Rhythm Modelling of Long-Term Activity Data

A dissertation approved of
TECHNISCHE UNIVERSITÄT DARMSTADT
Fachbereich Informatik

for the degree of
Doktor-Ingenieur (Dr.-Ing.)

presented by

Marko Antonio Matteo Borazio
Dipl.-Inform.
born in Frankfurt/M, Germany

Prof. Dr. Kristof Van Laerhoven, examiner
Prof. Dr. Antonio Krüger, co-examiner

Date of Submission: 18th of September, 2014
Date of Defense: 24th of November, 2014

Darmstadt, 2014
D17

ABSTRACT

Long-term monitoring for activity recognition opens up new possibilities for deriving characteristics from the data, such as daily activity rhythms and certain quality measures for the activity performed or for identifying similarities or differences in daily routines. This thesis investigates the detection of activities with wearable sensors and addresses two major challenges in particular: The *modelling* of a person's behaviour into rhythmic patterns and the *detection* of high-level activities, e.g., having lunch or sleeping. To meet these challenges, this thesis makes the following contributions:

First, we study different *platforms* that are suitable for long-term data recording: A wrist-worn sensor and mobile phones. The latter has shown different carrying behaviours for various users. This has to be considered in ubiquitous systems for accurately recognizing the user's context. We evaluate our findings in a study with a wrist-worn accelerometer by correlating with the inertial data of a smart phone.

Second, we investigate datasets that exhibit *rhythmic patterns* to be used for recognizing *high-level activities*. Such statistical information obtained over a population is collected with time use surveys which describe how often certain activities are performed by the user. From such datasets we extract *features* like time and location to describe which activities are detectable by making use of prior information, showing also the benefits and limits of such data.

Third, in order to *improve* on the recognition rates of high-level activities from wearable sensor data only, we propose the use of the aforementioned prior information from time use data. In our approach we investigate the results of a common classifier for several high-level activities, after which we compare them to the outcome of a maximum-likelihood estimation on the time use survey data. In a last step, we show how these two classification approaches are fused to raise the recognition rates.

In a fourth contribution we introduce a recording platform to *capture* sleep and sleep behaviour in the user's common environment, enabling the unobtrusive monitoring of patterns over several weeks. We use a wrist-worn sensor to record inertial data from which we extract sleep segments. For this purpose, we present three different *sleep detection approaches*: A Gaussian-, generative model- and stationary segments-based algorithm are evaluated and are found to exhibit different accuracies for detecting sleep. The latter algorithm is pitted against two clinically evaluated sleep detection approaches, indicating that we are able to reach an optimum trade-off between sleep and wake segments, while the two common algorithms tend to overestimate sleep. Further, we investigate the rhythmic patterns within sleep: We classify sleep postures and detect muscle contractions with a high confidence, enabling physicians to efficiently browse through the data.

ZUSAMMENFASSUNG

Langzeitaufnahmen von Sensordaten zur Aktivitätserkennung ermöglichen die Detektion von Merkmalen innerhalb dieser Daten, wie z.B. Rhythmen und Qualitätsmuster für die durchgeführten Aktivitäten. Diese Informationen können zur Bestimmung von ähnlichen oder unterschiedlichen Alltagsroutinen genutzt werden. Diese Arbeit untersucht die Erkennung von physischen Aktivitäten mit Hilfe von tragbaren Sensoren und konzentriert sich dabei auf zwei wichtige Herausforderungen: Das *Modellieren* von rhythmischen Mustern abgeleitet aus dem Verhalten der Personen und die *Detektion* von komplexen Aktivitäten, wie z.B. 'zu Mittag essen' oder 'schlafen'. Um diese Herausforderungen anzugehen, leistet diese Arbeit folgende Beiträge:

Zunächst untersuchen wir die Nutzung verschiedener *Plattformen* zur Langzeitdatenerfassung: Einen Handgelenksensor sowie Mobiltelefone. In Bezug auf Mobiltelefone gibt es verschiedene Trageverhalten, welche zur Kontexterkennung in ubiquitären Systemen berücksichtigt werden müssen. Wir evaluieren das Trageverhalten mittels der Korrelation von Beschleunigungsdaten des Handgelenksensors mit den Bewegungsdaten des Mobiltelefon-Sensors.

Zweitens untersuchen wir Datensätze, die wiederkehrende Aktivitätsmuster aufweisen, um *komplexe Aktivitäten* zu erkennen. Solche statistischen Daten beziehen wir aus Zeitbudgeterhebungen die beschreiben, welche Aktivitäten zu bestimmten Zeiten von Einwohnern eines Landes ausgeführt werden. Aus diesen Datensätzen extrahieren wir *Features* wie Zeit und Ort, um aufzuzeigen welche Aktivitäten mit Hilfe dieser Informationen erkannt werden können. Weiterhin untersuchen wir die Vorteile und die Grenzen der Nutzung solcher Daten.

Drittens verwenden wir diese statistischen Informationen zur *Verbesserung der Erkennungsraten* von komplexen Aktivitäten. Unsere Vorgehensweise besteht darin, die Ergebnisse der Erkennung von komplexen Aktivitäten eines üblicherweise verwendeten Klassifizierers mit denen aus der Bestimmung des Maximum-Likelihood von Aktivitäten der Zeitbudgetdaten zu vergleichen. Daraufhin vereinen wir die Ergebnisse beider Vorgehensweisen und zeigen wie die Erkennungsraten dadurch verbessert werden.

Der vierte Beitrag führt ein Aufnahmesystem ein, welches das Überwachen des Schlafes und des Schlafverhaltens über mehrere Wochen in der gewohnten Umgebung des Benutzers ermöglicht. Wir verwenden dabei einen Handgelenksensor zur Erfassung von Bewegungsdaten, in denen wir Schlafsegmente detektieren. Wir stellen drei verschiedene *Schlafdetektionsalgorithmen* vor: Eine Gauss-, eine auf einem Generativen Modell und eine auf stationären Segmenten basierende Vorgehensweise, die verschiedene Genauigkeiten in der Erkennung des Schlafes aufweisen. Der letztere Algorithmus wird mit zwei medizinisch evaluierten Algorithmen zur Schlafdetektion verglichen. Unsere Ergebnisse zeigen, dass Schlaf- und Wach-Segmente

mit unserem neuartigen Algorithmus gleich gut erkannt werden können, während die zwei üblicherweise eingesetzten Algorithmen die Schlafdauer überschätzen. Weiterhin zeigen wir, wie rhythmische Muster innerhalb des Schlafes detektiert werden können: Schlafpositionen und Muskelzuckungen werden mit einer hohen Genauigkeit erkannt, wodurch Ärzten die Möglichkeit geboten wird, die Schlafdaten effizient zu begutachten.

ACKNOWLEDGEMENTS

"You have to work faster" were the words my 2 year old son spoke when I told him that I had to work late again. Of course I tried to be fast, especially when writing this thesis but I guess in the end it was a good thing I took my time. Research has, is and will be my passion - one way or the other.

First, I am most grateful to my advisor, Prof. Dr. Kristof Van Laerhoven, for taking me in as a PhD student. Thank you for your constant advice and guidance throughout the past years. I am thankful for you always pushing me to the limit, for motivating me even though there were times when nothing seemed to work out (especially before paper deadlines). I could not have imagined a better advisor and mentor for my doctoral studies. Furthermore, I would like to thank Prof. Antonio Krüger for agreeing to be the co-examiner of this doctoral thesis.

My special thanks go out to my current and former colleagues, who accompanied me during this thesis adventure and engaged me in interesting talks and ideas. Thank you Eugen Berlin with whom I shared an office with for quite some time and who always provided me with additional ideas in our numerous discussions. The same gratitude I would like to express to Philipp M. Scholl who managed to shed light on parts of my thesis I was not aware of. Additional thanks go out to Agha Muhammad, Manuel Dietrich, Christian Seeger, Pablo Guerrero, Paul Schnitzspan, Tâm Huynh, Michael Stark, Mykhaylo Andriluka, Christian Wojek, Stefan Walk, Anton Andriyenko, Tobias Grosse-Puppendahl and, last but not least, Ulf Blanke. I will miss our long discussions and short trips to conferences all over the world. My sincere thanks go to Ursula Paeckel for her constant advice during my stay at the TU Darmstadt. She managed to keep me on track regarding paperwork or planning this thesis - I will really miss our coffee breaks with all of you guys!

Additionally, I am glad that I had the opportunity to guide students in conducting their Bachelor-, Master- and Diploma-thesis. Thank you for your contribution to this project: Holger Becker, Alexander Wöhnl, Patrick Frankenberger, Nagihan Kücükyildiz, Jan Hendrik Burdinski and Martin Jänsch.

I would also like to thank my former English teacher, Dr. Rolf Theis, for not only providing me with the proper language skills to write this thesis but as well for his helpful comments on the final version of this thesis. Additionally, I thank Philipp M. Scholl and Philipp Eidam for providing me with helpful comments on various chapters of this work.

Above all, I would like to thank my lovely wife Nicole for her personal support and for keeping up with my stress level during the course of researching for and writing the following pages. Without you I would not have been able to finish this thesis! I thank you with all my heart! Family is the most important thing in the world and while I was writing these lines I had to change the following sentence: I thank *my son* $\Rightarrow$ *my son and daughter* for being there for me whenever I needed a big hug before continuing writing! My wife's and kids' love gave me the proper strength to cope with this challenge. I would like to thank my family and friends for their support, either by keeping me amused or by taking my mind off research to regenerate the brain. I am glad that you are all around me and enrich my life! Although the last, not least, I would like to thank my parents for raising me the way they did, even though my dad could not be around to see me harvesting the fruits of their parenting. Wherever you are, the following thesis is dedicated to you as well...

CONTENTS

1 **Introduction** 1
1.1 Motivation . . . 1
1.2 Challenges . . . 3
1.3 Contributions . . . 5
1.4 Thesis Outline . . . 7
1.5 Previously Published Work . . . 7

2 **Related Work** 11
2.1 Activity Recognition - A Summary . . . 12
2.1.1 Applications . . . 12
2.1.2 Recognizing Physical Activities . . . 14
2.1.3 Rhythm Detection . . . 20
2.2 Time Use Surveys . . . 21
2.2.1 The American Time Use Survey . . . 22
2.2.2 The German Time Use Survey . . . 22
2.2.3 User Monitoring . . . 24
2.3 Sleep . . . 24
2.3.1 A Definition . . . 24
2.3.2 Polysomnography . . . 25
2.4 Sleep Detection . . . 28
2.4.1 Algorithms of Clinically Tested Devices . . . 29
2.4.2 Other Commercial Products . . . 30

3 **Platforms for Monitoring Activities** 33
3.1 Introduction . . . 34
3.2 Wrist-worn Sensor platform . . . 35
3.2.1 The HedgeHog . . . 35
3.2.2 Application Examples . . . 36
3.3 Mobile Recording Platform . . . 37
3.3.1 Usability of Smart Phones - Method Overview . . . 37
3.3.2 User Study . . . 42
3.3.3 Evaluation and Results . . . 46
3.3.4 Monitoring User-Behaviour: A Discussion . . . 50
3.4 Sleep Monitoring System . . . 52
3.4.1 Challenges for Sleep Monitoring . . . 52
3.4.2 Unobtrusive Monitoring System . . . 53
3.4.3 Method Overview . . . 54
3.4.4 Privacy Issues . . . 55
3.4.5 System Discussion . . . 56
3.5 Conclusions . . . 57

4 Rhythms from Time Use Surveys and Sleep 59
4.1 Introduction . . . 59
4.2 Rhythms within Time Use Surveys . . . 60
4.2.1 Comparison of Time Use Data . . . 61
4.2.2 ATUS vs. GTUS Demographic Analysis . . . 62
4.2.3 Feature Extraction . . . 63
4.2.4 Recognizing Activities . . . 64
4.2.5 Time Use Surveys: A Summary . . . 69
4.3 Sleep as a Key Rhythmic Activity . . . 69
4.3.1 Sleep Patterns and Sleep Behaviours . . . 70
4.3.2 Comparison of Sleep Rhythms . . . 71
4.4 Conclusions . . . 74

5 Sleep Detection from Accelerometer Data 77
5.1 Introduction . . . 77
5.2 Sleep Detection with Accelerometers Only . . . 79
5.2.1 Algorithm Details for the Comparison Study . . . 79
5.2.2 The ESS Approach . . . 82
5.2.3 Experiment Set-Up . . . 83
5.2.4 Evaluation . . . 85
5.2.5 Discussion . . . 89
5.3 Sleep Detection with Additional Information . . . 92
5.3.1 Experiment Set-Up . . . 92
5.3.2 Night Segmentation Approaches . . . 93
5.3.3 Evaluation . . . 96
5.4 Conclusions . . . 99

6 Sleep Characterization from Accelerometer Data 101
6.1 Introduction . . . 101
6.2 Detecting Sleep Postures . . . 103
6.2.1 Methodology . . . 103
6.2.2 Results . . . 106
6.2.3 Visual Inspection of Long-Term Data . . . 107
6.3 Visualizing Sleeping Trends . . . 110
6.3.1 Methodology . . . 110
6.3.2 Results . . . 111
6.3.3 Discussion . . . 112
6.4 Myoclonic Twitch Detection . . . 113
6.4.1 Methodology . . . 114
6.4.2 Results . . . 115
6.5 Conclusions . . . 116

7 Improving Activity Recognition with Time Use Data 119
7.1 Introduction 120
7.2 Methodology 120
7.2.1 Wearable System 121
7.2.2 Dataset Description 122
7.2.3 Evaluation Measures 124
7.3 Evaluation Technique 124
7.3.1 Wearable Sensors Only 125
7.3.2 Time Use Only 125
7.3.3 Ensemble: Wearable Sensors and Time Use 126
7.4 Time Use Survey vs. Common Approach 127
7.4.1 Results for Wearable Sensors Only 129
7.4.2 Results for Time Use Only 130
7.4.3 Results for the Ensemble 132
7.5 Discussion 132
7.6 Conclusions 134

8 Conclusions 135
8.1 Contributions 136
8.2 Conclusions 137
8.3 Outlook 139

List of Figures 141

List of Tables 143

Bibliography 160

Curriculum Vitae 161

Publications 163

1 INTRODUCTION

Contents

1.1 Motivation . . . 1
1.2 Challenges . . . 3
1.3 Contributions . . . 5
1.4 Thesis Outline . . . 7
1.5 Previously Published Work . . . 7

It has become trivial to equip oneself with sensors, either embedded in the mobile phone or directly worn on the body. The widespread trend to quantify oneself is driven by benefits that arise from the use of these devices, like being motivated during workouts or knowing how many kilometres one has covered. In such scenarios, we would benefit from the knowledge of what people are doing currently, i.e., what activities people perform. Additionally, it would be beneficial to capture when an activity reoccurs periodically, i.e., in a rhythmic way, so that we are able to predict what the person is going to do in the future. Analysing a person's activities enables us to place a person in a situational context, especially when considering *rhythmic behaviours*. Activities surround us in everyday life and define our nature. Therefore, an activity performed is the key input for a recognition system that is able to determine the activity carried out expressed by movement, body orientation or light intensity. Considering these environmental characteristics, we can define a persons behaviour.

1.1 MOTIVATION

Temporal patterns have been used in many scenarios in the past, with the most traditional example being rhythms in musical patterns. But other fields make use of periodicity as well: In the theory of civilization, the rhythms of history are commonly scrutinized to predict future events (Blaha, 2002). The idea is to investigate the social properties of the population to get a clear view on the challenges that civilization was facing in the past. In theory, a certain social behaviour recurs in time, which is why rhythmic events are depicted and discussed in Blaha (2002). The behaviour and thus the activities of one or many people and their ability to deal with crises describes how well people are coping with a situation. To this day we are determined to scrutinize past knowledge in order to make assumptions about future events. Unlike then, today we are able to gather even more information from various sources to asses what people have been doing and are going to do. These days we gather information from the environment by the use of different sensing modalities, e.g., accelerometers or gyroscopes that are embedded in *wearable devices*. With such data we can make assumptions concerning the user's activities and with that about the context in the field of activity recognition and ubiquitous computing. *Rhythmic patterns* obtained by a *wearable platform* are interesting for many reasons. When modelling human behaviour we support the *awareness* of applications, e.g., a system that knows the consecutive steps in a working process can *assist* by displaying current work steps to new recruits (Lukowicz *et al.*, 2007). Being aware of a person's rhythms can aid in the daily workplace by knowing when a colleague is usually available (Begole *et al.*, 2003). Such a *rhythm awareness* can be beneficial for an activity recognition system, especially for activities that are hard to detect with just the sensor data. Therefore, the use of past knowledge to improve on the recognition rates requires activities that exhibit *circadian frequencies*.

With a high variety of platforms the question is now: How can we make use of such data, most of which is today utilized for advertising purposes (Haddadi *et al.*, 2010). Is there a platform that can be used out of the box by everyone? One answer could be the use of wearable devices, as we experience a rise in the market for wrist-worn sensing devices that can be purchased to perform simple tasks, like tracking GPS coordinates while jogging. The abilities of such devices are limited though, since most of them cannot be used to extract raw sensor data. Another answer to a wearable platform could be the use of smart phones which cannot only log data but the framework gives the possibility to directly use such data and even download it for offline evaluation. Still, the question whether such a platform is suitable for activity recognition has not clearly been answered yet. We come to the conclusion that the optimum platform depends on the application for which we might use it.

In order to enable rhythm detection in the aforementioned scenarios, we need a system that provides (1) *empirical data* that has been gathered over a long period of time, (2) yields *reliable detection* results for the activity performed and (3) is *robust* for various recording scenarios, e.g., that can be worn while doing workouts or other sports. Therefore, this thesis deals with the detection of rhythms within inertial data that has been recorded with wearable motion sensors by making use of machine learning techniques. We will introduce several wearable recording systems that can be used for a long period of time by the user and sense specific user characteristics, such as the activity performed and patterns within a certain activity. With such systems it is possible to record data in a real-world scenario, instead of following scripted activity sequences in a lab set-up. One of these systems has already been used in a healthcare scenario, to motivate the user to move more whenever the sensor detects insufficient movement (Seeger *et al.*, 2011). As already mentioned, commercial devices can detect simple activities like walking or jogging but assessing their algorithms is almost always impossible, since the systems are closed-sourced. In such scenarios it may be problematic to have ones well-being assessed by a system that has not been thoroughly evaluated by physicians.

We identify one important field for rhythm detection which is the monitoring of daily activities in healthcare scenarios to determine if user behaviour and more specifically a patient's behaviour, is changing. People tend to follow certain routines in daily life, depicted by psychological measurements as described in Monk *et al.* (1990) by the social rhythm metric. Such a routine could be, for example, going to sleep at certain times of the day. Changes in such *rhythmic behaviours* might give insight into outliers and therefore valuable information about oneself. In order to *detect outliers* it is important to first *find similarities* in daily routines and more specifically *differences* to other people or to healthy persons. Thus, whenever we are not able to keep a log of our activities it is crucial to use a system that automatically logs what we are doing. In psychiatric research, patients usually have to keep a diary of their activities performed, which helps physicians to evaluate the severity of their illness as for example when suffering from a bipolar disorder. The problem here is that patients tend to forget keeping logs, especially patients that live through major mood swings. But we can think of other healthcare-related fields: The monitoring of elderly who are living on their own, where the detection of behavioural outliers are crucial in emergency situations.

This thesis presents approaches to improve on the recognition of healthcare-related activities. More specifically, we detect sleep and its characteristics by using only a wearable device. With sleep researchers steadily discovering new ways in which sleep impacts quality of life, it has already been shown that a healthy sleep is at least equally important for our well being as nutrition (Everson *et al.*, 1989), and that it contributes significantly to regeneration and recovery (Bonnet and Arand, 2003). Therefore, detecting sleep accordingly is important not only to assess sleep itself but daily life as well, since there is a reciprocal relationship between sleep and daily life, with shortcomings and problems in one area tending to easily influence the other. Since general awareness concerning the importance of sleep is increasing, personal sleep monitoring applications have recently undergone a surge in variety and commercial success. Many of these units are meant to be worn only during the night while being charged on the night stand during the day; other types of units are recording full circadian rhythm data for a number of successive days that give more insight into changes in the user's habits and rhythms. We will evaluate sleep-wake cycles by presenting an observation technique that describes a person's sleeping schedule.

In the following we will present important challenges that are met in this thesis and describe the outline of this work in more detail.

1.2 CHALLENGES

In this section, we will discuss the challenges addressed in this thesis. A first challenge for identifying rhythms is to collect data that have been recorded over a long period of time. Additionally, this is a major challenge for activity recognition tasks as well, since many algorithms need a significant number of training data to be able to classify activities. As pointed out earlier, the recurrence of the same activity constitutes a rhythm. We identify the following challenges:

Long-term Monitoring. Activity recognition tasks require a huge number of data that have been collected over a longer timespan in order to capture all possible variations of an activity represented by the sensor data. Gathering data over a longer stretch of time is not always trivial: Additionally to the data recordings, the ground truth, i.e., what really happened during the time recorded, has to be logged. For these scenarios the user usually has to keep a diary of his or her activities, either by pencil and paper, on demand or by other means like keeping a smart phone log. Especially in long-term monitoring scenarios this becomes annoying and highly inaccurate (Coughlin, 1990), since the burden to keep an accurate log lies on the user. Either the user simply forgets to log what happened or is not in the mood to do so. These scenarios have two possible solutions: (1) The user does not have to log any activities because an unsupervised recognition system is being used or (2) a priori knowledge is utilized which enables the prediction of what the user might be doing next. The latter scenario requires less annotations since the activity has to be known only for the training phase. When considering the first scenario, we would completely rely on the data patterns and guess what the user is going to do.

Prior Knowledge. A recognition system has to be trained from scratch on a lot of sensor data. This way, the system cannot be reused in other environments since it relies on the specific domain for which the data have been accumulated. One way to reduce the number of training data is to make use of prior knowledge. Additional information, such as an approximation of the user's typical schedule, could be combined with real-time sensor data. Previous work has already shown that such a transfer of knowledge is promising and could lead to better results in a recognition system that suffers from insufficient sensor data (Partridge and Golle, 2008).

Rhythm Modelling. As mentioned before, prior knowledge enables us to predict what the user is going to do. For this purpose, not only prior information is useful but also reoccurring representations of an activity in the sensor data - for example, a meeting scheduled for 9:30am every Monday morning - provides a recognition system with an assumption of the performed activity. The obtained occurrences of daily activities enable us to model a rhythm of these activities. Researchers in Van Laerhoven *et al.* (2008) proposed the use of reoccurring activities to improve on the recognition rates. Yet, such rhythmic data has to be obtained by a system that detects the activities with a high confidence.

High-level Activity Recognition. Tracking the user's activities over longer periods with a body-worn sensor device is in particular interesting regarding various healthcare issues (Pansiot *et al.*, 2007; Van Laerhoven *et al.*, 2008): Many mobile and wearable systems[1] that monitor the user's movements and fitness activities (Avci *et al.*, 2010) are commercially available. Mobile phone platforms are being used as hubs to gather such data or collect information from on-board sensors to determine what the user is doing (Brezmes *et al.*, 2009; Sun *et al.*, 2010). Recognition perfor-

[1] such as www.actigraphcorp.com/, last access 09/2014

mance for activities like walking or sitting from mobile sensors are already promising (Bao and Intille, 2004; Jean-Louis *et al.*, 2001b; Lee and Mase, 2002). More diverse activities, however, such as having lunch or vacuum cleaning, are less pronounced: The complex nature of high-level activities requires not just performant machine learning techniques and considerable numbers of training data but person-specific information as well. Especially in battery-driven and resource-limited wearable platforms, it is an additional problem that data acquisition comes with a substantial cost.

Limited Dataset. In many scenarios, especially when monitoring people while sleeping, little training data are available. Depending on the machine learning technique used, more training data lead to higher recognition rates for the monitored activity. On the other hand, it is desirable to score a high recognition rate although data are either missing or simply not available. In addition to that, a sleeping pattern is very similar to the pattern of a person just lying on the couch and watching TV. When it is desirable to obtain sleeping data, the data of a person lying around leads to a confusion of both classes *sleeping* and *lying*. Therefore, extracting significant features for the activity is not a trivial task and decreases the performance of a recognition system.

Recording System. For the activity recognition task we have to use a reliable wearable system that records the required data for several weeks without maintenance. Additionally, it is crucial that the placement of the sensor does not change, otherwise it will be almost impossible to detect the desired activity, since a change of the sensor placement leads to different sensor data. This is also the reason for many research groups to perform studies in a lab-environment instead of in the real world. When thinking of mobile phones, for example, those platforms are seldom carried in the same fashion, leading to different sensor recordings. Nevertheless, researchers have overcome the problem of mobile phone orientation in various ways (Sun *et al.*, 2010).

Furthermore, it is required that the used hardware does not consume too much power and is comfortable to wear. When thinking of mobile phones again, those devices tend to run out of battery power within one day or less when used excessively. Imagine now that an application is running on such a device and gathering sensor data in addition to that. In such a scenario the application has to be programmed in such a fashion that it does not draw too much power. The need for high-frequency measurements demands a hardware system that can be used without running out of battery power or disk space, especially in small-sized devices.

1.3 CONTRIBUTIONS

The main task of this thesis is to answer the following three questions: (1) How can we gather long-term activity data, (2) how can we detect rhythms in such data and (3) benefit from it when performing activity recognition? We contribute to the answers to these questions by first introducing rhythmic data that is already available and showing how we can make use of it. Additionally, we evaluate platforms that can be used to gather sensor data in order to assess rhythms within such data, not only in the field of classical activity recognition but in sleep medicine as well. In this context, we present a novel approach to detect sleep and rhythmic behaviour during sleep. We show how rhythmic data from wearable sensors can be used to improve the recognition of common activities.

Mobile Sensing Platforms. We first analyse the mobile phone as a platform for recording sensor data in the context of activity recognition. As most people own a mobile phone, it is essential to evaluate not only the proximity of the phone to the user but which portion of the day the phone is actually carried by the user. This topic has been discussed in different groups (e.g., Dey *et al.* (2011a) and Patel *et al.* (2006)) but only the proximity of the mobile phone has been scrutinized, not the carrying of the mobile phone on the body. Further, we introduce a mobile sensing platform that allows us to supervise the activity 'sleep': For this purpose, we make use of an infrared camera attached to a netbook to obtain video footage of participants going to bed and waking up. This system enables us to not only monitor sleeping times but sleep characteristics as well, such as sleeping postures or spontaneous muscle contractions during sleep.

Statistical Data. We introduce statistical data from which we extract rhythmic information that gives us insight into a person's habits. For this purpose we analyse, in a first step, statistical data that was obtained by the government. These time use surveys (TUS) usually contain a three-day diary of participants keeping log of what they did, where they were at the time and with whom. Such information is already available since many countries perform such inquiries to obtain statistical information about the population, yet such data has not been used in activity recognition. In a second step we evaluate which features perform best to recognize the activities that have been logged in the time use survey. We further introduce common activities that can be spotted by using only time use information.

Activity Recognition Using Prior Information. In this thesis we propose to improve the recognition rates of common classifiers for activity spotting by embedding statistical data obtained over a population in the classification process. For this purpose we make use of time use surveys and show how recognition rates increase when using a Support Vector Machine output in an ensemble with a maximum-likelihood output performed on the time use database only. Further, we derive common habits in sleep assessment by using a Hidden Markov Model that is trained on data obtained from many different participants over several weeks. With the use of features such as time of day, number of accelerometer movement and light sensor intensity, sleep can be detected in long-term recordings.

Sleep Detection and Rhythmic Behaviour. Sleep is a very important activity, not only because we sleep one third of our life but also because whenever it is disturbed, the human body performs poorly when, e.g., accessing power or studying by simply being unable to concentrate. For this purpose it is important to know when people are sleeping. Still to this day monitoring of sleep is a challenge, since most systems are obtrusive and disturb the normal way of sleeping. Many commercial products do detect sleep with body-worn sensors but sleep is usually overestimated in comparison to medically evaluated approaches (especially polysomnography). Additionally, a common procedure to detect sleep via an inertial sensor is not publicly available. In this thesis we will introduce a novel sleep detection algorithm which not only detects sleep but wake states as well, since most medical devices tend to neglect the detection of wake segments by overestimating sleep. With such information, not only can we gain insights into sleep patterns but also into the rhythmic behaviour of sleep itself.

Dataset Sharing. This thesis makes use of datasets that were recorded for different studies performed in this work. Several of these datasets are unique: One dataset consists of sleeping lab data obtained from polysomnography that has been synchronized with inertial data recorded on the wrist. To our knowledge, such data is not yet available and could give other researchers the opportunity to conduct their own experiments without going through the lengthy process of recording and annotating such data. Furthermore, we obtained long-term recordings of inertial data that were annotated with sleep segments and sleep postures of the participants. Another dataset includes smart phone sensor data and wrist sensor data that were correlated to spot when the user moved with his smart phone or not. The rich content of these datasets leaves room for further evaluation in different fields of research.

1.4 THESIS OUTLINE

The thesis is structured as follows: We begin with a detailed discussion of research work that relates directly to this thesis, in Chapter 2. Then, we introduce the evaluation of platforms which have been used throughout the thesis in the experiments in Chapter 3 after which we show circadian rhythms that are already present in different data sources in Chapter 4. We continue to introduce detection techniques to highlight a most rhythmic activity - sleep - in Chapter 5, followed by Chapter 6 in which classification of postures during sleep as well as muscle contractions that give a clue about the person's sleep habits are scrutinized. We then use prior knowledge to detect common activities, in Chapter 7. We conclude this thesis in Chapter 8 with a final discussion of future work.

1.5 PREVIOUSLY PUBLISHED WORK

In the past years, some aspects of this thesis have already been published in conference proceedings mostly in collaboration with other colleagues. In this section I will reference all published papers and mention my contribution to each scientific work.

1. Marko Borazio, Ulf Blanke and Kristof Van Laerhoven, *Characterizing Sleeping Trends from Postures*, Proceedings of the 14th IEEE International Symposium on Wearable Computers (ISWC 2010), Seoul, South Korea, IEEE Press, pp. 167-168, 10/2010.
   Inspired by my diploma thesis in 2008, I followed up the idea of scrutinizing sleep and the way we sleep by taking into consideration sleeping postures. With the assistance of Ulf Blanke in the evaluation process, we were able to visualize sleeping postures and to sketch a possibility of sleep quality assessment.

2. Holger Becker, Marko Borazio and Kristof Van Laerhoven, *How to Log Sleeping Trends? A Case Study on the Long-Term Capturing of User Data*, The 5th European Conference on Smart Sensing and Context 2010 (EuroSSC 2010), vol. 6446, Passau, Germany, Springer Verlag, pp. 15-27, 2010.
   With the help of my diploma student Holger Becker, I was able to set up a long-term recording system in the field of sleep studies. On the basis of my idea of monitoring people at home, Holger Becker recorded a dataset of sleeping postures that could be used for publication later on.

3. Marko Borazio and Kristof Van Laerhoven, *Predicting Sleeping Behaviors in Long-Term Studies with Wrist-Worn Sensor Data*, International Joint Conference on Ambient Intelligence (AmI-11), vol. LNCS 7040, Amsterdam, Springer Verlag, pp. 151156, 11/2011.
   This work focuses mainly on predicting how the sleeping behaviour varies on different days of the week. With the gathering of long-term recordings of people sleeping, I was eager to know how people change their sleeping schedule according to different observational features, e.g., weekends vs. weekdays.

4. Marko Borazio and Kristof Van Laerhoven, *Combining Wearable and Environmental Sensing into an Unobtrusive Tool for Long-Term Sleep Studies*, 2nd ACM SIGHIT International Health Informatics Symposium (IHI 2012), Miami, Florida, USA, ACM Press, 01/2012.
   This work combines the topics of papers 1.-3. I conducted the experiments that yielded a lot of data to asses when the user is sleeping - based on wrist-sensor data. Additionally, I took a different approach to cluster sleeping postures that might be used in medical studies in the sleeping lab. I implemented the algorithm that detects muscle contractions of the wrist while sleeping.

5. Marko Borazio and Kristof Van Laerhoven, *Improving Activity Recognition without Sensor Data: A Comparison Study of Time Use Surveys*, 4th International Augmented Human Conference , Stuttgart, Germany, ACM Press, 03/2013.
   The use of statistical data obtained over a population enabled us to extract user-specific information from a large variety of people. I obtained the database for this study directly from the Federal Statistics Office in Germany and conducted the further evaluation of extracting relevant features for activity recognition.

6. Marko Borazio and Kristof Van Laerhoven, *Using Time Use with Mobile Sensor Data: A Road to Practical Mobile Activity Recognition?*, 12th International Conference on Mobile and Ubiquitous Multimedia, Lulea, Sweden, ACM Press, 12/2013.
   As a contribution to paper 5., I obtained sensor data by monitoring many users with a wearable sensor. Additionally, the participants were asked to keep a diary of their activities. The scripts to classify the activities afterwards were implemented by myself.

7. Marko Borazio, Eugen Berlin, Nagihan Kücükyildiz, Philipp M. Scholl and Kristof Van Laerhoven, *Towards Benchmarked Sleep Detection with Inertial Wrist-worn Sensing Units*, ICHI 2014, Verona, Italy, IEEE Press, 09/2014.
   Additionally to the idea of monitoring people's sleep in their home, I established an algorithm of detecting sleep with an accelerometer-based device and put the results of this algorithm in contrast to commonly known algorithm results. Eugen Berlin supported the implementation of the new algorithm, while Nagihan Kücükyildiz and Philipp M. Scholl helped at the data acquisition stage.

8. Marko Borazio, Jan Hendrik Burdinski and Kristof Van Laerhoven, *Wear is Your Mobile? An Empirical Comparison between Wearable and Mobile User Monitoring*, in submission.
   My original idea to use the smart phone as a sensing platform was supported by the implementation of an Android application by my Master student Jan Hendrik Burdinski. His initial data acquisition was continued by myself, which led to a total number of 51 participants in this study. The evaluation of the data was conducted by myself.

RELATED WORK 2

Contents

2.1 Activity Recognition - A Summary **12**
  2.1.1 Applications 12
  2.1.2 Recognizing Physical Activities 14
  2.1.3 Rhythm Detection 20
2.2 Time Use Surveys **21**
  2.2.1 The American Time Use Survey 22
  2.2.2 The German Time Use Survey 22
  2.2.3 User Monitoring 24
2.3 Sleep **24**
  2.3.1 A Definition 24
  2.3.2 Polysomnography 25
2.4 Sleep Detection **28**
  2.4.1 Algorithms of Clinically Tested Devices 29
  2.4.2 Other Commercial Products 30

The spotting of complex human activities has been scrutinized from different angles, showing not only the progress over the past years but the limitations of activity recognition as well. The idea of using *observational* information from past activities in the field of activity recognition was proposed, as well as different data sources to obtain such information.

In the following sections we will introduce the concept of activity recognition and the advances in this field, highlighting their potential and their limitations in Section 2.1 as well. Further, we provide insight into the concept of rhythmic data and to which extent they are used in the field of activity recognition. Specifically, we will describe the general idea of rhythms and how such daily routines have been spotted in previous work. Then, we introduce statistical data that are gathered by the government and which already exhibit many daily routines of a country's population in Section 2.2. We will describe the datasets and highlight relevant studies that make use of *prior knowledge* like statistics. Additionally, in Section 2.3, we will introduce the rhythmic activity sleep and current methods of investigating sleep in medical trials. Section 2.4 describes the spotting of sleep with medical and commercial devices, giving an overview of commonly used sleep detection algorithms that are relevant for the rest of this thesis.

2.1 ACTIVITY RECOGNITION - A SUMMARY

The idea of sensing the user's context and therefore the user's current situation was first introduced by the term context-aware computing by Schilit *et al.* (1994). In order to sense the user's context, it is important to detect current activities performed. For the purpose of activity recognition wearable sensors are used which appear in different shapes and variations. For example, in medical research in the early 1970s physical activities were measured by the use of a questionnaire evaluating energy expenditure (Taylor *et al.*, 1978). With the advances in the technology of wearable devices, such questionnaires have been replaced by activity recognition systems that automatically capture the user's movements.

2.1.1 Applications

Recognizing activities accurately led to applications that are situated in many different scenarios in industry or in healthcare. We will describe the areas in which activity recognition systems are used and highlight their approaches.

Healthcare Applications. The use of new technology like inertial sensors or wireless sensor networks (WSN) opens up new ways of *observing* what people are doing without invading their privacy compared to using visually based systems like cameras. One of the overall goals is to monitor elderly persons or patients in order to provide the medical staff with sufficient information about their well-being (Lymberis and Olsson, 2003) or to monitor them over a longer timespan for diagnostic purposes (Wu *et al.*, 2008; Plötz *et al.*, 2012). This way, early indications for a disease or an emergency situation are provided, enabling us to immediately respond to such scenarios (Jafari *et al.*, 2007). The evaluation of well-being to prevent the user from situations that might lead to an unhealthy lifestyle cover many different scenarios like the monitoring of one's physical work-out activities to stay in shape (Seeger *et al.*, 2011). By using a body-worn sensor to estimate the user's current work-out activity (e.g., performing curls), the data is displayed on a mobile phone to determine if the user has reached his movement quota as prescribed by the physician. In Amft *et al.* (2007), feedback is given regarding one's lifestyle by monitoring nutrition habits. Dietary intake cycles are categorized into different groups by the use of body-worn sensors to classify eating and drinking events.

One prominent area of using sensor technology these days is in assisted living scenarios. The goal is to distinguish different activities in order to *assist* the residents while, for example, cooking to provide chronically ill with the possibility of living on their own (Bächlin *et al.*, 2010) or to predict what the user is going to do based on past observations (Kautz *et al.*, 2003). For this purpose location information and the surrounding noise are used to detect the resident's current whereabouts to track unusual patterns in their behaviour. With such a system, the performance of activities of daily living (ADL) are measured. Helping elderly persons to take their medication was explored in De Oliveira *et al.* (2010) by engaging people in a social game. Similarly, researchers in Jafari *et al.* (2007) detected emergency situations (such as falling) in the homes of elderly people in order to generate automatically an emergency call.

In such a domain it is important to monitor ones sleep-wake cycles as well to determine if the circadian rhythm is disturbed (Adami *et al.*, 2003; Jean-Louis *et al.*, 2001a). A damaged sleep rhythm has negative effects on other daily physical activities and on well-being (Liang, 2013) which is why more and more commercial products are available on the market. Detecting sleep in a home set-up will be discussed in more detail in Section 2.4 and in Chapter 5.

These examples show that the gathering of medical data provided by sensor data is a major challenge. Not only do the local hospital or other physicians have to be supplied with sufficient data but one has to find new ways of improving the quality of clinical research as well. Additionally, having a database containing information over, for example, a 5-year period might enable physicians today to re-evaluate previous data in order to detect the origin of certain deceases. For such a scenario it is necessary to make use of centralized databases that can be used by physicians all over the world to enable multi-center clinical studies (Sahoo *et al.*, 2011).

Industrial Applications. In industry activity recognition has been used to support workers in their tasks, to guide new recruits through the work-flow process or to avoid mistakes (Lukowicz *et al.*, 2007; Stanford, 2002). With such a system it is possible to overview the whole work process and to assess it, while being provided with sufficient information to automatically generate a manual of the performed steps. In such scenarios it is desirable that the user has his hands free and is guided by a system that is unobtrusive but still lets the worker perform his normal routine.

There are many examples of studies with the main goal of reaching a higher quality in the work process, e.g., when assembling cars (Maurtua *et al.*, 2007; Stiefmeier *et al.*, 2008). In such an automotive environment it is desirable to overview the quality of the car building process by making use of sensors embedded in the workers' clothing and environmental sensors. In other scenarios like a wood workshop the tracking of consecutive steps has been pursued, such as drilling or hammering (Ward *et al.*, 2006). With such a system, it is possible to provide the user with a proactive assembly instruction, for instance, when putting together furniture (Antifakos *et al.*, 2002).

Sports and Entertainment. The quality of activities performed plays a key role in many applications. Especially athletes benefit from knowledge about their own performance. In Ladha *et al.* (2013) for example, a climbing performance analysis system is presented as an automatic coaching system for climbers. The movements are captured by evaluating inertial data which are categorized according to power, control, stability and speed. Other activities can be recognized with the use of body-worn sensors as well: Researchers in Ermes *et al.* (2008) detected cycling, Nordic walking or rowing. Other work used sensor devices embedded in the protection suit of Taekwon-Do tournaments to assist referees in detecting a score (Chi, 2005). Additionally to detecting the activity performed, sensing units are designed to measure the energy expenditure level as well (Bonomi *et al.*, 2010).

In the entertainment sector embedded sensors have been used for gaming purposes: Heinz *et al.* (2006) introduced the approach of user interaction with martial arts movements (Wing Chun) to apply an expertise score. In the commercial field, Nintendo successfully introduced the Wii in 2006 which was further developed in the gaming console Wii U[2]. The controller of the Nintendo Wii uses motion sensors to determine the user's movements which is the direct user input for video games. Microsoft's XBox One[3] on the other hand makes use of the Kinect, a video camera that enables the capturing of full body-motions, similar to Sony's Playstation 4[4] camera.

Other Application Areas. Activity recognition research is found in many different domains. In the military, sensors are used for modern warfare which enable the gathering of environmental information for mission planning by embedding sensors directly in the soldiers' clothing (Minnen *et al.*, 2007b; Schlenoff *et al.*, 2010). Additionally, the soldiers' movements and actions are identified such as crawling or kneeling. For prevention of emergency situations people's location changes at a festival have been evaluated in Blanke *et al.* (2014). This way, predictions are made about public spaces and whether they will be overcrowded, the intention being to prevent such a scenario by the local police.

2.1.2 Recognizing Physical Activities

In this section we give an overview of common terms, characteristics used and machine learning approaches for activity recognition with mobile and wearable sensors. The term *physical activity* can be divided into two sub-terms, namely *low-level* activities and *high-level* activities. These terms are not standardized in the activity recognition community, which is why we introduce them here. Low-level activities are activities like *jogging*, *walking* or *standing*. They are described by one single, completed event that lasts only for a short period of time, i.e., a few minutes. In contrast to that, high-level activities are of a more complex nature and imply low-level activities that, strung together, describe a high-level activity. Examples of such are *eating lunch*, *assembling furniture* or *cooking*, which usually last longer than only a few minutes.

Many researchers take different approaches to recognize activities. Usually, a lot of sensor data have to be recorded in order to train a new model that recognizes activities. Depending also on the setting, such data have to be recorded again and again which is why it is difficult to compare approaches of different researchers. One way to minimize the need of excessive training data has been shown in Blanke

[2]http://www.nintendo.com/wiiu, last access 09/2014
[3]http://www.xbox.com/en-US/xbox-one/innovation, last access 09/2014
[4]https://www.playstation.com/en-us/explore/ps4/, last access 09/2014

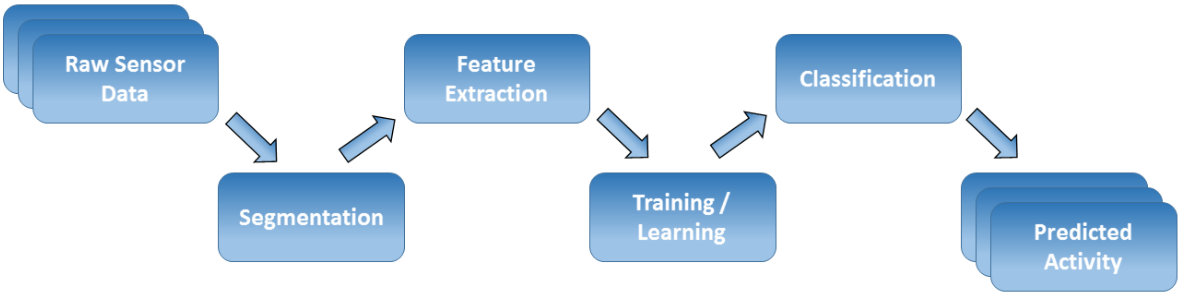


Figure 2.1: Activity recognition process.

and Schiele (2010). Here, high-level activities are defined as *composite activities* and researchers propose to transfer knowledge from one composite activity to others by a partonomy-based approach. It consists of single activities like *drilling* or *screwing* that are part of two different workshop scenarios *assembling a bookshelf* and *assembling a mirror*. The idea is to recognize activities in one scenario and use that information in the second scenario. The results in Blanke and Schiele (2010) show that such an approach is feasible and increases the recognition performance for composite activities when compared to standard approaches.

After the recording phase, the data has to be labelled accordingly to the activity that really occurred (which is called the *ground truth*). By extracting interesting properties from the data (the so-called *features*), we obtain a description of the activities and minimize the classification effort on the hardware side. The preprocessed data is then trained on and the classifier yields a confidence score for the recognized activity. Figure 2.1 shows the whole activity recognition process.

Feature Extraction

An important step in the activity recognition process is the extraction of relevant characteristics in the recorded data stream. Such *features* hold valuable information in distinguishing various activities. Usually, finding the right features to apply in the classification process is done by manual observation of the data. With such a *feature extraction* the basis is given for an input to pattern recognition or machine learning algorithms. The challenge in the algorithm design is to find the best feature that captures relevant motion characteristics and which is at the same time insensitive to inaccuracies that result from the limited resolution and sampling rate of the sensor used. Variations in sensor placement or other sources of noise need to be minimized as well like environmental magnetic fields. The most common features are described in the following paragraphs.

Time-Domain Features. The most prominent features used in activity recognition are *mean* and *variance* which express the user posture and acceleration of the user over a specific time-frame respectively (Bao and Intille, 2004; Huynh and Schiele, 2005; Lester *et al.*, 2006; Maurer *et al.*, 2006; Ward *et al.*, 2005). Huynh and Schiele (2005) assessed the use of such features by utilizing cluster precision to rank the features. Other features are *standard deviation* (Ravi *et al.*, 2005) or *peak extraction* where characteristic peaks are detected and used for classifying activities (Van Laerhoven *et al.*, 2003). In the medical field, *zero crossing* techniques are used as features to determine movement segments for sleep detection. In Chapter 2.4.1 we will discuss such algorithms in more detail, showing also how a movement or non-movement segment is detected efficiently in Chapter 5.

Other Features. Frequency-domain features such as *spectral energy*, *entropy* or *Fast Fourier Transformation* (FFT) are used for activities with recurring motion patterns (Huynh and Schiele, 2006b; Lester *et al.*, 2006). In Berlin and Van Laerhoven (2012a) the raw sensor data is first approximated by linear segments, then modelled by symbols in order to detect similar sequences that represent *motifs* of activities. Another type of feature that is applied in activity recognition is based on *wavelet transform* to characterize non-stationary signals for recognizing low-level activities, e.g., walking or running (Preece *et al.*, 2009).

Evaluation Measures

A common measurement for describing the recognition rate for an activity is accuracy, which can be derived from a confusion matrix. A confusion matrix is a $n \times n$ matrix, where n is the number of classes in the classification. The rows of the matrix represent the ground truth, i.e., the actual class, while the columns represent the predicted classes. A benefit of a confusion matrix is that it directly shows if the system is confusing two classes, i.e., commonly mislabelling one as another. Consider the confusion matrix in Figure 2.2 (left) with $n = 3$. Accuracy is the sum of the diagonal divided by the sum of all occurrences (here: $15/21 = 0.71$).

The overall accuracy is often not enough to reveal particular details of the system's performance. To gain this information, calculating per class performance values, namely precision and recall, are necessary. For this purpose, the confusion matrix of $n > 2$ has to be considered as a two-class matrix, by summing up the rows and columns outside the actual class. We obtain the confusion matrix in Figure 2.2 (right). *True positives* (TP) are the correctly predicted classes according to the ground truth and the $false\ positives$ (FP) are the wrongly predicted classes in regard to the ground truth. The $false\ negatives$ (FN) on the other hand are all the activities labelled as one class but do actually belong to another class. The $true\ negatives$ (TN) are the sum of all the $true\ positives$ of the other classes, i.e., the sum of all correctly predicted classes. Here, precision is the amount of correct labelled classes (TP) that was labelled as the activity in the ground truth:

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{2.1}$$

From equation (2.1) we calculate for class 1 in Figure 2.2: $6/(6 + (1 + 1)) = 0.75$. Recall is defined as the proportion of the data originally labelled as an activity that ground truth / predicted class

	C1	C2	C3
C1	6	1	1
C2	1	5	2
C3	0	1	4

ground truth / predicted class

	C1	$C2'$
C1	TP	FP
$C2'$	FN	TN

Figure 2.2: Left: Example of an confusion matrix with three classes $C1$, $C2$ and $C3$. Right: Confusion matrix shown as a two-class matrix, with classes C1 and $C2' = C2\&C3$.

was correctly classified as the activity:

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{2.2}$$

leading to a recall of $6/(6+1) = 0.86$ for class 1.

With such measurements, we are able to describe the performance of our algorithms, by either obtaining a high recall or precision value or a trade-off between these two, depending on the study being conducted.

