Center Validation Study. *J Clin Sleep Med*. 2011;07(03):301-306. doi:10.5664/JCSM.1078
155. Hermans L, van Meulen F, Anderer P, et al. Performance of cardiorespiratory-based sleep staging in patients using beta blockers. *J Clin Sleep Med JCSM Off Publ Am Acad Sleep Med*. 2024;20(4):575-581. doi:10.5664/jcsm.10938
156. Benjafield AV, Pepin JL, Cistulli PA, et al. Positive airway pressure therapy and all-cause and cardiovascular mortality in people with obstructive sleep apnoea: a systematic review and meta-analysis of randomised controlled trials and confounder-adjusted, non-randomised controlled studies. *Lancet Respir Med*. 2025;13(5):403-413. doi:10.1016/S2213-2600(25)00002-5
157. Yeghiazarians Y, Jneid H, Tietjens JR, et al. Obstructive Sleep Apnea and Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation*. 2021;144(3). doi:10.1161/CIR.0000000000000988
158. Norlock V, Vazquez R, Dunn A, Siegfried C, Wadhwa M, Medic G. Comparing the outcomes and costs of cardiac monitoring with implantable loop recorders and mobile cardiac outpatient telemetry following stroke using real-world evidence. *J Comp Eff Res*. 2024;13(6):e240008. doi:10.57264/cer-2024-0008
159. Krenk L, Jennum P, Kehlet H. Sleep disturbances after fast-track hip and knee arthroplasty. *Br J Anaesth*. 2012;109(5):769-775. doi:10.1093/bja/aes252
160. Cremeans-Smith JK, Millington K, Sledjeski E, Greene K, Delahanty DL. Sleep disruptions mediate the relationship between early postoperative pain and later functioning following total knee replacement surgery. *J Behav Med*. 2006;29(2):215-222. doi:10.1007/s10865-005-9045-0
161. Xie J, Fonseca P, van Dijk J, Overeem S, Long X. Assessment of obstructive sleep apnea severity using audio-based snoring features. *Biomed Signal Process Control*. 2023;86:104942. doi:10.1016/j.bspc.2023.104942
162. Stephansen JB, Olesen AN, Olsen M, et al. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nat Commun*. 2018;9(1):5229. doi:10.1038/s41467-018-07229-3
163. Sandhu RK, Healey JS. Screening for undiagnosed atrial fibrillation. *Expert Rev Cardiovasc Ther*. Published online August 3, 2018. Accessed May 19, 2025. https://www.tandfonline.com/doi/full/10.1080/14779072.2018.1496018
164. Moody GB, Mark RG. A new method for detecting atrial fibrillation using RR intervals. *Proc Comput Cardiol*. 1983;10:227-230.
165. Bonomi AG, Schipper F, Eerikäinen LM, et al. Atrial Fibrillation Detection Using a Novel Cardiac Ambulatory Monitor Based on Photo-Plethysmography at the Wrist. *J Am Heart Assoc*. 2018;7(15):e009351. doi:10.1161/JAHA.118.009351

Measuring Sleep and Respiration with Chest-Wall Accelerometry

Obstructive Sleep Apnea (OSA) can severely impact daily wellbeing and long-term health. Positive airway pressure -such as CPAP- remains the gold-standard OSA treatment, and devices routinely provide an estimation of the apnea-hypopnea index (AHI) as an indicator of treatment efficacy. However, this critical feedback loop is notably absent in other OSA treatments, such as positional therapy. Devices for positional therapy use chest-worn accelerometers to detect body position, and subsequently correct supine sleeping.

The work in this thesis advances the field of sleep monitoring by introducing novel ways to use chest-worn accelerometry for clinical-grade assessment. Advanced methods are presented to derive respiratory effort and instantaneous heart rate solely from the accelerometer signal. By tuning neural networks trained on large datasets, accurate sleep staging and AHI estimation algorithms were developed. This approach enables therapeutic devices to monitor residual OSA and thus increase clinical confidence in treatment outcome.