Platforms for Activity Recognition

Most research studies propose sensor platforms that are required to be unobtrusive and that find acceptance in everyday life. Therefore, sensor systems are used that can be easily integrated into existing equipment like watches (Lester *et al.*, 2006) or into garment (Dunne *et al.*, 2006; Paradiso *et al.*, 2005; Philipose *et al.*, 2004). Mobile phones are being considered as well since they already embed many sensors like accelerometers or gyroscopes (Kwapisz *et al.*, 2011). Apart from using one sensor to detect activities, another approach is the combination of sensors such as accelerometers with audio (Ward *et al.*, 2005) or video (Brashear *et al.*, 2003). For the latter purpose, acceleration is compared to video data to correct misclassified movements.

Sensor Types. In order to measure rapid movements, Kunze *et al.* (2006) used several accelerometers with gyroscopes to recognize basic martial arts movements. In garments, piezoresistive fabric sensors for acceleration, respiration sensors and ECG electrodes are used in the context of health care for patient monitoring (Paradiso *et al.*, 2005). Another sensor type worked into the back part of the garment is the plastic fiber sensor (POF), a light sensor used to measure the amount of light that is being transmitted (Dunne *et al.*, 2006). The light sensor detects electromagnetic radiation in a spectral range in order to monitor the sitting position to inform the user of a wrong sitting position. RFIDs are considered, for example, in gloves for identifying tagged items nearby (Patterson *et al.*, 2005). This way activities are recognized such as holding a cup which is equipped with an RFID-Tag. Textile pressure sensors (Meyer *et al.*, 2006) and force sensitive resistors (FSR) (Amft *et al.*, 2006) are used in experiments for detecting muscle activity. Two electrodes equipped with conductive yarn on both sides of a compressible spacer are worked into the textile. The electrodes establish a capacitor whose capacitance changes when the spacer is compressed (Meyer *et al.*, 2006).

Sensor Systems. Many different approaches and sensor systems have been pursued by research groups for capturing sensor data. One prominent example is the use of a wrist-worn device[5] that logs inertial and light data over a long timespan. With such a device it is possible to capture low-level activities (Seeger *et al.*, 2011) and high-level activities (Scholl *et al.*, 2013). The device will be introduced in more detail in Chapter 3. Other examples are the use of electroocculography to trace the eye movements to detect activities (Bulling *et al.*, 2011) or platforms that are equipped with a 2-axis accelerometer, a skin temperature, a galvanic skin response and a heat flux sensor (Sunseri *et al.*, 2009). The latter is used to monitor the activity performance during work-outs and while resting, e.g., sleeping.

Commercial platforms like the Nike+ Fuelband[6] or Fitbit One[7] are available for many researchers but scarcely used due to the closed-source nature that prevents the use of the raw sensor data. Such fitness devices raise an interest in the community since they can be used directly and display information about oneself, e.g., how many steps one has taken per day or how many kilometres one jogged. Although the Fitbit has been used in Montgomery-Downs *et al.* (2012) for assessing sleep, the device was found to misidentify sleep patterns.

Activity recognition with mobile sensors has been investigated for some years now (Brezmes *et al.*, 2009; Sun *et al.*, 2010), with researchers also analysing if it is feasible to use a mobile device for detecting activities (Patel *et al.*, 2006). In Brezmes *et al.* (2009), basic human movements (walking, sitting, standing, climbing stairs) are detected in real-time on a mobile phone by analysing the accelerometer data with a high confidence. Similar results are obtained by researchers in Sun *et al.* (2010), again detecting basic activities on a mobile phone but considering the orientation of the device as well. The performance of the system is high, stating that basic activities can be captured with a mobile device. Patel *et al.* (2006) on the other hand investigated for which portion of the day a mobile device (smart phone) is with the user. Interestingly, results indicate that half of the time the mobile phone is not with the user.

We will show in detail in Chapter 3 that a mobile phone can be used in certain ubiquitous computing scenarios for activity recognition.

Activity Recognition Approaches

Many machine learning algorithms are commonly used for activity recognition with wearable sensors. In the following we will highlight a few to show the achieved progress and to introduce methods that are used within this thesis.

[5]http://www.ess.tu-darmstadt.de/hedgehog, last access 09/2014
[6]nikeplus.nike.com/, last access 09/2014
[7]www.fitbit.com/one, last access 09/2014

Supervised Learning. *Supervised* learning or classification is a technique for creating a classification function from the training data. The training data consist of pairs of input objects, feature vectors, and class labels (ground truth). The task of the supervised learner is to predict the value of the function for any valid input object after having observed a number of training examples. A common method is the *Support Vector Machine* (SVM) (Anguita *et al.*, 2012; Cortes and Vapnik, 1995; Cristianini and Shawe-Taylor, 2000), for which a classification process looks as follows: (1) Gather sensor data and simultaneously the ground truth. (2) Label the data according to the ground truth and extract appropriate features. (3) Divide the dataset into *training* and *test* sets, whereat the training and test sets are interchanged before testing it again. (4) Training is repeated several times according to the new combination of data. (5) Validate the algorithm by using the test set. A so-called cross-validation (steps 3-4) ensures that by several combinations of data the classifier yields a high confidence of the recognized activity.

There are several other methods that follow the same procedure, such as *k-Nearest-Neighbour* (kNN) (Kunze *et al.*, 2006; Maurer *et al.*, 2006; Ravi *et al.*, 2005), *Naïve Bayes* (Maurer *et al.*, 2006; Ravi *et al.*, 2005) or *decision trees* (Bao and Intille, 2004; Maurer *et al.*, 2006; Ravi *et al.*, 2005). A *Hidden Markov Model* (HMM) (Brashear *et al.*, 2003; Lester *et al.*, 2006; Murphy, 2003; Patterson *et al.*, 2005), for example, is a statistical approach in which the system being modelled is assumed to be a Markov process with unknown parameters and the challenge is to determine the hidden parameters from the observable parameters. The extracted model parameters can then be used to perform further analysis, for example, for pattern recognition applications.

Ensemble learning is the idea of fusing two or more classifiers to improve the recognition rate for activity recognition and has been mentioned in several works (Alexandre *et al.*, 2001; Polikar, 2006; Su *et al.*, 2009; Tax *et al.*, 2000). Zappi *et al.* (2008), for example, use multiple body-worn sensors for activity recognition in the context of quality assurance in a car assembly factory. Using a discrete HMM, the results led to an improvement in the recognition rate. In Polikar (2006), different ways of classifier fusion or ensemble classifiers are being discussed, like *mixture of experts*, *bagging*, *boosting* or *algebraic combination rules*. The latter are usually *majority voting*, *sum* and *product rule*. Researchers in Alexandre *et al.* (2001) evaluated the two common combination rules (the sum and product rule) for fusing classifiers by their posterior probabilities.

In Section 6.2 we will use the kNN approach to classify sleeping postures. SVMs are used in Section 6.2 to classify exceptional movements during sleep and in Chapter 7 to detect common daily activities. An Ensemble of classifiers is presented in Chapter 7 as well to embed prior knowledge in the classification process in order to increase the recognition rate.

Unsupervised Learning. In contrast to a completely supervised learning method, *unsupervised* learning is the classification of activities without the knowledge of the user's activities. The common approach taken is completely different from the supervised learning approach, since the labelling and training phase are not required. Instead, a model of the data is created, such as when using a clustering routine like the *Kohonen Self Organizing Map* (KSOM) (Kohonen, 1990; Van Laerhoven *et al.*, 2003). The data is allocated to specific centroids because of their similarity to those centroids, which enables the detection of similar data.

Other examples for unsupervised learning techniques are *k-means* (Huynh and Schiele, 2005; Krause *et al.*, 2003; Kwon *et al.*, 2014), *motifs* (Fuchs *et al.*, 2009; Minnen *et al.*, 2006, 2007a; Murakami *et al.*, 2005) and *multiple Eigenspaces* (Huynh and Schiele, 2006a,b).

In Chapter 6 we will use the KSOM to allocate sleep postures to different color codes in order to illustrate similarities.

Semi-Supervised Learning. A *semi-supervised* approach is situated between the supervised and the unsupervised learning domain and is applied when the obtained data are only partially labelled. The gathering of sensor data usually comes with a high cost: For a supervised learning technique lots of training data is needed, which requires a lot of effort in first of all *obtaining* the ground truth and afterwards *annotating* the sensor data accordingly. A semi-supervised approach is beneficial when only small parts of the data need to be known and for real-world settings. *Transductive Support Vector Machines* (TSVM), for example, infer the correct labels for unlabelled data by using only a small portion of labelled data (Zien *et al.*, 2007). Other examples of semi-supervised approaches can be found in Longstaff *et al.* (2010); Stikic *et al.* (2008, 2009); Subramanya *et al.* (2006).

2.1.3 Rhythm Detection

An important research field in activity recognition is the detection of temporal patterns such as recurring activities at specific point in time. The daily routine of a person is defined by his or her behaviour. For example, before a co-worker starts to work he will probably follow his usual routine and first get a coffee and have small talk with his colleagues. As this is true for most of the colleagues in an office, such pattern knowledge can be used to describe the social context the user is currently in. In many different scenarios such information is helpful to determine *a priori* knowledge, i.e., knowing what a person is going to do in the future. Predictions have been used in various scenarios for activity recognition (Farrahi and Gatica-Perez, 2011; Krumm and Brush, 2011; Scott *et al.*, 2011; Tominaga *et al.*, 2012; Zhu *et al.*, 2013). We will summarize related work that motivates the topic of this thesis.

Behavioural Rhythms

Being able to predict when someone is available is beneficial in many different scenarios. For example, if it is known when a person is currently at his work-desk, distributed team members are aware of the availability of their co-worker. Hill and Begole (2003) observed for such a purpose the computer activity and established a namespace for different locations and associated with that location the activity, e.g., commuting from office to home. Such a model of patterns gives the possibility to not only show the availability of a person but reflects what the person is doing throughout the day.

Recently, researchers in Tominaga *et al.* (2012) predicted a person's going-out behaviour in order to asses if the person is going to leave the home or not. A rhythm of a person's habits has been established by using a camera to determine if the user is at home or not. This information was then used as a prior for a Hidden Markov Model (HMM). Here, time histories of people leaving or entering the home have been generated.

A similar social aspect has been monitored by Eagle and Pentland (2006) by observing the location and proximity of mobile phones in order to detect weekly patterns. The idea of using mobile phones not only for real-time activity recognition but also for gathering useful information about the user and the activities, has been investigated in Farrahi and Gatica-Perez (2008), in order to develop rhythmic behavioural data that can be used to detect daily routines. We will have a look at such data in Chapter 4, which has not been gathered by mobile devices but by other means.

Daily Routines

The work in Van Laerhoven *et al.* (2008) uses diary data to extract daily routines in order to improve activity recognition. The first step to capture routines without the user having to interfere is to gather environmental data, either from a wearable sensor, or from sensors that are installed, for example, in the home or workplace.

Especially in healthcare such scenarios have to be considered, when, for example, elderly people living on their own are monitored in order to be able to respond to emergencies. Daily routines can help here by predicting what the user will be doing most likely. A similar idea is being pursued in Krumm and Brush (2011), by gathering prior information by keeping a diary and additionally obtaining GPS information about a person being at home. The paper shows how the prior location improves the prediction.

In Farrahi and Gatica-Perez (2011), on the other hand, individual and group behaviours have been investigated in a large mobile phone dataset. A probabilistic topic model has been used on the data, detecting routines in the data to determine behavioural patterns that give an insight into daily routines. Overall, the research community is interested in what people are doing next, predicting the behaviour not only from the same persons but from various individuals as well. Such information could then be used in different models that deal with sequential data where such prior probabilities or even posterior probabilities might improve the classification results. An example of such a classifier is the *Conditional Random Field* (CRF) (Wallach, 2004), which is a temporal probabilistic model.

In the following section we will have a close look at daily routines and behavioural data from governmental databases.

2.2 TIME USE SURVEYS

The idea of using prior knowledge led to an investigation of databases that are already publicly available (Partridge and Golle, 2008). More specifically, several countries perform inquiries from which statistical information about the population can be derived. These *time use survey* data are usually obtained by keeping a diary for one or more days, which give, for example, an insight into a countries' activities. Such data has been gathered since the 19^{th} century, with Russian officials being the first to collect statistical information of peasant families starting from the 1860s (the zemvsto statistics, Gershuny (2011)). Over the years the idea of keeping track of the population became more and more prominent, especially for the purpose of capturing a nation's well-being. First, each country performed their own time use survey data collection but with the possibility to store data digitally in databases, time use surveys have become more and more standardised. This way the databases can be used internationally, enabling comparisons between different countries or fusing surveys to one big survey, like the Multinational Time Use Survey (MTUS[8]) or the Harmonised European Time Use Survey (HETUS[9]). The HETUS is being maintained by EuroStat, the statistical office of the European Union, and embeds time use data from 15 different European countries since 1993. Data analysis can be conducted directly on the HETUS web page, though displaying summary statistical information only. All participating countries are using the same database structure in order to be able to fuse the data afterwards. Additionally, activity and location descriptors are being logged in the database and abstracted: They are categorized in *tiers* and recorded for at least 24 hours. Participants are included according to a rigid selection process and financially recompensed, and for each a standard and anonymised set of demographic information is available.

The usage of time use data is not trivial: Although some databases are freely available, they are not useful for the purpose of deriving activity information about the population, because they contain only summarized statistical information. In the following, we will have a closer look on the German Time Use Survey (GTUS) 2001/2002 dataset that was used in Chapters 4 and 7 of this thesis, as well as the American Time Use Survey (ATUS[10]) 2006 database that was used in recent work (Partridge and Golle, 2008).

2.2.1 The American Time Use Survey

The ATUS dataset is one of the few time use study databases that is freely available without restriction and is being updated every year since 2003. Self-reported activities of US residents are logged for 24-hours. It contains 18 different (Tier 1) activity groups and in total a distinction is made between 462 activities (Shelley, 2005). In

[8]http://www.timeuse.org/mtus/, last access 09/2014
[9]https://www.h2.scb.se/tus/tus/, last access 09/2014
[10]http://www.bls.gov/tus/datafiles_2006.htm, last access 09/2014

addition to the activities, participants logged the time the activity started, how long it lasted and where it took place. The dataset is anonymised after collection and contains information about gender and age for all 12,943 participants.

The dataset was primarily used by Partridge and Golle (2008). The work contains a detailed description on how these data might be applied for activity recognition, and reports on a study using a 10-fold cross-validation analysis of the features, showing that the hour of day and location are the most useful features for activity estimation. The authors note in their paper that studies performed by different nations vary in terms of participant behaviour (observed, for instance, in response rates) and constructs (motives range from quantifying unpaid work to measuring exposure to environmental pollutants). This is exactly the motivation for the work conducted in Chapter 4, as it reports on studies and comparisons with the time use data from a large European country. For more information on the dataset we refer to the original work in Partridge and Golle (2008) that scrutinized the ATUS 2006 in depth. We will now introduce the time use survey that is the basis for the work in Chapters 4 and 7.

2.2.2 The German Time Use Survey

The GTUS was first surveyed in 1991/1992, being updated every 10 years and is only accessible by regional government employees, after going through a formal admission process. The data acquisition takes usually a year (therefore it is labelled 1991/1992), in order to compensate seasonal bias and also to capture certain population groups (e.g., a single mother or father). We used the survey from 2001/2002, since the data from the current measurement period (2011/2012) will be available the earliest in 2015.

The 2001/2002 GTUS consists of data from 13,798 participants, all older than 10 years, who kept a detailed diary for three days each. They wrote down which activity they performed in 10-minute slots. The diary keeps account of the location where the activity took place, as well as whether a secondary activity was performed (e.g., *watching TV* while *eating*) and who was present at the time (e.g., a household member). Additionally, personal information like relationships between household members are being logged. In total 272 single activities were distinguished and allocated to three hierarchical tiers, with Tier 1 containing generic descriptions such as *personal care*, *household activities* and *mass media*, Tier 2 including a more precise description of the activity, like *sleeping*, *cooking* and *reading*, while Tier 3 contains the highest specificity, such as *sewing clothes*, *doing laundry*, and *traveling on a bus*.

Table 2.1 depicts an example of such a dataset (of the household with the ID no.123), displaying the first two household members and their age (ph01b2x) and gender (ph01c), citizenship (ph01d), marital status (ph01e) and if they are in parental leave (p05). Within a household, each member is allocated to an ID (idpers), whereas each household is assigned to an unique ID (idhh). A second dataset shown in Table 2.2 holds for the same household the main performed activities (zhc76) for timeslot 76 (which corresponds to the time interval between 16:40 and 16:50), as well as the

idhh	idpers	ph01b2x	ph01c	ph01d	ph01e	...	p05	...
123	1	34	female	german	married	...	yes	...
123	1	34	female	german	married	...	yes	...
123	1	34	female	german	married	...	yes	...
123	2	36	male	italian	married	...	no	...
...	...	...	...	...	...	...	...	...

Table 2.1: The person-specific GTUS 2001/2002 dataset, displaying the household ID (idhh), person ID (idpers), age (ph01b2x), gender (ph01c), citizenship (ph01d), marital status (ph01e) and parental leave status.

idhh	idpers	idtag	...	zhc76	...	zvc76	...	zgc76	...
123	1	1	...	cooking	...	at home	...	listening to radio	...
123	1	2	...	going shopping	...	in car	...	listening to radio	...
123	1	3	...	cooking	...	at home	...	talking on the phone	...
123	2	1	...	eating	...	restaurant	...	talking	...
...	...	...	...	...	...	...	...	...	...

Table 2.2: Entries from the GTUS 2001/2002 dataset, displaying the household ID (idhh), person ID (idpers), recorded day (idtag = {1,2,3}), main activity in time-slot 76 (zhc76), location or means of transportation in time-slot 76 (zvc76) and simultaneous activity in time-slot 76 (zgc76).

locations (zgc76) and simultaneous activities (zgc76). In Chapter 4 we use these two datasets by fusing them both into one table for feature extraction in the context of activity recognition.

2.2.3 User Monitoring

Monitoring hundreds to thousands of participants over a longer time-span, as it is being done with the time use survey approach, has in recent years become easier especially in mobile activity research. Some large-scale studies involving many participants being monitored continuously over weeks to months have been reported that are relevant in the context of this thesis. For instance, Do and Gatica-Perez (2011) describe an experiment involving smartphone-based monitoring of 40 participants over a year to mine for human interactions. This study was widened recently in the framework of the Lausanne Data Collection Campaign[11] to 200 participants. Another work also uses data from mobile phones of 215 subjects over 5 months to analyse

[11] http://research.nokia.com/page/11367, last access 09/2014

dwelling times, places, and mobility patterns (Eagle *et al.*, 2009). As wearable and ubiquitous sensors are harder to deploy, similar studies in this area have had far less participants, though some studies have monitored their participants for several weeks (Van Laerhoven *et al.*, 2008). A different approach was taken in Berchtold *et al.* (2010) by crowdsourcing data annotation for a wearable activity and context recognition, using the mobile phones of the participants.

2.3 SLEEP

In this section we will introduce the activity *sleep* and give insight into important aspects, such as circadian rhythms and how sleep is being assessed in a medical set-up. This knowledge is particularly important for understanding how to detect sleep with an activity recognition system. After an introduction to sleep itself, we will have a closer look at sleep studies in a sleep laboratory.

2.3.1 A Definition

About one third of our life we spend sleeping which sustains the fact that sleep is as important as proper nutrition and common exercising. It is an essential part of our life and has been identified to be crucial to our health for a variety of reasons. Sleep deprivation is known to lead to stress, a disturbed circadian rhythm, weight loss and, eventually, to death (Banks and Dinges, 2007; Siegel, 2009; Van Dongen *et al.*, 2003). Yet little is known about sleep which is why researchers are determined to conduct more research in that field.

Sleep is a rhythmic activity and reoccurs every day, according to the circadian rhythm which is the self-regulation of one's 24-hour cycle, including sleeping behaviours. Sleep manifests itself in two different sleep phases. The sleep phases consist of Non-REM and REM (*Rapid Eye Movement*), whereat the latter is also called the dream phase. The Non-REM phase is divided into three different sleep stages (S1 - S3) which depict the transition from falling asleep to deep sleep. While stages S1 and S2 are light sleep stages, S3 describes the deep sleep stage. Note that traditionally, a 4^{th} sleep stage (S4) is sparsely present in the dataset's polysomnography section, although the concept of this sleep stage was abandoned in 2012. During the night we go through different sleep stages peaking in the REM phase, in which we show no movement at all due to the shut-down of all body muscle activities. Solely the brain itself is still active and lets one experience vivid dreams. The typical characteristics of each stage can be captured by muscle contractions especially around the eyes (which is why it is called rapid eye movement). This is also primarily the reason why sleep cannot accurately be evaluated outside a sleeping laboratory environment: Without proper sensing systems to capture certain characteristics, it is very hard to determine when a person has really fallen asleep or when one passes over into a different sleep stage.

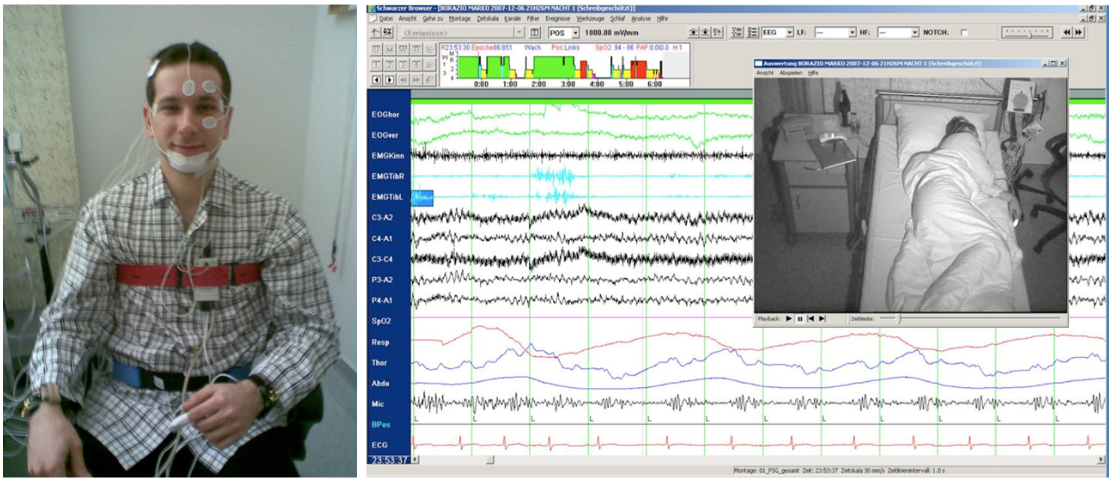


Figure 2.3: Left image: The author in a sleeping lab attached to 25 hard-wired sensors to capture sleep stages. Right image: The resulting sensor readings of one night, browsable in a sleeping lab tool, showing additionally to EEG and EMG a video feed of the patient while sleeping.

2.3.2 Polysomnography

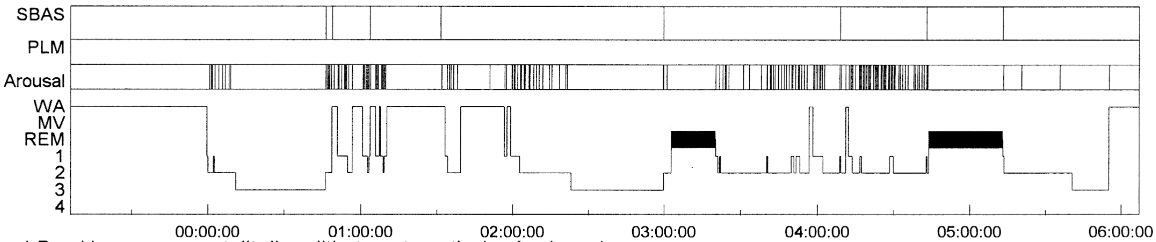
Polysomnography (PSG) is the golden standard for observing sleep-wake patterns. It incorporates multiple sensing modalities to capture relevant sleep information with typically 20, mostly wired sensors attached to the patient's face, torso and limbs (see Figure 2.3, left image, for an example of such a wired set-up). In addition, the patient is recorded by video for the entire night. As a result, relevant data can be accurately captured allowing in depth analysis. However, PSG is limited for short-term observations, that is, often a few nights are observed only, due to its cumbersome set-up. Such monitoring tends to be uncomfortable and less feasible over longer periods.

Sleep is assessed by using information that is obtained by electroencephalography (EEG), electrooculography (EOG) and electromyography (EMG) (Kushida *et al.*, 2005; Rechtschaffen and Kales, 1968). With such information and certain changes in the electrophysiological recordings it is possible to divide sleep into REM and Non-REM. Usually, such information is presented in a *hypnogram* which depicts time elapsed sleep stages for one night. Additionally, cardio-respiratory information, leg movements and a video feed are recorded during a stay in the sleep lab. Depending on the reason for assessing ones sleep, more or less sensors are attached to the body. Through the additional information gathered, other aspects of sleep during certain stages can be captured and evaluated. An example of such a recording is shown in Figure 2.3, right image. The figure depicts the analysed data after a stay in the sleeping lab, which can be browsed through with the standard software, showing for each timestamp the sensor information and the video feed of the patient sleeping.

Figure 2.4 shows an example of an hypnogram obtained by a healthy patient. It lists additionally information like *total sleep time* (TST), *sleep latency* or the percentage

Polysomnographischer Kurzbericht

Patient: Untersuchung vom: 17.6.2012
Aufzeichnung: 23:05:36 - 06:07:22 Bettzeit (TIB): 23:05:36 - 06:07:06
Aufzeichnungsprogramm: ( ) KRPSG ( ) KRP ( ) Sonstiges: ____________________



* Das Hypnogramm stellt die editierte automatische Analyse dar.

Schlaf Schlafzeit (TST): 302.5 min (72 %TIB) Anteil NREM1 : 10 %TST Anteil REM : 16 %TST
Schlaflatenz : 54.0 min Anteil NREM2 : 46 %TST Anzahl Arousal : 159
Latenz Schlafbeginn zu REM : 183.0 min Anteil NREM3&4 : 28 %TST Übergänge nach WACH : 10

PLM Index: PLM 0.0 n/h PLMS 0.0 n/h TSTe PLMS+Ar. 0.0 n/h TSTe Ausschlußzeit 00'00"

SaO2 Mittelwert 98 % Minimum 96 % 0 %TIB < 90% SaO2 0 %TIB < 80% SaO2
SaO2-Abfälle 0 Maximalabfall 0 % Maximaldauer 0 s Ausschlußzeit 10'18"

Herzf. Mittelwert 65 bpm Min.-Max. = 55-94 bpm 0.0 %TIB < 50 bpm 0.0 %TIB > 100 bpm
Bradykardien 0 (n) Tachykardien 0 (n) CVHR 3 (n) Ausschlußzeit 10'18"
Rhythmusstörungen: ____________________

CPAP Mittelwert 0.0 cmH2O Minimum 0.0 cmH2O Maximum 0.0 cmH2O Ausschlußzeit 421'30"

cmH2O	<3	4	5	6	7	8	9	10	11	12	13	15	>15	ausgeschlossen
%TIB	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
SBAS (n)	0	0	0	0	0	0	0	0	0	0	0	0	0	8

SBAS Index (n/h TIB): Apn.I= 0.0 Hyp.I= 0.4 ObS.I= 0.7 RDI= 1.1 Ausschlußzeit 00'00"
Anzahl in TIB (n): Apn.=0 (OA=0, ZA=0, GA=0, UA=0) Hyp.=3 (OH=3, ZH=0, UH=0) ObS.=5
Index (n/h TST): Apn.I= 0.0 Hyp.I= 0.6 ObS.I= 0.6 RDI= 1.2
Anzahl in TST (n): Apn.=0 (OA=0, ZA=0, GA=0, UA=0) Hyp.=3 (OH=3, ZH=0, UH=0) ObS.=3

Befund:

Figure 2.4: A typical hypnogram established usually after a few days of the stay in the sleep laboratory. It shows additional information such as total sleep time (TST), average heart frequency or arterial oxygen saturation (SaO_2). [With kind permission by the head of the sleep laboratory Dr. Steinmetz in the Marienhospital in Darmstadt, Germany and the patient.]

of the occurrence for each sleep stage during the night. In the hypnogram, the REM phase is marked by a thick bar to highlight this phase. Per night, each healthy person passes a fixed sequence of sleep stages. These *sleep cycles* are characterized by an average length of 90-110 minutes. Per night 4 to 6 sleep cycles occur for a healthy young male. Such a typical sleep cycle is shown in Figure 2.5. For more information on polysomnography in general we refer to Peter *et al.* (2007).

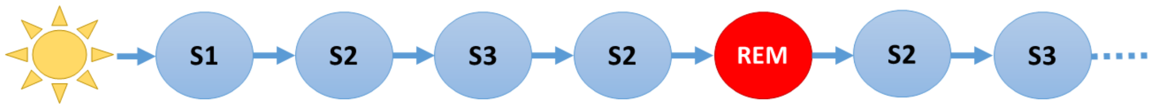


Figure 2.5: The typical sleep cycle for a healthy young male. The transition from awake (sun) to Non-REM1 (S1) usually lasts only a few minutes. The whole cycle of S2, S3, REM, S2 repeats itself 4 to 6 times throughout the night, with varying durations of each sleep stage.

The detection of sleep stages without the use of polysomnography is not trivial and has not yet reached a scientific standard. There exist portable polysomnographic appliances that are more comfortable to use since the patient can sleep in his or her common environment (Mykytyn *et al.*, 1999). Still, the electrodes and other sensors need to be attached to the body, usually by the patients themselves, which leads to faulty or incomplete data. This requires to repeat the polysomnography at home and is a burden for the patients. The patient still feels like a "marionette" (according to a sleep lab patient that has been interviewed during the course of this thesis) through the hard-wired sensors that keep him most of the time in one position. Additionally, the accuracy of a sleeping lab set-up is never reached. Such a portable solution is applied whenever the patient is not able to sleep in a laboratory or whenever sleep in the usual environment has to be captured.

Nevertheless, researchers try to find new ways not to replace polysomnography but to enable the investigation of sleep stages or certain sleeping characteristics by other means than polysomnography (Herscovici *et al.*, 2007; Kawamoto *et al.*, 2013; Walsh *et al.*, 2011). Other common techniques include non-wearable solutions that deploy sensors in the home of the patient. Bain *et al.* (2003) describe, for instance, a pressure mapping technology that could give an extremely detailed view on the total body posture of the patients throughout the night. Recent research in sensor networks (Hoque *et al.*, 2010) have offered similar fine-grained approaches to detect and monitor the patient's body positions and movements by active RFID-based accelerometers (WISPs) placed on the mattress. Camera-only methods have been explored as well (Liao and Yang, 2008), e.g., with the sleeper's motion detected by a night stand camera's frame-by-frame differences and estimates of body posture.

In Chapter 6 we will introduce the detection of sleeping postures which can be used as a sleep quality measure. In the following section we will give an overview of standard approaches to detect sleep with actigraphs for medical purposes and commercial devices for private use.

2.4 SLEEP DETECTION

The importance of finding out more about the way we sleep and whether our sleep is sufficient is not limited to traditional disciplines such as somnology, neurology or psychiatry: Providing a better picture on how well we sleep is relevant to everybody. Many off-the-shelf commercial devices can be bought for this purpose in a wristband form factor, from relatively compact devices such as the FitBit One[12], the Nike+ FuelBand[13] or the Jawbone UP[14] that are primarily aiming at fitness and activity tracking, to clinically evaluated devices such as the Actiwatch (Cambridge NeuroTechnology, Cambridge, UK, Kushida *et al.* (2001)). Some evaluations of commercial devices have shown that sleep information obtained from many such devices, like total sleep time (TST), is not sufficiently accurate for sleep disorder assessment, (e.g., Montgomery-Downs *et al.* (2012), which compares the FitBit to a commonly used actigraph for sleep evaluation). Such devices might be a benefit for private use but have been found to overestimate sleep by a large margin (Pollak *et al.*, 2001).

While PSG captures in-detail data during sleep over a single or a few nights only (see Section 2.3.2), this method is expensive, time-consuming, and the *first-night-effect* (Agnew *et al.*, 1966), i.e., a bad perception of sleep due to a novel environment, is inevitable. Therefore, alternative solutions, e.g., accelerometer-based wrist-worn devices that might provide additional long-term information on top of polysomnography, are being pursued and investigated as a complementary instrument. Actigraphy has been used in somnology as the de facto method for observing long-term trends that become only evident during weeks or months (Welk *et al.*, 2004). These devices are not necessarily deployed by every sleeping lab, since they tend to be expensive to acquire, to maintain, and to replace. Nevertheless, given such long-term capabilities, actigraphy then also captures activity levels during the day, resulting in a more holistic view on human activity.

In this section we will introduce the commonly used sleep detection algorithms and then describe commercially available products and their benefits in regard to sleep detection.

2.4.1 Algorithms of Clinically Tested Devices

Several research groups have investigated the use of actigraphy for sleep disorder assessment such as sleep-wake disorders, sleep-schedule disorder, periodic limb movement (PLB), narcolepsy and sleep apnoea (Jean-Louis *et al.*, 1999; Morgenthaler *et al.*, 2007; Sadeh *et al.*, 1995; Sadeh, 2011). The results indicate that actigraphs can be used in addition to polysomnography, especially if it is important to monitor the patient in his or her usual environment, over longer stretches of time, or in paediatric treatment. The actigraph approach typically captures levels of activity on a minute-by-minute basis. The data returned consist of *activity counts*, a measurement that has not been standardized across devices, and interpreted by each actigraphy manufacturer individually, making it challenging to compare the different algorithmic approaches with each other. It is most commonly implemented as a wrist-worn device that can be easily deployed and worn without any additional effort. However, in order to interpret the data captured, patients are usually required to annotate the sleeping and awake times (Littner *et al.*, 2003). Nevertheless, actigraphs are also

[12] http://www.fitbit.com, last access 09/2014
[13] http://nikeplus.nike.com, last access 09/2014
[14] http://www.jawbone.com, last access 09/2014

known to be less accurate in detecting wake segments during sleep and for sleeping disorders that exhibit vast amounts of such motionless periods such as in insomnia (Lichstein *et al.*, 2006).

Based on the activity count measure, two validated algorithms have been introduced in previous research that calculate sleep parameters as well as sleep-wake cycles in actigraphs: (1) Oakley (1997) is used for the Actiwatch (Cambridge NeuroTechnology, Cambridge, UK, Kushida *et al.* (2001)), and (2) Cole *et al.* (1992) is the basic approach for the Mini-Motionlogger actigraph (de Souza *et al.*, 2003). For both algorithms, the sleep-wake cycle is calculated offline, requiring the data to be downloaded after recording. Many sleep studies make use of these formerly mentioned devices, detailing how accurate these devices can detect sleep for a large variety of disorders (Benson *et al.*, 2004; Jean-Louis *et al.*, 2001a; de Souza *et al.*, 2003; Weiss *et al.*, 2010). These algorithms also form the basis of many novel devices that are equipped with a 3D MEMS (Microelectromechanical System) accelerometer, as opposed to the traditional actigraphs that contain an omni-directional accelerometer.

Cole et al.

Cole et al. in their approach make use of the *zero-crossing* technique (Jean-Louis *et al.*, 2001a) to calculate first the activity counts for a specified *epoch*, i.e., a time interval in which activity counts are being calculated. The activity counts per epoch are used to determine the total activity count D by considering a 7-minute window according to the following equation:

$$D = P \cdot \sum_{i=-4}^{2} A_i \cdot W_i \tag{2.3}$$

which essentially detects sleep whenever $D < 1$. In this formula, P is the scaling factor and W the weighting for each activity count, calculating the weighted sum over the epochs 4 minutes prior (A_{-i}) and 2 minutes after (A_{+i}) the current epoch (A_0). The common parameters according to Takeshima *et al.* (2014) are:
$P = 0.0033$, $W_{-4} = 1.06$, $W_{-3} = 0.54$, $W_{-2} = 0.58$, $W_{-1} = 0.76$, $W_0 = 2.3$, $W_{+1} = 0.74$ and $W_{+2} = 0.67$.

Oakley

Oakley presented a similar approach in his paper (Oakley, 1997) to detect sleep and wake phases, making use of amplitude-based activity counts. The algorithm examines the epochs in the 2 minutes before and the 2 minutes after a scored epoch:

$$A = \frac{1}{25}a_{-2} + \frac{1}{5}a_{-1} + 2a_0 + \frac{1}{5}a_{+1} + \frac{1}{25}a_{+2} \tag{2.4}$$

where A is the total activity count for the scored epoch a_0, a_{-x} is the activity count before the scored epoch and a_{+x} the one after, with $x \in [1,2]$ minutes. Each surrounding epoch is multiplied by a weighting factor ($\frac{1}{25}$ and $\frac{1}{5}$). The *sensitivity threshold* for A can be set to high (80), medium (40) or low (20) sensitivity, detecting sleep whenever $A <$ threshold. Low and medium sensitivity thresholds correlate to a high degree with sleep estimated by polysomnography (Kushida *et al.*, 2001).

We will focus in this thesis solely on the algorithms and their ability to accurately detect sleep and wake phases, from *any source* of inertial sensor data. In Chapter 5 the aforementioned algorithms are used to evaluate a novel sleep detection approach by using a raw 3D MEMS acceleration sensor.

Several studies are dedicated to the detection of body posture and movements during the sleep, motivated especially by sleep apnoea (Hedner *et al.*, 2004) and as a tool to measure for sleep quality (Liao and Yang, 2008). This thesis addresses in Chapter 3 the long-term challenges in particular by integrating methods for night segmentation, posture clustering, and myoclonic twitch detection, in order that these can be applied in behavioural monitoring.

2.4.2 Other Commercial Products

Accelerated developments in the use of inertial sensors in cars and personal computing devices has led to the introduction of many commercial inertial loggers that can be worn around the wrist for several weeks at a time, and which are used to monitor both sleep and physical activity of the wearer. Most of these products are intended to be used in preventive health care scenarios by the users themselves to track and quantify their lifestyle. Verification of these commercial devices for clinical trials is rarely a priority. These devices generally estimate the times when the user is asleep, and several also contain models of sleeping cycles and individual stages (such as REM and Non-REM). These models and their different hardware solutions are mostly closed-sourced, which makes the validation of the used algorithms challenging.

Several recent wearable products have been targeting sleep phase detection specifically in order to allow the wearer to wake up at a more convenient sleep stage, or display sleeping trends for the users so that they can keep track of their own circadian rhythms. The most prominent are summarized below:

- The Sleeptracker[15] is a wristwatch-shaped unit that apart from telling the time, also infers whether the user is in deep sleep, light sleep, or awake, using an accelerometer.

- The aXbo alarm clock[16] is packaged as a stand-alone application in the form of an alarm clock that wirelessly communicates with a wrist-band unit. It wakes you up in the optimum sleep phase by evaluating online the sensor data from the wrist-worn device.

- Somnus sleep shirt[17] is a "smart shirt" that embeds sensors in the garment to measure respiratory patterns. With such knowledge, the detection of REM and Non-REM are feasible according to the study in Lipoma *et al.* (2011).

[15]SleepTracker: http://www.sleeptracker.com, last access 09/2014
[16]aXbo: http://www.axbo.com, last access 09/2014
[17]Somnus: http://nyxdevices.com, last access 09/2014

- The BodyMedia SenseWear^{TM}Armband[18] is used to monitor ones activity, especially during workouts and while resting, e.g., sleeping (Sunseri *et al.*, 2009), and is worn on the upper arm. Additional information, such as how much activity has been performed or a step counter are accessible as well, enabling the user to keep track of his fitness status.
- The FitBit One[19] is an inertial sensor-based device that can be clipped to clothing or an arm strap, and comes with software to extract basic sleep information. Additionally, it provides fitness status and motivates the user by pre-defined goals like reaching a certain amount of steps per day.
- The Nike+ FuelBand SE[20] is designed as an activity tracker to be able to compare every day's activities to other people in the Nike+ community. It does not provide sleep analysis but can track your daily sleep duration.
- The Samsung Gear Fit[21] is not a classical sleep detection device but rather a fitness tracker with smartwatch abilities. With such, it provides the possibility to track fitness trends and sleep itself by the use of certain sleep apps on the device (e.g., S Health Sleep).

Unfortunately, a minority of the above products reveal details on how night segments are calculated from the basic actigraphy log, making their detection mechanism hard to reproduce. Additionally, many devices are not fit to be used in sleep studies since they tend to overestimate sleep, as researchers in Montgomery-Downs *et al.* (2012) have shown.

Not discussed here but interesting to note is the research (Chen *et al.*, 2013; Lane *et al.*, 2011; Natale *et al.*, 2012) and the vast number of applications (such as Sleep Cycle[22] or Sleep as Android[23]) for detecting sleep on the smart phone. Such an approach makes sense in regard to the progress in smart phone technology but it has still not been properly evaluated scientifically. Studies in Natale *et al.* (2012), for example, suggest the use of smart phone sensors to track total sleep time but reach only the same estimation accuracy for sleep parameters as common actigraphs.

[18] SenseWear: http://sensewear.bodymedia.com/, last access 09/2014
[19] FitBit: http://www.fitbit.com, last access 09/2014
[20] Nike+: http://nikeplus.nike.com, last access 09/2014
[21] Gear Fit: http://www.samsung.com/us/mobile/wearable-tech/SM-R3500ZKAXAR, 09/2014
[22] http://goo.gl/3SYG3z, last access 09/2014
[23] http://goo.gl/pP9EPs, last access 09/2014

3 PLATFORMS FOR MONITORING ACTIVITIES

Contents

3.1 Introduction **34**
3.2 Wrist-worn Sensor platform **35**
3.2.1 The HedgeHog 35
3.2.2 Application Examples 36
3.3 Mobile Recording Platform **37**
3.3.1 Usability of Smart Phones - Method Overview 37
3.3.2 User Study 42
3.3.3 Evaluation and Results 46
3.3.4 Monitoring User-Behaviour: A Discussion 50
3.4 Sleep Monitoring System **52**
3.4.1 Challenges for Sleep Monitoring 52
3.4.2 Unobtrusive Monitoring System 53
3.4.3 Method Overview 54
3.4.4 Privacy Issues 55
3.4.5 System Discussion 56
3.5 Conclusions **57**

WEARABLE SYSTEMS have gained an increasing amount of popularity since the use of sensors in modern devices increases. The usage of these devices for recording sensor data seems reasonable, since they are equipped with several different sensing modalities but is often limited due to the closed-source format, as described in Chapter 2.4.2. Therefore, researchers tend to establish their own sensor platforms that suit their study set-up.

In this chapter we first introduce a *wrist-worn sensing unit* that has been applied to many scenarios throughout this thesis. With such a system we are able to record long-term sensor data to detect activities and with such activity rhythms. Additionally, we determine the smart phone usage of users with the aforementioned device, in regard of having their phone *carried on the body* or not. Such an evaluation is crucial since the usage of smart phones in general for recognizing activities has not been fully explored yet. In addition to that, we show how a home set-up for sleep studies looks like, by again using the wrist-worn sensor and adding an infrared camera to the system to capture *sleep rhythms* and its *characteristics*.

3.1 INTRODUCTION

Activity recognition requires a certain amount of sensor data in order to detect an activity's characteristics. In this chapter we focus on finding suitable sensor platforms that can record the needed sensor data that is later on used to recognize activities. The choice of a sensor platform influences the quality of the obtained sensor data and therefore has to be chosen according to the task it should accomplish. Despite many research efforts into creating platforms or even environments for activity recognition, many studies use a laboratory set-up to record sensor data instead of making use of the sensor systems in real-world scenarios. The urge to perform such experiments "in the wild" has been described in Rattenborg *et al.* (2008), showing that animals behave differently when observed in the wild than when they are locked-in in a zoo.

The main contribution of this chapter is to evaluate sensing platforms that can be used in real-world set-ups. More specifically, we contribute an in-depth analysis of smart phone usage for activity recognition tasks. The usage of these general-purpose platforms is experiencing an unprecedented uptake: In 2013, almost 1 billion devices have been sold worldwide[24]; In the year before, a recent publication by the United Nations Organization claimed that more people worldwide had access to a mobile phone than to a clean toilet[25]. Additionally, approximately 6 billion people in the world have access to a mobile phone. Statistics show a steady increase in the number of smart phone owners all over the world, indicating also that user behaviour has been gradually changing over the past years, with smart phone usage topping that of desktop computers. At the same time the phones are being used more frequently by the user. Since the acceptance of these mobile devices has become so high, they have been targeted increasingly to be used as devices for self-monitoring (Bardram *et al.*, 2012; Bielik *et al.*, 2012; Oresko *et al.*, 2010). These trends have caused researchers to study to what degree these mobile devices have become suitable for activity recognition and user monitoring, in several studies over the past years (Dey *et al.*, 2011b; Patel *et al.*, 2006). The results of these studies suggest that the phone is within the user's arms reach about half of the time, indicating that the use of a mobile platform may not be suitable for any user monitoring application.

In addition to a mobile platform, we introduce a novel domestic sleep monitoring system that consists of an infrared (IR) camera and a wrist-worn device to determine when a person is sleeping. With such a set-up we are able to record data over a long period of time, which enables the detection of sleep postures that are interesting in many different areas like personality evaluation[26] or obstructive sleep apnoea (Oksenberg *et al.*, 2009). The emphasis here is put on the sensing modality and a high performance on recording data for a longer timespan.

The remainder of this chapter is structured as follows: In Section 3.2, we introduce a wrist-worn sensing platform that has been used not only in this chapter but throughout the whole thesis for various studies. Following that, in Section 3.3, we

[24]http://goo.gl/mcTmha, last access 09/2014
[25]http://www.un.org/apps/news/story.asp?NewsID=44452, last access 09/2014
[26]http://news.bbc.co.uk/2/hi/health/3112170.stm, last access 09/2014

use that wrist-worn sensor to determine when a smart phone is with the user, i.e., being carried on the body. Finally, in Section 3.4, we present our sleep recording system that does not only capture sleep itself but also certain sleep characteristics.

3.2 WRIST-WORN SENSOR PLATFORM

The use of wearable recording systems equipped with different sensors has been investigated for various scenarios. Usually, an expensive sensor platform has to be bought but other researchers tend to build their own devices that suit their current study. Throughout this thesis, we will make use of such a self-built sensor platform - called the HedgeHog[27]. The device was introduced by Laerhoven and Gellersen (2004) and revised ever since to meet the requirements of the activity recognition studies:

1. **Unobtrusiveness.** Users of a wearable platform prefer systems that are lightweight and small. Additionally, the system should be worn comfortably for a longer period of time without limiting the user in his or her movements.
2. **Long-term recording capabilities.** The wearable system should be able to record and store the sensor data over several weeks. Additionally, the battery should last at least that long, with the possibility of recharging it.
3. **Data recovery.** The possibility of extracting raw sensor data without any preprocessing should be given.

In the following sections we introduce the device and examples of the application of such a sensor in previous studies.

3.2.1 The HedgeHog

The HedgeHog is a self-contained wrist-worn device that records 3D acceleration data with a default sensitivity range of $\pm 4g$ that are sampled at $100Hz$, as well as ambient light readings and time and calendar information. The on-board accelerometer, the ADXL345 from Analog Devices, can be reconfigured from sensitivity ranges from $\pm 2g$ up to $\pm 16g$ and supports sampling rates of several thousands of samples per second. The whole unit fits in a plastic enclosure that protects the module to prevent damage from falls and accidental splashes of water. The device has to be removed when showering or swimming. It is small enough to be worn comfortably on the wrist and is attached with an elastic strap to it.

It can record sensor data on a local flash storage (a removable microSD card) for about 2 weeks before its rechargeable battery is depleted. Figure 3.1 depicts the prototype with and without enclosure and straps. With the OLED display mostly powered off (the wearer can request the current time by double-tapping the watch),

[27] http://www.ess.tu-darmstadt.de/hedgehog, last access 09/2014

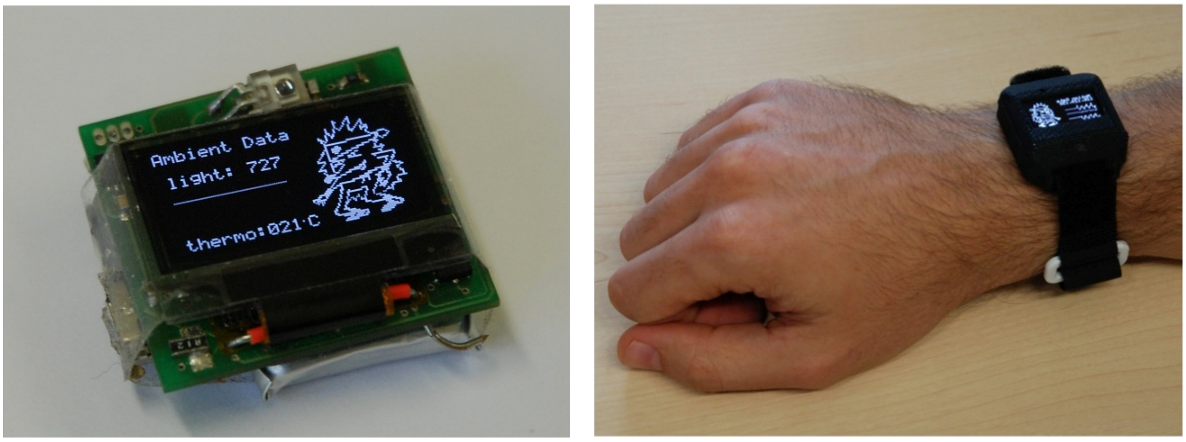


Figure 3.1: The wrist-worn prototype is intended to be worn continuously and is able to function as a basic wristwatch (via its OLED display and integrated realtime clock). It stores data from a 3D accelerometer at $100Hz$, as well as ambient light data, on a local microSD-based flash memory.

the sensor unit runs for about 2 weeks on a 600mAh Li-Ion rechargeable battery at $100Hz$, which can be extended significantly by lowering the sampling rate. When the data is uploaded, the USB connection provides additionally power to an on-board recharging circuit connected to the battery. With this prototype, continuous logging is feasible with 2 week intervals for charging and downloading the data. It is important to note that nothing more is required from the wearer of the sensor after starting and attaching it: Apart from its double function as a wristwatch, the user is normally not required to press buttons, annotate data, or fill in questionnaires.

3.2.2 Application Examples

The HedgeHog has been used in many different research scenarios. We will name a few here to underline the feasibility of the device to be used in long-term activity recognition studies. The most prominent ones cover research areas like leisure activity recognition (Berlin and Van Laerhoven, 2012a), for which the user was wearing the device while performing certain activities like Zumba (a martial-arts fitness mix), playing the guitar, or housekeeping recognition (Stikic and Van Laerhoven, 2007). In addition to that, the device was used to detect sleep postures (Van Laerhoven *et al.*, 2008), daily patterns in long-term recordings (Van Laerhoven *et al.*, 2008) and in the determination of train types (Berlin and Van Laerhoven, 2012b). For the latter purpose the sensor was attached directly to the railway track which yielded different train patterns, depending on the train that passed the track. The various fields of applications show the validity of the sensor platform, as well as the capability of using it in everyday activity recognition areas. In Scholl and Van Laerhoven (2012) for example, the HedgeHog was used to determine if a person was currently smoking or not in order to urge people to cease smoking for a better well-being.

In this thesis we use this sensor in two types of studies: First as a typical wrist-worn sensor unit to assess if the user is moving with his smart phone. Second in the detection and characterisation of sleep by modelling sleep postures and certain non-natural movements that occur while sleeping. In the following section we introduce another sensing platform and evaluate its usability in the context of activity recognition.

3.3 MOBILE RECORDING PLATFORM

Mobile phones have become generic and personal computers that fit in the user's pocket. They are used by an enormous number of people around the world. We use them to manage our schedules and appointments, as music players, navigation systems, gaming platforms and to obtain updates on daily news[28]. Their computing capacity is rising steadily, enabling the current generation of smart phones to be used for various research experiments (Berchtold *et al.*, 2010; Farrahi and Gatica-Perez, 2008; Kwapisz *et al.*, 2011; Sahami Shirazi *et al.*, 2013). Such mobile devices have been deployed especially for sensing the user's immediate surroundings and to recognize the physical activities of the user (Ashbrook and Starner, 2002; Brezmes *et al.*, 2009).

In this thesis we present a novel study approach that estimates how frequently the mobile phone is *on the user*, i.e., being carried by the user. For this purpose, we make use of a wrist-worn, accelerometer-based unit (see Section 3.2) which registers the user's physical movements, along with a custom-built Android application that enables us to log from the sensing modalities the smart phone has embedded (see Figure 3.2). Both were designed so that they would minimally impact the participant's phone use and maintenance behaviour (e.g., recharging or otherwise interacting with both devices), and that they would allow also continuous deployment with the user for up to two weeks.

With this system, we carried out a study in which 51 participants installed our Android application on their personal phone and were asked to continuously wear a wrist-worn accelerometer logger over the course of one to two weeks. This resulted in a total of 638 days / 15,300 hours worth of mobile and wearable accelerometer data, reflecting how often the phone and the wrist-worn unit were actively worn by the 51 users. The results are analysed in this section for the span of time the two device types were worn but our investigations also show more in-depth analysis on how different users manage their mobile and wearable devices, and how consistent (or variable) these different behaviours are between users.

3.3.1 Usability of Smart Phones - Method Overview

Whether a mobile device is suitable as a continuous sensing platform, and whether it stays in the user's proximity, has been investigated before by Patel *et al.* (2006)

[28] http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats, last access 09/2014

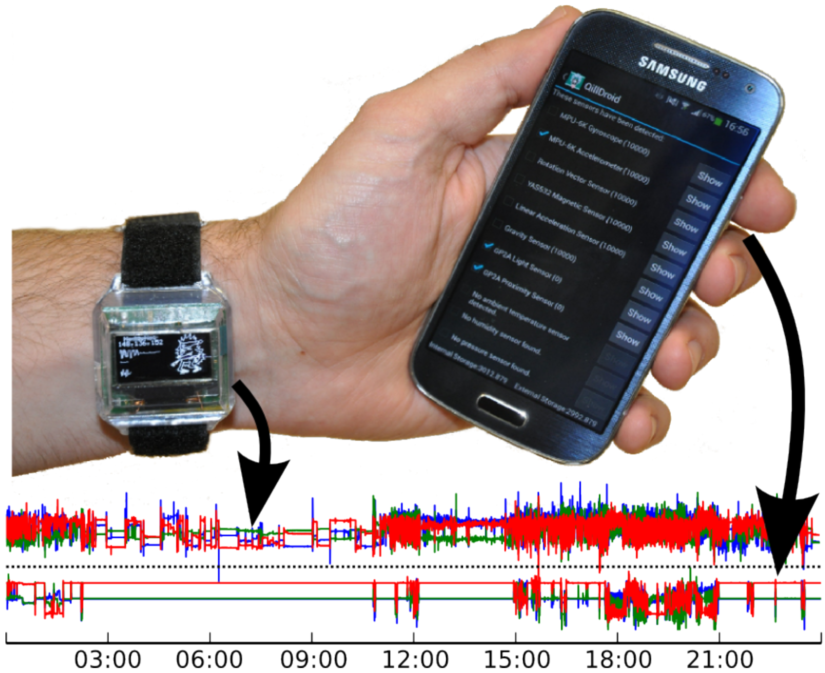


Figure 3.2: Our study uses a wrist-worn device (left, top plot) and an Android application running on the user's smart phone (right, bottom plot), that both record 3D inertial data. Our method compares these to estimate when the phone was *on the user*.

and followed up by Dey *et al.* (2011b). Both studies conclude that users are farther away from their phones than one might expect, by making use of the received signal strength of a neck-worn Bluetooth token to record its distance to the mobile phone as *within arms length*, *within room* or *no signal*, based on calibration data. The study's findings suggest that the phone is *within arms length* less than 50% of the day, *within the room* for about 65% of the time and switched off for most of the remaining time. Interestingly, the portion of the day for which the phone is *within arms length* seems to decrease from 2006 to 2011 while the amount of time the phone is *in the same room* has increased. In addition, both studies indicate that the proximity of smart phones to their users has not changed significantly in the meantime. Since both of these studies had a relatively small user base focused on North America, these findings may vary elsewhere and may have changed in the past years. Additionally to the proximity evaluation, both studies recorded a vast amount of sensor data from the phones and users were interviewed in order to produce a journal about their activities during the experiment. With such a journal, the user behaviour was structured into 15 to 20 classes of activities each related to one of the three smart phone distances. Using a decision tree, the recorded data were matched to one of these classes. An accurate prediction was reached by ranking the different features based on the ground truth data using the Bluetooth tokens. Interestingly, the study did not find a "one-fits-all" decision tree: The ranking of the single features for an accurate decision differed from participant to participant.

This thesis presents (1) an alternative method to those presented in Dey *et al.* (2011b) and Patel *et al.* (2006) to research user proximity to their phone, by requiring study participants to wear an accelerometer-based logger on the wrist and install an accelerometer-logging application on their Android phones. As a second contribution, we present (2) a study that uses this system with 51 participants (almost double the size of Dey *et al.* (2011b) and Patel *et al.* (2006)). Our proposed method depends on two sources of information that need matching: (1) A miniature wrist-worn sensor that records the user's motions, and (2) a sensor data recording application for Android. In this approach, the data measured by the wrist-worn unit serve as an indication on when the user was physically active, while the data recorded by the mobile phone characterize when the mobile phone was experiencing acceleration. A comparison of both could therefore result in estimating when the phone is experiencing the same acceleration as its user, and therefore when the mobile phone was on the user. This effectively means that the proximity measure for our method will be restricted to *on the user* or *elsewhere*, yet we argue that this measure in itself is already interesting for research, and that our method does have significant advantages over the wearing of Bluetooth transceivers.

Recording Platforms

In this section we will present the description of the respective information sources.

The wrist-worn unit was configured for our experiments to record at a sensitivity of $\pm 4g$ and a frequency of $100Hz$ (i.e., a 3D acceleration vector every 10 milliseconds). Once the data are uploaded after the study period (via USB), they have to be converted to acceleration values in g for comparing the data later on to the mobile phone values. An example of such raw values gives an impression on when participants have been moving or not, as depicted in the top plot of Figure 3.3 for a time period of 24 hours. While sleeping, for instance, (here between 03:00 and 12:00) the inertial data exhibits significantly less movement, with the data changing only whenever the user is transitioning between sleep postures.

The Android application has been developed based on the Android framework to compare the inertial data from the wrist to those from a phone, therefore recording similar data with a mobile phone, using its built-in sensors. The application is compatible with phones running Android 2.3.3 or higher (covering a majority of Android phones) and can record from all sensors available within the Android sensor framework. In Figure 3.4, the menu for selecting the sensors' channels that can be logged are shown with the possibility to define a maximum storage size that is reserved for the application to store the data directly on the internal memory of the phone. The sensor data is directly put into an SQLite database as provided by the Android framework.

The first challenge that is met when developing an Android application is that it has to be installable on various Android versions and should operate robustly. With that in mind, we encountered a first obstacle in the Android framework, since its

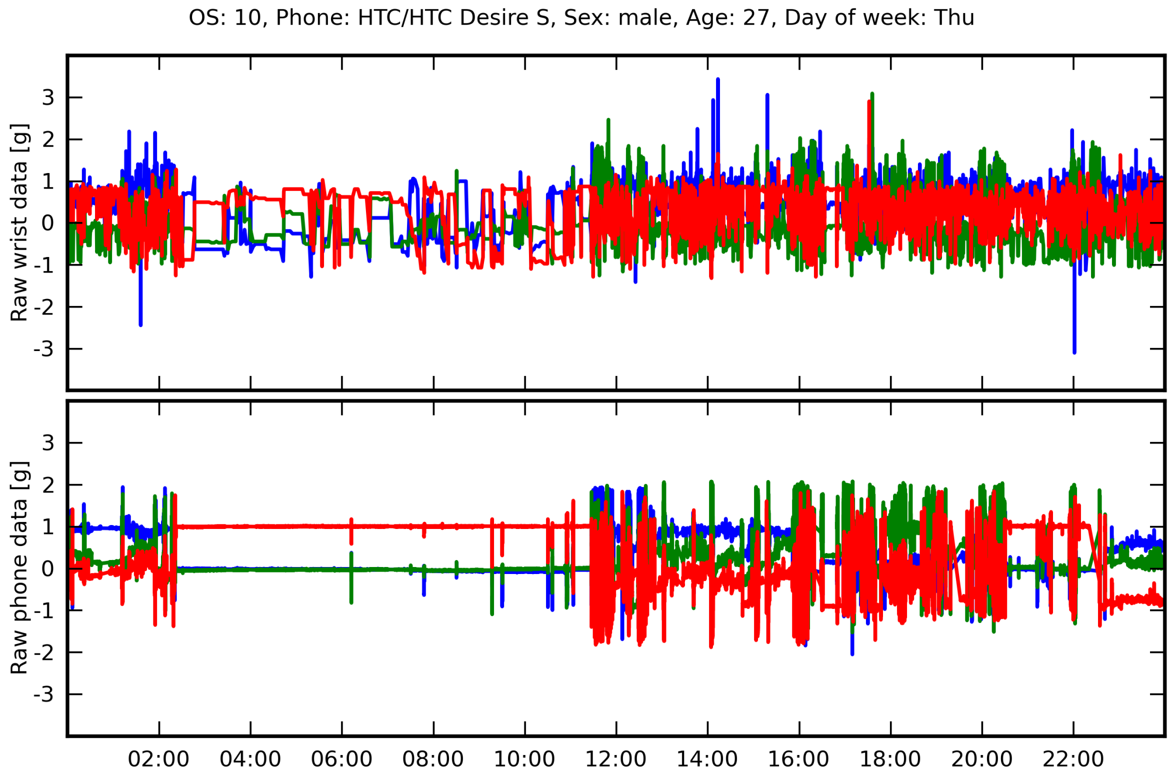


Figure 3.3: Raw acceleration data from a male participant (age 27), along with the acceleration data from his smart phone. Note here, that this participant put his phone on the mattress while sleeping, visible in the bottom plot by peaks in the phone data (6:00 to 11:45) that correlate with some of the sleep pose transitions in the wrist-worn's data (top plot).

policy for management of resources can have different priorities for processes and their threads. The policy provides several priority levels that will influence scheduling of the processes and threads. Additionally, Android distinguishes between processes in the foreground and processes in the background, which is important to consider for our application. A foreground process is, for example, an application that the user can currently interact with, like an on-screen application. A background process is, on the other hand, an application that is not actively being used, e.g., file downloads or other processes that do not need user interaction. In resource critical situations, the processes will be handled differently by the framework, such as being shut-down automatically. For this reason we implemented our recording software as a background service, so that it does not impact the recording software according to the user's phone behaviour by using a *partial wake lock*, which basically requires only the CPU to stay awake. With such a design of recording data in the background it is possible that periods of time exist when the application is not recording any data. This was especially the case with older phone models with limited processing resources (like single-core processors) and while users were talking on the phone. Therefore, we implemented our Android application to automatically restart

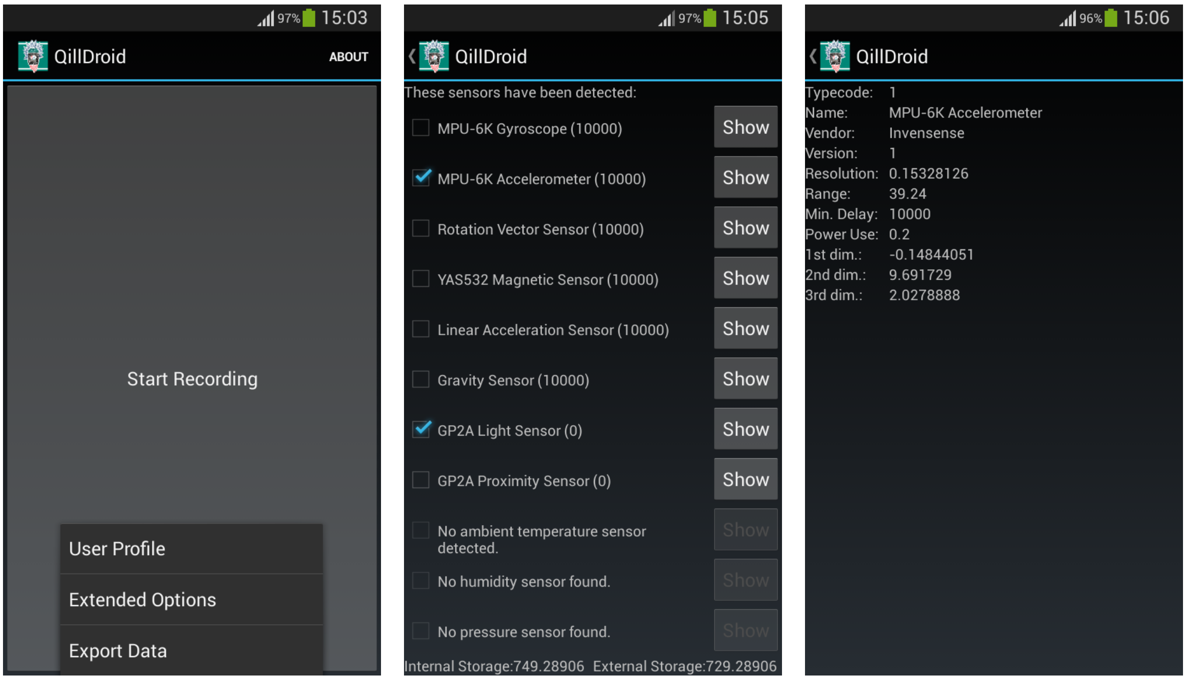


Figure 3.4: Android recording application for all sorts of sensors detected on the smart phone. Apart from starting the application (left), the menu allows to choose from different sensing modalities to be stored in the database internally (middle image, green check mark). Additionally, the current sensor readings can be displayed directly (right).

itself after an unexpected shut-down. This impacted only slightly the selection of participants in our study, as most participants owned newer Android phone models.

The Android sensor framework uses the International System of Units (SI, $\frac{m}{s^2}$) instead of g for the inertial data, which is stored directly in the database to ensure that no accuracy is lost. Preliminary tests showed that a reliable sampling rate, like the one of the wrist-worn sensor, is difficult to obtain in the Android framework ($100Hz$ being unobtainable). Some exemplary data recorded with the Android application are shown in Figure 3.3 bottom, along with the previously-discussed wrist-data. The comparison of both plots in Figure 3.3 already allows to make a coarse-grained inspection about when the phone might have been with the user or not. Immediately noticeable is that the phone data exhibits far more 'flat', motionless segments than the wrist data. This particular user was carrying the phone in the front pocket (as approximately 57% of males tend to: Ichikawa *et al.* (2005); Steinhoff and Schiele (2010)) during the day and on the mattress during the night but one could ask oneself whether this is representative for most phone users.

Data Comparison

For this study, we rely on motion from the smart phone and the participant's wrist. However, the data recorded by both platforms cannot be compared directly for the following reasons: (1) Both devices are carried in different positions and will therefore experience different motion patterns and force of acceleration. (2) With such, the axes of the sensor coordinate systems will unlikely be aligned to each other most of the time. (3) Data is recorded by both systems independently, since intervals between smart phone sensor readings (readings were time-stamped on the phone as they were obtained) tend to vary substantially, with the most robust rates obtained for $10Hz$, while the wrist-worn sensor records accurate equidistant 3D acceleration samples.

In order to compare the two datasets from both wearable and phone over longer stretches of time, we calculate the variance of the magnitude of the 3D acceleration vector over a one-minute time interval. By using the magnitude of the 3D acceleration we have to consider only one rotation-invariant scalar value. The variance has the advantage that it does not require calibration with respect to gravity, as the mean would have to. To estimate whether the wearable and mobile devices moved in conjunction, we defined different thresholds on the variances to detect movement segments. The impact of the threshold will be explored in detail in Section 3.3.3.

3.3.2 User Study

In order to illustrate our method, a study was held with 51 participants, in which they were asked to install our logging application on their Android phone and wear the wrist-worn sensor during a period of two weeks. This section will describe the details of the set-up for our experiment and will give an overview of the collected data.

Participants Recruitment

The 51 participants were recruited through a local poster advertisement campaign. As this was done in a university town, about half of them were university staff and students, with the other half including mostly housewives and office employees. The number of 51 was initially much larger but only one out of five persons that responded to the advertisement participated in the study. The reasons for not being part of the experiment were either because people did not respond after a first contact or they decided not to participate after a detailed explanation of it. Although the data in our method was stored locally on the device, especially students cited mostly privacy concerns as a reason for not participating. Interestingly, comfort concerns for wearing the wrist sensor were rarely mentioned for their decision not to take part in the study. Only two participants had to take off the sensor for a few hours due to being uncomfortable with it.

Observations. The participants' ages range from 14 - 62 years, including also one pupil and two senior citizens. In total, 10 female and 41 male participants participated in the study. The high number of male participants can be explained by the fact that we advertised the study at a technical university which is attended mostly by males.

user	age	gender	% wrist	% phone	#charges	#days
1	32	male	97.62	92.76	11	21
2	29	female	54.13	43.04	5	11
3	26	male	98.64	98.89	10	6
4	32	male	87.50	90.63	5	9
5	36	female	85.04	99.62	10	11
6	28	male	72.86	100.00	11	17
7	27	female	99.14	65.23	4	7
8	33	male	89.62	96.19	18	11
9	31	male	88.37	99.70	19	14
10	27	male	99.42	99.23	10	11
11	27	female	99.33	99.66	6	6
12	27	male	80.84	34.36	3	9
13	20	female	98.95	98.95	17	10
14	33	male	87.85	98.21	11	10
15	25	male	97.00	95.71	9	10
16	23	male	98.42	92.53	6	9
17	26	male	54.01	93.88	13	10
18	27	female	93.57	95.48	8	9
19	27	male	82.19	73.80	8	14
20	25	male	93.19	99.28	11	11
21	26	male	98.92	98.39	5	4
22	24	male	99.63	98.88	9	11
23	30	male	70.85	98.52	13	11
24	31	male	98.93	98.72	11	10
25	28	male	89.09	90.88	14	15
26	38	male	30.70	99.53	9	9
27	20	female	80.38	100.00	13	11
28	58	male	43.69	98.63	12	12
29	26	male	99.11	99.70	16	14
30	38	male	99.57	99.45	31	34
31	33	male	96.94	97.28	27	37
32	25	male	89.97	97.16	28	14
33	25	male	57.78	97.01	14	14
34	21	male	87.36	98.10	25	15
35	25	male	76.32	99.86	16	15
36	25	male	90.15	95.82	26	14
37	26	male	96.63	93.25	5	7
38	27	male	77.78	52.36	?	15
39	33	male	99.05	100.00	6	5
40	33	male	87.46	93.05	12	14
41	31	male	98.80	96.54	17	14
42	23	male	74.67	100.00	9	14
43	25	male	93.91	99.84	15	13
44	62	male	92.56	94.35	6	7
45	55	female	94.92	99.58	10	10
46	28	female	95.73	00.00	?	14
47	35	male	96.87	92.24	13	14
48	35	male	94.51	96.59	12	14
49	14	female	97.87	100.00	9	7
50	50	male	85.76	99.70	15	14
51	33	male	99.04	99.84	6	13

Table 3.1: Variety of participants for this study, including the information of how much data from both sides, wrist-worn sensor and smart phone, are missing in the dataset, as well as how often the phone was charged during the study.

The participants were asked to partake in this study with their personal (Android) phone. The study was advertised with the purpose of obtaining inertial data to detect daily activities afterwards, not telling the participants that we investigate the user's phone carrying habits, to avoid bias. We met with the participants three times during the study: An initial meeting explained the purpose of this study, showing the participants our privacy policy ensuring anonymization of the data after the trial would be completed. Additionally, the wrist sensor functionality was explained and the sensor handed out to the participants. In addition to wearing the sensor, we asked the participants to keep a journal of their sleeping times. A second meeting was held after one week to ensure that data had been properly recorded, followed a week later by a third meeting for returning the sensor, downloading all data from the smart phone. The participants' data was evaluated directly to show and explain the real purpose of the study. In addition, we conducted a post-study interview concerning wearing comfort of the sensor and their perception of how often, in their estimate, they carry their phone on the body. Additionally, we asked about the power consumption of the Android application.

Table 3.1 summarizes the demographic information on all the 51 participants that took part throughout the study, additionally showing the amount of data obtained from the wrist-worn sensor and the smart phone, as well as how often the user charged the smart phone during the study, plus the total number of days recorded. As expected, we gathered almost a continuous recording of smart phone data for all the participants. A few outliers (users 2, 7, 12, 19, 38 and 46) are visible, because either the application stopped recording due to the power saving mode of the phone (which is always switched on when battery power is low enough) or the phone running out of battery power. In these cases, participants switched off the application by themselves, unfortunately sometimes also forgetting to switch it back on again. Participant 46 represents an exceptional case: During the study she was cleaning up her phone storage and by accident deinstalled our Android application, which resulted in a swiping of the database entries. Nevertheless, we could obtain the wrist-sensor data, as shown in the table. Due to a not totally functioning smart phone, participant 38 had problems using his smart phone which is why it was switched off most of the time during the study (almost 50% of the time). This is probably also the reason why the charging status could not be logged by the Android application for this user.

Recording data with the wrist-worn sensor suffered from other obstacles: Obtaining almost 100% of the data over the recording time is almost impossible, since whenever the sensor is taken off long enough, mostly for showering or swimming, the recording is interrupted. For 27 participants, we nevertheless obtained almost 100% of recording over the study period. 5 participants (users 2, 17, 26, 28 and 33) found it uncomfortable to wear the wrist sensor during most nights, which is why we obtained such a significantly smaller portion of inertial data from their wrist-worn sensors. Additionally, most of the participants tended to take off the sensor on weekends for leisure activities or family celebrations (wearing the sensor with a shirt seemed too uncomfortable). Participant 26 was also on holidays while wearing

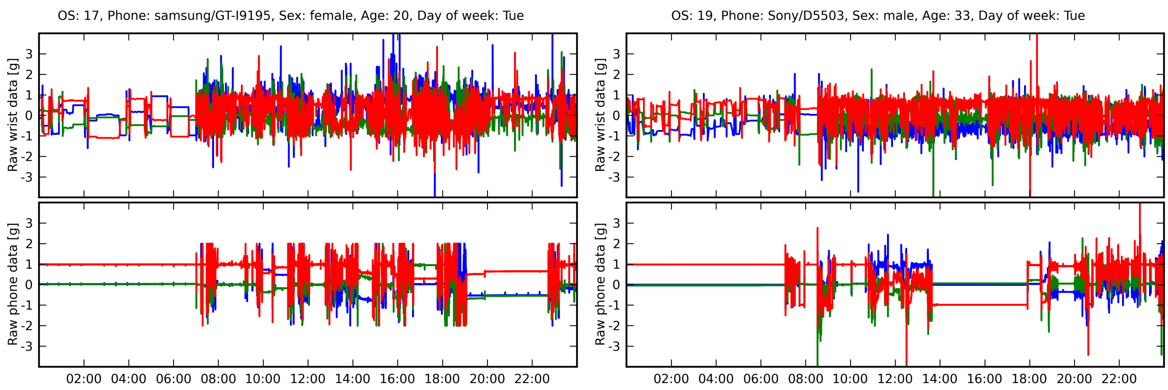


Figure 3.5: Raw acceleration data from a female (age 20, left plot) and male (age 33, right plot) participant for both wrist sensor data (top) and smart phone data (bottom). Both plots show the diversity in behaviours: The female participant carries her phone intermittently throughout the working day but less in the morning and evening; The male participant carries his phone mostly throughout the morning and in the evening.

the sensor, which resulted in data from the wrist-sensor of 30.7 % for the whole recording period of 9 days, showing again that wearing devices on the body is hard to accomplish when travelling for private reasons. Two of the study participants were willing to wear the wrist-sensor and log phone data for over five weeks, resulting in a recording time of 34 and 37 days for users 30 and 31 respectively. The coverage of obtained data from both modalities for these two participants was also remarkably high: For participant 30 we gathered over 99% of wrist and smart phone data and for participant 31 around 97% from both recording platforms. On average, we obtained 87.31% of wrist data and 91.22% of smart phone data from all participants.

Figure 3.5 shows examples of the raw acceleration data from two participants (left: female, age 20 and right: male, age 33). The top plots depict the wrist-sensor data and the bottom plots represent the smart phone's inertial data. We observe here two different phone carrying behaviours: The female participant carries her phone mostly during the day from approximately 7 to 19 o'clock while the male participant carries it mostly from the morning until the early afternoon and in the evening. Note here that the smart phone was used shortly before going to bed and again immediately after waking up by both participants. Most participants used their phones as an alarm clock.

Variety in Mobile Phones

The fact that participants were using their own phones throughout the study led to a high variety of different models, for which not all of the built-in accelerometer modules were previously known. Additionally, it was not guaranteed that the Android application would be running properly on all devices, since manufacturers tend to modify the Android OS according to their needs, leading to possible problems for recording continuously without interruption. Fortunately, we obtained most of the data with the Android application, as indicated in Table 3.1 by the amount of data recorded by the smart phone. The majority of the phones were from Samsung (23), followed by HTC (11), Sony (8), LG (6), Motorola (2) and Huawei (1). The Android OS installed on the smart phones varied from 10 to 19, with platform version 10 corresponding to Android OS 2.3.x (9 participants), platform versions 14 and above to Android 4.x.x (OS 15 = 5, OS 16 = 9, OS 17 = 8, OS 18 = 7 and OS 19 = 13 participants). Note here that Android OS versions 11 to 13 are only installed on Android Tablet models and therefore are not present in this study. None of the above mentioned Android phones or OS versions caused problems in the application, which is why recording data over a two week timespan was feasible. As already mentioned, two participants used the application for over a month without any drawbacks in their smart phone usage.

3.3.3 Evaluation and Results

To estimate whether the phone is *on the user*, our method compares each of the detected motion segments from the wrist sensor to the motion segments present in the phone data. For this purpose, we first needed to obtain a proper threshold for both datasets to detect these motion segments. We first discuss the chosen parameters and their effect, and then present the wearable-phone correlations obtained in the data.

Threshold Selection

Key to our method is the choice of a proper variance threshold for the detection of motion in the one-minute windows. We believe that there should be one best threshold that detects accurately all motion segments over all devices and OS Android versions. Therefore, we aimed to set a threshold that applies for all models, with which we primarily filter out noise and artefacts due to the unstable logging frequency which the Android framework delivers[29].

Essentially we made the following assumptions to determine the thresholds for the entire study:

1. **A phone never moves without its user.** We assume this to be true most of the time when using large and long-term datasets. Nevertheless, a user might lend his phone to someone else or leave the phone where it experiences motion (e.g., a stationary phone that vibrates due to a received message or email might generate phone motion without the user moving). This might, however, lead to a bias in small studies, though we haven't found such occurrences in our dataset.

2. **The phone does not move while charging or during the user's sleep.** Although it is imaginable that users sleep in means of transport (e.g., bus or

[29] Android only takes a *desired* delay between sensor readings as a configuration parameter.

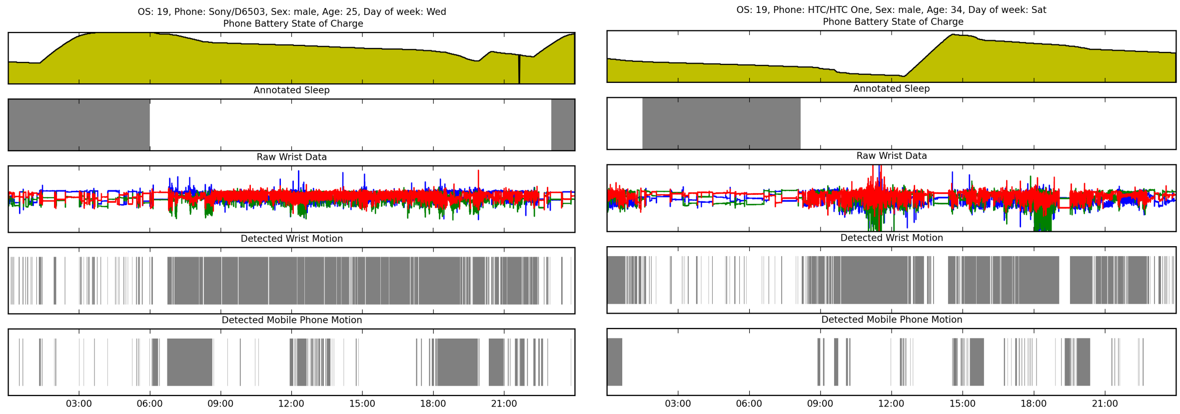


Figure 3.6: Results for two male subjects for one day showing from top to bottom: Battery charge status, sleep as annotated by the participant, raw acceleration data from the wrist-worn sensor, motion as detected by the wrist-worn sensor and the thresholded detection of movement according to the smart phone's acceleration data. The left plot shows a participant who put his phone on the mattress while sleeping as depicted by the black bars that represent the movement per minute.

plane) and could be charging their phone in transit (e.g., in the car or in a train), none of our study participants was found to have done so.

With these assumptions, we found by experiment that a threshold of $0.000012g$ delivered the best results among all participating devices: Higher thresholds do not increase the precision but lead to a further drop in recall, meaning that motion of the phone might have been missed. Additionally, lower thresholds tend to detect phone motion where there is assumed to be no motion, e.g., when the participant is sleeping.

Mobile Phone vs. Wearable Data

We applied the threshold on the variances of the inertial data using a window size of one minute. A first glimpse at estimated movements on both the wrist and the phone data is given in Figure 3.6. It shows for two male participants in each plot from top to bottom: The battery charging status, the sleep segments as annotated by the participant, the raw inertial data from the wrist sensor, the detected movement segments from this sensor and the detected movement intervals of the smart phone data for one day. In the left plot, several motion segments are immediately visible during the night, although one might expect that the phone should not be moving. During the study, this participant put the phone on the mattress while sleeping, which is why we observe motion during sleep in the smart phone data. Additionally, we see for both participants that the wrist sensor recorded far more motion segments during the day in contrast to while sleeping, which makes sense since we move mostly during the night when we change postures. Note here, that two different charging behaviours can be observed: On the left, the participant was charging his phone during the night, while on the right the participant charged only during the day while at home. In total, over 46% of the charging events from all 50 participants happened during the night. 31% of the charging occurred during the day between 9:00 and 18:00, mostly while people are at work or in the university. The other 23% happen in the morning or when people are at home in the evening.

On average, participants carried their phone with them for 22.19% of the time throughout the whole study. Per participant the results vary a lot however: From 3.87% up to 52.91%, highlighting also that the results depend on the individual user's habits. The dataset includes participants who excessively used their phone, most noticeably for playing Ingress[30] (3 users: 13, 21 and 22), an augmented-reality game which correlates with high smart phone usage. For these users the *carrying phone* results are substantially higher than the average (up to 52%). Some participants forgot to charge their phone on a few occasions during the study, which led to a shut-down of the recording application and therefore to data loss. 5 participants found it uncomfortable to wear the wrist-worn sensor during the night, which is why they took off the sensor on most nights.

Smart phones are on the user at different times of the day. The results listed in Table 3.2 depict when the users had the smart phone on them at defined times of the day (using the divisions as in Dey *et al.* (2011b)): *in the morning* (7:00 to 9:00), *during the day* (9:00 to 18:00), *in the evening* (18:00 to 23:00) and *during the night* (23:00 to 7:00). Additionally, we calculated the amount of wearing time during the whole day (from 9:00 to 23:00, last column). As expected, users carried their phone mostly during the day. In contrast to the overall results of 22%, participants, on average, carry their phone for 35.59% of the time from 9:00 to 23:00. Depending on the participants time schedule (especially between students and office employees), the individual results vary highly. Participant 1, for example, uses his smart phone almost only while being at the office (9:00-18:00), while user 49 seems to be carrying his phone almost all day long. These findings suggest that not only the type of user monitoring application but also the habits on phone usage of the targeted user need to be known in advance. Note here that user 46 was excluded from the evaluation step, since the phone data was deleted by accident.

We further investigated if there are differences in phone carrying between genders and age-groups. For the first case, we obtained a daily smart phone usage of 22% for female and 34% for male participants. Interestingly, male participants tend to use their phone in equal shares during work (9:00-18:00) and in the evening (18:00-23:00). This matches the female participants' perception who stated that during the day the smart phone is in the purse most of the time. Males are the classical "quick trip" participants, which according to Dey *et al.* (2011b) describes the behaviour of taking the smart phone with them when leaving the desk or for coffee breaks. For the second case we created four different age groups according to their occurrence in the study: (1) 10-19, (2) 20-29, (3) 30-39 and (4) $\geq$ 50. The most active mobile phone users during the day are age groups (1) and (3) with around 60% and 35% respectively. Unfortunately, age-group (1) is not representative, since only one participant is in

[30]https://www.ingress.com/, last access, 09/2014

user	7-9	9-18	18-23	23-7	9-23
1	3.82	**25.43**	11.26	1.18	20.98
2	6.67	**11.30**	3.78	0.83	9.57
3	3.02	28.00	**30.62**	4.60	29.43
4	34.86	30.92	**55.22**	7.64	44.68
5	5.31	**24.98**	18.42	1.04	23.45
6	5.19	**36.48**	28.80	7.40	34.54
7	22.38	15.95	**23.19**	0.30	21.79
8	18.71	**27.91**	25.70	5.83	29.87
9	10.77	**50.23**	40.38	6.07	48.35
10	9.77	30.88	**31.39**	6.50	32.52
11	**25.42**	16.42	13.94	2.50	19.25
12	15.83	**19.26**	12.57	7.71	19.15
13	24.58	**37.33**	21.23	6.64	35.23
14	24.33	**29.94**	7.87	8.15	25.61
15	0.28	18.64	**23.15**	4.00	20.29
16	31.20	32.82	**33.30**	9.93	37.55
17	5.50	48.22	**66.20**	19.00	55.55
18	10.65	**19.11**	16.56	2.29	19.75
19	14.38	**46.10**	45.00	7.54	47.84
20	5.14	35.08	**37.86**	3.23	36.85
21	35.42	36.11	**38.00**	9.74	41.90
22	30.83	**70.89**	68.61	27.88	74.58
23	7.00	51.63	**52.00**	7.15	52.86
24	15.95	**55.11**	33.05	13.21	49.62
25	13.97	**19.97**	15.69	0.71	20.51
26	**39.06**	25.90	4.17	16.74	23.75
27	3.67	**35.96**	32.03	6.25	35.13
28	24.92	**36.18**	15.57	0.19	32.44
29	11.25	28.89	**35.36**	12.96	32.84
30	25.15	**40.42**	31.16	1.11	40.83
31	18.03	**41.14**	39.58	4.16	43.28
32	5.12	**33.93**	16.55	7.69	28.50
33	6.73	28.94	**56.08**	21.64	39.68
34	16.89	**43.25**	24.00	7.19	38.86
35	13.00	35.79	**37.40**	9.30	38.30
36	17.08	52.22	**53.93**	16.12	55.40
37	22.78	**28.06**	26.50	1.95	30.89
38	3.43	**27.28**	14.37	5.42	23.17
39	20.00	30.00	**52.08**	17.66	40.89
40	21.79	28.52	**40.05**	7.01	35.89
41	13.99	**38.85**	33.07	8.12	38.88
42	5.75	22.46	**36.47**	21.85	28.37
43	52.14	27.42	**55.55**	6.44	45.02
44	6.46	15.56	**28.58**	0.57	21.19
45	4.25	15.83	**16.04**	4.06	16.52
47	30.89	**50.71**	43.55	8.93	52.67
48	33.27	**47.36**	11.44	3.35	39.42
49	40.00	**73.07**	58.14	7.65	73.62
50	48.01	**52.82**	36.13	23.25	53.82
51	6.67	**17.26**	17.13	5.98	18.22
avg	**17.43**	33.93	31.37	7.93	35.59

Table 3.2: Results for all users showing in percent if they were wearing their smart phone at specific segments of the day with maximum values highlighted.

that group. Nevertheless, we suspect that the smart phone usage among teenagers is higher than in other age-groups[31]. In the following section we will discuss the benefits and drawbacks of this study and especially the results.

3.3.4 Monitoring User-Behaviour: A Discussion

We identify many different topics with regard to using a smart phone as a sensing platform in this section and highlight our most interesting findings in the following:

Demographic User Behaviour. In contrast to the work of Dey *et al.* (2011b) and Patel *et al.* (2006), which evaluated the proximity to the phone in the North American region, we investigated the carrying behaviour in a European country. We believe that cultural differences exist which is being reflected in the behaviour of a countries' population as well. To our best knowledge such an experiment as presented in this section has not been conducted yet outside North-America.

Acceptance of the Wrist-Worn Unit. The wrist sensor used for this study was perceived as comfortable to use by most of the participants and most reported that they quickly forgot about it while wearing it. The fact that the device's battery charge lasts more than 14 days, meant that participants did not have to charge or manage the device themselves. Although for five participants the device was not comfortable enough to wear during sleep, these users did remember to wear their unit again after waking up. The data was reliably recorded for the day-time periods.

Platform Differences. We encountered a difficulty in the evaluation process that has to be considered in future studies that make use of the accelerometer on the smart phone: Due to the vast number of different mobile phones that participated in this study, the algorithm for detecting motion segments had to be insusceptible to noise, jitter or other sources of disturbance. Additionally, different OS versions led to an unbalanced priority for our application to obtain sensor data, since it is handled differently by the OS. Such scenarios should not occur, especially if continuous data is needed to detect, for example, activities with a smart phone. A benefit of using the accelerometer though, is the fact that it does not require any security permissions, which is why most of the applications available in the Google Play store use it.

Power Consumption of the Android Application. According to the participants, their behaviour of charging the smart phone did not change significantly, since most of the users charge their phone overnight (almost 47%), as depicted in Table 3.1. The smart phone's battery lasts for typically a minimum of 24 hours under normal usage while having the app running. This was an early design constraint for the application since an application that drains too much power will quickly be deinstalled by the

[31]http://www.reuters.com/article/2013/03/13/us-usa-internet-teens-idUSBRE92C04C20130313. The article states that the ownership of smart phones among teenagers is steadily increasing, which supports the idea in a rise of phone usage amongst teenagers.

user. In this study, participants had to charge their phone only once. An exceptional case were the two senior citizens in our study: Since both of them are using their smart phones only to make phone calls and take pictures every now and then, our application forced them to charge the phone every day, instead of every three days as reported by the participants. In total, two users contacted us during the trial to discuss the higher power consumption. We suspect that the power consumption depends on the CPU as well, which provides the application with the sensor values when requested, as well as the Android OS version.

Proximity to the Phone. Our study shows that participants have their phone on them for 36% during the day time on average, and 22% over the whole study on average. This is over a half of the time reported by Dey *et al.* (2011b); Patel *et al.* (2006) that the phone is *within arm's length* of the user (58% and 53% respectively). Interviews showed that indeed many participants put their phone on the desk or at a table nearby, especially while working and during the night. Additionally, some participants put their phone on the mattress while sleeping, either because they used it before falling asleep or to have it immediately on hand when waking up. Exceptional cases were met for the 14-year old participant that was supposed to switch off the smart phone while attending class but rather muted it during the whole day. During that time, the smart phone was always in the front pocket and therefore being carried on the body most of the time. The senior citizens are an exception to the rule: As low profiled phone users a mobile application should motivate them to use the phone more often, e.g., by including a healthcare application. We argue that the *on the user* proximity information could give extra insight in future studies, especially those that explore the use of smart phones as wearable devices.

The User's Perception. Especially when we interviewed the participants after the study, we asked them about their perceived smart phone behaviour. Interestingly, many participants underestimated their smart phone usage. The most common answer was "I almost never use the phone while being at work.". This was proved wrong for most of the participants. Even though the smart phone is lying on the table while the user is at his desk, whenever the user leaves his desk the smart phone is put in the pocket. Many participants became aware of that fact after the study. For context-based systems such knowledge is crucial, since it shows that the smart phone is a suitable platform for sensing the environment. In their conclusion, Dey *et al.* (2011b) and Patel *et al.* (2006) stated that the mobile phone is farther away from its user than expected. In contrast to that, we conclude that the mobile phone is being carried on the body more often than assumed by the users.

Rich Dataset. The recorded dataset consists of almost 638 days of sensor data from two modalities that can be used in future studies: (1) The smart phone data implies not only acceleration data and battery status information but also light intensity values and the proximity sensor values. (2) The wrist-worn sensor logged the inertial data and, at the same time, the light intensity. Additionally, the dataset contains a sleep diary from almost every participant. Such information enables future studies that aim at detecting, for example, sleep segments with a mobile phone only. The dataset is publicly available for download at http://www.ess.tu-darmstadt.de/smartphone14.

3.4 SLEEP MONITORING SYSTEM

In contrast to using a mobile phone based platform to sense the user's activities as shown in the previous section, we will introduce here a system that records sleep as an example of an especially important rhythmic activity. The monitoring of sleeping behaviour is an important challenge, especially when monitoring the user outside of a sleep laboratory. Many different approaches are commercially available (see Section 2.4.2) but other researchers try to apply their scenario to the real-world of a patient (see Section 2.3.2). In this section, we will describe the requirements of a sleep monitoring system that is unobtrusive and can be deployed over several weeks without the user having to interfere with it.

3.4.1 Challenges for Sleep Monitoring

Three challenges are in particular important to enable the long-term monitoring in domestic environments:

Reliability. A system to be deployed over longer stretches of time (from months to even years) without maintenance needs to be reliable in collecting data. For that it has to detect the observed characteristics accurately. In such real-world deployments, the capturing of specific phenomena is often impossible, due to the presence of noise and irrelevant data, as well as limitations on the sensor's behalf. This, for example, is the case when direct video footage is covered and body-worn sensors cannot be worn continuously (because the device is not waterproof and has to be taken off when showering).

Privacy. By recording everyday activities, privacy is a critical factor when larger sections of our lives are recorded. The safety of the data needs to be guaranteed since the sensors register data in a most sensitive context. Therefore, ethics' guidelines need to be kept to ensure the privacy of the user. For the case study of observing one's sleep (see Chapter 6), the used sensing modality could automatically record and reveal any activities related to the user's most sensitive environment: The bedroom and bathroom (Hong *et al.*, 2004).

Deployment. A monitoring system needs to be easy to deploy and has to be modular. With such a system, less time is spent on the installation of the system and the moving of it to another environment, lowering the costly installation procedure. Dependability and usability of the system are additional key factors, i.e., not having to maintain the system due to hang-ups, or avoiding failures in critical medical applications. In the case study of detecting sleep, the recording system needs to remain running for at least several hours per day, also without interfering with the user's daily routine. Causes for interfering could be frequent battery changes or required system interactions.

With these requirements, we describe in the next section such a monitoring system, which has been used in the sleep studies in Chapters 5 and 6.

3.4.2 Unobtrusive Monitoring System

This thesis sets out to investigate first and foremost the detection of sleep related characteristics to capture its rhythmic nature and to show outliers of the obtained rhythms. We will focus on the wrist-worn sensor from Section 3.2 which is used for these detections. For visual feedback of the detections, an active IR night-vision unit is used that is synchronized with the wrist-worn sensor, which is introduced in the following.

IR Camera. For this study, the TrendNet TV-IP422W was chosen for deployments as it provides an adequate resolution at a frequency of 30 frames per second, and is equipped with an array of IR Light Emitting Diodes (LEDs) which are powerful enough to sufficiently illuminate an area from up to 5 meters away (see Figure 3.7). It can be configured to provide the recordings on a network drive via ethernet or the local wireless network but on local flash storage via an USB connector as well. A pan-tilt motor allows the camera to re-adjust itself to fully focus on the sleeping subject. At deployment, the camera's embedded real-time clock is synchronized to that of the wrist-worn sensor so that data can be merged afterwards.

The IR camera is powered from a wall-socket and as such can be activated for longer periods of time. The camera is scheduled to automatically switch off between 11:00 and 19:00, and is by default configured to store still pictures on the local flash storage and movies in 10-minute chunks to an ethernet-attached netbook. Since the data produced by the camera for one single day sizes to about 9 Gigabytes on average, this recording set-up can remain unmaintained for longer stretches of time. The wireless capability of the camera was turned off.

The studies in this thesis used one camera as a means to obtain the ground truth concerning (1) whether the test subject was sleeping, (2) the posture the test subject was lying in, and (3) excessive movement characteristics spotted in the data. The intended scenario in this section uses the camera for visual inspection by a somnologist: By filtering out everything but the relevant sleeping postures and short movies of possible limb twitches, a sleep expert has the required material for visual inspection on a PC.

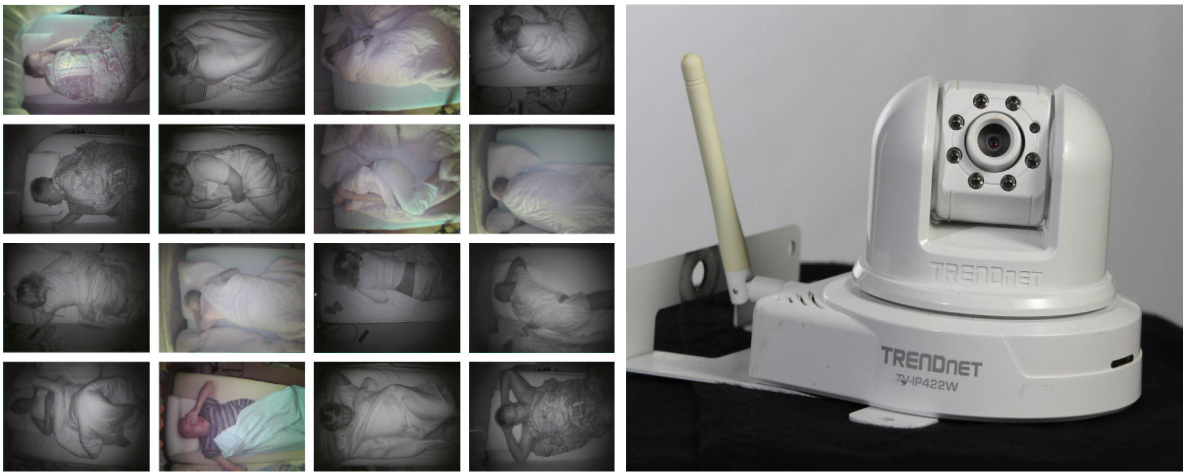


Figure 3.7: The IR camera (shown right) is an off-the-shelf logging device, illuminating the observed area with an array of 7 strong IR LEDs and producing high-quality 640x480 images, tagged with second-resolution timestamps. Left are some examples of test subjects.

3.4.3 Method Overview

This thesis' proposed method to find significant events in sleep, based on motion from a highly deployable system, consists of three major steps. First, the continuously recorded information from the wrist-worn sensor is automatically processed and segmented into awake and sleep. After that, non-motion data in the resulting sleep segments that occur and re-occur are automatically clustered into postures in order to visualize general trends over several nights. In a last step, excessive movements are segmented out of the remaining motion segments. Figure 3.8 illustrates how the original raw data from the wrist-worn sensor is processed.

The envisioned scenario of our method comprises the following steps, from preparation to data analysis:

1. The user obtains the camera and wrist sensor.
2. The user places the camera in the bedroom, synchronizes with the wrist sensor, and wears the wrist sensor.
3. Throughout the monitoring phase of multiple weeks, both camera and wrist sensor record their data continuously.
4. After the monitoring phase, the wrist sensor is synchronized with the camera, and the proposed method provides:
    - The sequence of clustered postures per night.
    - The detected myoclonic twitches per night.
    - Video footage of extracted events.

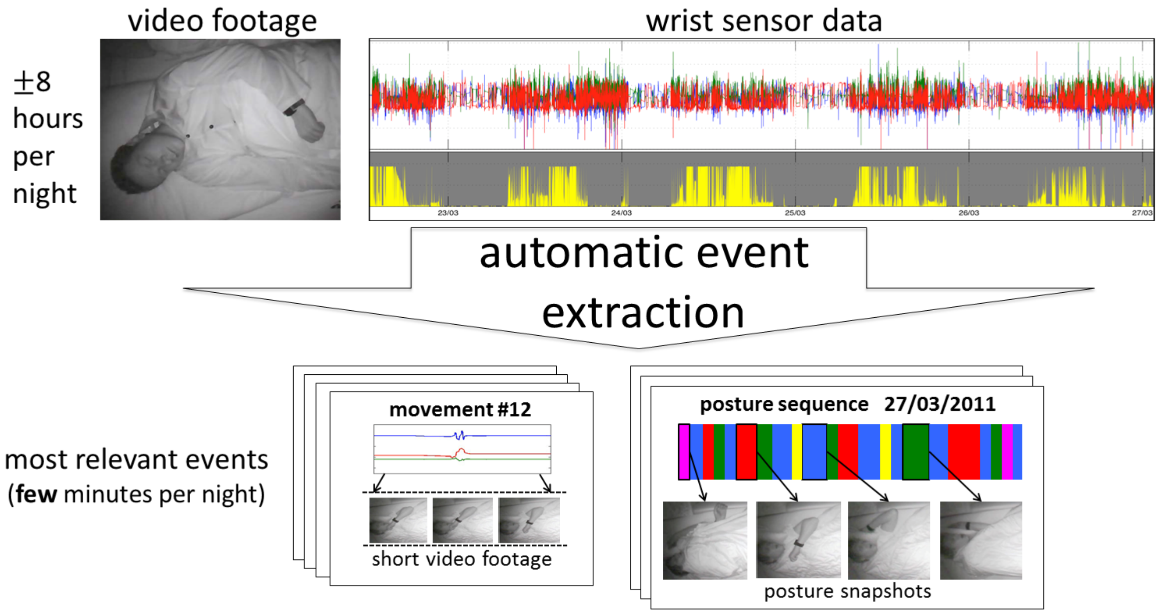


Figure 3.8: Overview of the detection system: Long-term inertial, time, and ambient light data from a wrist-worn unit are analysed for sleep events of interest, and synchronized with camera footage for easy visual inspection.

3.4.4 Privacy Issues

Recording subjects via a video camera in a most private room - the bedroom - is a delicate matter and requires a privacy consent signed by the subject. For this purpose we used privacy guidelines that were elaborated with the local data protection officer of the university. It was ensured that all the data was anonymized, stored in a secure place and used only by researchers with a permit. From our experience with the test subjects, ensuring the privacy with a privacy policy is the minimum requirement for conducting such studies.

Additionally, the data collection process has to be evaluated in regard to security and privacy insurance. We identified three different steps for this: (1) The *configuration* implies the user's set-up of the camera in order to capture the right angle of the recording scope. For this purpose, the camera has to be accessed via the web-interface using either a connection via LAN or WLAN. (2) The *recording* comprises the storage of video footage on an external device, by either streaming it via the network (LAN or WLAN) or via an attached USB drive. (3) The *transfer* consists of the transmission of the recorded data from the storage device to another computer for later on post-processing.

More specifically, the data has to be transferred in a secure way. For this purpose, we established three different possible scenarios that are scrutinized in regard of privacy. The *first scenario* is the use of the camera's wireless abilities, which transmits data only with WEP encryption in an ad-hoc mode which is highly insecure (Tews

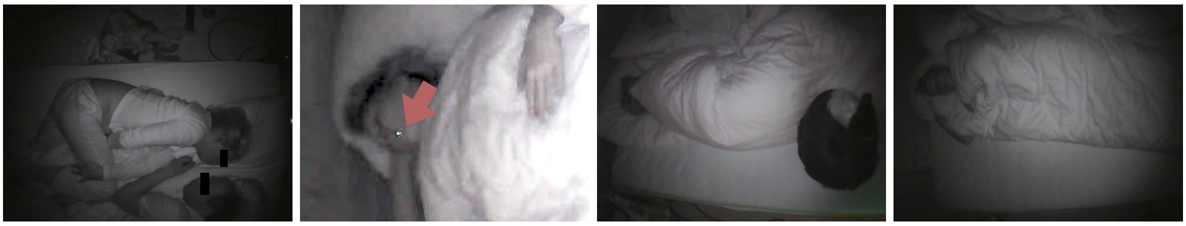


Figure 3.9: Potential difficult situations for vision-based recognition that were observed during the study (from left to right): Multiple participants in the scene, reflecting metallic objects such as piercings, pets moving into the bed and thick blankets covering the entire participant.

et al., 2007). A *second scenario* would imply the use of a WLAN router that supports a more secure encryption technique like WPA2. However, studies show that a person's sleep can be disturbed by the radiation of a wireless transmission (Hyland, 2000), which is why we decided to apply a *third scenario*: We make use of a wired connection to meet privacy concerns, as well as to increase the reliability of the system, which requires a nearby PC or laptop. With the latter set-up we can not only decrease the user's interaction with the system but also increase trust in the system by having the camera physically disconnected from any open network connection.

3.4.5 System Discussion

During the process of building and perfecting the prototypes for this system and instructing subjects on its usage, several surprising issues and obstacles had to be solved. Apart from the predictable difficulties that come with designing robust measurement prototypes and deploying ubiquitous technology in domestic environments, especially those worn 24/7, other interesting lessons were learned. The following sections will discuss these in more detail.

Video Footage. One advantage of the continuous logging that has not been specifically focused on so far, is the ability to correlate activity during the day with possible effects for the following nights. This requires an activity recognition component, however, which might be added at a later stage. Similarly, one might expect the video footage to be exploited more and analysed for steady body postures and sudden motions instead of relying on a body-worn sensor. An initial study investigating this avenue met with a lot of obstacles during the first datasets, as illustrated in Figure 3.9. The same limitations apply to the visual inspection: Especially in case of occlusion by blankets, exact body postures and twitches are sometimes hard to verify. In our experiment that was mostly conducted during wintertime, these problems concerned about 18% of all twitches observed in the wearable data.

Timing Considerations. As monitoring periods will be extended to the scale of months, one issue that can be expected to become prominent is a possible larger drift between the wrist sensor's timestamps and those embedded in the pictures and movies. For the study data recorded for this thesis, an extra 2 seconds were attached before and after the signal, and a variable offset was built in the visualization tool to match the $100Hz$ accelerometer signal and the recorded movie more perfectly. As drifts due to temperature and humidity fluctuations continue over longer periods, this approach might become less scalable when aiming at year-long logs.

3.5 CONCLUSIONS

In this chapter we investigated the use of different wearable sensor platforms that were applied in the course of this thesis to detect activities. The focus of this chapter is on the challenges of the presented recording systems regarding the *hardware requirements* and the *user acceptance* of the systems. More specifically, we showed that a wrist-worn data logger can be used in two different scenarios: (1) For detecting movement of a person which is used to correlate inertial data with accelerometer data from a user's smart phone and (2) to obtain inertial data from the wrist and ground truth data with an IR camera to determine sleep and its parameters. Both scenarios are crucial for the rest of this thesis, since they pave the way for the experiments conducted in Chapters 5, 6 and 7 to obtain data that can be scrutinized for the rhythmic nature of the performed activities.

In the first scenario we presented a novel approach to estimate how often the smart phone is *on the user*, by using an additional wrist-worn sensor and comparing the inertial data from both the user's smart phone and the wrist sensor. The approach can be combined with the previously suggested methods (Dey *et al.*, 2011a; Patel *et al.*, 2006), and allows for a characterization of when the phone's built-in sensors could be expected to monitor the user, for instance, to detect the user's physical activities (sedentary, walking, running, etc.). We performed a 51-participants study using this method over two weeks, resulting in a dataset of over 638 days (more than 15,300 hours) of recorded data from both modalities used. The analysis of these data indicate that the users' smart phones are on the user on average for 22% of the time (day and night). This figure is considerably higher for some users (up to 52%) and considerably less for others (4%). During day time (9:00-23:00), our results show that users have their phone on them on average 36% of the time. The study also suggests that users have very different habits in phone charging behaviour and usage, stressing the importance of knowing the target users when designing monitoring applications for smart phones that require the user to be carrying their phone with them. We argue that the method of investigating phone use through the comparison of inertial data from phone with a wrist-worn sensor is particularly interesting for long-term studies.

In the second scenario a monitoring tool is presented, which combines low-effort deployment with the capturing of sleep. Given automatically detected events during sleep, data is synchronized with a camera to highlight relevant sections and extract them from a large corpus of data which cannot possibly be skimmed by humans.

Given these properties it enables video analysis at home and additionally uses a power-efficient wrist-worn activity sensor for long-term recordings.

4 RHYTHMS FROM TIME USE SURVEYS AND SLEEP

Contents

4.1 Introduction **59**
4.2 Rhythms within Time Use Surveys **60**
4.2.1 Comparison of Time Use Data 61
4.2.2 ATUS vs. GTUS Demographic Analysis 62
4.2.3 Feature Extraction 63
4.2.4 Recognizing Activities 64
4.2.5 Time Use Surveys: A Summary 69
4.3 Sleep as a Key Rhythmic Activity **69**
4.3.1 Sleep Patterns and Sleep Behaviours 70
4.3.2 Comparison of Sleep Rhythms 71
4.4 Conclusions **74**

In this chapter, we statistically evaluate data with regard to its rhythmic nature. Our method is based on (1) a governmental database and (2) sleeping patterns obtained from inertial data. The novel approach of using statistics shows how crucial such information is for an activity recognition system in the context of ubiquitous computing. By using large datasets in the experiments, we are able to perform an in-depth analysis of features that we extracted from these statistical data. The features are particularly interesting for future studies, in which prior knowledge is used to improve the recognition rate of activities.

4.1 INTRODUCTION

An activity recognition system is usually prototyped and evaluated by first recording typical data from one or many different sensing modalities to then create an optimal classifier with distinctive features from such data. The results have been found satisfying when classifying low-level activities with clear links to the sensed modalities: Motion and posture sensors have, for instance, shown to be suitable for detecting activities like *walking*, *standing*, *sitting* (Huynh *et al.*, 2008; Lee and Mase, 2002; Lester *et al.*, 2006) or *sleeping* (de Souza *et al.*, 2003; Jean-Louis *et al.*, 2001b).

When considering high-level activities like *having lunch* or *commuting to work*, finding distinctive sensing modalities and features is often as hard as designing an appropriate classification system. In these cases, having good and sufficient training data available is crucial to the performance of the system. Especially the variability among study participants in how they execute the target activities, as well as in how long these activities tend to last and when they occur at different times, has been shown to be quite an obstacle as large sets of exemplar data are required (Bao and Intille, 2004).

In this chapter we investigate a topic that has not been addressed thoroughly so far: The use of statistical data obtained over a population in the activity recognition process. The first contribution of this chapter is therefore a new type of region specific information source that describes what people are usually doing. For this purpose we consider all activities a person performs in order to determine certain rhythmic activities. To achieve this, we use different features that can be extracted directly from the database to describe the activities. We evaluate these features for a specific European region and assess their value for activity classification. The results indicate that the combination of features yields the highest results for classifying activities.

Further, we pick out an exemplary activity that in itself can contribute to its rhythmic occurrence: By investigating sleep patterns from several weeks, we obtain features describing recurring patterns for sleep-related characteristics. More specifically, we show how sleeping behaviour changes when comparing weekend days to working days and also single days of the week with each other. Results indicate that it is possible to model a subject's nights by using past observations to categorize new nights into regular or irregular ones.

This chapter is organized as follows: First, in Section 4.2, we introduce time use survey data as a statistical database obtained over a population that is used to extract certain features like *location* to classify activities, using the database's information only. Then, in Section 4.3, a sleeping pattern is presented that allows us to extract sleep related features to model a subject's nights according to the normal sleeping habits.

4.2 RHYTHMS WITHIN TIME USE SURVEYS

To make use of the information from large time use surveys was suggested in Partridge and Golle (2008) as a valuable instrument for designing activity recognition systems for which it is hard to obtain training datasets with sufficient variability. For a substantial proportion of a population these national surveys capture which activities are executed at which times during the day, and in which locations (see Section 2.2). This would allow, for instance, a classifier to use the current time and location to estimate what activity the user is most likely to perform, based on matching entries in the time use database. For wearable sensing systems one can assume that further user properties such as gender, profession or age, are readily available and can be used to refine this search. Although such estimates could be accurate enough for some applications, they would at the very least be promising as priors in a sensor-based activity recognition system.

In this section, we investigate the possibilities of using data from a recent European time use survey with activity logs from 13,798 participants for activity recognition in general, and compare the results to a previously published analysis from a similar-sized North-American time use study. Three contributions are made in particular:

- We give a qualitative and quantitative comparison of the US-based data analysis given in Partridge and Golle (2008) with that of comparable time use data from across the Atlantic - showing cultural differences. This includes a discussion on the use of large datasets and time use studies in wearable computing and the challenges in extracting data from these large datasets.
- We argue for a leave-one-household-out cross-validation methodology and perform such an evaluation in order to study the time use survey features for activity classification.
- We discuss results from single activities for a specific European region and evaluate different features that can later on be used for activity recognition.

In Section 2.2 we already introduced time use surveys as a promising data source which has to be investigated in more detail. In the following we will compare the database used in Partridge and Golle (2008) with the German Time Use Survey (GTUS) to emphasize the need of inspecting such data for specific regions.

4.2.1 Comparison of Time Use Data

The GTUS can be compared easily to other European countries' time use surveys since the data is similarly structured. The American Time Use Survey (ATUS) on the other hand is built up differently, not logging activities, location and simultaneous activities in 10-minute slots but recording when the activity started and how long it lasted. Nevertheless, research groups like the Centre for Time Use Research (CTUR[32]) are maintaining the Multinational Time Use Survey (MTUS, see Section 2.2) in order to create a huge time use database, including to this day ATUS and Harmonised Eutopean Time Use Survey (HETUS, see Section 2.2).

Table 4.1 lists the basic properties in terms of the type and the amount of data that was included for the ATUS and GTUS sets: Both datasets are similar in regard to the number of participants and the time-period monitored. GTUS identified more activity episodes (the number of activities occurring in the dataset), which can be explained by the fact that more days per participant are included and that activity episodes are logged in 10-minute intervals. A previous research study on the ATUS dataset (Partridge and Golle, 2008) considered hour-of-day (60 mins) as a time interval for simplicity.

A further property of the GTUS is that it keeps track of the time use for all members in a household above the age of 10, whereas ATUS explicitly chooses single participants from one household. We keep in mind that ATUS is updated every year, by interviewing participants over the phone and keeping track of their activities for

[32]www.timeuse.org, last access 09/2014

Property	ATUS 2006	GTUS 2001/2002
participants	12,943	13,798
households	12,943	5,160
# activities per tier (1 / 2 / 3)	18 / 110 / 462	10 / 48 / 272
# locations	27	8 / 21
time interval (mins) dataset / study	1 / 60	10 / 10
monitoring period	1 day	1-3 days
activity episodes	263,286	356,910

Table 4.1: Comparison of the basic properties of the ATUS 2006 taken from Partridge and Golle (2008) and GTUS 2001/2002. The time interval for ATUS was remodelled in Partridge and Golle (2008) from minutes to hour-of-day, displaying therefore not the original intervals (duration of an activity in minutes).

one day. Therefore, the dataset is always up-to-date in contrast to the GTUS, which is refreshed every 10 years. The following section will display quantitative results for the GTUS, as well as the ATUS, highlighting promising features from the datasets.

4.2.2 ATUS vs. GTUS Demographic Analysis

The contributions of this section are threefold: (1) We will discuss a demographic comparison between GTUS and ATUS. (2) We show quantitative results for the GTUS dataset and (3) we highlight important activities from the GTUS for probable mobile and wearable research.

In order to perform a graphic comparison, a few conversion steps had to be taken. A first hurdle is the difference in categorization and hierarchy of activities for both datasets: The most relevant tier 1 and tier 2 activities of the GTUS were translated to the corresponding 18 tier 1 activities of ATUS. Furthermore, to reflect the 10-minute segments in the GTUS dataset, all entries from the ATUS were converted to 10-minute time slots.

The resulting more in-depth visualization of both datasets is shown in Figure 4.1. Displayed are all 18 tier 1 activity groups from the ATUS 2006, showing the performed activity of the participants in percent per time-of-day. The figure shows some differences that exist in the activity reporting: Where the ATUS dataset contains significant digit bias (i.e., a bias of participants rounding off start and stop times of reported activities toward full or half hours, visible as jagged edges in the plot), this is less pronounced in the GTUS even though the reporting time intervals are relatively small for both. A second, rather cultural, difference which can be identified from this visualization is that the GTUS dataset shows stronger time dependencies for particular activities (see the larger increases around the times for breakfast, lunch and dinner for instance), including a sharp rise in leisure activities after 20:00. The plot displays significant differences in both US and European regions, pointing to the importance of using time use data for specific regions only, since

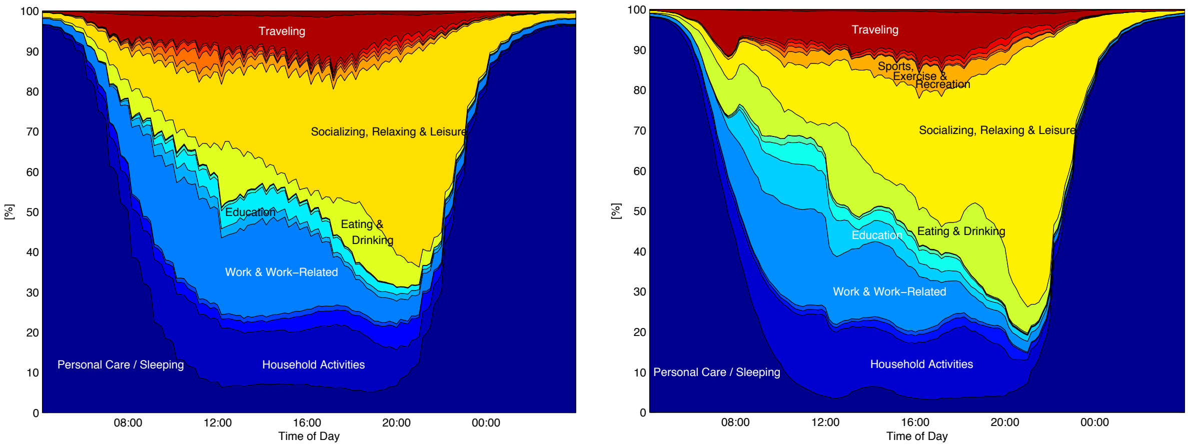


Figure 4.1: A visual comparison of both the ATUS 2006 (left) and the GTUS 2001/2002 (right), depicting per time-of-day the normalized occurrences of common tier 1 activities over all participants. Slight cultural differences appear in digit bias (jagged edges on the left) and activity clusters (e.g., breakfast, dinner on the right).

demographic characteristics have to be considered. We will now have a closer look at the GTUS data, describing how time use data is used to infer which activity has been performed.

4.2.3 Feature Extraction

Methodology. We use all 5,160 households from the GTUS for a leave-one-household-out cross-validation analysis: Each of the 5,160 households is left out to calculate the maximum likelihood of an activity for the rest of the 5,159 households. For each member of the left-out household we compare the activity to the most likely one from all the other households. For each household a confusion matrix is stored, from which precision, recall and accuracy are calculated for later analysis. We believe that within a household, participants tend to engage in the same activities when spending time together, leading to biased results. Therefore, we exclude a complete household and observe the activity distributions for the rest of the households, constructing maximum likelihood classifiers. For this we use the maximum conditional probability $P(A|f_1, f_2, ..., f_n)$ which the classifier derives for a target activity A given different features $f_1, ..., f_n$ as input.

Features. Many different features can be extracted from the GTUS, like gender or even location. In Section 4.2.2, we remodelled the GTUS to fit the ATUS. Here, we use the GTUS as it is, changing only minor things (the changes will be discussed later in this section). To evaluate the features for activity recognition from time use data, we consider different aspects similar to Partridge and Golle (2008). Therefore, as features we use:

1. **Time.** *Time* is a significant feature for activity recognition as derived in Eagle

and Pentland (2006) and Farrahi and Gatica-Perez (2008), since people tend to stick to their behaviour patterns. The fine granularity as given by the GTUS for logging activities in 10-minute slots will confirm that. Therefore, we will not use hour-of-day as Partridge and Golle (2008) did.

2. **Prev. act.** A *previous activity* is an activity that took place prior to a different activity. We are not considering the activity prior to a time-slot, which would lead to a biased result like for sleeping: Prior to sleep we most likely would be sleeping.

3. **Location.** The idea that knowing the *location* might lead to the activity that is being performed, was mentioned in Eagle and Pentland (2006) and Ashbrook and Starner (2003). For our work, this feature needs some remodelling, fusing all 20 different means of transportation (e.g., *in the bus*, *by foot* or *in car*) into one *transportation* variable to simplify the use of this feature and to minimize classification runtime. We obtain 10 different locations out of 29 from the original dataset.

4. **Gender.** Recently, researchers in Maekawa and Watanabe (2011) and Krassnig *et al.* (2010) used *gender* to infer which activities are being performed by participants of the same sex. Other physical characteristics have been added as well but we will focus on gender itself as a feature.

5. **Age.** The *age* in the GTUS varies from 10 to 80 years, which is why we divided the datasets into groups each covering 5-year spans (10-14, 15-19, ..., 75-80), just as in previous work (Partridge and Golle, 2008). This was done not only to simplify the calculation but also to preserve a significant number of participants per age group when using the classifier.

In the process of evaluation, we do not only use single features but also several combinations of the features, e.g., by combining *time* and *location*. A list of these combinations can be observed in Figure 4.2, left side.

In the course of this work, we considered *travelling* (e.g., *travelling between the home and the office* or *from school to the home*) as an important activity, which is present in all of the 10 tier 1 activity groups of the GTUS as individual activities (e.g., within the tier 1 group *work*, *going to work* would be an activity). To detach *travelling* from the tier 1 groups, we create an 11^{th} tier 1 activity group *travel* to which we relocate all travel codes from the other tier 1 groups. A complete list of the GTUS tier 1 activity groups as used in this section is shown in Table 4.2, left side.

4.2.4 Recognizing Activities

In the following sections we will present our evaluation results for suitable features and their combinations to detect activities efficiently.

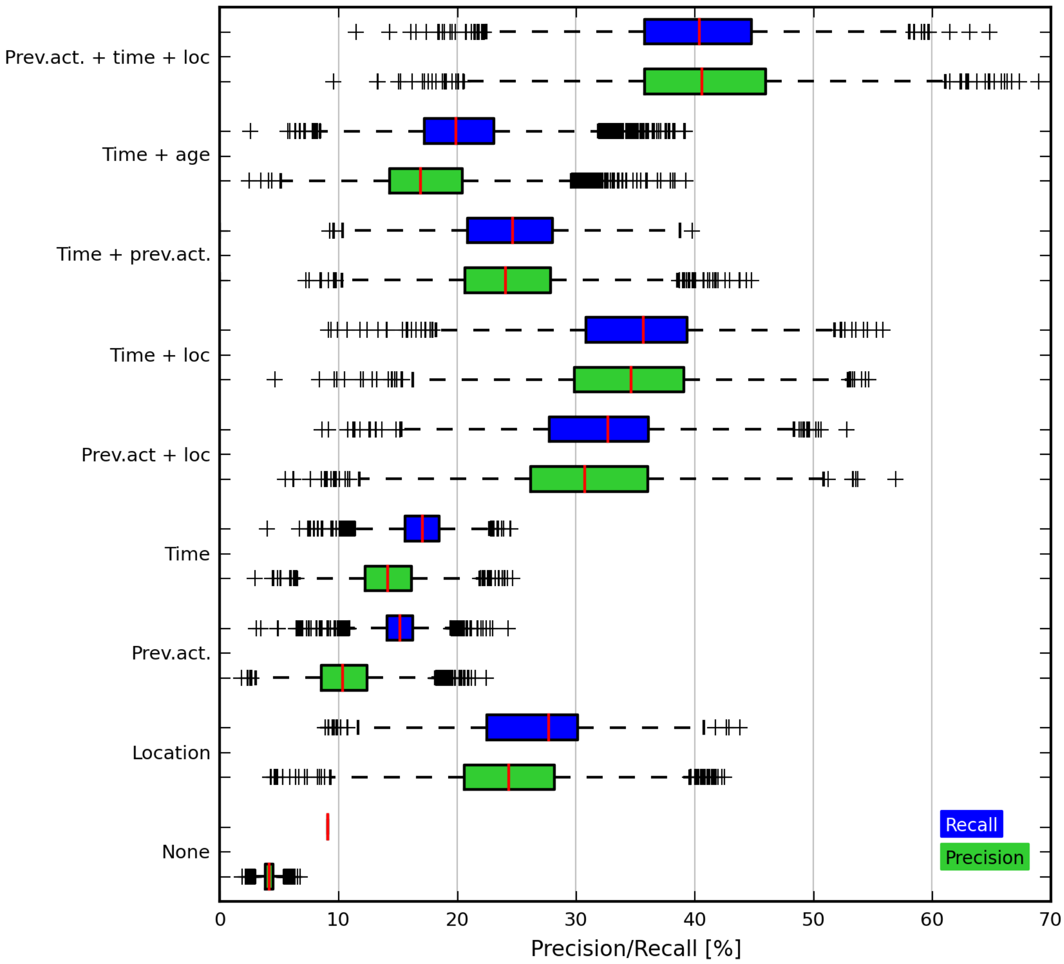


Figure 4.2: Precision and recall results for different features and their combinations from the GTUS, plotted as box plots with the median (red line), upper and lower quartile (75% and 25% respectively, end of boxes), whiskers for highest and lowest values and fliers ('+') as outliers. The combination of *prev.act.*, *time* and *location* yields the highest results.

Feature Evaluation

Figure 4.2 shows precision and recall for different features and feature sets as box plots. We see the median being displayed within the boxes as a red line, the upper and lower quartile (75% and 25% respectively), the whiskers which show the whole range of the data, as well as the fliers ('+'), displaying the outliers within the results. Starting with *none* we can observe how the values for precision and recall vary as we add more and more features to the classifier. *None* yields a very low precision and recall (4.17% and 9.09% respectively), since the results are biased by the activity *sleep*, which is the most probable activity (this is the case also for *age* and *gender*, leading to the same overall results as *none*). For single features, precision and recall are below 25% and 28% respectively, whereas the highest results are gained by using *prev.act.+time+location* (precision of 40.79% and recall of 40.25%), showing also outliers in the 60% region.

Location has the highest impact on the results: For *location* alone the results are already quite high but can even be increased by adding *time* as a feature, resulting in 34.58% precision and 35.24% recall. Some improvement can still be achieved by adding *prev.act.* to the feature set. However, *time* seems to have a higher impact here, since the results are not far apart, whereas *prev.act.* contributes little but significantly to the results.

We showed that a combination of features yields the highest probable result for inferring the right activities over a large-scale dataset from the GTUS, highlighting especially *time*, *location* and *prev.act.* as proper features. We will now have a look at the results taken from Partridge and Golle (2008) for the ATUS dataset and compare them to the GTUS results, also showing accuracy for different features.

GTUS vs. ATUS

In order to perform a quantitative comparison of the GTUS and ATUS, we display the results for different features for both in Figure 4.3. Note here, that from Partridge and Golle (2008) only the results presented in a barplot graph were available. Unfortunately, both of the authors did not have the original result sets, which is why we read off the approximate results from their paper.

For different feature combinations Figure 4.3 shows the distributed accuracy per household as box plots, highlighting the median (red line in the box). Additionally, a red 'X' marks the accuracy from the ATUS dataset for the same features and their combinations. We can observe that the accuracy results for the GTUS exceed the ATUS results, especially for the highest accuracy, when *prev.act.* is combined with *time* and *location* (the median is always above the ATUS accuracy).

Prev.act. showed promising results in Partridge and Golle (2008), which is why we use it in combination with *location*, resulting in an accuracy of 67.7%. Compared to *location* on its own, a slight increase of +6.54% is achieved. The same can be observed for the ATUS. Adding now also *time* to the feature set *prev.act.+location*, accuracy rises to 74.78%, leading to the highest results. In contrast to that, if we use *prev.act.* with *time*, we achieve a better result than using *time* on its own (51.76%). We can conclude that *previous activity* is rather a weak feature but combined with *time* and *location*, the results are promising. Compared to previous research, similar findings have been reported for the ATUS. *Time* combined with *location* yields an accuracy of 67.72%, again increasing the accuracy by +15.96% compared to *time* itself. Adding *age* or *gender* to *time*, a marginal increase of 1% to 2% in contrast to *time* by itself is observable. The results indicate that activity inferences perform better for the GTUS.

Note here, that the duration of an activity has not been considered. As shown in Partridge and Golle (2008), a duration weight for the activity can increase the accuracy. Since we consider a 10-minute interval of the given activities as sufficient, we left out a duration weighting. In the next section we will evaluate the results

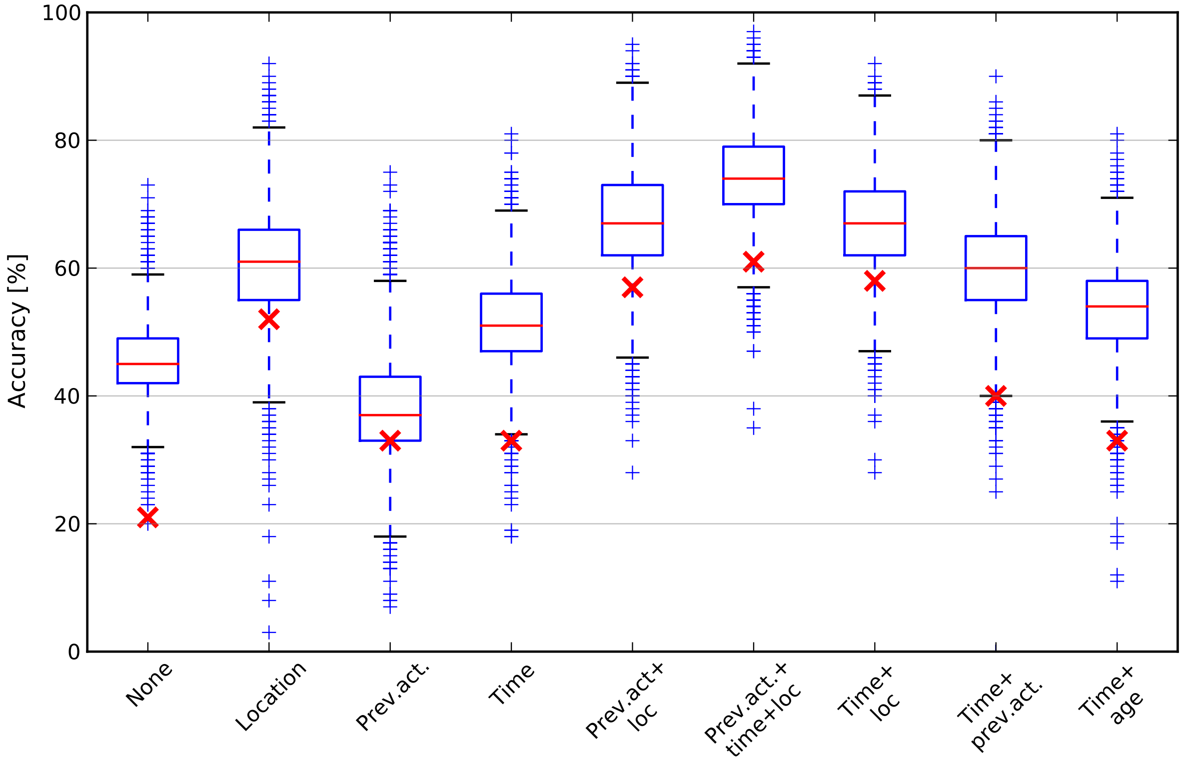


Figure 4.3: Box plots of the accuracy results per household from the GTUS and overall average results from the ATUS (values taken from Partridge and Golle (2008)) marked as a red 'X' for different features and their combinations. Here, the GTUS results exceed the ATUS results, marked by the median box plot (red line in each box).

for single activity groups of the GTUS dataset, highlighting important features for recognizing certain individual activities.

Results for GTUS Tier 1 Activities

In order to perform a more in-depth analysis, we again use the maximum likelihood classifier in a leave-one-household-out cross-validation (the same method which was used in Section 4.2.3). We calculate for the best performing feature combinations from Figure 4.2 the precision and recall for single tier 1 activity groups of the GTUS. We include *travel* here, thus resulting in 11 groups in total. Table 4.2 displays the results for precision and recall. Additionally shown are *average duration per day* for the activities, describing how participants spend their time during the day on average. The hyphens in the table indicate that an activity group was never chosen by the classifier, because other activities were preferred. Results for the ATUS, which are taken from Partridge and Golle (2008), are also shown in the table. Activity groups *hobbies*, *mass media* and *unknown* do not occur as single tier 1 activity groups in the ATUS, which is why 'n/a' was put in the corresponding rows of the table. Note here that we are not comparing the ATUS to the GTUS but rather displaying

Activity	**Average hh:mm per day**		**time**		**time + age**		**time + prev.act.**		**location**		**location + time**		**location + prev.act. + time**	
	GTUS	*ATUS**	*GTUS* Pre Rec	*ATUS* Pre Rec	*GTUS* Pre Rec	*ATUS* Pre Rec	*GTUS* Pre Rec	*ATUS* Pre Rec	*GTUS* Pre Rec	*ATUS* Pre Rec	*GTUS* Pre Rec	*ATUS* Pre Rec	*GTUS* Pre Rec	*ATUS* Pre Rec
Personal care	11:00	9:23	68.8 86.7	74.5 87.2	72.5 78.3	77.8 85.8	80.8 89.2	84.0 88.2	60.2 97.7	56.8 100	80.3 86.6	82.3 89.0	85.7 92.1	86.6 89.1
Work	2:00	3:27	22.6 2.6	30.2 61.3	22.9 32.2	36.7 68.5	32.0 42.8	44.8 74.5	53.3 60.8	93.7 87.9	53.3 60.8	93.6 87.9	56.1 60.4	95.0 87.6
Education	0:40	0:27	-	-	10.7 15.2	32.9 39.0	-	44.1 4.4	-	70.0 64.7	-	70.0 64.3	6.9 1.2	72.2 59.8
Houshold activities	2:50	1:49	23.9 53.9	-	26.4 38.0	-	34.7 52.5	28.6 11.6	37.8 16.0	52.3 0.1	37.1 55.4	31.5 14.2	50.2 71.9	34.0 22.2
Volunteer activities	0:10	0:07	-	-	-	-	-	15.7 0.8	-	-	0.1 0.0	2.9 0.0	6.2 1.8	24.7 2.9
Socializing and pleasure	1:30	4:31	-	39.3 52.1	4.3 0.4	39.8 57.5	24.5 11.3	41.6 61.0	-	47.8 14.3	36.9 23.5	47.1 71.2	45.2 28.5	49.0 73.8
Sports	0:30	0:18	-	-	-	-	-	16.8 0.4	19.4 23.8	48.5 36.4	24.9 19.0	47.2 35.5	35.8 29.6	48.8 31.3
Hobbies	0:30	n/a	-	n/a	4.2 5.5	n/a	-	n/a	-	n/a	0.1 0.0	n/a	3.4 0.7	n/a
Mass media	2:30	n/a	40.7 43.7	n/a	38.2 44.2	n/a	54.1 54.2	n/a	-	n/a	49.7 43.0	n/a	60.2 57.1	n/a
Travel	1:20	0:11	-	-	5.6 0.4	-	41.0 19.4	49.4 31.8	98.0 99.3	97.0 96.0	98.0 99.3	96.8 95.9	98.4 99.3	96.1 94.8
Unknown	0:00	n/a	-	n/a	-	n/a	-	n/a	-	n/a	-	n/a	0.5 0.2	n/a

*The values and results for the ATUS are taken from Partridge and Golle (2008)

Table 4.2: Our tier 1 precision and recall results for the GTUS database, alongside the results from the ATUS tier 1 activity results for reference (as mentioned in Partridge and Golle (2008)). The best performing features from Figure 4.2 are used for displaying the results for the 11 activity groups. A hyphen indicates that the activity was never predicted, and 'n/a' indicates that the activity group is non-existent in the ATUS database.

the results as a reference in the table.

Immediately visible in Table 4.2 is that *location* yields high results for specific activities. *Personal care* achieves a high recall of 97.7% and a precision of 60.2%. *Location* results are again biased by the activity *sleeping*, which is a tier 2 activity of *personal care*, and since participants usually sleep at their home, this activity is preferred by the classifier. This explains the poor results for *household activities* for *location* as well, although one might expect that this activity is mostly performed at home. Adding *time* raises the results for *household activities*. Note here, that while *household activities* gains in recall, for *personal care* it drops, showing that misclassifications are being corrected. On the other hand, recall drops for *personal care*.

For some activities adding features leads to an increase in precision and recall, the highest results again achieved by the combination of *location*, *prev.act.* and *time* in almost all the activity groups. Thus, *volunteer activities* can now be predicted, whereas a drop in the results of *education* is observable. Therefore, adding features does not always yield better results. *Travel* is a striking example here, since precision and recall exceed both 90%, although *prev.act.* is not adding much to the results. *Location* and *time* are the preferred features here. *Travel* as well as *sports* are an example of activities that cannot be inferred from the feature *time*. Additional features correct that, especially *location* has an impact on the results for *sports*, which seems to be performed in specific locations.

Rather poorly detected are *volunteer activities*, *education* and *hobbies*, for which none of the features yield high results in recognizing these activities. For the latter two activities we believe that activity recognition systems can help raise the recognition rate, whereas *volunteer activities* could be hard to predict. A key factor here is the time spent on performing the activity: *Volunteer activities* are carried out 10 minutes per day on average. The 10-minute slots of the GTUS could lead to other activities preferred by the classifier. In addition to that, not every participant of the GTUS is engaged in *volunteer activities*.

We conclude that for specific activities, a combination of features yields the highest results for recognizing the activities. Such information could be used as a prior for activity recognition systems. Note here that with the GTUS a demographic analysis for the activities is being performed. We believe that using such information in combination with a classification system that considers a user's activity pattern, even higher recognition rates could be achieved.

4.2.5 Time Use Surveys: A Summary

Having analysed the GTUS 2001/2002 dataset in detail, we conclude the following from the results shown throughout the previous section:

- The two time use survey databases experimented on (ATUS 2006 and GTUS 2001/2002) have fundamental differences in composition and structure, like how to log activities in the survey (i.e., start and stop time versus 10 minute slots). These make it challenging to apply time use surveys across regions. Especially for activity recognition systems, it would be important to have more unified datasets with identical activities and survey data collection approaches.

- Demographic differences were found between the data from North-America (ATUS 2006) and Germany (GTUS 2001/2002) concerning when and for how long activities are carried out. Therefore, it is important to use time use data from the same region one would like to study.

- We detected *location* and *time* as the best performing features in the GTUS dataset, which is in line with previous work in this area. Activities such as *personal care* or *travel* can best be represented by *location* and *time*. *Prev.act.* was also highlighted during the evaluation but made only minor contributions to the overall results. Similar results were found in recent work (Partridge and Golle, 2008), with the exception that *prev.act.* had a higher impact on the results.

The idea of observing past knowledge will be applied in the next section, highlighting how sleep patterns in long-term recordings enable the portraying of sleep rhythms in regard to different working days or specific days of the week.

4.3 SLEEP AS A KEY RHYTHMIC ACTIVITY

While the significance of sleep was identified as an important factor in the medical field, it also attracted interest from psychology due to its impact on daytime behaviour, and from an increasing group of consumers who want to track their sleeping behaviour (see Section 2.4). Scientists also analyse sleep to assess its quality and to discover potential irregularities (Gregory *et al.*, 2004).

This section contributes with a behavioural sleep model based on actigraphy-like motion data from a wrist-worn sensor (see Section 3.2), collected in a long-term and continuous dataset of 141 days. From the original $100Hz$ sensor samples, user-specific data from nights is categorized into four different features, which forms the behavioural model. These features are: *Amount of motion*, *duration of sleep*, *sleep start time* and *sleep stop time*. In this preliminary study, our model is capable of capturing regularities from working days, weekends, as well as from individual days of the week, enabling it to predict future sleeping behaviour by observation of past nights, and discover irregularities that deviate from prior observations.

4.3.1 Sleep Patterns and Sleep Behaviours

After multiple discussions with sleep experts in the medical field, we identified several characteristic features based on inertial-only measurements for characterizing sleeping behaviours. We will first describe these features and then discuss their value for current behavioural research.

Sleep Duration. Observing a person's sleeping hours gives an insight into the habits as well as into irregularities, i.e., shows whenever a person does not reach the usual number of sleeping hours. Healthcare systems benefit from such information, since it enables physicians to detect nights that are out of the ordinary. Although there are different theories on how much a person should sleep, a deviation in the daily routine makes irregularities immediately visible.

Start and Stop Time of Sleep. Persons tend to have regular habits, especially during working days, and therefore this feature is an important characteristic in a person's sleep and sleep behaviour. People tend to go to bed earlier on working days in contrast to weekends.

Motion. Movement during the night is an indicator for sleep quality: Increased movement is considered as a sign for a qualitatively bad night. Sleep scientists, for example, relate the number of posture changes to a normal (15-20 posture changes) or abnormal night (over 30, Gordon *et al.* (2004)). Here, motion not only appears while changing postures but also during spontaneous movements in the same posture. In order to describe a person's night, we use the number of motions detected between non-motion segments.

Discussion. Sleep quality estimation is an interplay of the previously mentioned features. Extracting these features from long-term and especially continuous motion data is therefore important in such assessments. This work focuses on prediction of sleeping behaviours based on previous nights for following nights.

In interviews with sleeping specialists from the local sleeping lab, the value of the chosen features was discussed. Since sleep is usually investigated in sleeping labs, long-term studies are only conducted to assess a patient's circadian rhythm, which is the self-regulation of one's 24-hour cycle that includes sleeping behaviours. Sleeping disorders like *Delayed Sleep Phase Syndrome* are usually immediately visible in such long-term observational studies, since the start and stop times of sleep are logged. With the focus in this study on sleeping behaviour for night prediction, the importance in further studies is apparent and strengthens the selection of our features.

4.3.2 Comparison of Sleep Rhythms

The main reason for building a sleep model based on past behaviour is to first predict a user's upcoming night to discover regularities as well as irregularities and evaluate a person's sleep on a long-term basis. We will illustrate how data is collected and further processed, resulting in a preliminary evaluation on night prediction.

Method and Dataset

The dataset used for this purpose was obtained from a healthy 30 year old male. The sensor was worn on the non-dominant wrist, recording inertial data at $100Hz$ from the embedded 3D accelerometer, which resulted in an almost continuous dataset of 141 days. In the beginning of the recordings minor problems were experienced in the hardware, leading to a few gaps in the dataset. The last 105 days were then continuously recorded 24/7.

The data was further processed by extracting night segments with a threshold-based algorithm, which uses motion, light intensity and sleep time for classifying potential night segments. Ground truth is provided by a time diary maintained by the test subject. The resulting night segments are used to calculate the sleep duration given by start and stop time of sleep.

The motion segments are identified as follows: Over a window of $2sec$ the variance of the acceleration is calculated. Whenever the variance exceeds a threshold

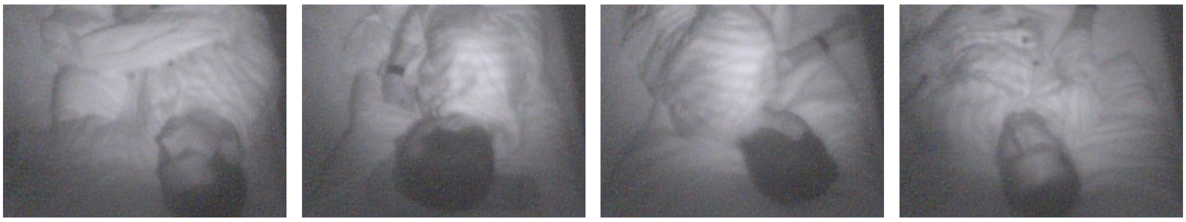


Figure 4.4: Screenshots of the obtained video footage of the test subject.

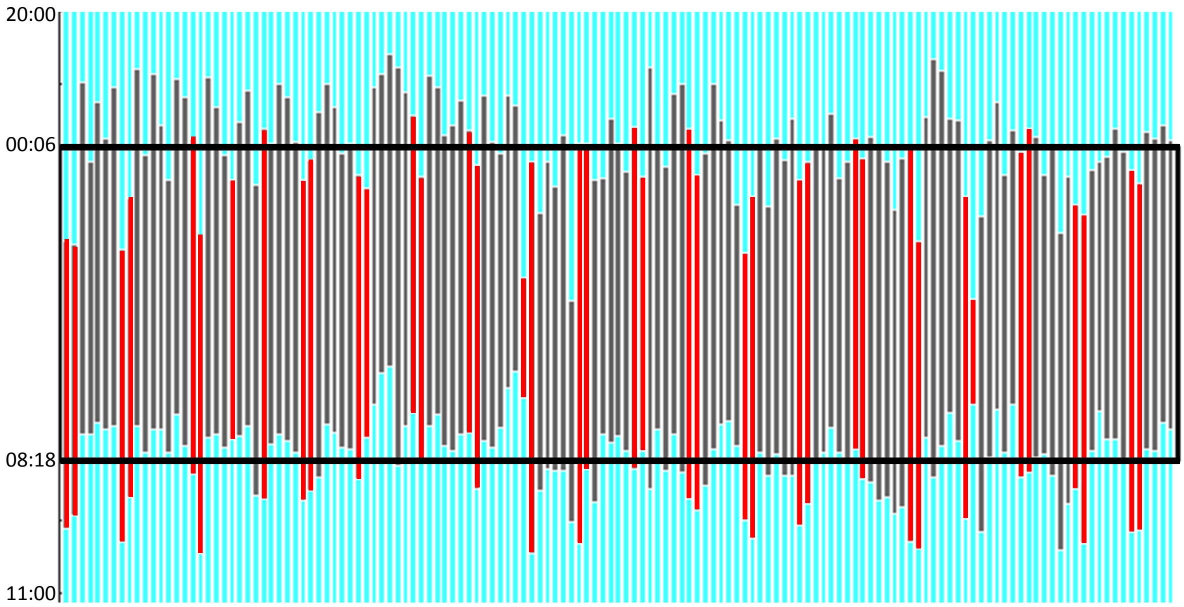


Figure 4.5: All 141 nights (gray = workdays, red = weekends) are displayed with portions of day data (blue) prior and after sleep. The black lines describes the average falling asleep and waking up times calculated over the whole dataset.

vthresh of 1, a motion segment is detected, until the variance is below *vthresh*. We experimentally estimated *vtresh* by video observation. An example of the obtained video footage is shown in Figure 4.4.

Evaluation

The obtained features are used in two different scenarios. The first depicts how well the features describe usual sleep habits in the weekend (Friday and Saturday night) and during workdays (Sunday to Thursday night). The second scenario displays what sleep habits the same weekdays exhibit.

Weekend vs. Workdays. We state that there is a significant difference between a person's sleeping behaviour on weekends and that on working days. The weekends are defined by Friday and Saturday, due to the fact that the person is not working the next day.

In Figure 4.5, all examined nights (gray = workdays, red = weekends) are displayed with the average falling asleep and waking-up times depicted as black lines. By visual inspection, differences between weekends and working days are already visible, showing different patterns in both environments. It seems as if the subject goes to bed later on weekends and also wakes up later, while following a more strict behaviour of going to bed before midnight during the week. The average falling asleep and waking-up time confirm our observation when comparing the black lines to the start and stop times of sleep for weekends and working days.

The overall results from the features extracted for this scenario are displayed in Table 4.3. The difference in all features depicts the deviation between both, weekends and workdays, showing almost none in sleep duration and number of movements but a significant one in falling asleep and wake-up times. As expected, the person falls asleep later on weekends, exhibiting a completely different pattern in contrast to workdays. Interestingly, the user shows an average sleep duration of approximately 430 minutes for both environments, strengthening the theory that a person uses a typical amount of sleep. We conclude that it is possible to detect differences in sleeping behaviours between weekends and working days. The knowledge gained from the features provides us with the information needed to categorize the subject's night into weekend or workday nights. In the next section a more fine-grained approach is performed by comparing same weekdays to each other, giving more insight into a user's sleep.

	sleep duration	**start**	**end**	**#movements**
weekends	428 min	01:55	08:02	61
workdays	434 min	23:47	07:01	59
diff	**6 min**	**2hrs 8min**	**1hr 1min**	**2**

Table 4.3: Average sleep duration, falling asleep time, waking-up time, and number of movements for all nights, observed during weekdays and weekend days.

Same Weekdays. Only a dataset covering about six months makes it possible to gather a sufficient amount of data to observe which sleeping behaviour exists on same weekdays. On average we obtained data on 20 days for each weekday and then built the model. In order to examine the features for these days, we performed a leave-one-day-out cross-validation, by calculating the average for all other features and by comparing the results to the day that was left out. For this, we used a threshold for all features, which displays how well the model fits to the night. The thresholds are: Duration of sleep ± 45 minutes, wake-up or falling asleep time each ± 45 minutes and number of motion-detections ± 15.

The results are displayed in Table 4.4, which shows how the allocation of the same weekday to the model of the other same weekdays performed. Overall, we can state that crucial parameters for similar nights are *falling asleep* and *waking-up times*. These are regular during all scrutinized nights, which strengthens the assumption

accuracy for...	sleep duration	start sleep	end sleep	#movements
Sundays	79%	95%	89%	79%
Mondays	62%	86%	90%	76%
Tuesdays	71%	76%	86%	76%
Wednesdays	76%	76%	100%	67%
Thursdays	53%	79%	63%	68%
Fridays	62%	81%	76%	52%
Saturdays	53%	89%	89%	42%
total	65%	83%	85%	66%

Table 4.4: Accuracy results for the leave-one-day-out cross-validation over all weekdays (Sunday, Monday, ..., Saturday). The last row lists how well the four features were captured overall by the prediction model.

that a person tends to follow a time-critical sleep habit. Although the sleep duration and the number of movements vary a lot on same weekdays, these features need more individual analysis in the context of sleep quality. As stated in Section 4.3.1, movement is an indicator for sleep quality, as well as sleep duration.

4.4 CONCLUSIONS

An important finding of this chapter is the use of statistical data obtained over a population to be applied in activity recognition scenarios. For this purpose we investigated two different information sources: First, we evaluated the feasibility of using time use databases for wearable activity recognition systems, contributing with the analysis of the GTUS specifically. We showed for different population groups (e.g., arranged in age, gender, etc.) what types of activities they perform at specific times of day, reflecting how common these activities are. We compared the results from this study with previous work in Partridge and Golle (2008) and discussed the use of these time use data in wearable activity recognition. The comparison of a US and European time use survey revealed that demographic differences need to be considered. Results of feature analysis for both datasets show that *time* is a very important feature, and that even when considering more fine-grained time slots of 10 minutes, the activity estimation with just *time* is still 50% accurate on average, without using any sensor datasets to train from. Another import feature is *location*, which shows strong affiliation to certain activities, for example, when considering *travelling* or *sleeping*. The best results were achieved when using a combination of the features *location*, *prev.act.* and *time*, showing an overall accuracy across all tier one activity classes of almost 75%.

Such knowledge transfer derived from statistical data can be used for other activity recognition systems. Making use of prior knowledge was introduced in Van Kasteren *et al.* (2008) as well, showing how a sensor network system can profit from knowledge of other sensor network systems in a similar environment. Time use data reflect how common activities are performing for their residents, additionally exhibiting rhythmic behaviour in their activities. Our finding is limited to the same region in which data will be recorded for activity recognition.

Further, this chapter investigates the usage of sleep related information to characterize trends such as workdays, weekends and individual days. For this purpose, we use sleep data features to describe sleeping trends, on a continuous dataset of 141 days from a 30 year old healthy (not suffering from sleep disorders) male. Furthermore, we could assess the test subject's night by our prediction model as a prior, using new nights as input to the model. With such information the rhythmic nature of sleep is depicted, especially by its regular occurrence on, for example, same working days. We illustrated the feasibility of our technique on the obtained dataset, modelling the user's normal nights, which can be put into contrast to a model of an irregular night.

The results presented in this chapter are a first approach to make use of prior knowledge for recognizing activities. As pointed out earlier, sleep is an important activity and can be modelled with such prior knowledge. Therefore, in the next chapter, we will focus on detecting sleep accurately and evaluate novel approaches by comparing standard sleep detection algorithms to ours.

5 SLEEP DETECTION FROM ACCELEROMETER DATA

Contents

5.1 Introduction 77
5.2 Sleep Detection with Accelerometers Only 79
5.2.1 Algorithm Details for the Comparison Study 79
5.2.2 The ESS Approach 82
5.2.3 Experiment Set-Up 83
5.2.4 Evaluation 85
5.2.5 Discussion 89
5.3 Sleep Detection with Additional Information 92
5.3.1 Experiment Set-Up 92
5.3.2 Night Segmentation Approaches 93
5.3.3 Evaluation 96
5.4 Conclusions 99

In this chapter we rely on the extracted features from the previous chapter of two different data sources: Time use surveys and a person's inertial wrist data. We concluded that such features from the aforementioned sources can be used in the activity recognition process. With sleep being a very important activity that exhibits a most prolific rhythm, we employ such information for detecting it. First, we use inertial data only and then we add additional information on top of the accelerometer data. For these scenarios, we propose three different sleep detection algorithms that are based on (1) a novel approach that considers immobility segments, being comparable to traditional sleep detection algorithms, as well as (2) a Gaussian and (3) a generative model-based approach. We give a thorough evaluation of these procedures and highlight important findings in the field of sleep research.

5.1 INTRODUCTION

The importance of evaluating sleep in healthcare scenarios has been presented in Section 2.3, highlighting especially traditional techniques to detect sleep segments. Such techniques are used in actigraphy devices for the diagnosis of sleeping disorders like sleep apnoea or to detect shifts in sleeping patterns. They tend to overestimate sleep, which is why such devices are not necessarily deployed by every sleeping lab. The devices' algorithms are based on the activity count measure, which describes the activity level per minute, to determine sleep segments. Additionally, in order to interpret the data captured, patients are usually required to annotate the sleeping and awake times (Littner *et al.*, 2003).

In this chapter, we evaluate three different sleep detection algorithms that can be used in long-term monitoring scenarios to capture the user's sleeping patterns. More specifically, we argue that wearable wrist-worn activity loggers can be deployed as an additional instrument to complement the traditional polysomnography (PSG) observation method, on the premise that the internal algorithms that estimate sleep and sleep stages are verified to work on benchmark data. For this purpose we obtain two different datasets to evaluate our algorithms:

(1) We first recorded a dataset which consists of PSG patient and inertial data from 42 subjects to determine the feasibility of our novel sleep detection algorithm based on immobile segments. The sleeping lab scenario is used to obtain the actual *sleep ground truth*, i.e., it shows when people are really sleeping or lying awake in bed. Comparing the traditional algorithms to detect sleep and wake cycles (as for example being used by the Actiwatch - see study in Weiss *et al.* (2010) - and the Mini-Motionlogger - see study in de Souza *et al.* (2003)), we show that the novel algorithm for *sleep-wake phase detection* yields a mean detection accuracy of 79%. For this purpose, we explicitly detect segments of idleness, as opposed to being based on detected activity counts, which are simple representations of accelerometer data, as is prevalent in related work. Additionally, with the contribution of a dataset consisting of raw acceleration data from a wrist worn device with ground truth as PSG outputs, detection algorithms can be pitted against each other and results can be reproduced. To our knowledge, such data are not publicly available. One exception is a dataset that includes inertial data (Tilmanne *et al.*, 2009), though the authors focused in their paper on a device that was worn around the waist, which is a common procedure for detecting the body posture outside a sleeping lab.

(2) Then, we use our sleep monitoring system (see Section 3.4) to obtain data from 8 subjects monitored in their home environment over a long timespan to evaluate a generative model- and Gaussian-based approach to detect *night segments*, i.e., when a person is lying down in bed and getting up again. As ground truth we use video footage obtained from a camera that is set up in the user's bedroom. In contrast to the first study, which relies on inertial data only, we embed additional information, like time use data and light sensor readings. We will show the detection of night sleep segments with a high confidence, given a dataset that illustrates a subjects common sleep routine.

For the remainder of this chapter, we introduce the following terminology in regard to the presented sleep detection algorithms: (1) We will use the terms *sleep-wake cycles*, *phases* or *segments* to describe intervals in which a person is actually sleeping, i.e., being in a sleep stage or being awake by lying in bed without moving. (2) In contrast to that, *night/ly sleep segments* or *night segments* refer to intervals in which a person is lying down to fall asleep and getting up after awakening at night-time.

In Section 5.2, we will first present the novel algorithm based on accelerometer values only, comparing our results to traditionally used approaches that detect sleep with data provided by actigraphs, i.e., activity counts. Further, in Section 5.3, we describe the evaluation of two approaches to detect night segments on long-term datasets, which rely not only on inertial data but take into consideration other information like time of day and light intensity.

5.2 Sleep detection with accelerometers only

In order to capture sleep segments in sensor data streams, we established a novel sleep detection algorithm that uses detailed information from the accelerometer data. We rely on the knowledge that we move a lot less when sleeping than when being awake. We evaluate our approach on a benchmark dataset that we obtained in a sleeping lab, allowing to determine the exact time of a person falling asleep. Additionally, we use two traditional algorithms (Oakley, 1997; Cole *et al.*, 1992) from medical sleep trials which we compare directly to our approach.

5.2.1 Algorithm Details for the Comparison Study

In Section 2.4.1, we introduced two clinically tested algorithms to detect sleep-wake cycles. In this section, we discuss the reproduction of both algorithms, Oakley (1997) (which is used for the Actiwatch) and Cole *et al.* (1992) (which is the basic approach for the Mini-Motionlogger actigraph), that we apply on inertial data. In order to compare their results to ours, it is necessary to calculate an activity count that is similar to the ones provided by the two aforementioned actigraphs. The calculation of the activity count can vary substantially, depending on the device that is used. Several research efforts, e.g., Virkkala (2012), have performed comparison studies to map raw accelerometer data to activity counts, such as those from the Actiwatch 7 (CamNtech Ltd.), to show that it is possible to use a 3D accelerometer to calculate the activity counts. This has been evaluated on sleep data that have been recorded with an accelerometer and the Actiwatch in parallel, and is based on the findings of van Hees *et al.* (2010), where accelerometer data during the day were matched to an actigraph output. More recent work (te Lindert and Van Someren, 2013) took a similar approach to derive the activity count solely from inertial data to be able to use traditional sleep detection and sleep parameter algorithms. The algorithms were evaluated by data obtained from 15 healthy subjects in their home environment, showing high agreement rates between epochs (i.e., observed time intervals in sleep research, typically 30 seconds long) for an actigraph and a Microelectromechanical System (MEMS). We will pick up these approaches and show how accelerometer data obtained from our wrist-worn device has been processed to apply the sleep detection algorithms.

Data Processing

The data processing chain for all algorithms can be described in four distinct steps: (1) Obtaining the raw 3D accelerometer data, (2) band-pass filtering the data, (3)

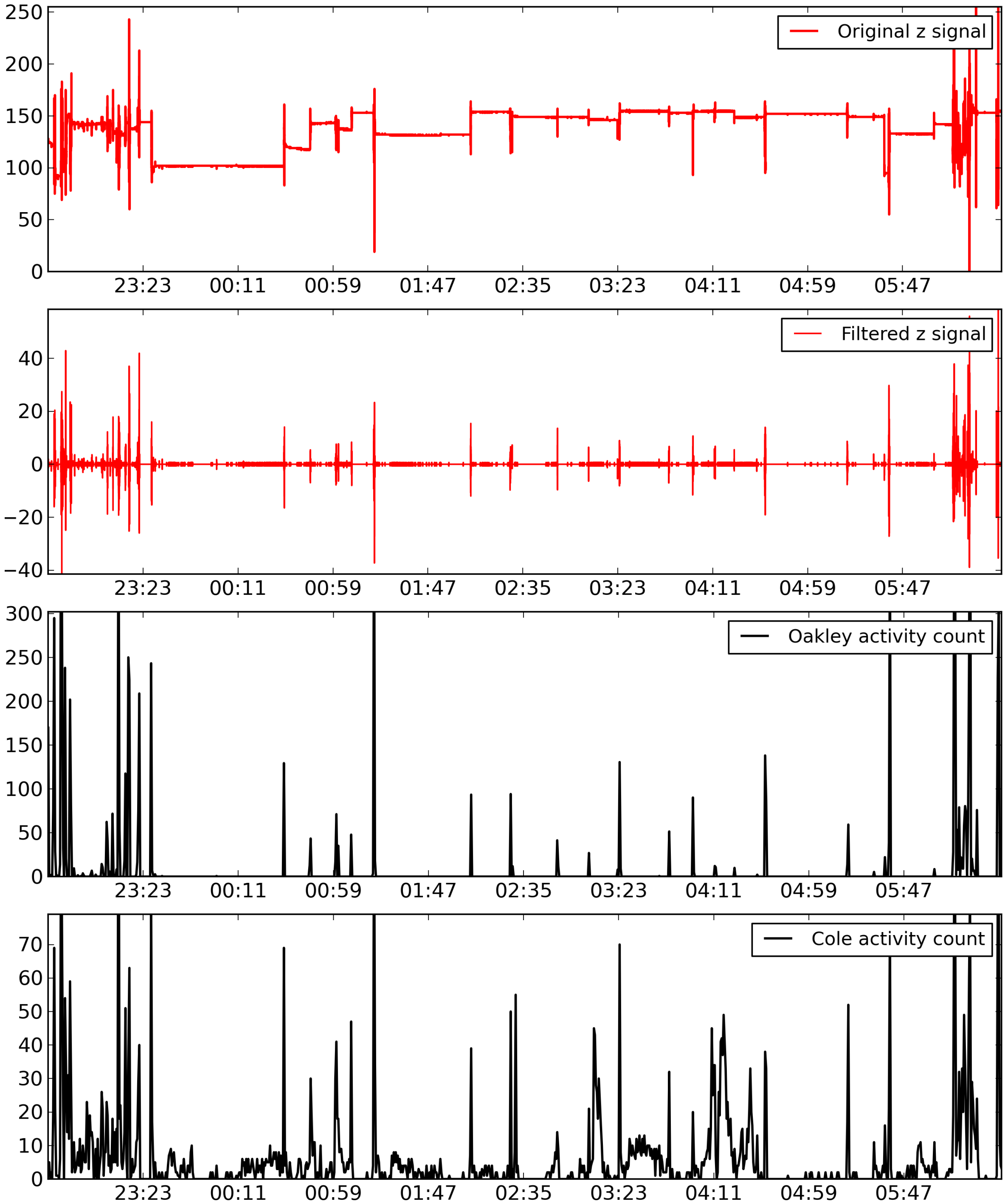


Figure 5.1: Time series of the data abstraction steps performed in this study, to compare methods that detect sleep-wake phases based on 3D acceleration. The raw accelerometer data (top) is first treated with a band-pass filter (middle, in red), after which method-specific features, called activity counts, are computed per epoch (bottom, in black) to detect the sleep and wake phases.

calculating the algorithm-specific features per epoch[33] and (4) applying the sleep detection algorithms on this feature set. Figure 5.1 depicts plots from this process up to the feature extraction for the algorithms by Cole et al. and Oakley. Before any of the algorithms are being used, we filter the raw acceleration data to remove any noise present in the raw data, as well as any high-frequency motion artefacts, as argued for in (Ancoli-Israel *et al.*, 2003). For this purpose we use a 1st-order Butterworth band-pass filter with different low-cut (Oakley: $lowcut = 3.0$, Cole et al.: $lowcut = 0.5$) and same high-cut (both $= 11.0$) frequencies that have been experimentally verified for sleep and wake phase detection in Virkkala (2012). The activity counts for both methods have been implemented as follows:

Oakley. In order to be able to evaluate against the algorithm by Oakley, we use the approach presented in Virkkala (2012) to obtain activity counts for Oakley's algorithm equivalent to the Actiwatch output. The activity count is estimated by using the z-axis[34] only to determine the maximum absolute value inside 1-second windows. These per-second values are accumulated over the observed epoch length and scaled by two parameters, x and y, accordingly: $A = x \cdot G + y$, where A = total activity count equivalent to the Actiwatch activity count, G = activity count over the epoch length derived from inertial data, $x = 66$ and $y = -3.3$. In the experiments of Virkkala (2012) the scaling factors have been estimated for the commonly-used epoch length of 30 seconds for a wrist-worn, 3D accelerometer-based device, which is why we will use this same epoch length for our comparison study to calculate the activity count. The use of a different epoch length will require the reassessment of the scaling factors.

Cole et al. For the algorithm by Cole et al. we implemented a windowed zero-crossing count on the inertial data to obtain the activity count. For this purpose we conducted the zero-crossing on the accelerometer's z-axis as detailed by Karlen *et al.* (2008). We replicated this approach with these parameters, counting the zero-crossings on the filtered data for every 1-second interval.

The activity counts for both algorithms were accumulated over epochs of 30 seconds, enabling the use of these two algorithms with the exact same parameters as introduced in previous work. It is important to note that the results from these re-implementations might still deviate slightly from the algorithms' designs in the way they are embedded in the Actiwatch and Mini-Motionlogger devices, since they operate on essentially different sensor modalities. However, the two independent studies that our implementations are based upon (Virkkala, 2012; van Hees *et al.*, 2010) report encouraging approximation results between the respective activity count methods (embedded in hardware) and their 3D MEMS accelerometer-based reproduced variants, indicating that differences can be expected to be small.

[33] A time interval in which activity counts are being calculated.

[34] This is taken to be the axis that is perpendicular to the hand palm.

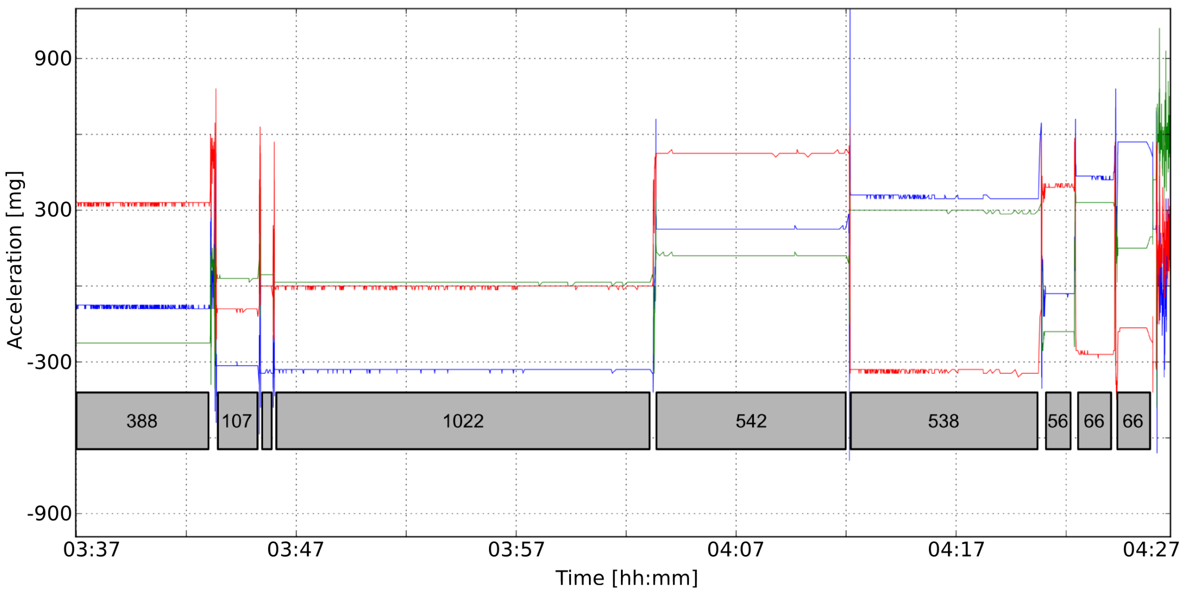


Figure 5.2: Our approach focuses on the detection of sustained periods of idleness during sleep, in which the 3D acceleration signals remain flat. We use these segments and their duration (grey bars at the bottom of the plot, duration in seconds) as a basis for sleep detection, as opposed to sliding a fixed-width window over the data as in traditional activity count-based analysis methods.

5.2.2 The ESS Approach

We present in this section our alternative method for detecting sleep segments, called ESS (Estimation of Stationary Sleep-segments), that is inherently different from the methods of Oakley and Cole et al. since it does not rely on activity counts (whether produced by amplitude or zero crossings) over pre-defined timespans. Instead, it relies on the presence of long periods of idleness that in 3D acceleration data manifest themselves as flat horizontal signals. These are typically interchanged now and then with short transitions where the patient changed his or her sleeping posture. These segments are then used similarly to the epochs in the previous two methods, along with their duration (in seconds) as weights. Figure 5.2 illustrates this concept on typical sleep data from a 3D acceleration sensor (in mg) recorded at $100Hz$ over a timespan of 50 minutes before awakening (at $4:27am$): The intervals between motion segments are typically quite long during normal sleep. The detection of these segments is based on the following method:

$$S_\delta = \begin{cases} 1, & \text{if } \sqrt{\frac{1}{99}\sum_{i=1}^{100}(z_i - \overline{z})^2} > \delta, \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

Our approach consists of two steps: The first one applies a strong low-pass filter to the data to identify the segments in which there are no movement patterns present in the accelerometer data, as detailed in the formula above. Similar to the

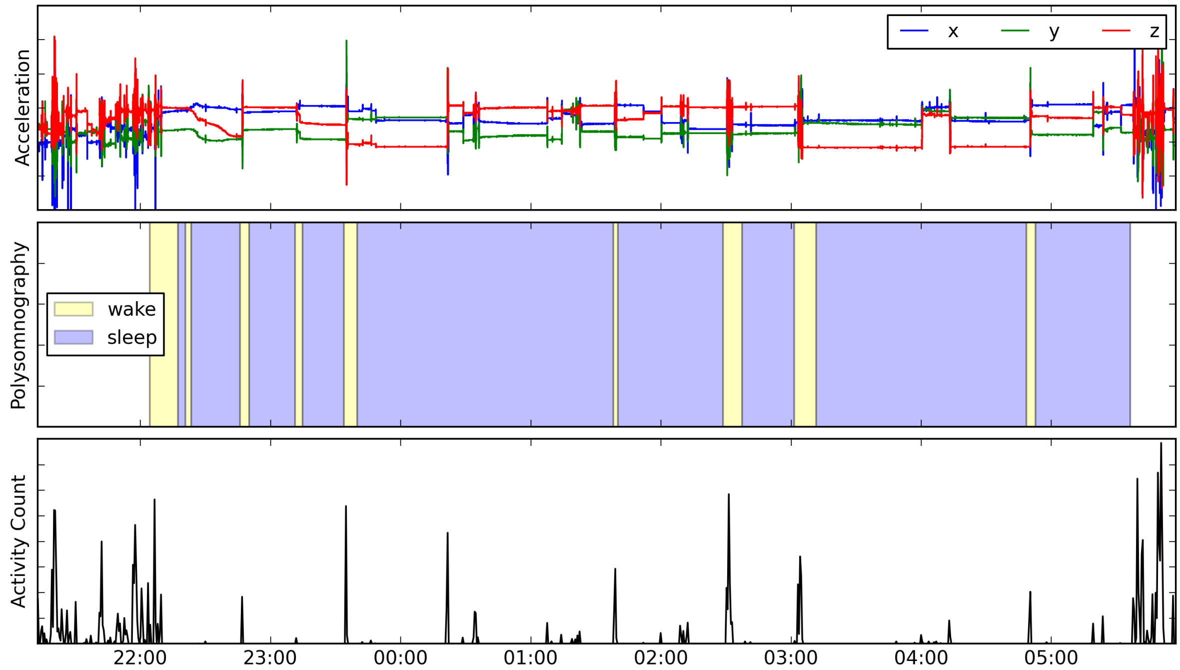


Figure 5.3: An illustration of time series data from a wrist-worn 3D accelerometer (top) and polysomnography (PSG), suggesting that there is a strong correlation between sleep-wake phases (middle) and the number of activities (bottom).

implementation for Oakley and Cole et al., we use solely the z-axis readings to which the filter is applied to. This is achieved by a sliding window approach in which the standard deviation (STD) is calculated over a 1-second interval, after which the resulting value is thresholded for δ. The second step then performs the identification of entire segments and collects these segments' start and stop times along with their length in seconds in a lookup table for later reference. A second threshold parameter is required here to select the minimal length (in seconds) for such intervals in which the accelerometer values remain unchanged.

5.2.3 Experiment Set-Up

For a comparative study between the methods from Oakley, Cole et al. and our proposed ESS sleep-wake phase detection method, we collected 3D acceleration data from 42 patients spending a night in a sleeping lab, while being supervised by somnologists and monitored with PSG. The latter is used as so-called *ground truth* for the patients' actual sleep-wake phases, as well as to provide more details on the individual sleep phases (such as REM vs. Non-REM). For logging 3D acceleration data we use our wrist-worn device, set to a sensitivity of $\pm 4g$ by logging data at $100Hz$. An example of the obtained data is shown in Figure 5.3, illustrating the inertial data at the top, the obtained PSG ground truth in the middle and the number of activities in the bottom plot. Just by visual inspection there seems to be a strong

number of nights:	45
number of patients:	42
gender distribution:	22 male, 20 female
age distribution:	24 - 86
disorders (diagnosed):	insomnia, narcolepsy, SAS, RLS
data (in minutes):	24,475

Table 5.1: Some key properties of the collected benchmark data (SAS = sleep apnoea syndrom, RLS = restless leg syndrom).

correlation between sleep-wake phases and the number of activities.

Study Participants. We gathered data from 42 sleeping lab patients aged between 28 and 86 years, suffering from a variety of sleeping disorders (though most were later diagnosed with primarily sleep apnoea syndrome (SAS), restless leg syndrome (RLS), or narcolepsy). In total, we recorded 45 nights worth of data, whereby three patients wore the sensor for two nights in a row, attesting for over 409 hours of $100Hz$ acceleration samples, annotated with ambient light readings (which are not used in this study but have been included in the benchmark dataset for incorporation in later algorithms) and the PSG details. Table 5.1 summarizes the main details of the patients, indicating how much data we obtained from the wrist sensors as well.

The patients were recruited by staff of the sleeping lab and monitored for at least one night via the standard PSG method, as well as with the wrist sensor. After a short introduction on how the wrist sensor works and what type of data it captures, each patient was asked to start wearing the sensor unit at least one hour before going to sleep and to take it off one hour after waking up. The patients signed a privacy policy and a consent form, allowing the scientific use of the obtained anonymized data from both the PSG as well as from the wrist sensor. Furthermore, they were given documents that describe the experiment in detail and stipulate how the data will be anonymized afterwards in order to enable sharing of the dataset with other researchers for future studies.

In general, the acceptance of wearing the wrist-worn device in addition to the PSG set-up was high: Many patients expressed interest in future studies of the device and responded positively to the idea of having such devices complement PSG for recording in their usual home environment.

Dataset. Data obtained in this study consists of over 409 hours of inertial and PSG data. The sensor was instructed to be worn on the dominant wrist, although previous studies have shown that the wrist placement is not crucial in sleep studies (Sadeh *et al.*, 1994). Before the sensor distribution, the real-time clock embedded on the accelerometer-based device was configured to be aligned to the clock of the PSG system in the sleeping lab, in order to obtain a synchronized dataset. On return of the wrist-worn sensors, their data logs were visualized to the patients as part of the privacy policy (see Figure 5.4, the two bottom plots, for an example of such a

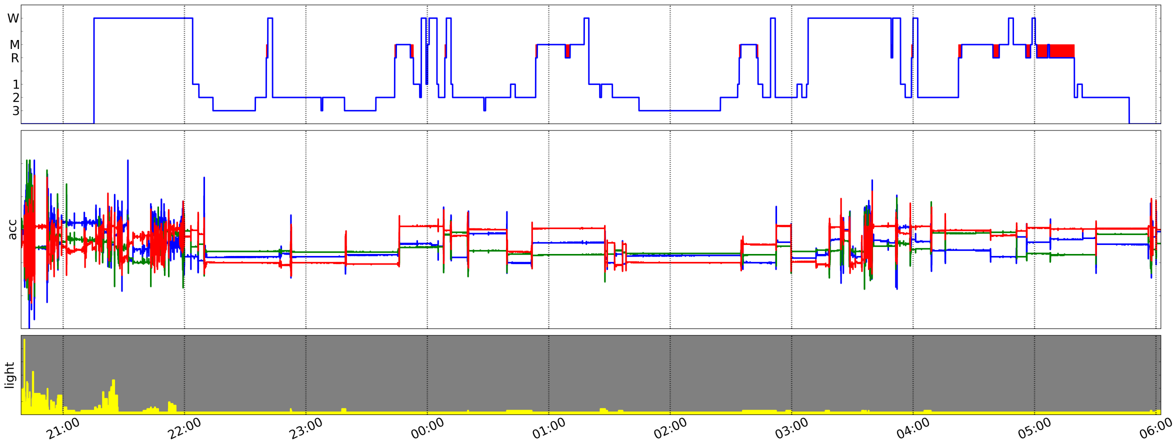


Figure 5.4: TOP: Sleep phases from a 24-year-old female subject, showing awake (W), movement (M), REM (R, red rectangles in the plot) and the Non-REM phases (1, 2, 3). MIDDLE: Inertial data from the sensor worn at the dominant wrist. BOTTOM: Light sensor values for the entire recording period.

visualization).

The patients' PSG data was scored in 30 second epochs by standard procedure obtained after a few days of monitoring, since the medical staff had to analyze the data first, after which the doctor had to summarize these diagnostic findings. The data consists of the patients' demographic information, the wake and sleep stages, with sleep being displayed by REM and Non-REM (sleep phases 1-3) and sleep characteristics, e.g., total sleep time (TST). We refer to Section 2.3.2 for an example of such a diagnosis.

The sleep phases are divided into three different stages, labelled '1', '2' and '3'. Additionally, periods of particularly high amounts of limb movement are marked in the dataset ("M"), as well as when the PSG detected a wake phase ("W").

5.2.4 Evaluation

In order to find the optimum minimum length for the intervals of non-movement, we defined six different interval thresholds (300, 360, 480, 600, 720 and 900 seconds) and evaluated their performance in regard to sleep detection, using the PSG dataset as ground truth. We take the accuracy for detecting sleep and wake phases into account, and use the precision and recall to investigate further differences between the individual parameters' performances. We observe that the mean accuracy is best for the 600 interval which is why we choose this interval to determine immobile segments. Additionally, we set the standard deviation (STD) threshold to 6, as derived by experimental evaluation of different thresholds.

For each of the three sleep estimation algorithms we compare the results to the PSG output. Figure 5.5 shows the visual results for all three algorithms together with the raw data and the PSG estimation for sleep and wake (blue and yellow

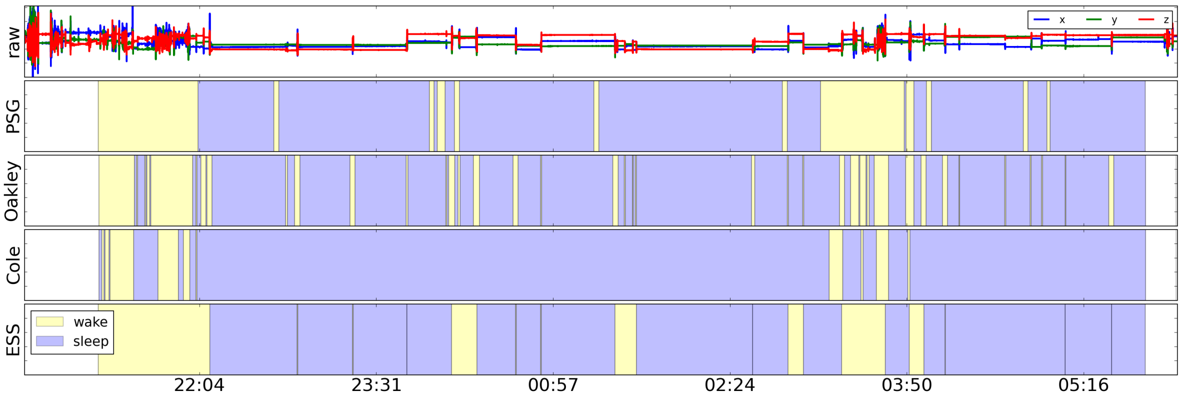


Figure 5.5: Sleep-wake estimation results for a 24-year-old female suffering from narcolepsy. Displayed are the evaluation of the activity count-based algorithms (Oakley and Cole et al., middle plots) and the ESS algorithm (bottom plot) compared to the PSG output. Additionally, we see the raw 3D inertial data of the wrist-worn sensor (top plot). Precision and recall are similar for all three algorithms (87%-99.9%).

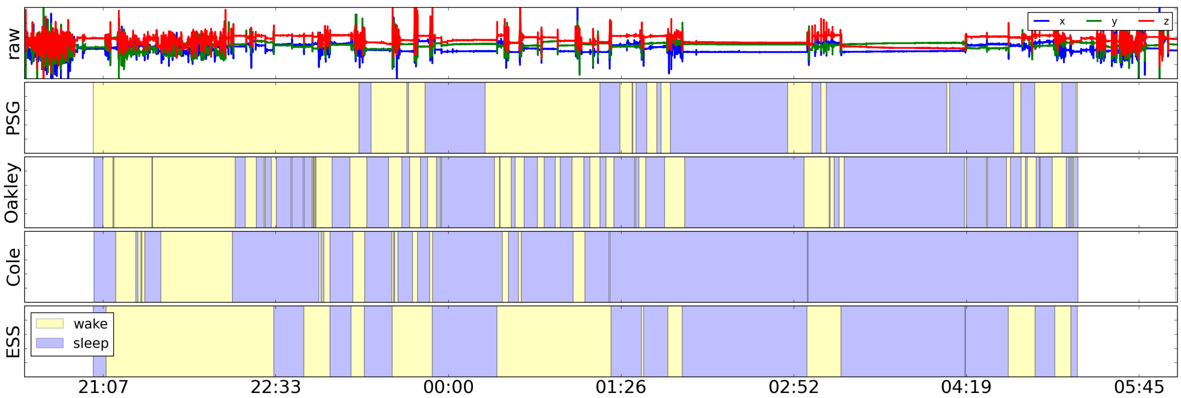


Figure 5.6: Sleep-wake estimation results for a 69-year-old male suffering from SAS. Displayed are the evaluation of the activity count-based algorithms (Oakley and Cole et al., middle plots) and the ESS algorithm (bottom plot) compared to the PSG output. Additionally, we see the raw 3D inertial data of the wrist-worn sensor (top plot). Especially at the beginning of the recording a sleeping segment is detected, which is due to immobility of the patient while starting the PSG. ESS shows clearly detected wake segments, outperforming the other two algorithms by 10%-20% in accuracy.

respectively). For Oakley's algorithm we use the most sensitive threshold of 20[35] to mark the epoch as 'sleep'. By visual inspection we see that Cole et al. overestimates sleep and fails to detect small wake segments. Oakley exhibits many, mostly short wake segments within sleep intervals but detects most of the sleep. Interestingly, ESS detects the initial wake segment which is almost identical to the PSG output, while sleep is being detected accurately. Quantitative results confirm the observations: Accuracies vary for all three in the range of 82% - 85%. More visual results are

[35] As described in Section 2.4, the sensitivity can be set to 20, 40 or 80.

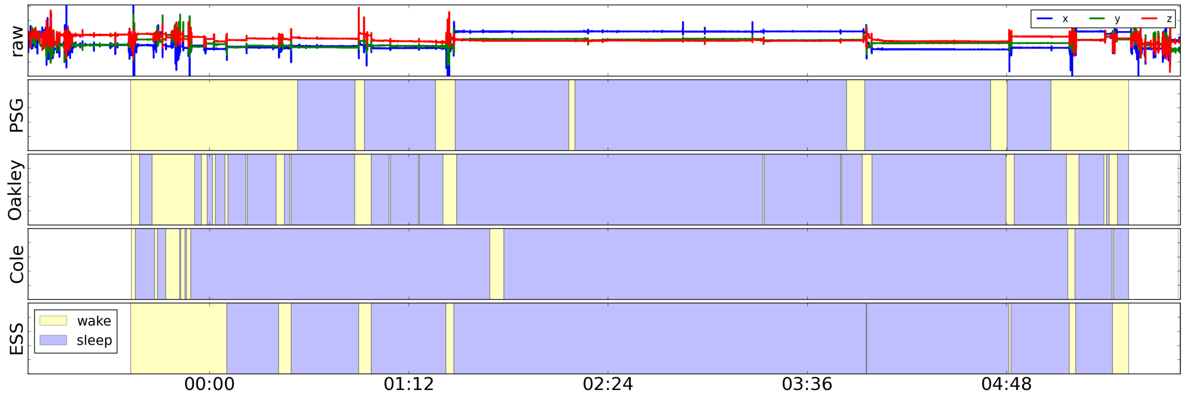


Figure 5.7: Sleep-wake estimation results for a 72-year-old male patient suffering from SAS. Displayed are the evaluation of the activity count-based algorithms (Oakley and Cole et al., middle plots) and the ESS algorithm (bottom plot) compared to the PSG output. Additionally, we see the raw 3D inertial data of the wrist-worn sensor (top plot). ESS clearly detects the initial wake segment while Oakley and Cole et al. tend to detect short sleep intervals.

Parameter	**PSG**	**Oakley**	**Cole et al.**	**ESS**
TST	287 min ±92	413 min ±42	406 min ±44	328 min ±88
SE	66% ±20%	94% ±3%	93% ±4%	76% ±18%

Table 5.2: Total sleep time (TST) and sleep efficiency (SE) for all nights according to PSG, the two activity count-based algorithms and our approach (ESS). ESS estimations are closer to the PSG results as opposed to Oakley and Cole et al.

shown in Figures 5.6 and 5.7, indicating a better performance for the ESS algorithm in contrast to Oakley and Cole et al.

Additionally to the visual inspection, we investigated some sleep parameters to complete the dataset's description. We calculated total sleep time (TST) and sleep efficiency (SE) for each algorithm and compared these values. Total sleep time is the amount of sleep in minutes being detected by the algorithm. Sleep efficiency is the quotient of TST and total recording time (here: PSG start and PSG end). Table 5.2 shows the results for these parameters, indicating a very low mean value for TST as determined by PSG. In comparison to that, Oakley and Cole et al. tend to overestimate TST, while ESS represents a TST value between PSG and Oakley. Here, the inevitable problem can be observed: Activity count-based systems tend to overestimate sleep in general, as depicted in Montgomery-Downs *et al.* (2012). The same is shown here in SE: Oakley and Cole et al. exhibit high values of 94% and 93% respectively, while ESS is very close to the PSG SE. We can state here that all three algorithms differ from PSG, ESS less than Oakley and Cole et al.

Accuracy results for all algorithms are shown in Figure 5.8 as boxplots. We observe that the median (red line in the box) for ESS is slightly higher than for Oakley and Cole et al. Median accuracy values for the three approaches ESS (78.83%), Oakley (74.94%) and Cole et al. (73.75%) are all close to each other. Overall,

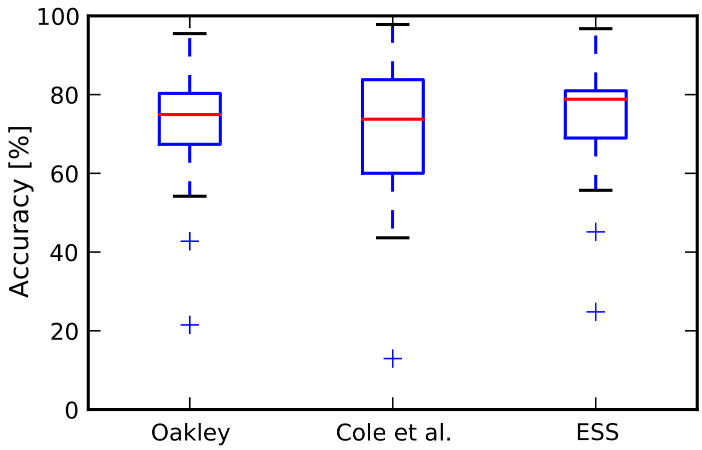


Figure 5.8: Accuracy results for the ESS (median: 78.83%) algorithm compared to Oakley (median: 74.94%) and Cole et al. (median: 73.75%). Both median (red line in the box) and interquartile results indicate that the proposed ESS outperforms the traditional actigraphy-based approaches, while all approaches still leave ample room for improvement.

however, ESS' performance is slightly better over all participants in the study, which can also be seen by observing the interquartile borders. Note here that for each algorithm an outlier is visible (lowest '+' for each boxplot): This is explained by a 69-year-old female patient, who lay awake most of the time during recording, while her inertial data log exhibited almost no movements. For this strong outlier case, all three algorithms estimated sleep instead of wake.

Additionally, we show in Figure 5.9 precision and recall results for both sleep and wake segments (left plots: precision and recall for sleep, right plots: precision and recall for wake). Recall is the portion of sleep (or wake) segments that were correctly identified as sleep (or wake) during the classification. We observe for sleep that precision results are slightly better for ESS (median: 78.95%) and Oakley (median: 78.29%), whereas Cole et al. rests at 72.91%. We can highlight here that all three algorithms perform similarly in retrieving sleep segments from the given dataset. Cole et al. exhibits a high recall for sleep (median: 98.74%), which can be explained by the fact that Cole mostly detects sleep throughout the whole dataset, while Oakley (median: 92.93%) and ESS (median: 94.12%) highlight wake states more often. Interestingly, wake is being detected with a high variety in precision for all three algorithms, showing a higher recall for Oakley and ESS (both over 20% higher than Cole et al.).

The approach suggested by Cole et al., while detecting almost all the sleep intervals, fails to detect the relevant wake segments resulting in a much lower recall. The problem for detecting wake segments in this dataset in particular from a sleeping lab is visible in the results and is a challenge for most sleep-wake detection algorithms. It is also important to note here that the ESS algorithm keeps an adequate balance of detecting sleep and wake segments in the dataset, as opposed to Oakley and Cole et al., which tend to neglect wake segments (see Figure 5.9).

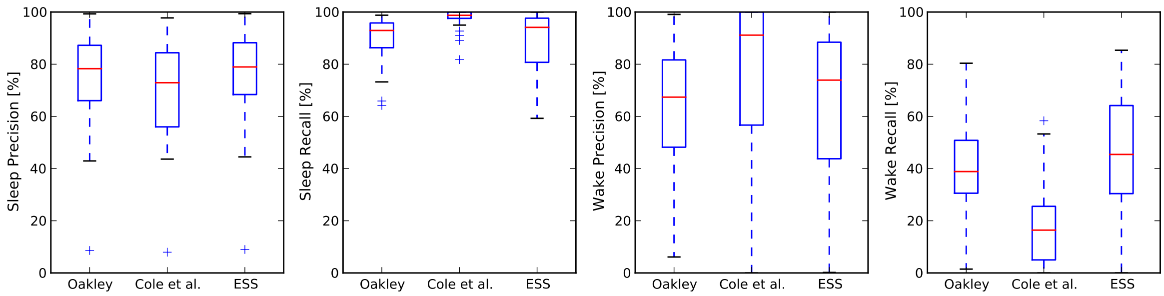


Figure 5.9: Precision and recall results (left: sleep, right: wake) for ESS compared to Oakley and Cole et al. Sleep precision (leftmost plot) for Oakley (median: 78.29%) and ESS (median: 78.95%) are on the same level, while Cole et al. exhibits a better recall (median: 98.74%). For wake the results are inverted: Oakley (median: 38.86%) and ESS (median: 45.42%) show a higher recall, while Cole et al. yields a higher precision (median: 91.14%).

5.2.5 Discussion

This study compares two commonly used sleep detection algorithms to the results of the ESS approach on sleep-wake detection with data from sleeping lab patients. The accuracies for detecting sleep and wake segments are very promising already, as shown in the previous section, yet we believe that parameters can be optimized to improve on the detection of such sleep and wake intervals. We will focus here on finding good candidates for these parameters, as well as what impact the dataset has on future studies.

Dataset. The dataset recorded for this study is a challenging one: First of all, most of the subjects observed suffer from a sleep disorder (diagnosed after their visit to the sleep lab), which makes it difficult to determine when the patient is really awake or just exhibiting spontaneous muscle contractions. This we observed, for example, for a 69-year-old female subject, suffering from sleep apnoea syndrome (SAS). According to PSG the patient was sleeping but this sleep was interrupted by various incidents which let the sleep-wake algorithm detect wake segments even though the patient was sleeping. Second, the dataset includes also healthy patients (though a minority at 5/42 in total), which makes it a rich dataset not only on various sleeping disorders. All sleeping lab patients were diagnosed several days after their stay in the sleeping lab, we did not include healthy patients in the dataset on purpose. Additionally, three patients had to spend two consecutive nights in the sleeping lab, which produced particularly useful data as it should minimize the *first-night-effect* drastically on the second night.

Our dataset with PSG data is enriched with acceleration data from a wrist-worn sensor and enables follow-up research to use data that is not only based on activity counts similar to actigraphs but also on more fine-grained and relatively high-resolution (100*Hz*) signals. Especially in the paediatric sleep research (Insana *et al.*,

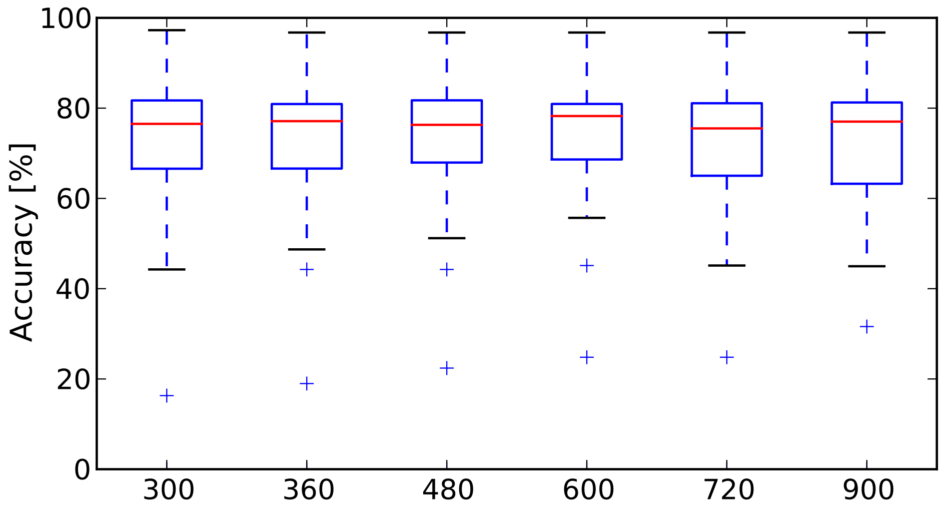


Figure 5.10: Different idleness segment thresholds (x-axis, in seconds) compared among each other, indicate that 600 seconds (10 minutes) is the optimum interval length threshold for the ESS algorithm.

2010) such a set-up could lead to a more reliable detection of sleep-wake segments, since children have been observed to move more during sleep.

Some limitations of the dataset have to be considered here as well: (1) The benchmark dataset contains recordings that are mostly from patients spending one night in the sleeping lab, which is bound to have a drastic impact on the dataset. Normal sleep that approximates that of the patient's home environment usually is achieved on the second night in the sleeping lab. According to the medical staff, the patients' diagnoses could be obtained by spending only one night in the sleeping lab. A minor part of the dataset was obtained from patients who did spend two subsequent nights in the sleeping lab but most of the data can be regarded as atypical and challenging for detecting sleeping patterns. Further evaluation needs to be conducted on how such an effect is influencing the overall results. (2) We assessed only few healthy patients, which is why the behaviour of the ESS algorithm still has to be investigated under "normal" circumstances, i.e., observing sleep in the home environment. This obviously becomes a challenge when ground truth data (PSG recordings) are needed to asses and evaluate the algorithm. Due to this focus on sleeping lab patients, the results on this dataset are therefore likely not to be representative for healthy subjects.

Parameters. As mentioned before, we determined an immobile segment length of 600 to mark the segment as sleep. This immobility threshold when varied yields different accuracy results on our dataset, as depicted in Figure 5.10 for the aforementioned thresholds (300, 360, 480, 600, 720 and 900). We observe an increase of the median until 600, after which it slightly drops again. Whether other thresholds can improve on the results has still to be investigated. For this purpose we have to take into consideration the STD threshold for detecting immobile signals within

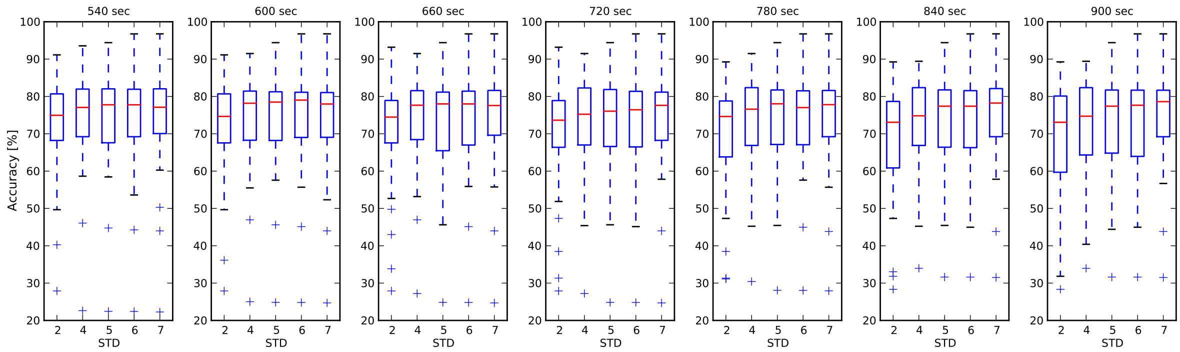


Figure 5.11: Accuracy results for different STD thresholds (2, 4, 5, 6 and 7) for idleness interval length thresholds (540, 600, 660, 720, 780, 840 and 900 seconds). These two parameters were varied to achieve the optimum recognition rate for detecting both sleep and wake segments across all patients.

the inertial data. In Figure 5.11 we show the accuracy results for each immobile threshold length and the individual STD thresholds (2, 4, 5, 6 and 7). A too small STD leads to a lower accuracy, while the highest depicted here (7) stays steady over all immobile lengths. Not shown here is that for lower thresholds, we receive a very high precision for sleep but a very low one for wake, which would contradict a system that should detect both sleep and wake segments. We can conclude that depending on the scenario, the thresholds can be varied and that for this study, the optimum thresholds of 6 for the STD and 600 seconds for the idleness interval length return optimal results.

Activity count-based methods such as the ones by Oakley and Cole et al. take the surrounding epochs into consideration to smooth out the sleep detection over longer intervals of time. Such a step is not implemented in the ESS algorithm. One possibility to embed this is to determine all the stationary segments and detect smaller movement segments in-between these segments. These could be filtered out by setting a specific windowed threshold for these movement segments (e.g., 2-3 seconds) that are additionally marked as sleep. Nevertheless, these are considerations for future studies that need to be evaluated more thoroughly.

This section's topic was limited to the identification of sleep and wake phases present in the 3D accelerometer data. Since the presented dataset contains more fine-grained sleep phase annotations as well, an interesting further line of investigation would be to take a deeper look into algorithms that not only determine sleep-wake intervals but also estimate further phases such as REM and Non-REM, based on wrist-worn accelerometer data. As an initial investigation on how indicative the presence of activity in the data is for particular sleep phases, a histogram was constructed that shows the number of occurrences for each sleep phase per variance bin, as depicted in Figure 5.12 by a distribution of variances over 1 second for all sleep phases (SP1-3 and REM) including wake segments. As can be observed, REM (red) occurs only on the low ranges of variance, indicating that this phase exhibits low variances only (which is in support of what is know about limb motion

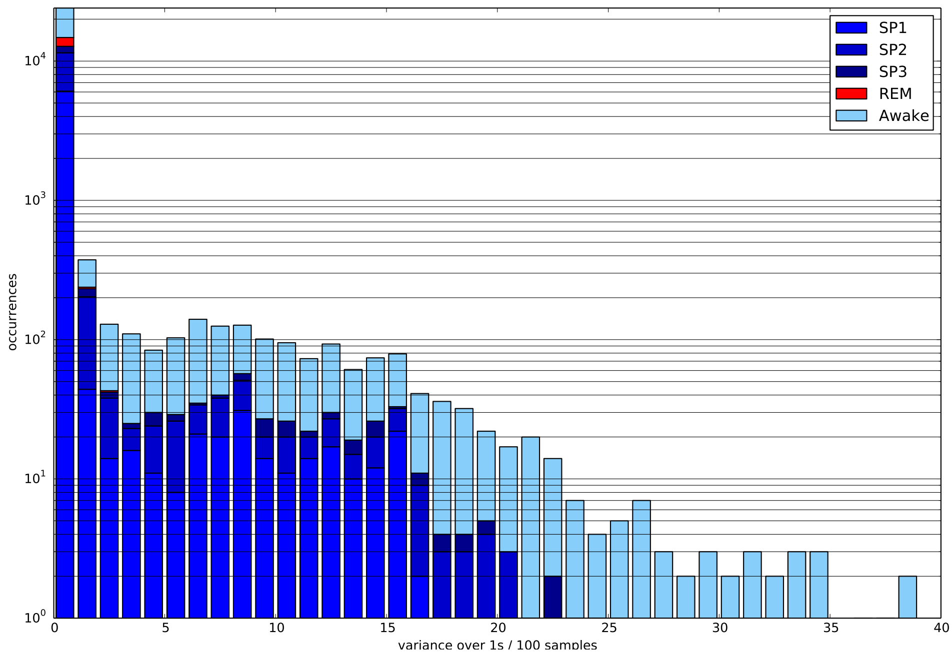


Figure 5.12: Histogram of variance occurring in the different sleep phases REM and Non-REM (SP1-SP3) and wake phases. REM occurs only in low variance intervals.

during the REM phase). However, how well any algorithm could manage to estimate REM/Non-REM phases needs much more investigation on the dataset itself.

5.3 SLEEP DETECTION WITH ADDITIONAL INFORMATION

In the previous section we introduced a sleep detection algorithm that is based on inertial data only. In contrast to that, we present in this section an adaptive method that extracts nightly sleep segments from continuous recordings of inertial data, using an approach that can be bootstrapped from time use data and personal recordings, combined with measured light intensity and physical motion. Additionally, we compare our approach to one baseline algorithm that performs simple histogram-based thresholding on the training data. For this purpose, we make use of the sleep monitoring system presented in Section 3.4 to obtain 8 datasets from study participants that have been recorded over a longer period of time. We will first describe the dataset used for the evaluation, then the approach to detect night segments.

5.3.1 Experiment Set-Up

From our sleep monitoring system set up in the bedroom of 8 participants, we obtained video footage and 3D acceleration data from the wrist-worn device which was set to a sensitivity of $\pm 4g$ to log data at $100Hz$. The data from the IR camera is used as ground truth, showing the exact timestamps of the participants going to bed and getting up again.

In total, we obtained over 4,416 hours / 184 days of continuous recordings from the wrist-sensor with ground truth data from a variety of test subjects as summarized in Table 5.3: These were selected to ensure that several different types of subjects were included and that the segmentation algorithm can be properly stress-tested with sufficiently contrasting types of sleeping behaviours. Two of the test subjects were included that are diagnosed with a specific sleep disorder, one female subject was monitored being 5 months pregnant, and one elderly test subject was included. All subjects were recruited outside the sleep lab environment to test the set-up within a real-world scenario at the participant's home.

Subject	Gender	Age	Hours of data	Comments
1	female	33	336	at 5th month of pregnancy
2	male	30	1,344	normal sleep
3	male	30	432	normal sleep
4	male	28	694	irregular night segments
5	male	35	360	periodic limb movement disorder
6	male	35	672	delayed sleep phase syndrome
7	male	61	648	early morning awakening
8	male	26	576	irregular night segments

Table 5.3: The group of participants used in the evaluation, specifying gender, age, the length of their dataset in hours, and additional factors which are likely to be sleep-relevant. Two of the participants were diagnosed with a sleep disorder, the six others have no known sleep disorders.

5.3.2 Night Segmentation Approaches

The method to detect sleep consists of using as features the *time of day*, *ambient light* and *physical amount of motion* to estimate the start and stop times, as well as the duration, for the night segments in continuously recorded long-term data. First we discuss the chosen features in more detail, followed by a discussion of the two multi-modal classifiers.

Features for Sleep Detection

Time of Day. The time of day is generally a strong clue for estimating the night segment in a day, as most people tend to adhere to a strict circadian rhythm with regular bed- and wake-times. Furthermore, with the help of time use databases, it is feasible to start off with a prior estimate that is generated from a sizeable amount of questionnaire results as introduced in Section 4.2.

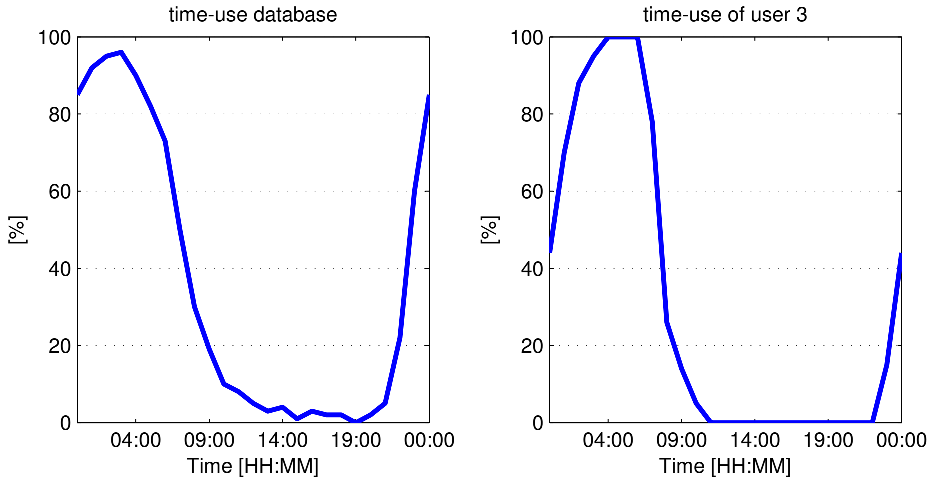


Figure 5.13: An example of typical sleeping times, taken from the GTUS (left) for males between 30-34 years of age, and a one-person (32-years-old) dataset over 30 days (right). The sleeping times of the user match to a high degree to the ones in the GTUS (e.g., shortly after 4:00 both diagrams reach their maximum). The feature time of day is used in the classification process.

The wrist-worn sensor logs its data with timestamps that are provided by the on-board real-time clock. With an expected deviation of approximately 2 seconds per week, this is sufficient to fuse the recorded data with those from other modalities such as the IR camera. For characterization of the night segment, the minute of the day[36] is extracted from the complete timestamp (which holds other information such as year, month, day, and day-of-week as well).

A user-specific model for linking time of day to the nightly sleep segment is trained from recorded data, as will be done for the other two features. However, in the case of time of day, it is also possible to use existing information as presented in Section 4.2 in the form of time use surveys. We illustrated the use of these often large-scale datasets to create informed activity classifiers, which we found particularly promising in the special case of sleep. For this activity an exceptionally large number of samples are available in time use study data. More specifically, we will make use of the German Time Use Survey (GTUS) for one night segmentation approach (see later in this section). In Figure 5.13, the left plot depicts a minute-by-minute normalized histogram of sleep collected from the GTUS and, in comparison to that, the right plot shows the same information from one of our test subjects from the dataset obtained for this study. The similarity indicates that such time use survey data reflects the common sleep habits of a German citizen.

Ambient Light. A second feature candidate for detecting nightly sleep segments is the absence of light in the environment, as most people tend to sleep in darkened

[36] In more detail: Midnight corresponds to minute 0, while 23:59 to minute 1439

environments. Studies have shown that the gradually dimming light at dusk and increasing light at dawn are used to regulate sleep and wake up times, and that these could be used in sleep therapy (Partonen and Magnusson, 2001). The mean of several light sensor values of a specified time frame is used as a feature.

The wrist-worn sensor has been designed so that the ambient light sensor is directed outward, to avoid as much as possible an occlusion by long sleeves or jackets. Furthermore, the sensor, a TSL250 photodiode, has been chosen and configured for maximal sensitivity to light while providing a low dark (offset) voltage. The prototype reads the voltage at full 12-bit resolution, providing a sensor reading that is capable of detecting particularly small lighting changes under poorly-lit environmental settings.

Physical Activity Intensity. A large body of research, including many studies using actigraphy, indicates that activity intensity levels tend to be more elevated during the day and fairly low during the night for subjects with normal sleeping behaviours. As such, this can be used as a discriminant feature for the recognition of the nightly sleep segments. Since the wrist sensor in our studies is worn on the dominant hand, and since the sampling of the on-board accelerometer is set at a frequency of $100Hz$ in order to capture even slight movements, the calculation of standard deviation provides a robust value to represent the wearer's activity level.

Classification Techniques

Before we apply the following algorithms, we split up the dataset for each participant into separate subsets with a duration of 24-hours each, from noon on one day to noon on the next, so that each timespan would contain exactly one nightly sleep segment. The classification performance of the segment was then measured in (1) a leave-one-day-out and (2) a leave-one-user-out cross-validation experiment. The purpose of these two experiments is to first see how well the algorithm performs on previously trained data from the same user, and second, how well the classification does on data from a new user.

Classifier 1: Gaussian model-based approach. The first algorithm is a Gaussian model-based classifier that calculates the variance and mean parameters for the *light intensity* and *motion data* from the training data, and uses a *likelihood per minute of the sleep state* from the GTUS. New data is used as input to the model and a thresholded vote among all values over a sliding window of several (5, 10, and 15) minutes is then used to classify the night sleep data in a robust way. The threshold is experimentally determined and set to 0.1 for both mean light values and variance of motion.

Classifier 2: Generative model-based approach. The second algorithm is based on two two-state discrete Hidden Markov Models (HMMs), which enables capturing changes in sleep habits for new training data efficiently. As features the *variance of acceleration*, *mean of light* and *time of day* (in minutes) are used. The first HMM models the data taken during the awake state, and the other for the sleep state. After

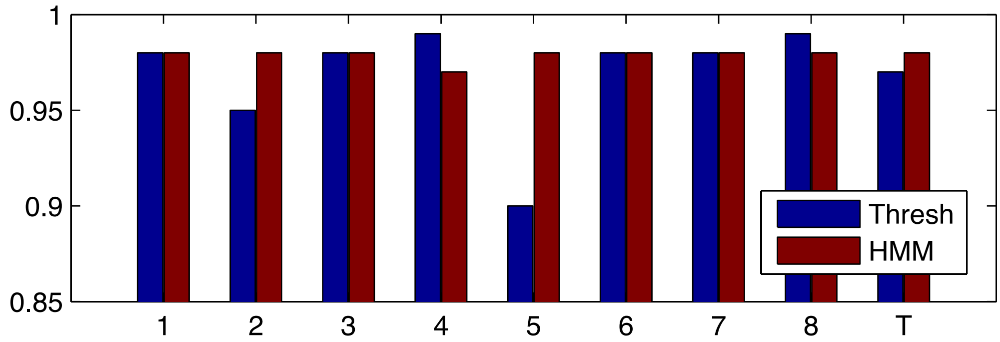


Figure 5.14: The recall values for the Gaussian-based (Thresh) and HMM-based classifiers for the best performing window size of 10 minutes, obtained through leave-one-night-out cross-validation and displaying the total recall (T). Both classifiers perform similar but the HMM approach is able to capture person-specific sleep habits (which is why the recall for subject 5 is higher).

training the HMMs, the highest likelihood for new sequences for each classifier of several minutes (5, 10 and 15 minutes) from the testing data decides whether the method assigns the latest observation in this sequence as asleep or awake data. In the following we present the results for both classifiers.

5.3.3 Evaluation

Results for recall are summarized in Figure 5.14, depicting the cross-validation results for person-dependent training. A comparison between the two classifiers resulted in a set of close detection scores, both peaking at a window length of 10 minutes. However, in case of subject 5, where a lot of motion is present during the night, larger differences can be observed in the results for both person-dependent and cross-subject training. The Gaussian-based classifier uses as prior for sleep the time use database, leading to lower recall in contrast to the HMM-based classifier, which trained on the participant's data. In the following we will scrutinize the impact of an HMM model being trained on a person's data or on all the participants' data.

Table 5.4 lists the results for the HMM-based method using a 10-minute window, which performed best over all tested parameters. In general, the results perform well for most subjects, with cross-user training performing slightly better for most test-subjects. A significant outlier is the result set from subject 7, where the recognition results are hurt from cross-user training. This is explained by the severe early morning wake-up times (4*am* on average) and the overall higher incidence of motion throughout nights, hurting especially the recall of night segments.

Figure 5.15 shows an example where the night segments for several weeks of data from one person (subject 2) are classified. Gaps in the data correspond to sections of the dataset between recordings or sections where the device was intentionally

Subject	Per subject [%] pre	rec	acc	Across subjects [%] pre	rec	acc
1	79	98	91	88	98	95
2	92	98	97	92	96	96
3	87	98	94	95	98	97
4	87	97	94	92	97	96
5	82	98	92	95	89	94
6	72	98	86	96	97	97
7	93	98	97	*80*	*75*	*84*
8	89	98	95	96	98	98
total	**85**	**98**	**93**	**92**	**94**	**95**

Table 5.4: The HMM precision (pre), recall (rec), and accuracy (acc) results for all subjects under per-subject training (left half of table) and cross-subject training (right half of table). Results are from leave-one-day-out and leave-one-subject-out cross-validation respectively. The results perform slightly better for the per-subject training, since the cross-subject approach lacks of sufficient training data to capture all possible variations of sleep habits (see results for subject 7).

turned off. Typical false positives can be seen as small sections in the late evenings, where the test subject was often watching television in a darkened environment. Such false positives were discarded by selecting the largest segment only as the most likely night sleep segment for further analysis.

Discussion. From the night sleep segmentation approach we follow that a high recall classifier can extract most of the night sleep segments from continuous activity data, using additionally information on when sleep occurred and the illumination of the environment. A suitable scenario for the HMM classifier is the detection of advanced or delayed sleep phase syndrom (Weitzman *et al.*, 1981), since we train the classifier on personal sleeping habits, which is identified by the classifier in the shift of night segments. This sleeping disease is characterized by a patient's habit to go to bed late and wake up late (delayed) or go to bed early and wake up early (advanced). Deriving a prior from such a sleeping habit is feasible with our system and requires a person dependent training.

Although all the test subjects were asleep only during the night, the HMMs should be able to detect sleep during the day. Data from a person napping during the day could be used as training data for the HMMs, which is able to adapt to this scenario. Further studies need to be conducted to confirm this theory.

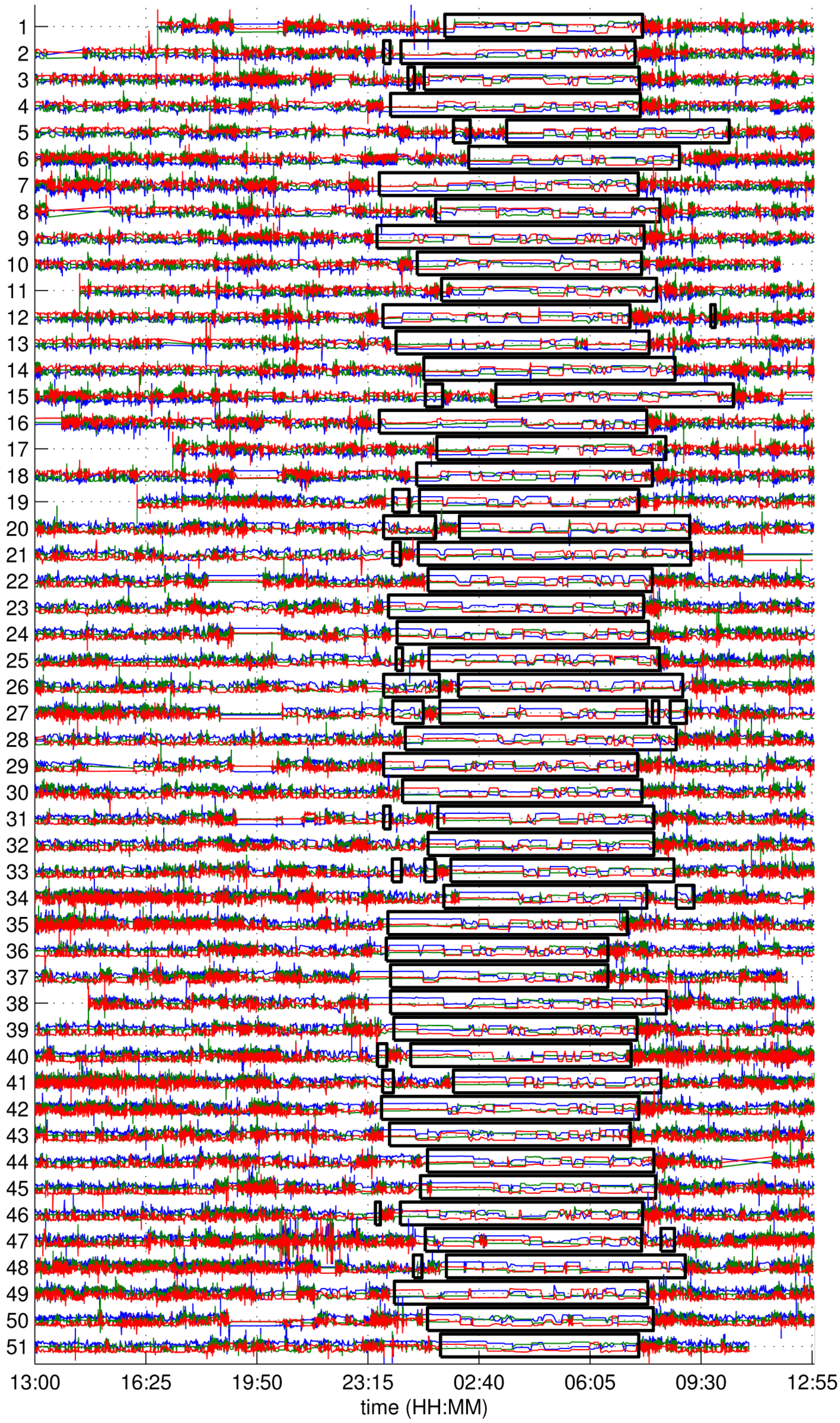


Figure 5.15: The raw classification output from night segmentation (black rectangles) for 51 days of test subject 2 (with each row representing a day's acceleration data) for the HMM-based classifier.

5.4 CONCLUSIONS

With the advent of 3D accelerometer MEMS chips that are both small and power-efficient enough to be included in wearable devices, long-term monitoring of sleep and wake phases has become an attractive and cost-effective instrument to complement traditional sleep lab studies using PSG. The systematic evaluation of algorithms that detect sleep and wake phases in such accelerometer data is still lacking, since current personal sleep devices and systems on the market are closed-source and not meant to be clinically deployed.

This chapter contributes to such systematic evaluation of detection algorithms by presenting two challenging and publicly-available datasets: (1) Over 409 hours worth of PSG-annotated 3D acceleration data logged at $100Hz$ for 42 sleep lab patients and (2) over 4,416 hours (184 days) of continuous recordings from the same wrist-worn sensor for 8 subjects, including ground truth data for night segments. The datasets enable the extraction of sleep rhythms for subjects with different properties, for instance, exhibiting certain sleeping disorders or normal sleep.

We furthermore presented a novel method to detect sleep and wake phases evaluated on the 1^{st} dataset by correlating between sleep stages and inertial data: The ESS algorithm is compared to two traditional activity count-based methods on the PSG dataset. Results show that the ESS algorithm achieves an overall median accuracy of almost 79% for detecting sleep and wake intervals. Compared to the other two methods of Oakley and Cole et al., relevant wake segments are detected with a higher confidence.

Using the 2^{nd} dataset consisting of data from 8 test subjects with a high variety of sleeping patterns, we detected night segments by adding more information to the inertial data: We used light intensity and the information of sleep times to detect these segments. With such additional modalities, evaluation shows that night segmentation with *high recall* (i.e., almost all sleep segment data is retrieved) can be achieved by using an HMM-based method. Compared to a Gaussian model-based approach, the HMM classifier adapts to person-specific sleep habits and yields high results when trained on all data available, due to the high variety of test subjects in the dataset. In addition, the sleep monitoring system has been evaluated which enables video analysis in the wild for long-term recording. Visual inspection is built in the tool by means of an IR camera, which together with detection techniques make the output available for scrutinizing by sleep experts. As a result, the time to analyse long-term video footage can be downsized significantly, since only the relevant sleep segments are made available to physicians for further analysis.

6 SLEEP CHARACTERIZATION FROM ACCELEROMETER DATA

Contents

6.1 Introduction 101
6.2 Detecting Sleep Postures 103
6.2.1 Methodology 103
6.2.2 Results 106
6.2.3 Visual Inspection of Long-Term Data 107
6.3 Visualizing Sleeping Trends 110
6.3.1 Methodology 110
6.3.2 Results 111
6.3.3 Discussion 112
6.4 Myoclonic Twitch Detection 113
6.4.1 Methodology 114
6.4.2 Results 115
6.5 Conclusions 116

In the previous chapter, we showed how night segments are detected within inertial sensor data. Once we extracted the night segments, we discovered mostly rhythmic behaviours in a person's sleep. Such rhythms reflect recurring sleep postures or spontaneous but periodic limb movements. Knowing what is happening during sleep is most important to assess, not only for spotting sleeping disorders but also to derive a sleep quality measure from the obtained data.

In this chapter, we will introduce two different approaches to model sleep data and to visualize its rhythmic nature: By dividing a night segment into non-motion and motion intervals, we (1) present two techniques to visualize sleep postures and their recurrence during sleep. For this purpose, we rely on the sensor data only and describe each posture individually. (2) Further, we inspect sleep data to spot periodic limb movements, so-called myoclonic twitches, that could be linked to possible health-related issues, like restless leg syndrom (RLS) (Lugaresi *et al.*, 1985) or Parkinson's disease (Caviness and Brown, 2004).

6.1 INTRODUCTION

This chapter focuses on visualizing rhythmic patterns within night segments. For this purpose, long-term data has to be obtained and evaluated to detect such patterns. We identified two areas of interest which were arrived at after discussions with the leading somnologist of the local sleeping lab:

(1) Persons without sleeping disorders tend to change postures between 12 and 20 times during relaxed nights (Gordon *et al.*, 2004). This can rise significantly in other situations and this can also change drastically over the course of a night. A view on how often the patient changed posture and between which reoccurring postures this happened is therefore helpful, not only to asses quality of sleep (De Koninck *et al.*, 1983) but especially for particular sleeping disorders such as obstructive sleep apnoea (where certain postures increase apnoea, see Oksenberg and Silverberg (1998)) and restless leg syndrome (where frequent posture changes are witnessed). Therefore, our first contribution in this chapter is the presentation of an effective low-cost posture detection method. Further, we discuss how the visualization of postures may give insights into their recurrence and the assessment of sleep quality in long-term data. As a second contribution, we visualize sleep posture trends without the knowledge of the actual ground truth of a participants' postures. Early research has shown that sleeping quality manifests itself in a low incidence of motion (e.g., few awakenings) during the nights (Peter *et al.*, 2007). When regarded over a longer timespan, these motion models can be used to identify changes and abnormalities in sleeping behaviours. With the modelling of sleep postures into sequences, the spotting of such irregularities is possible. Additionally, such a visualization provides the possibility to observe postures that reoccur periodically.

(2) A second area of interest are involuntary muscular contractions made during sleeping, commonly known as myoclonic twitches, that some people tend to experience when drifting off to sleep but also during the REM phase, or while dreaming. Their detection is not straightforward as these range from subtle and short flexings of (mostly limb) muscles to violent shakes that can last over a few seconds (Caviness, 1996; Roze *et al.*, 2009). As a third contribution of this chapter, we present a method to detect twitches with high precision for additional analysis by physicians. Based on such events, video data is pre-filtered in order to extract and show only relevant parts to the expert and speed up the analysis significantly.

These areas were selected as a trade-off between (a) the modalities of interest for several sleeping disorders and diagnosis types and (b) the attempt to ensure minimal involvement of the patient. Furthermore, with a visualization of sleep postures and the detection of myoclonic twitches we contribute to the spotting of rhythms within sleep, enabling the categorization of such patterns as 'normal' or 'deviant'.

In Section 6.2, we first introduce a posture visualization technique that is based on the classification of sleep postures by training a model with obtained ground truth. In Section 6.3, we illustrate posture sequences without using ground truth data, to visualize sleep posture trends over long periods of time that enable the spotting of rhythmic posture behaviour. In Section 6.4, we analyse movement data and present a detection technique for myoclonic twitches, discussing their meaning in sleep research and the value of the approach within this domain.

6.2 DETECTING SLEEP POSTURES

Sleep specialists have studied sleeping postures for monitoring disorders and revealing a subjects personality[37]. The BBC article is based on the work of Idzikowski (2000), who categorized sleeping postures according to Figure 6.1, which define a person's personality depending on the preferred sleeping posture.

In this section we use inertial data to detect sleep postures similar to Idzikowski (2000). For this purpose, we deployed our sleep monitoring system (see Section 3.4) to obtain data from two healthy subjects monitored in their home environment. The goal is to classify their postures by using video footage as ground truth and to visualize postures sequences. Additionally, we describe a procedure that might be used to assess sleep quality.

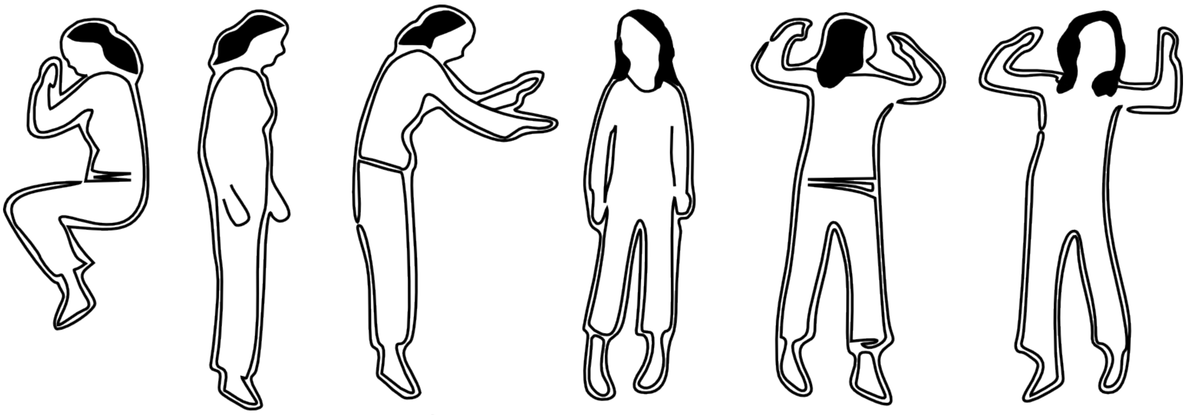


Figure 6.1: Which sleeping posture is your favourite? Personality clues can be derived from sleep postures as described by Idzikowski (2000). [The images have been recreated by kind permission of Prof. Chris Idzikowski]

6.2.1 Methodology

Definition of Sleep Postures. The postures detected in this section are defined in previous work (Van Laerhoven *et al.*, 2008) and are described in Table 6.1. They are divided into four basic (1-4) and four extended sleep postures (5-8). The first two, *left lateral (1) and right lateral (2)* are lying on the left and right respectively. For *supine (3)* the subject is lying on the back. When the body is slightly tilted to one side in this position, a *right supine (5) or left supine (6)* posture is present. Finally, *prone (4)* means that the subject is lying on the chest. Again a slightly tilted body results in *right prone (7) or left prone (8)*. Throughout this section we will make use of the color-codes representing a posture, as depicted in Table 6.1.

[37] http://news.bbc.co.uk/2/hi/health/3112170.stm, last access 09/2014

Posture ID	Posture	Colour	Description
1	left lateral		lying on the left
2	right lateral		lying on the right
3	supine		lying on the back
4	prone		lying on the chest
5	right supine		supine slightly tilted to right
6	left supine		supine slightly tilted to left
7	right prone		prone slightly tilted to right
8	left prone		prone slightly tilted to left

Table 6.1: Sleep postures as described in Van Laerhoven *et al.* (2008). The colours are used throughout this section to differentiate between the different postures.

Dataset Description. In this study, two test subjects at the ages of 29 (male) and 32 (female) were observed during five nights each, neither of the subjects suffering from a sleeping disorder. The sensor was worn on the dominant wrist and data was sampled at $100Hz$ with a resolution of $\pm 4g$. Recording of sleep started approximately one hour prior and stopped one hour after sleep.

For this set-up, the monitoring system was configured to store an image of the person sleeping each minute with the corresponding timestamp. Usually, people tend to move very little during sleep, which is why the image sampling rate is sufficient for posture annotation. Consequently, a night results in about 400 to 500 pictures. Experimental studies revealed that posture transitions happen on average 12 to 20 times per night (Gordon *et al.*, 2004). Therefore, after a night's recording, the subject was able to browse the recorded images and annotate them with the postures. To minimize the annotation effort of the images, a routine first analyses consecutive images for differences and stores only the images that are different. The threshold required for this is calculated by going over all images first, after which the difference between Gaussian filtered images is compared against this threshold. The subject is therefore presented with the final set of these images which consists of typically 10-15 frames that were determined as different. With such an approach, the annotation of one's own sleep postures takes only a little time, making it feasible on the subject's side.

The postures are extracted by mapping the sensor values directly to the annotation by relating the timestamp of the images to the timestamps of the sensor readings. It happened twice that pictures were annotated as unknown because the test subject was performing a movement in the picture. These unknown marked areas in the dataset are cut out since they would lead to wrong results. An unknown marking could contain a known posture and would have a negative influence on our results.

Extracting Non-Motion Segments. Movement during sleep happens sporadically, which allows to detect and gather the phases when the person does not move. A sleeping segment starts with the subject lying down to sleep and stops with the

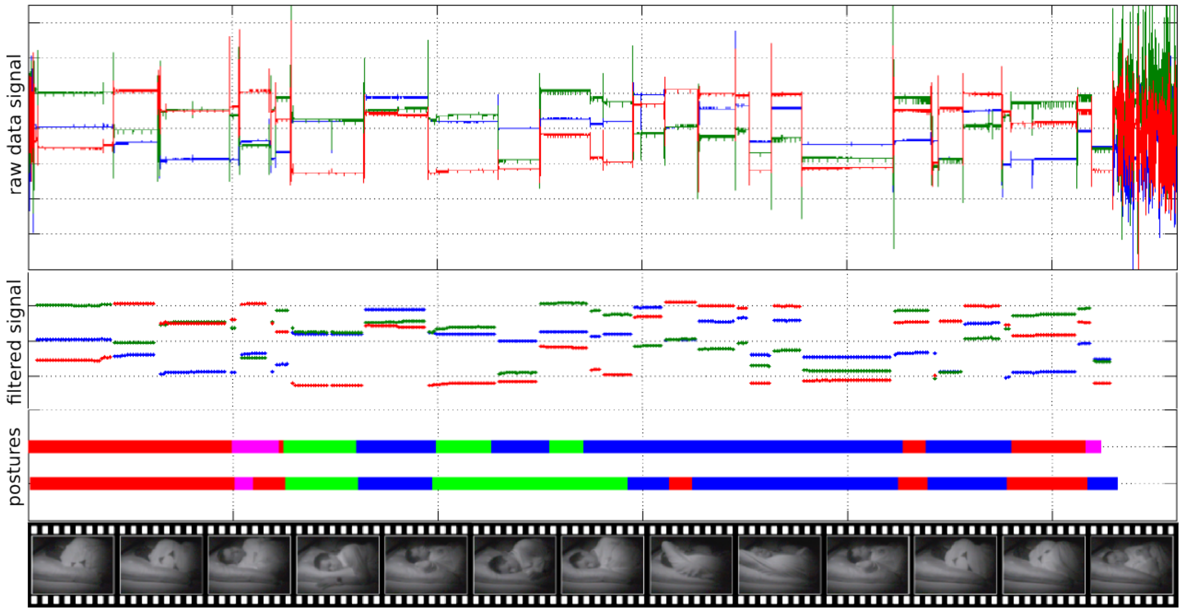
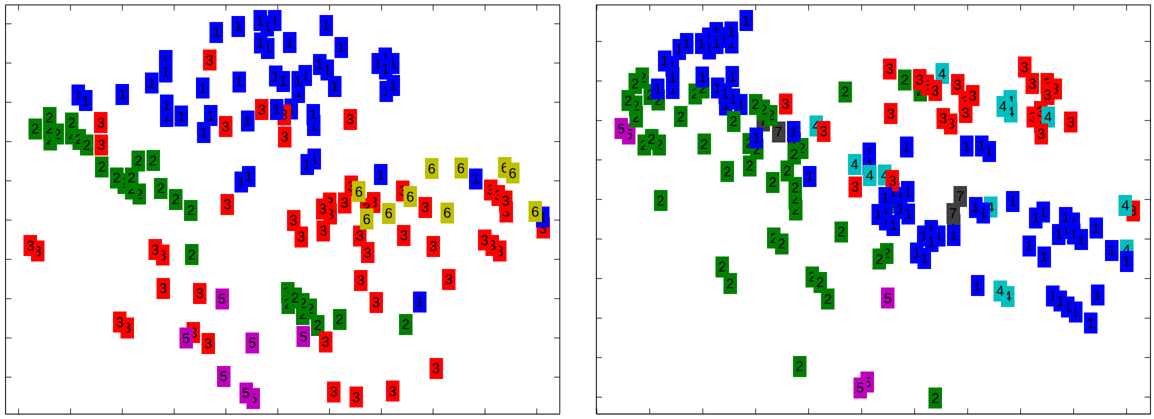


Figure 6.2: Example of (from top to bottom): Raw data, filtered values, postures as obtained by ground truth and estimation for one night and images of different postures taken with the IR camera. The estimated postures are very similar to the ground truth, as can be seen in the second to last plots (top barplot = ground truth, bottom barplot = estimated sleep postures).

subject getting up. A typically huge difference in standard deviation (STD) of the acceleration data between being asleep or being awake (see Figure 6.2 top plot) allows a threshold t in our accelerometer data to be set to detect movement vs. non-movement. Therefore, we determine the STD over 1 minute worth of sensor readings (since we obtained ground truth for every minute), and mark the window as *non-motion* if the STD is below $t = 5$ (see Figure 6.2, first two plots).

A sleeping posture manifests itself in a time series plot as a constant value over longer timespans. A transition is typically associated with lots of movement over a small timespan, and tends to appear quite regularly as well as the wrist-worn sensor picks up hand repositioning and twitches during sleep. By setting the t parameter, these transitions are removed from the data and isolated from the postures.

Analysing Non-Motion Segments. In order to investigate what types of algorithms can be applied most successfully to classify the sleeping postures, we have followed a two-phase process. First, visual inspection of the data was performed to investigate the nature of the data, with both scatter plots of the 3D data and 2D visualizations using principal component analysis (PCA). In a next phase, the data was clustered with k-means and afterwards these clusters were used in a standard k-Nearest-Neighbour (KNN) classifier that was trained with posture labels and accelerometer readings using 5-fold cross-validation, where one fold corresponds to one night.



(a) Subject 1

(b) Subject 2

Figure 6.3: PCA visualization of sleep posture data as clusters to a 2-dimensional space. Most of the same postures are clustered close to each other.

Typical results for this procedure are shown in Figure 6.2 in the second to last bottom plot. The ground truth is plotted as coloured bars at the top (with red, green, blue and magenta depicting supine, right lateral, left lateral and right supine, respectively), while the bottom shows classifier results using the same colors. Additionally, the image taken with the IR camera are put at the bottom, as examples of the captured postures.

6.2.2 Results

Visual Inspection of the Obtained Data. In 3D visualizations cluster analysis showed clear areas that are marked for each posture, with a minimum of overlap and outliers. For better visibility, the data dimension was reduced from 3D to 2D by using PCA on all available data per subject. In Figure 6.3, the results of this reduction are displayed for both subjects. The different postures are coloured and numbered individually and, although they show more overlap than in the 3D plots, also illustrate: (1) That postures that happen frequently (in particular the lateral and supine postures) stretch out over a larger area. (2) That these posture clusters are not Gaussian in nature but rather tend to appear as a mixture, and (3) that there is a large distinction between subjects concerning what postures tend to occur and in which quantities.

Clusters can be observed for several postures per subject (1, 2, 3 and 5 for subject 1). Only posture 3 for subject 1 and posture 1 for subject 2 seem to be widely spread, although this is less prominent in the 3D plots. The clusters for each posture can still be identified. Interestingly, the two subjects shared their most common postures (1, 2, 3 and 5) while other postures occur rarely or never in the datasets (4 and 6), which is in line with the theory of subjects favouring certain postures (Idzikowski, 2000).

classes

groundtruth	1	2	3	4	5	6	7	8	
1	**612**		46			6			0.92
2		**403**	5						0.99
3	61	48	**436**		2	14			0.78
4									
5		5	15		**84**				0.81
6	11		11			**76**			0.78
7									
8									
	0.89	0.88	0.85		0.98	0.79			

(a) Subject 1: Overall accuracy 88%

classes

groundtruth	1	2	3	4	5	6	7	8	
1	**804**	3	10				3		0.98
2	22	**390**	14		6				0.90
3	29	13	**172**	39					0.68
4	8	32	6	**108**					0.70
5		11			**30**				0.73
6									
7	9						**41**		0.82
8									
	0.92	0.87	0.85	0.73	0.83		0.93		

(b) Subject 2: Overall accuracy 88%

Figure 6.4: Confusion matrices for two subjects showing very high results for certain postures (e.g., recall of posture 2 for subject 1 and recall for posture 1 for subject 2).

Quantitative Results. Although visual inspection of the data samples were promising already, we performed a quantitative study of our approach as well to ensure that classifier approaches would work on unseen data in cross-validation. Figure 6.4 shows the confusion matrices for both subjects. The last row provides the posture's precisions, whereas the last column displays the recall figures. KNN (with $K = 5$) achieves a precision of 88% and 86% for the respective subjects, and a recall of 86% and 80%. The overall accuracy is 88% for both subjects. For postures 1 and 2 a recall over 90% and precision above 87% was obtained for the subjects.

This confirms findings from other work with similar but less precise sensor modalities (Van Laerhoven *et al.*, 2008) and shows that, with using training data from just 4 nights per subject, acceptable sleep posture recognition can be achieved. We present a visualization of these postures in the following section for one subject to show the recurrence of postures over a month worth of nights.

6.2.3 Visual Inspection of Long-Term Data

In addition to the 5-days dataset of the male participant in this study, 25 nights were recorded, resulting in a dataset of 30 nights. The dataset contains each day of a week in order to capture different sleeping habits of weekdays and weekends. The purpose of this study is to visualize sleeping trends over several weeks, enabling the spotting of recurring postures to be able to detect 'irregular' nights.

Sleep Questionnaires. For the purpose of detecting sleep abnormalities, a sleep questionnaire is still the most widely used instrument. We will introduce sleep questionnaires here as a complementary approach to highlight irregular nights. Such questionnaires are applied to obtain a subjective evaluation of sleep by the test subject and characterizes its quality. If a person thinks she is feeling well prior

in the evening
1. How do you feel right now (1=very good,...6=very bad)?
2. Did you sleep during the day? (time and duration)
3. How tired are you right now (1=very,...6=not at all)?
4. When did you go to sleep?
in the morning
5. How tired are you right now (1=very,...6=not at all)?
6. How do you feel right now (1=very good,...6=very bad)?
7. Did you wake up during the night? How often?
8. When did you get up?

Table 6.2: Eight basic sleep quality related questions that were answered by the participant before going to bed and after awakening.

to sleep, she most certainly will sleep better than when in a bad mood and very pessimistic about sleeping well (Buysse *et al.*, 1989). Therefore, sleep questionnaires are needed to detect this mood which cannot be detected by any kind of sensor.

A minimal sleep questionnaire was used for our study to be able to obtain the ground truth for nights that differ from other, regular nights. Although a state-of-the-art sleep questionnaire used by all medical sleep research does not exist yet, various questionnaires are being recommended by sleep specialists (Carpenter and Andrykowski, 1998; Johns *et al.*, 1991). The sleep questionnaire used in this study is based on the sleep diary described by Dr. Tilmann H. Müller[38], which is recommended to be applied for home sleep quality assessment. Some questions, for example, whether alcohol had been consumed or medication was taken, were deemed not relevant for this study at this stage and left out. It is important to register healthy sleep and find irregularities when monitoring a person over several days by considering as little information as possible. Therefore, we used only relevant questions for our study (see Table 6.2). Since the diary has to be used for at least 14 days, we prolonged our study to 30 days.

Discussion. In Figure 6.5, all nights and the detected postures of the subject are shown. The plot on the left displays all sensor values obtained over 30 nights, whereas in the middle the detected postures with the classifier from Section 6.3 are presented. The plot on the right displays the results from the questionnaire of the questions 1 (column 1), 3 (column 2), 5 (column 3) and 6 (column 4), color-coded by having more red for higher scores and more green for lower scores.

Although this figure presents just a one-subject study that is limited to one month, certain patterns and observations are striking when analysing the posture's visualization in Figure 6.5. Several overall patterns can be witnessed, such as the fact that this subject tends to start the night by lying on his left (red) or back (blue), and

[38] http://www.schlafgestoert.de/site-49.html, last access 09/2014

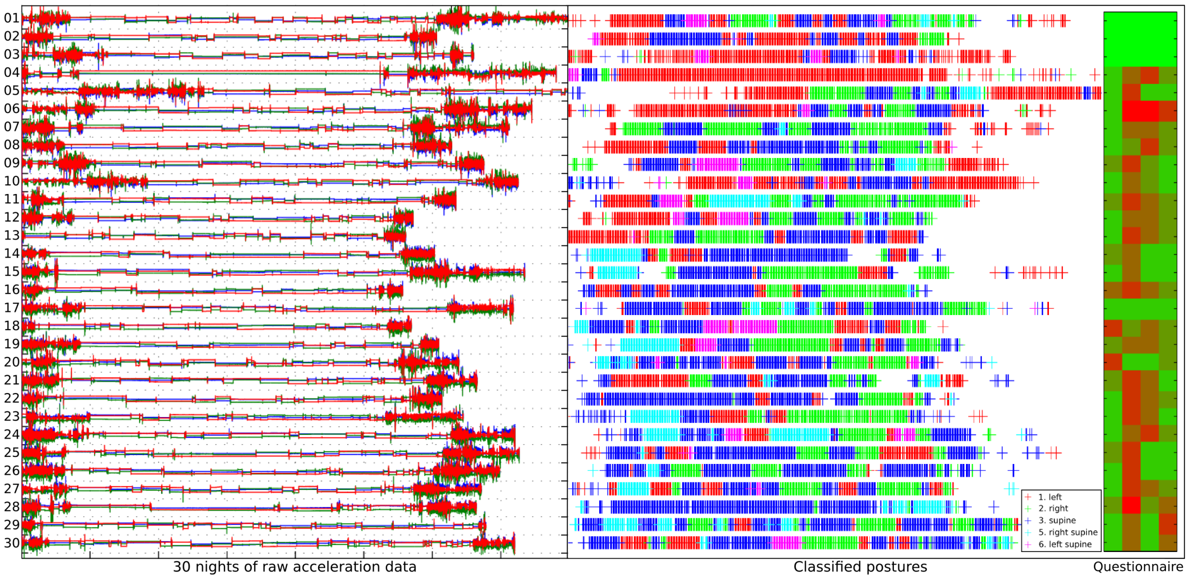


Figure 6.5: Thirty nights (one per row) of subject 1: (left) Raw 3D accelerometer data, (middle) posture classification, (right) questionnaire scores (per column from left to right: Questions 1,3,5 and 6, with green for low and red for high values). Favourite postures recur not only during one night but over consecutive nights as well.

gradually is positioned more on his right towards the morning. Most nights in the first half of the month tend to start on the subject's left, while the second half (from the 14th day onwards) tend to start from the supine (blue) or right-supine (cyan) posture. Additionally, we observe the rhythmic nature of sleep postures: Lying on the right reoccures every night, mostly starting in the middle and towards the end of the night. The subject tends to keep his usual postures and might change the posture he is falling asleep in but will retain this behaviour in the following nights, which was confirmed in several discussions with the sleep lab staff. Therefore, visualizing known posture sequences is beneficial for physicians to spot irregularities.

In addition to posture rhythms, we investigated the correlation between the sleep questionnaire and the different postures. It can be noticed that especially nights 29 and 30 are marked as poor in terms of sleeping quality by the subject. The subject described these nights as feeling good and a bit tired before going to sleep and woke up feeling very bad but not tired anymore. The high number of posture transitions between left- and right-based postures during these nights is also visually clear. In contrast to that, night 8 seemed to have been better, with less posture transitions and following the overall regular sequence.

Investigating certain patterns in postures and modelling such patterns would give more insight into this study's theory and would lead to more precise findings. To this point, the visual inspection drives these findings into the direction of being able to model sleep by examining sleeping postures. While in this section's scenario we are dependent on the ground truth, we present in the next section an approach to visualize sleeping postures without knowing the actual posture.

6.3 VISUALIZING SLEEPING TRENDS

For this section, we utilize the night-segmented data from Section 5.3 to present a coarse-grained visualization method based on sleeping postures over several nights. The dataset includes data from five different subjects monitored in their home environment over long stretches of time, where candidate night segments were obtained. While the previous section classified sensor data to visualize known postures, we focus here on the visualization of unknown postures to open up a venue to describe their rhythmic occurrence in a single night and over several nights. With such an approach, the spotting of outliers in long-term posture visualizations is possible, enabling physicians to categorize a person's sleep in regard to quality and health without the need of obtaining ground truth data.

This section will look at non-movement data within night segments. Although the filtering of this data is straightforward (see Section 6.2 for the applied technique), the main question that will be answered in this section relates to how these postures can be visualized most appropriately. We will first introduce the methodology used for this purpose and then show how we evaluated our approach with an inquiry.

6.3.1 Methodology

Dataset Description. The dataset used contains five test subjects between the age of 26 to 61, as described in Table 6.3. Additionally, the age, gender and information about the participants sleep is displayed, as well as how many hours of night segment data we obtained. In total, we use 1,192 hours worth of inertial data that is scrutinized for different postures.

Subject	Gender	Age	Hours of sleep data	Comments
1	male	61	226	early morning awakening
2	male	26	208	irregular night segments
3	male	28	213	irregular night segments
4	male	35	33	delayed sleep phase syndrome
5	male	30	407	normal sleep

Table 6.3: The group of participants used in the evaluation, specifying gender, age, the total length of their night segments and additional information that might be sleep-relevant. In total we obtained 1,192 hours of night sleep data.

Clustering Technique. The reoccurring sleep postures are modelled by a clustering method approach which facilitates an optimal visualization of the posture sequences later on. Similar to the k-means method, the *Kohonen Self-Organizing Map (KSOM)* (Kohonen, 1990) is a clustering algorithm that holds a fixed number of cluster centroids, to which new data samples are clustered in an iterative way by selecting and updating the cluster for which the centroid is closest to this new input. 36 cluster centroids are chosen for that, keeping the range of the 6*x*6 map small, which are allocated on a semantic map created by the KSOM. Similar values are mapped close together, while dissimilar are mapped apart. The choice for the Kohonen map brings in this case an added value in terms of visualization: By requiring that clusters are structured along a tight topology, neighbouring clusters obtain centroids after training that are close in Euclidean space. By assigning a gradual color map to the Kohonen clusters, the clustering of posture data will result in similar postures being assigned a similar color. Thus, even if not the same cluster is assigned to 2 similar postures, the visual representation for both will look very much alike.

The color-code is obtained by using the night segments from all subjects as training input which leads to a unique posture colouring for each subject. The map grid coordinates are normalized to a unit square, and each coordinate is mapped to a color. As output we receive a grid allocation, which is used as input for a Hidden Markov Model (HMM) classifier that is trained to show similarities through the dataset per subject, resulting in a typical posture sequence which was used in the following paragraph.

Evaluation Technique. The method to evaluate the posture representation is to conduct a straightforward survey by multiple users. The participants were recruited in the research facilities but also among family and friends. The goal of this survey is to evaluate the uniqueness of clustered postures for each subject, by using posture representations of several nights for all five subjects which are then shown to each of the 60 participants (Figure 6.6). Additionally, a single night from each of the subjects is shown to the respective participant (Figure 6.7, *a-e*). They are asked to assign each of these nights to the night-collection of subjects from Figure 6.6.

The results of the survey were obtained as follows: The number of correct answers per typical night (Figure 6.7 *a-e*) is divided by the number of participants, resulting in an accuracy value per question. The overall accuracy is then calculated by the sum of each individual accuracy divided by the number of questions.

Note here, in a first version of the survey we displayed the postures from all eight subjects of the original dataset from Section 5.3 with their typical nights. Again, people were asked to allocate the nights, which proved to be difficult, since too many plots with colors were displayed at once. Therefore, the optimum representation of five subjects was chosen, since the feasibility of assigning postures correctly is the goal of this study.

6.3.2 Results

An overall accuracy of 92% is reached for correct allocation of the typical night to the subjects overview plots. Figure 6.6 shows how the test subjects differ in their postures. A suitable scenario of such a representation is to compare a new night of a patient to the previous nights just by the color encoding which leads directly to outliers that can be scrutinized further.

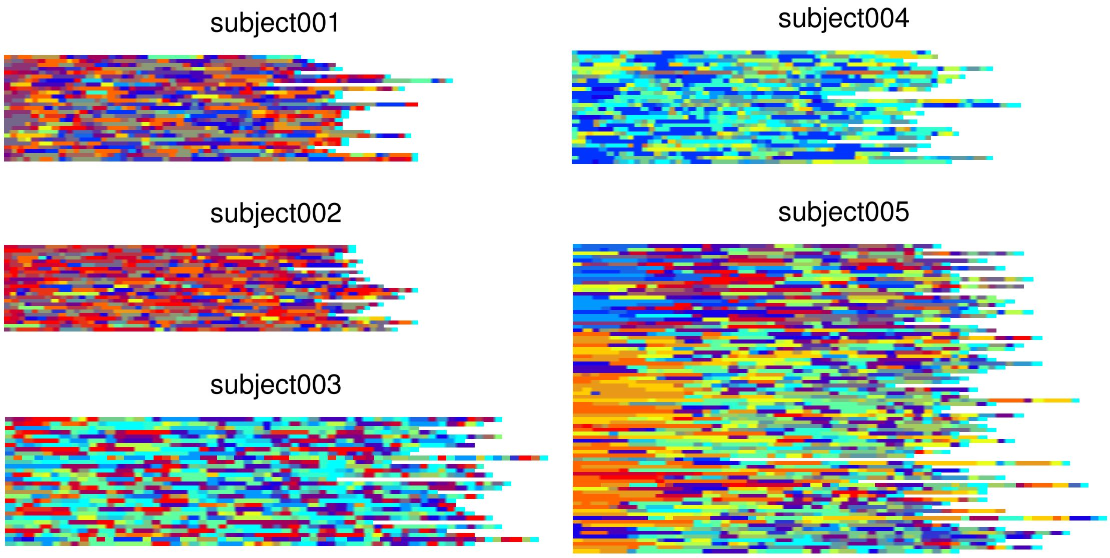


Figure 6.6: The color coded postures used in the survey for allocating typical nights to subjects. Displayed are all night-collections from subjects 1 to 5.

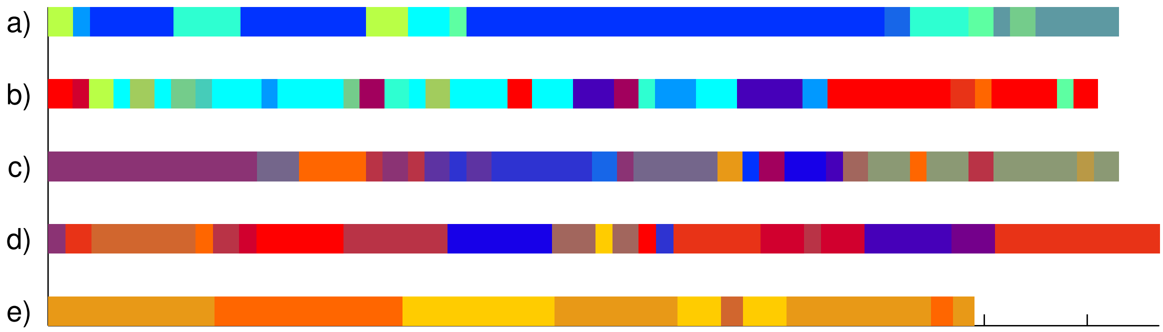


Figure 6.7: Visualization of the typical nights of the 5 subjects (*a-e*). The correct combinations to the datasets in Figure 6.6 are: *1c*, *2d*, *3b*, *4a* and *5e*.

While conducting the inquiry, some users had problems allocating the typical night to subjects 1 and 2, since they exhibit similar colouring. The postures of subjects 3 and 4 contain a lot of turquoise-like colouring but differ in occurrence of red. Assigning the night to subjects 4 and 5 was almost perfect with an accuracy of about 99%, showing explicit postures for each subject. With such an approach, it is possible to display sleep posture trends that exhibit a rhythmic behaviour.

6.3.3 Discussion

We have shown that it is feasible to classify postures with inertial data by using a KNN classifier and ground truth from our sleep monitoring system. Additionally, sleep postures have been clustered with a KSOM approach to find similarities between postures. With such an approach it is possible to illustrate rhythmic patterns of sleep postures in long-term studies. Interestingly, the typical nights can be assigned almost perfectly to individual subjects within our study. The results are interesting for a different reason as well: When observing a single user only, the same method can be used to identify non-typical nights which can serve as an indicator for medical sleep analysis.

During the survey we displayed color-coded postures of five subjects to different users. Without explaining for which purpose this inquiry was conducted, participants immediately began answering the questions. This can be explained by the way the information was presented: Using colors as representations of the posture characteristics simplifies analysis. Such an approach is welcomed by physicians as well, since such data is easy to scrutinize and outliers can be identified accordingly.

However, the usage of both approaches differs essentially: The classification requires ground truth data that has to be obtained by either using a monitoring system or letting the user annotate sleep and awake times. While the clustering approach does not rely on ground truth data, the application scenario is essential here. Knowing the sleep postures is especially important in the research of sleep apnoea. If a person lies on the back most of the time and suffers from poor sleep, this is a hint for such a disease. Whether our approach can be applied to this scenario has still to be evaluated. But in a discussion sleep lab staff supported the use of our classifier: Clustering of posture data is especially interesting for long-term deployments to highlight posture changes and related to that, sleep quality assessment. Summing up, we believe that both approaches are valid to be used in medical sleep analysis and that the application scenario determines the approach that should be used.

After the posture data has been removed from the night segments, the data that remains is a mixture of motion data from the subject moving between postures, the subject moving during wakeful periods, and a third type of *involuntary* motion which requires a specific detection step. This step is described in the next section.

6.4 MYOCLONIC TWITCH DETECTION

Myoclonic twitches describe spontaneous muscle contractions and occur by sporadic limb movements. Note, that it is not the number of twitches that is relevant to detect but how such twitches are expressed. We will introduce a classifier to detect such twitches but we propose to optimize the precision, leading to a low false positive rate, to identify such events. Given few but correct twitches, the video data can be pre-filtered and facilitate analysis for experts significantly. Our classifier is thus optimized for high precision at the cost of lower recall.

6.4.1 Methodology

We again make use of the dataset obtained in Section 5.3, along with the ground truth data of video footage to detect events during sleep. The data consist of only motion data that was left out in Section 6.3, where we extracted non-motion segments. Since not all subjects exhibited myclonic twitches, the evaluation of the classifier was performed on an one-person subset of the dataset only. We identified three subjects in the original dataset of Section 5.3 that showed obvious myocloni in the video footage. Unfortunately, only one participant, aged 35 years and suffering from a periodic limb movement disorder, delivered a significant number of twitches. In total, 5 nights of the test subject could be used for further evaluation.

The following two-step process is pursued to detect myoclonic twitches from the data that was not assigned as posture: First, filtering is done on the duration of the motion in question. Since it is known that these twitches are generally not longer than mere seconds, any segment over three seconds is discarded. Then, we calculate three different features, which describe a twitch pattern, from the remaining motion data. An example of the features is shown in Figure 6.8: At first, we calculate the *length of the motion* data. Further, we determine the *euclidean distance of the start and stop values* of the motion segment since a twitch mostly results in almost the same wrist position as before twitching. Finally, the *distance between the minimum and maximum* of the motion data is computed since non-twitching motion data exhibits larger peaks in contrast to twitches.

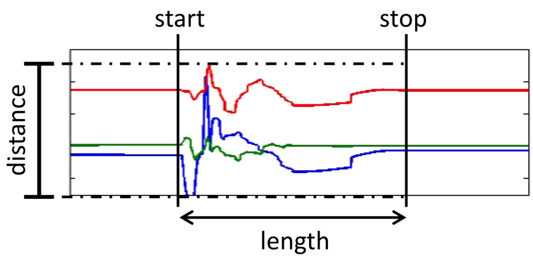


Figure 6.8: Key features used for the myoclonic twitch detection approach: The *euclidean distance* of start and stop time, the *length* of the motion and the *distance* between minimum and maximum.

These three features are used as input to a Support Vector Machine (SVM), a common linear classifier which is trained on sample data from both myoclonic twitches as positive and posture changes as negative examples, obtained by annotations where these events were also clearly visible in the video footage. Figure 6.9 displays different types of myocloni which occurred in the dataset, including multiple ones from the test subject, appearing in short intervals of several seconds. Evaluation was performed using a 5-fold leave-one-night-out cross-validation, using video footage to denote ground truth for the twitch detection.

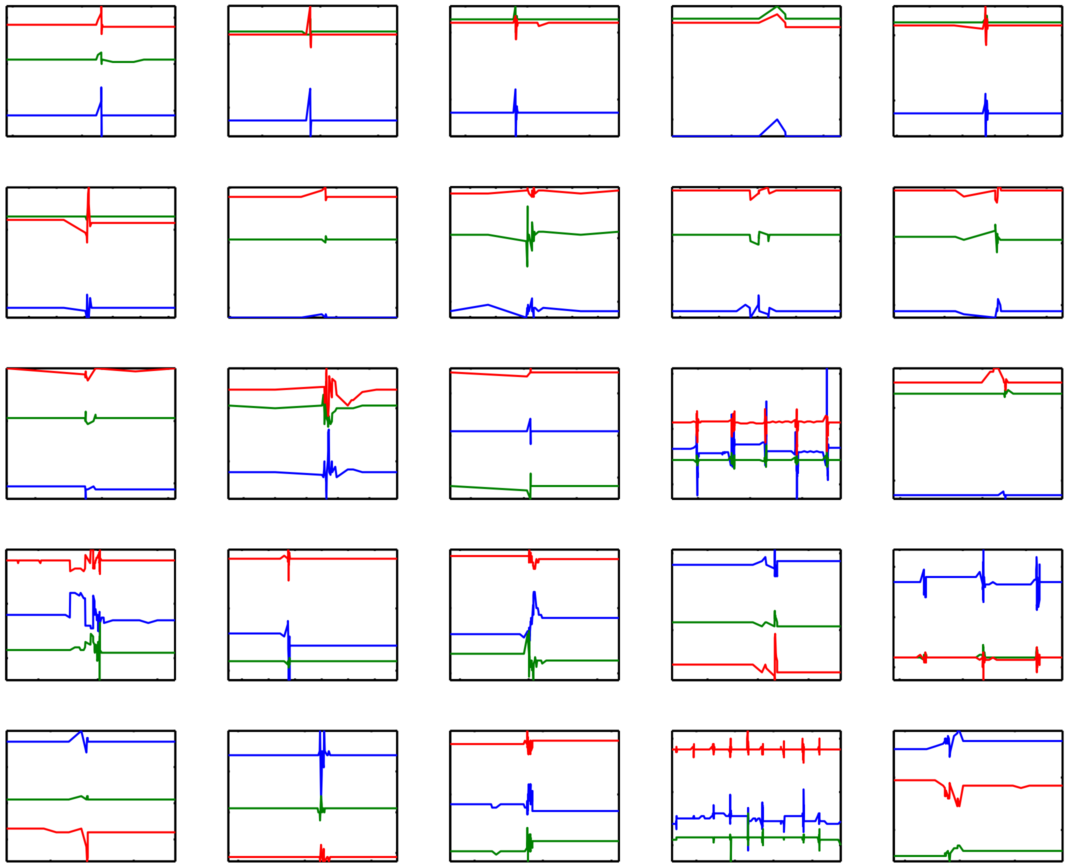


Figure 6.9: Some examples of the observed myoclonic patterns, including consecutive multiple twitching, which occurred during nights of mostly one subject over less than 3 seconds.

6.4.2 Results

Table 6.4 shows the precision and recall values for the SVM classifier based on different initial windows (1, 2 and 3 seconds), showing an ideal window size between 2 and 3 seconds. The 2 second window size exhibits for four nights the highest precision results. For the third night in Table 6.4, the 3 second window is an ideal size but a low recall is perceptible. From visual inspection of the false positives, many could be attributed to external factors that were neither myocloni nor posture changes but rather conscious short motions during awake periods. Missed detections largely constitute very slight contractions.

The results are preliminary but demand to be further evaluated within a cooperation at a sleeping lab, where myocloni can be detected accurately and without painstakingly browsing the data for possible twitches, as was done in this study.

night	1 second [%]		2 seconds [%]		3 seconds [%]	
	pre	rec	pre	rec	pre	rec
1	58	96	**62**	75	55	48
2	71	94	**76**	71	74	51
3	59	70	74	57	**88**	41
4	77	67	**81**	67	80	51
5	50	78	**71**	59	71	47
total	63	81	73	66	**74**	48

Table 6.4: The precision (pre) and recall (rec) results for a detection window of 1, 2, and 3 seconds on a subset of five nights by one subject for which clear video footage was used to obtain ground truth. Additionally, the total results for precision and recall are shown, indicating that a 2 second window yields the optimum results for detecting myoclonic twitches.

6.5 CONCLUSIONS

This chapter presented first of all a simple posture detection technique with data obtained from a wrist-worn sensor in long-term studies. From long-term observations, a sleep model was used to extract those postures, giving a basic insight into a person's sleeping trends. These first studies have shown that sleeping postures are estimated with relatively high accuracies (88% for two subjects respectively), and that visualizing these postures over multiple days could offer an insight into posture rhythms and trends observed over a longer timespan. Furthermore, we used the posture classification technique to introduce an approach to model sleep quality. In order to obtain a sleep quality measure, the approach required a sleep questionnaire that depicts how sleep has been perceived by the subject. Additionally, by considering the number of posture transitions, the proposed system provides an approximation to how well a person slept, with a high number of transitions being an indicator for poor sleep.

Additionally, we proposed the visualization of postures without knowing the actual posture. We derived a coloured representation of the night segmented sensor data, which displays the sequence of postures that occurred in the data of 5 participants' night segments. We showed that such a posture visualization is unique for each subject, by conducting an inquiry with 60 participants, resulting in an overall accuracy of 92% of recognized common postures for individual subjects. The approach provides physicians with the possibility of scrutinizing long-term sleep data to identify candidate nights that show an irregular sleep posture pattern. The proposed method can be applied to any inertial sensor data and does not require the ground truth of sleep postures.

By segmenting night sleep into motion parts, we display the detection of myoclonic twitches, optimized for a *high precision* (to avoid swamping the system with false positives) which was able to achieve precision and recall results of 73% and

66% respectively for one subject over 5 nights. Previous sleep research involving myoclonic twitches indicates that they are more likely to occur in data from people with an irregular sleep schedule like insomnia or in data from people with neurological disorders, such as Parkinson's disease. The tool resulting from this study enables the detection of myoclonic twitches over long monitoring periods with an inexpensive set-up of devices.

7 IMPROVING ACTIVITY RECOGNITION WITH TIME USE DATA

Contents

7.1 Introduction **120**
7.2 Methodology **120**
7.2.1 Wearable System 121
7.2.2 Dataset Description 122
7.2.3 Evaluation Measures 124
7.3 Evaluation Technique **124**
7.3.1 Wearable Sensors Only 125
7.3.2 Time Use Only 125
7.3.3 Ensemble: Wearable Sensors and Time Use 126
7.4 Time Use Survey vs. Common Approach **127**
7.4.1 Results for Wearable Sensors Only 129
7.4.2 Results for Time Use Only 130
7.4.3 Results for the Ensemble 132
7.5 Discussion **132**
7.6 Conclusions **134**

We carry our wearable devices with us while performing our daily routines and tasks. Having a wearable device that is aware of our activities would result in a variety of benefits for its user. For instance, many commercially available devices support the user with additional information about how much they moved or detects sleep segments (see Section 2.4.2 for more examples). The overall problem is that high-level activities, which give insight into the user's current context, are difficult to grasp.

Therefore, we use data from a wearable device to take a first step towards detecting high-level activities. For this purpose, we employ the findings from Chapter 4, i.e., using the knowledge of activity rhythms present in time use surveys, in the activity recognition process. We have obtained inertial data from 17 participants who additionally kept a diary of high-level activities over a two week period each. We utilize three different algorithms to detect activities which imply an approach that indicates how prior information is used to achieve higher accuracy rates for high-level activity recognition.

7.1 INTRODUCTION

For the purpose of improving the recognition rates of high-level activities, we presented the use of prior knowledge obtained from time use surveys as described in Section 4.2. We discovered that such databases can be utilized to infer the user's activity based on the activities performed by similar users within the survey. Results have thus far shown that time use survey data, when used from within the same geographical region as the user, could enhance activity recognition systems considerably. Additional information, such as an approximation of the user's typical schedule, could be combined with real-time sensor data from a wearable sensor.

The goal of this chapter is to investigate a method that makes use of activity rhythms that are present in the time use survey for recognizing activities with a wearable device. Therefore, it is necessary to collect data of high-level activities in a real-world set-up. For this purpose, we obtained inertial data from 17 subjects over a recording period of 14 days each, resulting in a dataset of 228 days in total. We collected data for 11 activities performed by the users, additionally gathering the ground truth by the users keeping a diary.

We apply three different approaches to detect these activities in the data, showing which features from time use databases are important to consider when applied on real-world data. We present a system that outperforms the Support Vector Machine (SVM) classifier, resulting in high precision and recall values for activities like 'sleeping' and 'working'. Results indicate that certain activities can profit from prior information extracted from time use surveys. An additional contribution of this chapter is a long-term dataset which is publicly available and can be obtained on request. The rhythmic nature of certain activities, e.g., 'having lunch', is observable in the dataset. With our proposed method of detecting these activities with a higher confidence, we provide the possibility to model the rhythmic nature of certain activities.

The remainder of this chapter is structured as follows: In Section 7.2, we describe the methods used to determine the activities of the test subjects, additionally describing the sensor platform used. Section 7.3 presents three different algorithms we apply to detect high-level activities. In Section 7.4, we show the results obtained from these three different classification approaches. A short discussion about the experiment's findings will follow in Section 7.5. We conclude this chapter in Section 7.6 with a summary of this chapter's contributions.

7.2 METHODOLOGY

In order to detect high-level activities, we believe that sensor data is required that are recorded in a real-world set-up. For this purpose, we decided to use a system that is already commercially available, is comfortable to wear over a long timespan and enables researchers to get direct access to the raw sensor data. We will first introduce the sensor used for this study and then explain the classification process.

Figure 7.1: BodyMedia SWA worn by a male (left) and female (right) participant. The Armband is usually worn on the upper arm.

7.2.1 Wearable System

For this study we use the SenseWear^{TM}Armband (SWA) from BodyMedia[39]. The SWA monitors one's activity levels, especially during workouts and while resting, detecting 'sleeping' (Sunseri *et al.*, 2009) as well. It is worn comfortably on the upper arm, as shown in Figure 7.1, and can rest there permanently day and night. A graphical tool displays useful information to the user, like showing how data channels change over time. Additional information such as energy expenditure or a step counter are accessible, enabling the user to keep track of his fitness status (Shuger *et al.*, 2011). The device is splash waterproof, which is why it can be used during work-outs.

The SWA embeds a 2-axis accelerometer, a skin temperature, a galvanic skin response and a heat flux sensor. Sensor values can be stored in the internal storage in different intervals, from 32 samples per minute up to one sample every 10 minutes. Depending on the log frequency, the storage lasts for 2 hours only (32 samples per minute) or a little more than two weeks (one sample per minute). The power source is a common AAA battery, which needs to be replaced by the user after approximately one week, depending on how often the sensor was worn and how much the user moved. The sensor automatically starts logging when skin contact is detected and stops logging as soon as the user takes off the unit. The recording frequency for our study was set to one minute, being the optimal trade-off for recording for a long timespan and not loosing too much sensor information. Had we increased the sampling rate, the device would have run out of memory after a few days only instead of being able to record for 14 days straight.

We decided to use the SWA for this study since it is able to record various

[39]http://www.bodymedia.com/, last access 09/2014

other sensing modalities in contrast to the wrist-worn unit introduced in Section 3.2. Additionally, we were able to collaborate with the company that builds the device, enabling us to get direct access to the raw sensor data. Therefore, we are given the possibility to evaluate a commercially available body-worn sensor-unit which is already being used by many individuals around the world.

7.2.2 Dataset Description

The SWA was worn day and night by 17 test subjects for 14 days each, resulting in a dataset of approximately 228 days. We chose to use a high variety of test subjects as summarized in Table 7.1 to capture different data and especially different activity behaviours specific to their profession to stress-test our algorithm. The subjects were between 21 and 48 years old, 12 male and 5 female. The majority of the participants were white-collar employees, working either at the university or in an office. Students participated as well, who are known to have a completely different daily routine than employees. The number of data recorded for each participant is shown in Table 7.1.

While wearing the device, the subjects were asked to keep a diary of daily activities for the entire recording period, usually recalling at the end of the day what kind of activities they had performed. Some subjects kept a diary by writing down the activity immediately after performance. The established list of the activities as used in this study is shown in Table 7.2, being in accordance with the activities from the time use survey, with one exception: The activity 'personal care' from the time use data includes 'sleeping', 'eating' and other activities in the area of 'personal care', such as 'showering' or 'dressing'. We decided to split up these activities into the first three activities as shown in Table 7.2, especially to be able to catch 'sleeping' and 'eating' on its own.

We obtained inertial data such as the average, longitudinal and transversal acceleration, as well as the Mean Absolute Difference (MAD) of the acceleration. An exemplary dataset of 14 days for a male participant is displayed in Figure 7.2, showing in the top plot the MAD and the bottom plot the average of the longitudinal (blue) and transversal (red) acceleration. The dataset for each day is visualized, starting at the point where the number of the day is fixed on the x-axis. The next dataset starts at the next number. The nights are immediately visible, characterized by segments where the acceleration signal is low.

Additional sensor information, like skin temperature, was logged but was not considered in this study. For the evaluation, we used only the accelerometer data to infer the performed activity. Note here that the logged activities are not equally distributed. Especially the number of 'personal care' events appears rather low with 88 hours in total as can be observed in Table 7.2. This can be explained by the fact that the sensor was taken off for showering. Similar, sports occurs only 40 times throughout all of the datasets, either because the device was taken off during work-outs or because the participants were not very sportive. When considering 'eating', this activity occurs very often but does not take up as much time as 'sleep'.

subject	gender	age	data [hrs]	comments
1	male	32	338	employee
2	female	28	334	student
3	male	31	333	employee
4	male	27	319	employee
5	male	28	284	employee
6	male	32	334	employee
7	male	30	320	employee
8	male	27	328	student
9	male	28	315	employee
10	female	35	346	housewife
11	female	29	321	employee
12	male	31	340	employee
13	male	26	274	employee
14	female	28	292	employee
15	male	21	316	student
16	male	25	334	student
17	female	48	352	housewife

Table 7.1: The test subjects that participated in this study, along with additional information like gender and age, as well as the amount of data obtained per participant. Most of the participants were white-collar employees but also students were recorded.

ID	activitygroup	occurences	duration [hrs]
1	Sleeping	247	1868
2	Eating	373	257
3	Personal care	223	88
4	Working	297	1047
5	Studying	67	215
6	Household work	215	284
7	Socializing	111	249
8	Sports	40	41
9	Hobbies	72	172
10	Mass media	205	397
11	Travelling	366	864

Table 7.2: Activities taken from the time use survey that were logged by the test subjects, additionally displaying how often each activity occurred and for how long it lasted in total. Interestingly, 'eating' occurred far more often than 'socializing' but lasted just as long as 'socializing'.

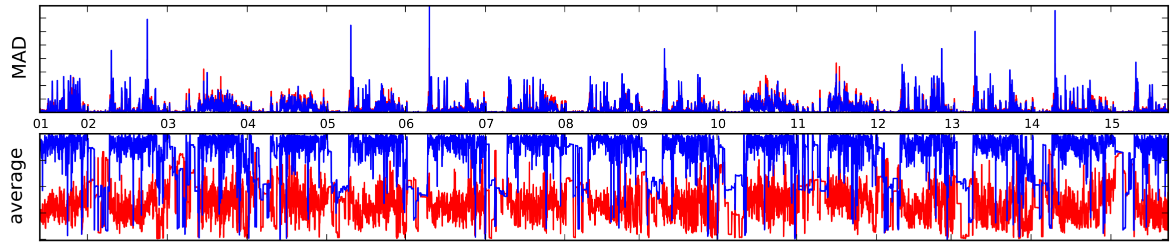


Figure 7.2: Inertial data from the SWA's 2-axis accelerometer from one participant over 14 days before normalization. The top plot displays the Mean Absolut Difference (MAD), while the bottom plot depicts the average of the longitudinal (blue) and transversal (red) acceleration over 20,312 samples, each sample taken every minute.

7.2.3 Evaluation Measures

Most activity classes are recognized by our system (see Section 7.3), which we observed in the various confusion matrices we established and in Figure 7.5 as well. Nevertheless, when considering classes that have been very poorly or not at all detected, the overall accuracy for recognizing a participants activity is still pretty high, while precision and recall are rather low. In order to depict how well the data describes each class, we will focus on the precision and recall values in this chapter.

With a wearable sensing platform like the SWA, we are able to record low-frequented sensor data over a long time-span. For 17 participants in this study we will investigate how activities can be sensed with a wearable system, evaluating the results by applying precision and recall on the obtained classifications. We will now have a closer look at the classification techniques and how the data was prepared for classification, explaining how probabilities have been calculated from the wearable sensor approach.

7.3 EVALUATION TECHNIQUE

The method proposed in this chapter can be described in three steps: (1) The inertial data is first being evaluated with a common classifier to determine the activities, (2) after which we use the time use dataset only to infer the activity. (3) Then, the results from the common classifier and the time use probabilities are fused to improve on the results from the first two steps. For Sections 7.3.1 and 7.3.2 we evaluate the activities by dividing the dataset into five equally distributed folds per participant's dataset, performing a leave-one-fold-out cross-validation per test subject to enable user-specific activity recognition.

Note here that we will use the term *wearable sensors only* to depict the approach of applying a common machine learning algorithm to the inertial data and *time use only* to describe the method of classifying activities by using likelihoods derived from the time use survey database.

7.3.1 Wearable Sensors Only

For detecting the activities within the sensor data, we used a Support Vector Machine (SVM, see Cristianini and Shawe-Taylor (2000)). The implementation of the classifier was done completely in Python. For this purpose, we used the sklearn[40] package, which embeds an SVM library that is based on LIBSVM (Chang and Lin, 2011). We utilized a linear SVM, since we were dealing with large datasets and had a multi-class problem at hand. As a strategy, the one-vs-the-rest method was applied, which basically trained a SVM for one class and tested it against the rest of the classes. Before starting the SVM training, we normalized the dataset. Then, we balanced the training set by randomly choosing data rows from labelled features and duplicated them to receive a conform dataset with an equal number of samples per activity. As shown in Table 7.2, the occurrence of activities was unbalanced, especially sleep and work dominated the datasets. We performed a five-fold cross-validation to estimate the optimal penalty parameter C on a small subset of the training data, which has been used in the training phase of the classification process. After having trained the SVM, we estimated the classes for the test set. Additionally, we calculated the *softmax* output for the test set which enabled us to derive a likelihood estimation for the input data. The softmax output is defined as

$$\sigma_{prob} = \frac{1}{1 + e^{-2 \cdot d}} , \tag{7.1}$$

where d is the decision function from the SVM, described by the Support Vectors of the dataset. For each data point $x_1, ..., x_i$ the SVM provides the decision function $f(x_1), ..., f(x_i)$, describing the distance to the calculated hyperplanes of the SVM. The likelihoods will be used later on when fusing the *wearable sensors only* and *time use only* results.

7.3.2 Time Use Only

The time use classification technique uses a maximum-likelihood estimation to determine which activity took place. For this purpose, we made use of features f within the time use survey (*time*, *age* and *gender*). In Section 4.2, different features have been evaluated for their use in activity recognition, identifying *time* and *location* as useful features. For this study, *location* information was not logged, since the used sensor is not equipped with a GPS module to infer the location where the activity took place. Having the user also log where the activity occurred would have increased the effort of keeping a diary. Therefore, we considered *time* combined with other, available information, adding *age* and *gender* as features.

We calculated a histogram from the time use dataset for each of the given features *time*, *age* and *gender* and the 11 activities, obtaining a 4D histogram of the shape $[144, 5, 2, 11]$. The shape corresponds to 144 10-minute time-slots per day, 5 age-

[40] http://scikit-learn.org, last access 09/2014

groups[41], 2 gender types (male and female) and our 11 activities. According to the 3-tuple $(time, agegroup, gender)$ we summed up all the occurrences for each of the 11 activities. We then calculated the distribution of the 11 activities for each 3-tuple to obtain the probabilities.

The maximum-likelihood estimation calculated the probability $P(c_i|f_1, ..., f_n)$ for a target class c_i, $i \epsilon [1, ..., 11]$ and the features $f_1, ..., f_n$, classifying the activity by the highest probability according to the 3-tuple given in the sensor dataset.

7.3.3 Ensemble: Wearable Sensors and Time Use

In this section we describe the combination of likelihoods from the *wearable sensors only* and the *time use only*, resulting in a new likelihood-table that is used to determine the activities. The equation

$$c_i = argmax_i \left(\frac{P(c_i|x) + P(c_i|TUS)}{2} \right) \tag{7.2}$$

describes the procedure of estimating class c by applying the mean rule (Alexandre *et al.*, 2001) on both likelihoods from the SVM output $P(c_i|x)$ and the time use approach $P(c_i|TUS)$. We scale the likelihoods from both *wearable sensor only* and *time use only* for each activity class c_i, i.e., we calculate the scaling for all likelihoods of activity c_i by the equation

$$P_{c_i} = \left(\frac{P_{c_i} - abs(min(P_{c_i}))}{max(P_{c_i}) - abs(min(P_{c_i}))} \right) \tag{7.3}$$

to avoid the domination of bigger likelihoods over those in smaller numeric ranges. Note that the likelihoods could be weighted additionally, depending on how the overall classification behaves. The weighting would be applied to equation (7.2). The overall likelihood $P(c_i)$ would be calculated for the wearable data after the training phase and multiplied with each class probability $P(c_i|x)$ in the test set. The same would be done for the time use dataset, obtaining the overall probabilities $P(c_i|TUS)$. At this stage of the study, we did not consider to apply a weighting technique but we will consider this approach for future studies.

Figure 7.3 shows the likelihoods for one participant at the age of 32 after scaling of the wearable data (top plots) and time use (bottom plots) likelihoods. Displayed are the activities 'sleeping' (blue) and 'eating' (red) over approximately two days, showing how likely they are to occur at a specific point in time. The rhythmic nature of the probabilities for the time use data can be observed here, as well as for the likelihoods for the wearable data which exhibit much more noise.

Being able to allocate likelihoods to the classification results of using only wearable data offers the possibility of combining the likelihoods with the probabilities

[41]Note here that for downsizing reasons and to obtain a representative histogram, we divided the time use survey into 5-years age-groups (20-24, 25-29, 30-34, 35-39, 45-49), according to our participants.

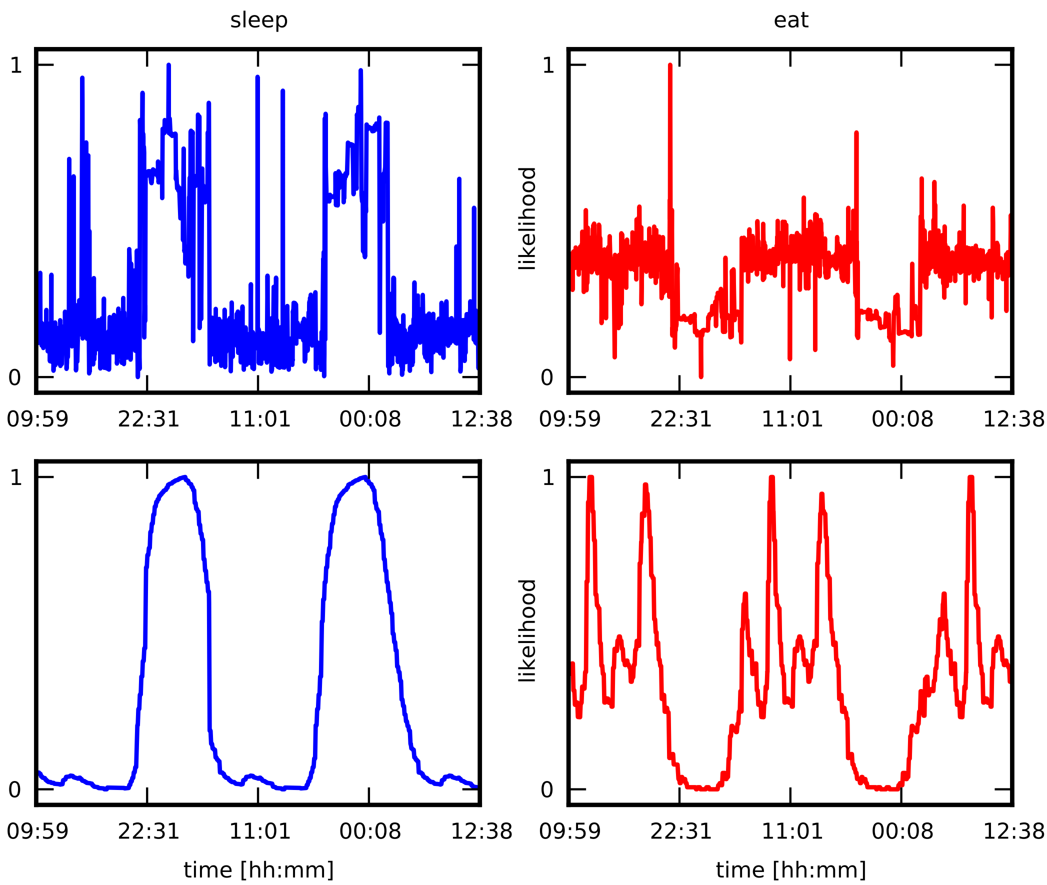


Figure 7.3: Example of likelihoods estimated from the sensor (top plots) and the time use (bottom plots) approach for a male subject in his thirties, displaying the activities 'sleeping' and 'eating'.

calculated from the time use survey. This way, we are able to not only consider information from the participant's wearable sensor data but additional information from the time use survey. We will now take a look at our classification results in the next section.

7.4 TIME USE SURVEY VS. COMMON APPROACH

In this work we compare three different classification techniques applied to the same dataset. We will be discussing the results individually, highlighting the differences for each approach. We will first visually inspect the outcome, then we are going to discuss the results quantitatively.

Visual Inspection. In Figure 7.4 we display qualitative results of all three approaches. Shown are all the activities (in different colors) of a male participant in his early thirties for the entire recording period of 14 days, describing from top to

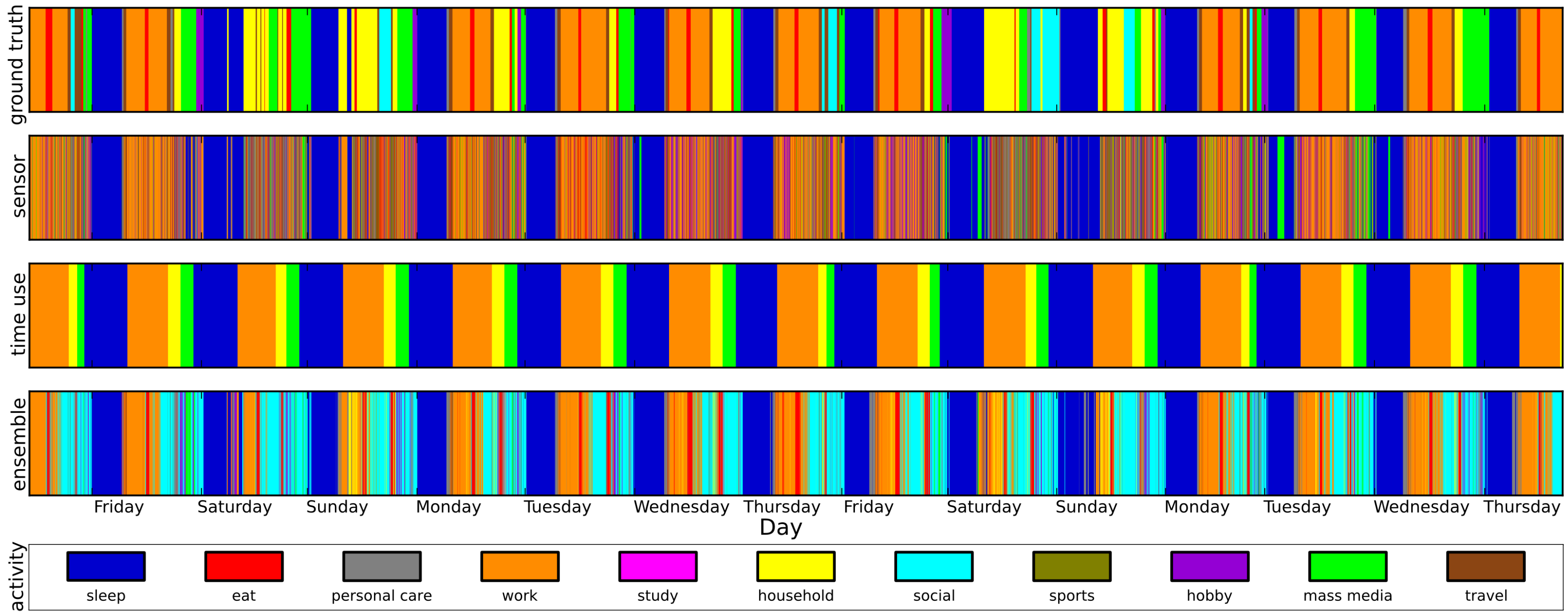


Figure 7.4: Example of classification results for a male subject (32), showing from top to bottom: The ground truth, estimated activities from the sensor, the time use and the ensemble of both modalities. Displayed are all 14 days from left to right, with barplots showing the activities in different colors. Immediately visible is the improvement when fusing both sensor and time use likelihoods in the ensemble plot, which exhibits a higher variety of detected activities over the whole dataset.

bottom: (1) The ground truth, i.e., what the user was actually doing, (2) the estimated activities from the *wearable sensor only* approach, (3) the *time use only* results and (4) the *ensemble* of the two modalities. When observing the ground truth, we discover that a certain pattern or even rhythm is visible in the recurrence of the activities throughout the 14 days (e.g., 'working' and 'sleeping', which occur mostly during the day, or at night respectively). Such a rhythm awareness could help in the classification process.

A rhythmic behaviour is visible in the sensor-based plot as well, which is riddled with small detection episodes of different activities. Remarkably, this approach shows a high variety in the detected activities, except for the activities which last longer during a 24-hour period, such as 'sleeping' and 'working'. The inertial data offers a high number of diverse sensor readings for each activity class, leading to a rich dataset for the training process, thus laying the burden of accurately detecting the activity on the classifier.

The time use plot exhibits a clear rhythmic activity representation by displaying the same activity sequence for each day. The time use classification takes into consideration only the information about *time*, *age* and *gender*. Therefore, we observe here the daily routine of a male person between 30-34 years of age. If we were to take other features like *day-of-the-week* into account, the plot would surely differ.

The ensemble results for this participant show longer periods for detected activities, e.g., the range of classified work episodes are wider than compared to the *wearable sensors only* approach. Additionally, the plot offers a view on how other activities are being favoured in the classification process, as for example 'socializing' and 'eating'. For the latter it seems that the ensemble quite accurately detects 'eating' episodes when compared to the ground truth. To confirm the visual findings, we will now summarize the quantitative results for each approach, starting with the *wearable sensors only*.

7.4.1 Results for Wearable Sensors Only

When using a SVM classifier to detect the 11 activities from the sensor data, we observe that it is feasible to catch all of the activities but mostly with low recognition rates. Overall, the most confident precision and recall results are obtained for 'sleeping' (88.7% and 85.35%) and for 'working' (30.3% and 43.68%), which is followed by 'travelling' (24.32% and 23.2%) as can be observed in Table 7.3. The rest of the activities perform rather poorly. One reason being that the training data for some activities is not distinguishable, since we balance the sensor values for each class as described in Section 7.3.1. Classes with a larger occurrence within the dataset do have an advantage over the other classes. The data of these classes offer a higher variety of features, which is advantageous for the training process. Note here that the classification technique is completely independent from the duration of each activity, which could lead to different results when considered in a sequential classification model.

In Figure 7.5(a) we can observe the precision (top) and recall (bottom) values

activity	**sensor**		**time use**		**ensemble**	
	precision	**recall**	**precision**	**recall**	**precision**	**recall**
sleeping	88.59	85.36	**92.87**	82.64	83.22	**95.62**
eating	5.98	13.91	0.0	0.0	**30.45**	**19.76**
personal care	13.67	6.29	0.0	0.0	**15.11**	**14.10**
working	30.00	43.65	55.35	46.06	**56.38**	**48.74**
studying	3.77	**9.34**	0.0	0.0	**11.5**	9.13
household work	11.34	**17.71**	**34.18**	14.25	7.76	20.87
socializing	12.30	9.43	0.0	0.0	**30.43**	**13.84**
sports	**12.65**	4.03	0.0	0.0	2.39	**7.73**
hobbies	5.11	6.93	0.0	0.0	**11.92**	**9.49**
mass media	20.74	14.42	**55.2**	31.06	30.97	**44.53**
travelling	**24.31**	23.24	0.10	0.0	6.68	**35.61**

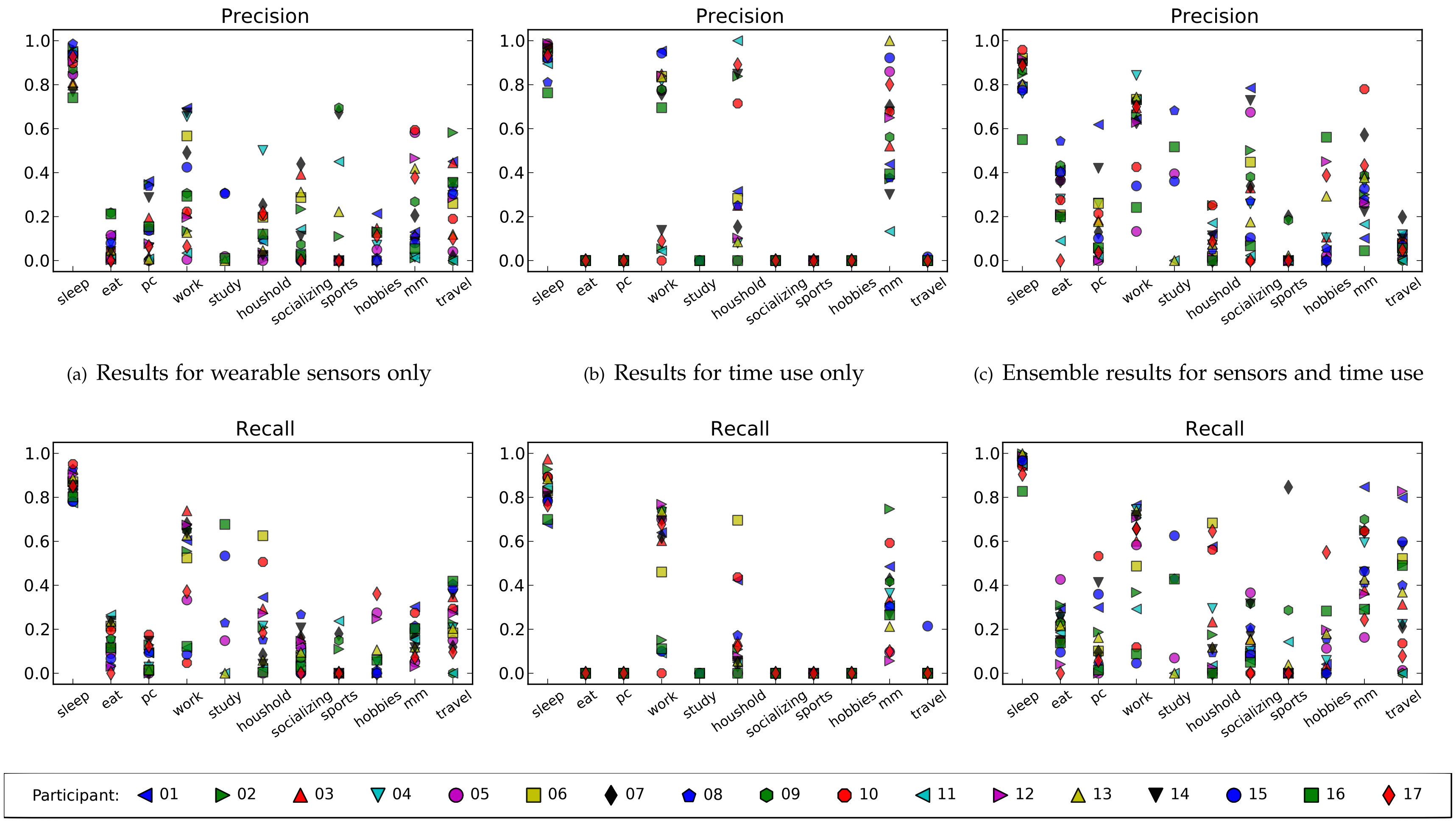
Table 7.3: Overall precision and recall results for each activity for all three classification methods: Sensor, time use and ensemble. The highest score for each activity is displayed in bold, showing how the ensemble method exceeds the results from the sensor- and time-use-only-based approaches for most of the activities.

and how they distribute over all activities for each user. Some of them were not performed by the user and therefore never detected, since there was no training data for these activities. Unsurprisingly, 'sleeping' is detected with a high confidence for all the participants. 'Working' varies quite a lot, depending on the participants and the number of 'working' phases having been logged in the diary as well. Students, for example, do work but only a few hours per week, like participant 15. Precision is high (43.4%) while recall degrades to 8.25%, which means that many of the 'working' events were unidentified. Participant 1, for example, is an employee working 8 hours every day which is being displayed in the results of precision and recall both being in the range of 60%. We note here that the data representation plays a significant role for the classification process.

The overall results of all the participants for *wearable sensors only* lead to a precision of 20.42% and a recall of 21.11%. We can conclude that it is possible to detect certain activities with a high confidence in a dataset which contains low-frequent inertial data (i.e., one minute intervals), using a common classifier like the SVM. However, since the activities are discriminated rather poorly, we need to improve the recognition rates with the help of additional information added to the classification process.

7.4.2 Results for Time Use Only

In Table 7.3 and Figure 7.5(b) we observe that with *time use only*, we detect four out of 11 activities from the dataset, namely 'sleeping', 'working', 'household work' and 'mass media'. The precision and recall scores for the time use based approach for



(a) Results for wearable sensors only

(b) Results for time use only

(c) Ensemble results for sensors and time use

Figure 7.5: Precision (top) and recall (bottom) results for three different classification techniques from left to right: *Wearable sensors only*, *time use only* estimation using the features *time*, *gender* and *age* and the combination of both modalities (*ensemble*). Each participant is displayed in a different sign and color. Immediately visible is that the *ensemble* exhibits a high variety of detected activities and is improving on the recognition rates in comparison to *wearable sensors* and *time use*.

each activity and user are depicted here individually. As features for the time use dataset we use as much information as possible, the features being *time, gender* and *age*. Overall, we reach a precision and recall of 23.76% and of 17.4% respectively. The overall results for 'travelling' can be neglected, since precision and recall are below 1%. Remarkably, 'travelling' was detected for a 21-year-old male participant only, who travelled home on the weekends, having to use public transportation for several hours. This coincides with a small segment of the estimation from the time use dataset. Even though the other participants travelled as well, the activity lasted only for a few minutes, which is why it is difficult to detect it. Furthermore, depending on the time use histogram, 'travelling' is often not the most likely activity to occur.

Overall, precision for the four detected activities range from 34.18% to 92.87%, exceeding the results from the sensor-based approach for these activities (see Table 7.3). Nevertheless, the rest of the classes are not recognized by the system, since the likelihoods are too small and therefore exceeded by the likelihoods of other activities. Our findings show that the participants' most common activities are detected with a high confidence by the *time use only* approach which is in accordance with our findings in Section 4.2.

7.4.3 Results for the Ensemble

Combination results for the *wearable sensors only* and *time use only* approaches are displayed in Table 7.3 and Figure 7.5(c), showing again for all participants (coded each with a different symbol and color) the classification results for the ensemble classifier. In contrast to *time use only*, we now detect all the activities appearing in the datasets. In regard to *wearable sensors only*, we see certain improvements for precision and recall but degrading results for some activities as well. For 'sleeping', we obtain a lower precision than when applying each of the other two models to participant 16's data (green cuboid in Figure 7.5(c)). Even though the results for precision are lowest for *wearable sensors only* and *time use only*, we still wonder why it drops to 55%. After investigating the ground truth and estimated classes, we discovered that participant 16 exhibits a very unusual sleeping pattern, e.g., sleeping from 8pm to 10pm, watching TV until midnight and then going to bed again, waking up quite early the next morning. It seems that the ensemble is confusing too many classes here, which is why precision is dropping. For the overall results of all activities we obtain precision and recall of 28.01% and of 28.38% respectively.

The results in Table 7.3 indicate that adding time use data after the classification process of the wearable sensor data is feasible and improves on the recognition rates of the *wearable sensors only*. We will now discuss our findings in detail, highlighting important results that have been observed during the evaluation process.

7.5 DISCUSSION

The evaluation of 17 datasets, consisting of a total of 228 days of wearable sensor data, leads to several interesting findings which are summarized and discussed in the following paragraphs:

Time use surveys are highly useful to improve on the recognition rates for activities that cover a significant portion of the user's day. In this study we identified the activities 'eating', 'working', 'socializing' and 'mass media'. The ensemble approach leads to better results than just using the *wearable sensor only* or the *time use only* approach to infer the activity. These results confirm the idea of time use data enhancing certain activity recognition systems as mentioned in Partridge and Golle (2008). Additionally, the recognition rate benefits from activities that occur more regularly in the time use survey for the inspected features, e.g., a male subject in his late twenties will, most likely, be working in the afternoon.

Time use statistics fit wearable devices. We benefit from the size of the time use database, which is below 1MB. Therefore, the data can be immediately pre-loaded on a ubiquitous wearable device. Combined with a common classifier, activity recognition can be improved directly on such an environment. Additionally, time use data incorporates information about the habits not only from the current user but from other participants in the time use survey as well. A large number of people (here: over 10,000) are represented in the time use database, along with their daily routines.

It is important to note that we inherently exploited the knowledge from the time use survey data as we classified low-frequent sensor data only. It is not trivial to detect activities with such low-frequent data using a common classifier but we nevertheless recognized certain activities such as 'working' with a high confidence. Additionally, 'sleeping' can be detected very accurately with 2D inertial data sampled over 1 minute.

Time use surveys are less useful for detecting activities that occur only for a short time each day, e.g., 'travelling'. Although the sensor classifier was quite confident in detecting 'travelling' (3rd best recognition score for *wearable sensors only*, see Table 7.3), adding the time use survey information led to a drop in precision. Note here that 'travelling' in our study includes activities like 'going home', 'going for lunch', 'taking the bus', etc., which do not occur regularly and lasted only a short time.

Even though our results are very promising, we believe that some aspects could still be improved. First, the ground truth was gathered by participants keeping a diary, where it is not clear how accurate the participants entered the activity events in the notebook. Another way of gathering the ground truth could be achieved by letting the user enter the activity episodes directly on a mobile device (a smart phone would be well-suited for that task). Secondly, we might benefit from knowing the participant's location, which is an evaluated feature of the time use survey, as shown in Section 4.2. GPS information is already available on most of the wearable devices, which is why recording this information too is feasible.

7.6 CONCLUSIONS

This chapter presented a novel approach to improve activity recognition with inertial data from a wearable platform (the SenseWear^{TM}Armband, which embeds a 2-axis accelerometer) by combining rhythms present in the time use survey with a common classifier in an ensemble approach. Making use of additional information in the classification process has many advantages, one being that the sampling rate of the sensor data can be reduced. We obtained a dataset of 228 days of inertial data from 17 participants. The dataset has been annotated with ground truth information of 11 high-level activities and is publicly available.

We showed on this dataset how recognition rates of activities varied when using a common classifier and a time use survey approach. Additionally, we improved on the results for certain activities with an ensemble model that combines the two aforementioned methods. Precision for activities like 'eating', 'socializing' and 'hobbies' increased by approximately 25%, 18% and 6% respectively in contrast to using a common approach like the SVM. We discussed certain advantages of using time use survey data and identified the limits of embedding such data in the classification process. The approach is a first step towards the recognition of high-level activities in order to visualize the rhythmic nature of these activities.

Further studies could embed time use databases on wearable platforms, such as smart phones or smart watches, to perform real-time activity recognition on the devices. The feasibility of using a smart phone as a sensing platform has been shown in Section 3.3, indicating that specific knowledge about the users is necessary. Such knowledge could as well be provided by time use databases.

8 CONCLUSIONS

Contents

8.1 Contributions **136**
8.2 Conclusions **137**
8.3 Outlook **139**

Rhythmic behaviour is an important aspect in long-term activity recognition systems, since it provides insight into a person's daily routine. Such information is a key ingredient of any system using data as a prior for an activity recognition system to achieve higher recognition rates. In this thesis several approaches to detect activities have been investigated, in order to allow the modelling of rhythmic behaviour.

Suitable recording platforms have been presented that enable the gathering of sensor data like movement or light intensity. We show the feasibility of using these platforms in the fields of activity recognition and of sleep monitoring (Chapter 3). Furthermore, we introduce data obtained (1) by a government agency and (2) through long-term monitoring (Chapter 4). For both datasets, we extract important features from wearable sensor data that are used in the activity recognition process for high-level activities, by employing the probability distribution of these activities as a prior (Chapter 7). The results indicate that a combination of a common classification approach with such prior information yields higher recognition results than without. Having extracted features to describe a most rhythmic activity - sleep - we focused on detecting sleep with a Gaussian-, generative model- and stationary segments-based approach (Chapter 5). For this purpose, we use first inertial data only and then additional information like time and light intensity from long-term data recordings. By partitioning data into night segments, we focus on more fine-grained activity rhythms that occur during sleep: We detect sleeping postures within non-motion and myclonic twitches within motion segments extracted from the sensor data that are presented to physicians to support the evaluation of sleep and to visualize the rhythmic nature of these events (Chapter 6).

In the following sections we summarize the main findings and the contributions of this thesis before giving an outlook on future studies.

8.1 CONTRIBUTIONS

This thesis introduces a set of challenges in Section 1.2 that include the feasibility of *recording* sensor data over a longer period and extracting *activity rhythms* from such data. For this purpose, the activity recognition system has to provide accurate results to *model patterns* for *high-level activities*. Furthermore, appropriate *features* have to be extracted from various data sources to enable activity recognition. In the following paragraphs we will address these challenges and show how we have contributed to solving them.

Evaluation of the Smart Phone as a Sensing Platform. We have answered the question for which scenarios the smart phone is a suitable sensing platform, taking into consideration how often the phone is *on the user*. For this purpose, we collected 638 days of smart phone data (accelerometer, proximity and light sensor values) and wrist-worn sensor data from 51 participants with additional ground truth information of sleep segments.

Usage of Statistical Data obtained over a Population. Time use surveys are scrutinized to extract important *features* from, which are used in the activity recognition process to improve on the recognition rates of activities.

Novel Sleep Detection Algorithm. We present a novel sleep detection algorithm (Estimation of Stationary Sleep-segments - ESS) that, in contrast to traditional approaches, does not overestimate sleep and therefore detects wake phases with a higher confidence. In the course of evaluating the ESS sleep detection approach, we observed 42 sleep lab patients and obtained polysomnography (PSG) data to annotate wrist-worn sensor data, that includes not only acceleration but also light intensity readings.

Datasets for High-level Activity Recognition. In the course of this thesis, we have presented many different long-term datasets that were recorded to show the feasibility of the proposed methods. The datasets are publicly available to enable other researchers to not only reconstruct our methods but to establish their own algorithms without the need of recording the data themselves. The following datasets were obtained additionally to the previously mentioned datasets in real-world scenarios: (1) Long-term data from a wearable sensor to detect high-level activities, containing inertial data and other information like body temperature. (2) For long-term sleep detection studies, we obtained inertial data with ground truth from 8 participants resulting in 184 days of continuous recordings.

We believe that all of the aforementioned datasets are of benefit for future studies and they are therefore publicly available at www.ess.tu-darmstadt.de/datasets or at www.borazio.de.

8.2 CONCLUSIONS

The conclusions drawn in this thesis are summarized in the following:

Appropriate platforms for activity recognition enable long-term monitoring. In Chapter 3, we present the evaluation of different sensing platforms that are deployed in various scenarios. We introduce a wrist-worn sensor that is able to record sensor data over several weeks, depending on the sensor's set-up. The sensor has been used in two further evaluations:

(1) In a feasibility study to use the smart phone as a sensing platform with 51 participants, our results indicate that the smart phone is *on the user* for 22% of the day. The results differ per user, hinting at a phone usage behaviour varying between below 10% and over 50%. Therefore, the choice of activities to be recognized has to be scrutinized by taking into consideration the user's phone carrying habits.

(2) Furthermore, we present a sleep monitoring system that can be deployed over several weeks to capture sleeping behaviours. We use a night vision camera that is set up in the user's bedroom to record important sleeping events that can later on be browsed by the user himself and physicians to detect interesting segments for further investigation. Such an evaluation of sleep patterns has been shown in Chapter 6, in which we detect different sleep postures and myoclonic twitches to evaluate the night. All three platforms gather sensor data over weeks or even months, enabling the long-term monitoring of users in their common environment, instead of in a laboratory set-up. Additionally, all these recording systems are of low maintenance, i.e., requiring almost no user interaction.

Statistical data provide suitable features for activity recognition. Statistical data obtained over a population is already available, since many countries gather information of their inhabitants in databases that are mostly publicly accessible. Such data is scrutinized in Chapter 4, resulting in a set of features that are suitable to recognize activities. We identify the features *location* and *time* as the best-performing features to detect activities like *sleeping*, *personal care* and *travelling*. When combining the two aforementioned features with *previous activity*, the results are even higher, with an overall accuracy of 75%. The proposed approach is limited to the region in which the experiments are conducted due to the demographic differences in time use surveys from different countries. Such statistical information reflects how often common activities are carried out.

Statistical information can be derived from other information sources as well: We evaluate sleep data to extract features that hint at a rhythmic pattern of sleep for observed individuals. Features like *time of day*, *amount of movement* and *start and stop time* of sleep are used to detect similar nights in order to be able to categorize new nights. Based on a long-term dataset of 141 days we showed the rhythmic pattern of sleep.

Rhythmic behaviour as a prior increases the recognition rate. This thesis takes a first step to recognize high-level activities by using wearable data and embedding prior information in the recognition process in Chapter 7. For this purpose, we evaluate a traditional approach to detect activities, in order to compare the results to an evaluation which uses only prior information. We make use of statistical data obtained over a population as described in Chapter 4. By fusing the traditional approach of a Support Vector Machine (SVM) with the probability results from the time use survey evaluation, we improve on the recognition rates of several high-level activities like *eating*, *socializing* and *hobbies*. The results for all three are 25%, 18% and 6% higher, respectively, than just using the wearable sensors' features. In the course of this evaluation, we also identify for which scenarios time use survey data is not useful: When trying to detect activities that last only for a small time each day, the results drop in precision, since other, more common activities, dominate the time-frame. Such information is crucial since the approach of using prior information is not generally valid for all activities.

Sleep and sleep pattern detection is feasible with inertial data only. In Chapter 5 of this thesis, we present three approaches to detect sleep with the use of a wrist-worn sensor only. The first algorithm relies on inertial data only, by detecting immobile segments over a 10-minute window. To properly evaluate our algorithm, we conduct a study in a sleeping lab with 42 patients. The obtained data includes polysomnography (PSG) output, the ground truth for sleep, and inertial data from the wrist-worn device. With such a benchmarked dataset, we are able to compare our sleep detection results from the ESS approach to traditional algorithms established by Oakley and Cole et al. These clinically evaluated algorithms are the basis for several actigraphy devices which are used by physicians in sleep assessment. The results indicate that our algorithm detects sleep and wake phases efficiently, while standard approaches tend to overestimate sleep and neglect wake segments.

Further, we evaluate the use of additional sensor data than inertial data to detect night segments. A Gaussian model-based approach relies on the current time, the presence of motion (variance) and light (mean). The results indicate that sleep can be detected with such an approach but a higher accuracy can be reached by using a Hidden Markov Model (HMM). The HMM applies the same features for sleep detection and is able to capture the user's sleeping routine. By training the HMM on a long-term dataset from 8 participants, the different sleep rhythms provide a suitable basis for capturing different sleeping habits. For such a classifier, long-term data is needed to provide the system with sufficient training data.

Additionally, in Chapter 6, we scrutinize sleep segments and detect posture changes within immobile and myclonic twitches within mobile segments. We are able to detect specific sleep postures with a high accuracy (up to 88% for two subjects) and we are able to visualize sleeping posture trends without knowing the actual posture. In addition to that, myclonic twitches are detectable with a confidence of 73% precision and 66% recall, which is interesting especially for physicians since such twitches can hint at certain neurological diseases like Parkinson's disease.

8.3 OUTLOOK

This thesis focuses on two major challenges, which are the modelling of a person's behaviour into rhythmic patterns to extract prior information and the detection of high-level activities with such data. In the following we will point to possible future work that can continue and extend our studies.

Sources of Prior Information. We introduced time use surveys and observational data from sleep diaries as sources for rhythmic behaviour modelling. With such data, we were able to improve high-level activity recognition, by embedding such information in the classification process. For this purpose we mainly used the German Time Use Survey (GTUS) from 2001/2002, since the current version is not yet available for research purposes. Since the GTUS 2011/2012 will be released in 2015, the main research question aims at the differences between the current and previous GTUS dataset. Interesting results are expected, indicating in how far the personal routines of Germans have changed or not.

We believe that there are far more platforms that provide valuable information about the user. Social networks, for example, contain activity patterns that hint at common behaviours, with each user posting the current location and what he or she is currently doing there. Such information can be extracted from other platforms as well, e.g., Foursquare[42] or Yelp[43] to be embedded in a context-aware system.

Furthermore, such prior information could be included in the classification process before training. For this purpose, a sequential probabilistic model could be used, where the sensor data is weighted with prior probabilities derived, for example, from the time use database or other modalities that exhibit personal routines.

Sleep Detection in the Wild. We showed the feasibility of detecting sleep with a Hidden Markov Model (HMM) and various thresholds on sensor readings in the home environment. The third approach has been evaluated in the sleeping lab only, which is why the algorithm needs to be stress-tested in a real-world set-up. The challenge here is the gathering of ground truth data, since polysomnography (PSG) cannot be performed as confident as in a sleeping lab. For this purpose an additional ground truth monitoring system has to be added to our sleep monitoring system, which could be a mobile PSG system. The feasibility of using such a system over several weeks could be a challenge, since the user has to attach hard-wired sensors to the body, again resulting in an unusual sleep environment.

Improvement possibilities can be incorporated directly in the new study: We believe that performance could be improved for our ESS approach by using additional information, gathered, for example, from on-board sensors (like light readings that are already recorded by the wrist-worn device). Another possibility is to include patient-specific models on sleeping disorders and personal routines, such as usual sleep times.

[42]https://foursquare.com/, last access 09/2014
[43]http://www.yelp.co.uk, last access 09/2014

Smart Phone-based Activity Recognition. Using the smart phone as a sensing platform is feasible and depends on the user's habits as we have shown in this thesis. Activity recognition could be conducted directly on the smart phone, since state-of-the-art devices incorporate powerful CPUs and provide the system with sufficient RAM. Therefore, online activity recognition on a smart phone is a future challenge that can be encountered with the usage of prior information. One scenario would embed time use data directly in the classification process, using as a prior the current most likely activity such as the time use survey to extract probabilities for performed activities. Additionally, such a platform is suitable to be used by the user as a diary for main activities that can be logged and interpreted just as the time use data. We would like to investigate the trade-off of using such data in the recognition process in a long-term smart phone based study.

LIST OF FIGURES

Fig. 2.1	Activity recognition process.	15
Fig. 2.2	Examples of confusion matrices.	17
Fig. 2.3	Sleeping laboratory set-up.	26
Fig. 2.4	Example of a hypnogram.	27
Fig. 2.5	Typical sleep cycles for a healthy young male.	28
Fig. 3.1	The wrist-worn sensor with OLED display.	36
Fig. 3.2	Smart phone study overview.	38
Fig. 3.3	Raw acceleration data from a male participant	40
Fig. 3.4	Android recording application for all sorts of sensors	41
Fig. 3.5	Raw acceleration data from a female and male participant	45
Fig. 3.6	Results for two male subjects for one day	47
Fig. 3.7	The IR camera	53
Fig. 3.8	Overview of the detection system	54
Fig. 3.9	Potential difficult situations for vision-based recognition	56
Fig. 4.1	Visual comparison of the ATUS 2006 and the GTUS 2001/2002	63
Fig. 4.2	Precision and recall results for different features	65
Fig. 4.3	Box plots of the accuracy results per household from the GTUS	67
Fig. 4.4	Screenshots of the obtained video footage of the test subject.	71
Fig. 4.5	All 141 nights for our test subject	72
Fig. 5.1	Time series of the data abstraction steps performed in this study	80
Fig. 5.2	Our approach focuses on the detection of sustained periods of idleness during sleep	82
Fig. 5.3	An illustration of time series data from a wrist-worn accelerometer	83
Fig. 5.4	Sleep phases and inertial data from a 24-year-old female	85
Fig. 5.5	Sleep-wake estimation results for a 24-year-old female	86
Fig. 5.6	Sleep-wake estimation results for a 69-year-old male	86
Fig. 5.7	Sleep-wake estimation results for a 72-year-old male	87
Fig. 5.8	Accuracy results for the ESS algorithm compared to	88
Fig. 5.9	Precision and recall results for ESS compared to	89
Fig. 5.10	Different idleness segment thresholds compared among each other	90
Fig. 5.11	Accuracy results for different STD thresholds	91
Fig. 5.12	Histogram of variance occurring in the different sleep phases	92
Fig. 5.13	An example of typical sleeping times, taken from the GTUS	94
Fig. 5.14	The recall values for the Gaussian-based and HMM-based classifiers	96
Fig. 5.15	The raw classification output from night segmentation	98
Fig. 6.1	Personality clues can be derived from sleep postures	103
Fig. 6.2	Example of: Raw data, filtered values, postures	105
Fig. 6.3	PCA visualization of sleep posture data	106
Fig. 6.4	Confusion matrices for two subjects showing	107

Fig. 6.5 Thirty nights (one per row) of subject 1 109
Fig. 6.6 The color coded postures used in the survey 112
Fig. 6.7 Visualization of the typical nights of the 5 subjects 112
Fig. 6.8 Key features used for the myoclonic twitch detection 114
Fig. 6.9 Some examples of the observed myoclonic patterns 115
Fig. 7.1 BodyMedia SWA worn by a male and female participant 121
Fig. 7.2 Inertial data from the SWA's 2-axis accelerometer 124
Fig. 7.3 Example of likelihoods estimated from the sensor 127
Fig. 7.4 Example of classification results for a male subject 128
Fig. 7.5 Precision and recall results for three different classification techniques 131

LIST OF TABLES

Tab. 2.1 GTUS dataset examples with personal information. 23
Tab. 2.2 GTUS dataset examples with activity information. 23
Tab. 3.1 Variety of participants for the smart phone study. 43
Tab. 3.2 Smart phone study results for all users. 49
Tab. 4.1 Comparison of the basic properties of the ATUS and GTUS . . . 62
Tab. 4.2 Our tier 1 precision and recall results for the GTUS database . . 68
Tab. 4.3 Average sleep duration, falling asleep time, ... 73
Tab. 4.4 Accuracy results for the leave-one-day-out cross-validation 74
Tab. 5.1 Some key properties of the collected benchmark data 84
Tab. 5.2 Total sleep time (TST) and sleep efficiency (SE) for all nights . . . 87
Tab. 5.3 The group of participants used in the evaluation 93
Tab. 5.4 The HMM precision, recall, and accuracy results for all subjects . 97
Tab. 6.1 Sleep postures used in this study 104
Tab. 6.2 Eight basic sleep quality related questions 108
Tab. 6.3 The group of participants used in the evaluation 110
Tab. 6.4 The precision (pre) and recall (rec) results for 116
Tab. 7.1 The test subjects that participated in this study 123
Tab. 7.2 Activities taken from the time use survey 123
Tab. 7.3 Overall precision and recall results for each activity 130

K. Adam and I. Oswald (1984). Sleep Helps Healing., *British Medical Journal (Clinical research ed.)*, vol. 289(6456), pp. 1401–1402.

A. M. Adami, T. L. Hayes, and M. Pavel (2003). Unobtrusive Monitoring of Sleep Patterns, in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003. 2003*.

H. W. Agnew, W. B. Webb, and R. L. Williams (1966). The First Night Effect: An EEG Study of Sleep, *Psychophysiology*, vol. 2(3), pp. 263–266.

L. A. Alexandre, A. C. Campilho, and M. Kamel (2001). On Combining Classifiers Using Sum and Product Rules, *Pattern Recognition Letters*, vol. 22(12), pp. 1283–1289.

O. Amft, H. Junker, P. Lukowicz, G. Tröster, and C. Schuster (2006). Sensing Muscle Activities with Body-Worn Sensors, *4th Internation Conference on Pervasive Computing*.

O. Amft, M. Kusserow, and G. Tröster (2007). Probabilistic Parsing of Dietary Activity Events, in *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007) 2007*.

S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak (2003). The role of actigraphy in the study of sleep and circadian rhythms., *Sleep*, vol. 26(3), pp. 342–92.

D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz (2012). Human Activity Recognition on Smartphones Using a Multiclass Hardware-friendly Support Vector Machine, in *Ambient Assisted Living and Home Care 2012*, pp. 216–223, Springer.

S. Antifakos, F. Michahelles, and B. Schiele (2002). Proactive Instructions for Furniture Assembly, in *UbiComp 2002: Ubiquitous Computing 2002*, pp. 351–360, Springer.

D. Ashbrook and T. Starner (2002). Learning Significant Locations and Predicting User Movement with GPS, in *Proceedings of the Sixth International Symposium on Wearable Computers (ISWC). 2002*.

D. Ashbrook and T. Starner (2003). Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users, *Personal and Ubiquitous Computing*, vol. 7(5), pp. 275–286.

A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga (2010). Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports

Applications: A Survey, in *Proceedings of the 23rd International Conference on Architecture of Computing Systems (ARCS), 2010 2010*.

M. Bächlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster (2010). Wearable Assistant for Parkinson's Disease Patients with the Freezing of Gait Symptom, *IEEE Transactions on Information Technology in Biomedicine*, vol. 14(2), pp. 436–446.

D. Bain, M. Ferguson-Pell, and A. McLeod (2003). Evaluation of Mattresses Using Interface Pressure Mapping, *J Wound Care*, vol. 12(6):231-5.

S. Banks and D. F. Dinges (2007). Behavioral and Physiological Consequences of Sleep Restriction, *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, vol. 3(5), p. 519.

L. Bao and S. Intille (2004). Activity Recognition from User-annotated Acceleration Data, *Proceedings of the 2nd International Conference on Pervasive Computing*, pp. 1–17.

J. E. Bardram, M. Frost, K. Szántó, and G. Marcu (2012). The MONARCA Self-assessment System: A Persuasive Personal Monitoring System for Bipolar Patients, in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium 2012*.

J. B. Begole, J. C. Tang, and R. Hill (2003). Rhythm Modeling, Visualizations and Applications, *Proceedings of the 16th annual ACM symposium on User interface software and technology - UIST '03*, vol. 5(2), pp. 11–20.

K. Benson, L. Friedman, A. Noda, D. Wicks, E. Wakabayashi, and J. Yesavage (2004). The Measurement of Sleep by Actigraphy: Direct Comparison of 2 Commercially Available Actigraphs in a Nonclinical Population., *Sleep*, vol. 27(5), pp. 986–9.

M. Berchtold, M. Budde, D. Gordon, H. Schmidtke, and M. Beigl (2010). Actiserv: Activity Recognition Service for Mobile Phones, *14th International Symposium on Wearable Computers (ISWC)*, pp. 1–9.

E. Berlin and K. Van Laerhoven (2012a). Detecting Leisure Activities with Dense Motif Discovery, in *Proc. of the 2012 ACM Conference on Ubiquitous Computing (UbiComp). 2012*.

E. Berlin and K. Van Laerhoven (2012b). Trainspotting: Combining Fast Features to Enable Detection on Resource-constrained Sensing Devices, in *Ninth International Conference on Networked Sensing Systems (INSS), 2012 2012*.

P. Bielik, M. Tomlein, P. Krátky, Š. Mitrík, M. Barla, and M. Bieliková (2012). Move2Play: An Innovative Approach to Encouraging People to be More Physically Active, in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium 2012*.

S. Blaha (2002). *The Rhythms of History: A Universal Theory of Civilizations*, Pingree-Hill Publishing.

U. Blanke, T. Franke, G. Tröster, and P. Lukowicz (2014). Capturing Crowd Dynamics at Large Scale Events Using Participatory GPS-localization, in *The 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing 2014*.

U. Blanke and B. Schiele (2010). Remember and Transfer what you have Learned - Recognizing Composite Activities based on Activity Spotting, *International Symposium on Wearable Computers (ISWC), 2010*.

M. H. Bonnet and D. L. Arand (2003). Insomnia, Metabolic Rate and Sleep Restoration, *Internal Medicine*, vol. 254, pp. 23–31.

A. G. Bonomi, G. Plasqui, A. H. Goris, and K. R. Westerterp (2010). Estimation of Free-Living Energy Expenditure Using a Novel Activity Monitor Designed to Minimize Obtrusiveness, *Obesity*, vol. 18(9), pp. 1845–1851.

H. Brashear, T. Starner, P. Lukowicz, and H. Junker (2003). Using Multiple Sensors for Mobile Sign Language Recognition, in *Proceedings of the 7th IEEE International Symposium on Wearable Computers 2003*.

T. Brezmes, J. Gorricho, and J. Cotrina (2009). Activity Recognition from Accelerometer Data on a Mobile Phone, *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, pp. 796–799.

A. Bulling, J. A. Ward, H. Gellersen, and G. Troster (2011). Eye Movement Analysis for Activity Recognition Using Electrooculography, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33(4), pp. 741–753.

D. Buysse, C. Reynolds, T. Monk, S. Berman, and D. Kupfer (1989). The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research., in *Psychiatry Res. 1989*.

J. S. Carpenter and M. Andrykowski (1998). Psychometric evaluation of the Pittsburgh Sleep Quality Index, *Journal of Psychosomatic Research*, vol. 45(1), pp. 5–13.

J. N. Caviness (1996). Myoclonus., in *Mayo Clinic Proceedings 1996*.

J. N. Caviness and P. Brown (2004). Myoclonus: Current Concepts and Recent Advances, *The Lancet Neurology*, vol. 3(10), pp. 598–607.

C.-C. Chang and C.-J. Lin (2011). LIBSVM: A Library for Support Vector Machines, *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27.

Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell (2013). Unobtrusive Sleep Monitoring Using Smartphones, in *7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 2013*.

E. H. Chi (2005). Introducing Wearable Force Sensors in Martial Arts, *Pervasive Computing, IEEE*, vol. 4(3), pp. 47–53.

R. Cole, D. Kripke, D. Mullaney, and C. Gillin (1992). Automatic Sleep/wake Identification from Wrist Activity, *Sleep*, pp. 461–469.

C. Cortes and V. Vapnik (1995). Support-vector Networks, *Machine learning*, vol. 20(3), pp. 273–297.

S. S. Coughlin (1990). Recall Bias in Epidemiologic Studies, *Journal of Clinical Epidemiology*, vol. 43(1), pp. 87–91.

N. Cristianini and J. Shawe-Taylor (2000). *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*, Cambridge University Press.

J. De Koninck, P. Gagnon, and S. Lallier (1983). Sleep Positions in the Young Adult and their Relationship with the Subjective Quality of Sleep, *Sleep*, vol. 6(1), p. 52.

R. De Oliveira, M. Cherubini, and N. Oliver (2010). MoviPill: Improving Medication Compliance for Elders Using a Mobile Persuasive Social Game, in *Proceedings of the 12th ACM International Conference on Ubiquitous Computing 2010*.

L. de Souza, A. Benedito-Silva, M. Pires, D. Poyares, S. Tufik, H. Calil, *et al.* (2003). Further Validation of Actigraphy for Sleep Studies, *SLEEP*, vol. 26(1), pp. 81–85.

A. K. Dey, K. Wac, and D. Ferreira (2011a). Getting Closer: An Empirical Investigation of the Proximity of Users to their Smart Phones, *ACM Ubicomp*.

A. K. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos (2011b). Getting Closer: An Empirical Investigation of the Proximity of User to their Smart Phones, in *Proceedings of the 13th International Conference on Ubiquitous Computing 2011*.

T. Do and D. Gatica-Perez (2011). GroupUs: Smartphone Proximity Data and Human Interaction Type Mining, in *15th IEEE International Symposium on Wearable Computers (ISWC) 2011*.

L. Dunne, P. Walsh, B. Smyth, and B. Caulfield (2006). Design and Evaluation of a Wearable Optical Sensor for Monitoring Seated Spinal Posture, in *10th IEEE International Symposium on Wearable Computers 2006*.

N. Eagle and A. Pentland (2006). Reality Mining: Sensing Complex Social Systems, *Personal and Ubiquitous Computing*, vol. 10(4), pp. 255–268.

N. Eagle, J. Quinn, and A. Clauset (2009). Methodologies for Continuous Cellular Tower Data Analysis, *Proceedings of the 7th International Conference on Pervasive Computing*, pp. 342–353.

M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen (2008). Detection of Daily Activities and Sports with Wearable Sensors in Controlled and Uncontrolled Conditions, *IEEE Transactions on Information Technology in Biomedicine*, vol. 12(1), pp. 20–26.

C. A. Everson, M. Bergmann, and A. Rechtschaffen (1989). Sleep Deprivation in the Rat: III. Total Sleep Deprivation, in *Sleep 1989*.

K. Farrahi and D. Gatica-Perez (2008). What did you do Today?: Discovering Daily Routines from Large-scale Mobile Data, in *Proceedings of the 16th ACM International Conference on Multimedia 2008*.

K. Farrahi and D. Gatica-Perez (2011). Discovering Routines from Large-scale Human Locations Using Probabilistic Topic Models, *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2(1), p. 3.

E. Fuchs, T. Gruber, J. Nitschke, and B. Sick (2009). On-line Motif Detection in Time Series with SwiftMotif, *Pattern Recognition*, vol. 42(11), pp. 3015–3031.

J. Gershuny (2011). Time-use Surveys and the Measurement of National Well-being, *Centre for Time-use Research. Department of Sociology. University of Oxford*.

S. Gordon, K. Grimmer, and P. Trott (2004). Self Reported versus Recorded Sleep Position: An Observational Study, *Internet Journal of Allied Health Sciences and Practice*, vol. 2(1), pp. 1–10.

J. Gregory, X. Xie, and S. Mengel (2004). SLEEP (Sleep Loss Effects On Everyday Performance) Model, *Aviation, Space and Environmental Medicine*, vol. 75(Supplement 1), pp. A125–A133.

H. Haddadi, P. Hui, and I. Brown (2010). MobiAd: Private and Scalable Mobile Advertising, in *Proceedings of the Fifth ACM International Workshop on Mobility in the Evolving Internet Architecture 2010*.

J. Hedner, G. Pillar, S. D. Pittman, D. Zou, L. Grote, and D. P. White (2004). A Novel Adaptive Wrist Actigraphy Algorithm for Sleep-Wake Assessment in Sleep Apnea Patients, *Sleep*, vol. 27(8), pp. 1560–1566.

E. A. Heinz, K. S. Kunze, M. Gruber, D. Bannach, and P. Lukowicz (2006). Using Wearable Sensors for Real-time Recognition Tasks in Games of Martial Arts - An Initial Experiment, in *Computational Intelligence and Games, 2006 IEEE Symposium on 2006*.

S. Herscovici, A. Pe'er, S. Papyan, and P. Lavie (2007). Detecting REM Sleep from the Finger: An Automatic REM Sleep Algorithm Based on Peripheral Arterial Tone (PAT) and Actigraphy, *Physiological Measurement*, vol. 28(2), p. 129.

R. Hill and J. Begole (2003). Activity Rhythm Detection and Modeling, *CHI '03 extended Abstracts on Human Factors in Computer Systems*, p. 782.

J. Hong, J. Ng, S. Lederer, and J. Landay (2004). Privacy Risk Models for Designing Privacy-Sensitive Ubiquitous Computing Systems, in *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques 2004*.

E. Hoque, R. F. Dickerson, and J. A. Stankovic (2010). Monitoring Body Positions and Movements During Sleep Using WISPs, in *Wireless Health 2010 2010*.

T. Huynh, M. Fritz, and B. Schiele (2008). Discovery of Activity Patterns Using Topic Models, in *Proceedings of the 10th International Conference on Ubiquitous Computing 2008*.

T. Huynh and B. Schiele (2005). Analyzing Features for Activity Recognition, in *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-aware Services: Usages and Technologies 2005*.

T. Huynh and B. Schiele (2006a). Towards Less Supervision in Activity Recognition form Wearable Sensors, in *Proceedings of the 10th IEEE International Symposium on Wearable Computing (ISWC) 2006*.

T. Huynh and B. Schiele (2006b). Unsupervised Discovery of Structure in Activity Data using Multiple Eigenspaces, in *2nd International Workshop on Location- and Context-Awareness (LoCA) 2006*.

G. Hyland (2000). Physics and Biology of Mobile Telephony, *The Lancet*, vol. 356(9244), pp. 1833–1836.

F. Ichikawa, J. Chipchase, and R. Grignani (2005). Where's The Phone? A Study of Mobile Phone Location in Public Space, in *2nd International Conference on Mobile Technology, Applications and Systems, 2005 2005*.

C. J. Idzikowski (2000). *Learn to Sleep Well*, Duncan Baird.

S. P. Insana, D. Gozal, and H. E. Montgomery-Downs (2010). Invalidity of one Actigraphy Brand for Identifying Sleep and Wake Among Infants., *Sleep Medicine*, vol. 11(2), pp. 191–6.

R. Jafari, W. Li, R. Bajcsy, S. Glaser, and S. Sastry (2007). Physical Activity Monitoring for Assisted Living at Home, in *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007) 2007*.

G. Jean-Louis, D. F. Kripke, R. J. Cole, J. D. Assmus, and R. D. Langer (2001a). Sleep Detection with an Accelerometer Actigraph: Comparisons with Polysomnography., *Physiology & Behavior*, vol. 72(1-2), pp. 21–8.

G. Jean-Louis, D. F. Kripke, W. J. Mason, J. A. Elliott, and S. D. Youngstedt (2001b). Sleep Estimation from Wrist Movement Quantified by Different Actigraphic Modalities, *Journal of Neuroscience*, vol. 105(2), pp. 185–191.

G. Jean-Louis, F. Zizi, H. Gyzicki, and P. Hauri (1999). Actigraphic Assessment of Sleep in Insomnia: Application of the Actigraph Data Analysis Software (ADAS), in *Physiology & Behavior Vol.65 No.4-5 1999*.

M. W. Johns *et al.* (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale, *Sleep*, vol. 14(6), pp. 540–545.

W. Karlen, C. Mattiussi, and D. Floreano (2008). Improving Actigraph Sleep/wake Classification with Cardio-respiratory Signals, in *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008. (EMBS) 2008*.

H. Kautz, O. Etzioni, D. Fox, D. Weld, and L. Shastri (2003). Foundations of Assisted Cognition Systems, *University of Washington, Computer Science Department, Technical Report*.

K. Kawamoto, H. Kuriyama, and S. Tajima (2013). Actigraphic Detection of REM Sleep Based on Respiratory Rate Estimation, *Journal of Medical and Bioengineering*, vol. 2(1).

T. Kohonen (1990). The Self-Organizing Map, *Proceedings of the IEEE*, vol. 78(9), pp. 1464–1480.

S. Krachman, G. Criner, and G. D'Alonzo (1995). Sleep in the Intensive Care Unit, *Chest*, vol. 107(6), p. 1713.

G. Krassnig, D. Tantinger, C. Hofmann, T. Wittenberg, and M. Struck (2010). User-friendly System for Recognition of Activities with an Accelerometer, in *4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) 2010*.

A. Krause, D. P. Siewiorek, A. Smailagic, and J. Farringdon (2003). Unsupervised, Dynamic Identification of Physiological and Activity Context in Wearable Computing, in *16th International Symposium on Wearable Computers 2003*.

J. Krumm and A. J. B. Brush (2011). Learning Time-based Presence Probabilities, in *Proceedings of the 9th International Conference on Pervasive Computing 2011*, pp. 79–96, Springer.

K. Kunze, M. Barry, and E. A. Heinz (2006). Towards Recognizing Tai Chi - An Initial Experiment Using Wearable Sensors, *Applied Wearable*, pp. 1–6.

C. a. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. C. Dement (2001). Comparison of Actigraphic, Polysomnographic and Subjective Assessment of Sleep Parameters in Sleep-Disordered Patients., *Sleep Medicine*, vol. 2(5), pp. 389–96.

C. a. Kushida, M. R. Littner, T. Morgenthaler, C. a. Alessi, D. Bailey, J. Coleman, L. Friedman, M. Hirshkowitz, S. Kapen, M. Kramer, T. Lee-Chiong, D. L. Loube,

J. Owens, J. P. Pancer, and M. Wise (2005). Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005., *Sleep*, vol. 28(4), pp. 499–521.

J. R. Kwapisz, G. M. Weiss, and S. A. Moore (2011). Activity Recognition Using Cell Phone Accelerometers, *ACM SIGKDD Explorations Newsletter*, vol. 12(2), pp. 74–82.

Y. Kwon, K. Kang, and C. Bae (2014). Unsupervised Learning for Human Activity Recognition Using Smartphone Sensors, *Expert Systems with Applications*, vol. 41(14), pp. 6067 – 6074.

C. Ladha, N. Y. Hammerla, P. Olivier, and T. Plötz (2013). ClimbAX: Skill Assessment for Climbing Enthusiasts, in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing 2013*.

K. V. Laerhoven and H.-W. Gellersen (2004). Spine versus Porcupine: A Study in Distributed Wearable Activity Recognition., in *Proceedings of the 8th IEEE International Symposium on Wearable Computers 2004*.

N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. Campbell (2011). Bewell: A Smartphone Application to Monitor, Model and Promote Wellbeing, in *5th International ICST Conference on Pervasive Computing Technologies for Healthcare 2011*.

S. W. Lee and K. Mase (2002). Activity and Location Recognition Using Wearable Sensors, *IEEE Pervasive Computing*, vol. 1(3), pp. 24–32.

J. Lester, T. Choudhury, and G. Borriello (2006). A Practical Approach to Recognizing Physical Activities, *Proceedings of the 4th International Conference on Pervasive Computing*, pp. 1–16.

W. Liang (2013). Sleep Duration and Its Links to Psychological Distress, Health Status, Physical Activity and Body Mass Index among a Large Representative General Population Sample, *International Journal of Clinical Medicine*, vol. 04(01), pp. 45–51.

W. Liao and C. Yang (2008). Video-Based Activity and Movement Pattern Analysis in Overnight Sleep Studies, in *19th International Conference on Pattern Recognition, 2008. ICPR 2008. 2008*.

K. L. Lichstein, K. C. Stone, J. Donaldson, S. D. Nau, J. P. Soeffing, D. Murray, K. W. Lester, and R. N. Aguillard (2006). Actigraphy Validation with Insomnia, *Sleep - New York*.

T. Lipoma, P. Bello, C. Darling, and M. T. Bianchi (2011). SOMNUS: A Sleep-Measuring Shirt Based on Chest Expansion and Respiratory Patterns, *Journal of Medical Devices*, vol. 5(2), p. 027509.

M. Littner, C. Kushida, D. Dennis Bailey, R. Berry, D. Davila, and M. Hirshkowitz (2003). Practice Parameters for the Role of Actigraphy in the Study of Sleep and Circadian Rhythms: An Update for 2002, *Sleep*, vol. 26(3), p. 337.

B. Longstaff, S. Reddy, and D. Estrin (2010). Improving Activity Classification for Health Applications on Mobile Devices Using Active and Semi-supervised Learning, in *4th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 2010*.

E. Lugaresi, F. Cirignotta, G. Coccagna, and P. Montagna (1985). Nocturnal Myoclonus and Restless Legs Syndrome., *Advances in Neurology*, vol. 43, pp. 295–307.

P. Lukowicz, A. Timm-Giel, M. Lawo, and O. Herzog (2007). wearIT@work: Toward Real-world Industrial Wearable Computing, *Pervasive Computing, IEEE*, vol. 6(4), pp. 8–13.

A. Lymberis and S. Olsson (2003). Intelligent Biomedical Clothing for Personal Health and Disease Management: State of the Art and Future Vision, *Telemedicine Journal and e-health*, vol. 9(4), pp. 379–386.

T. Maekawa and S. Watanabe (2011). Unsupervised Activity Recognition with User's Physical Characteristics Data, in *15th International Symposium on Wearable Computers (ISWC) 2011*.

U. Maurer, a. Smailagic, D. P. Siewiorek, and M. Deisher (2006). Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions, *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pp. 113–116.

I. Maurtua, P. T. Kirisci, T. Stiefmeier, M. L. Sbodio, and H. Witt (2007). A Wearable Computing Prototype for Supporting Training Activities in Automotive Production, in *4th International Forum on Applied Wearable Computing (IFAWC), 2007 2007*.

J. Meyer, P. Lukowicz, and G. Tröster (2006). Textile Pressure Sensor for Muscle Activity and Motion Detection, *10th IEEE International Symposium on Wearable Computers*, pp. 69–72.

D. Minnen, T. Starner, I. Essa, and C. Isbell (2006). Discovering Characteristic Actions from On-body Sensor Data, in *10th IEEE International Symposium on Wearable Computers, 2006 2006*.

D. Minnen, T. Starner, I. Essa, and C. Isbell (2007a). Improving Activity Discovery with Automatic Neighborhood Estimation, *International Joint Conference on Artificial Intelligence*.

D. Minnen, T. Westeyn, D. Ashbrook, P. Presti, and T. Starner (2007b). Recognizing soldier Activities in the Field, in *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007) 2007*.

T. K. Monk, J. F. Flaherty, E. Frank, K. Hoskinson, and D. J. Kupfer (1990). The Social Rhythm Metric: An Instrument to Quantify the Daily Rhythms of Life, *Journal of Nervous and Mental Disease*.

H. E. Montgomery-Downs, S. P. Insana, and J. a. Bond (2012). Movement Toward a Novel Activity Monitoring Device., *Sleep & breathing*, vol. 16(3), pp. 913–7.

T. Morgenthaler, C. Alessi, L. Friedman, J. Owens, V. Kapur, B. Boehlecke, T. Brown, A. Chesson, J. Coleman, T. Lee-Chiong, J. Pancer, and T. J. Swick (2007). Practice Parameters for the Use of Actigraphy in the Assessment of Sleep and Sleep Disorders: An Update for 2007, *Sleep*, vol. 30(4), pp. 519–529.

K. Murakami, S. Doki, S. Okuma, and Y. Yano (2005). A Study of Extraction Method of Motion Patterns Observed Frequently from Time-series Posture Data, in *IEEE International Conference on Systems, Man and Cybernetics, 2005 2005*.

K. Murphy (2003). Dynamic Bayesian Networks, *Probabilistic Graphical Models*.

I. J. Mykytyn, D. Sajkov, A. M. Neill, and R. D. McEvoy (1999). Portable Computerized Polysomnography in Attended and Unattended Settings, *CHEST Journal*, vol. 115(1), pp. 114–122.

V. Natale, M. Drejak, A. Erbacci, L. Tonetti, M. Fabbri, and M. Martoni (2012). Monitoring Sleep with a Smartphone Accelerometer, *Sleep and Biological Rhythms*, vol. 10(4), pp. 287–292.

N. Oakley (1997). Validation with Polysomnography of the Sleepwatch Sleep/Wake Scoring Algorithm used by the Actiwatch Activity Monitoring System, *Technical Report*.

A. Oksenberg, E. Arons, S. Greenberg-Dotan, K. Nasser, and H. Radwan (2009). The Significance of Body Posture on Breathing Abnormalities During Sleep: Data Analysis of 2077 Obstructive Sleep Apnea Patients, *Harefuah*, vol. 148(5), p. 304.

A. Oksenberg and D. Silverberg (1998). The Effect of Body Posture on Sleep-Related Breathing Disorders: Facts and Therapeutic Implications, *Sleep Medicine Reviews*, vol. 2(3), pp. 139–162.

J. J. Oresko, Z. Jin, J. Cheng, S. Huang, Y. Sun, H. Duschl, and A. C. Cheng (2010). A Wearable Smartphone-based Platform for Real-time Cardiovascular Disease Detection via Electrocardiogram Processing, *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14(3), pp. 734–740.

J. Pansiot, D. Stoyanov, D. McIlwraith, B. P. L. Lo, and G.-Z. Yang (2007). Ambient and Wearable Sensor Fusion for Activity Recognition in Healthcare Monitoring Systems, in *Proceedings of the 4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007) 2007*.

R. Paradiso, G. Loriga, and N. Taccini (2005). A Wearable Health Care System Based on Knitted Integrated Sensors, *IEEE Transactions on Information Technology in Biomedicine*, vol. 9(3), pp. 337–344.

T. Partonen and A. Magnusson (2001). *Seasonal Affective Disorder: Practice and Research*, vol. 66, Oxford University Press Oxford.

K. Partridge and P. Golle (2008). On Using Existing Time-Use Study Data for Ubiquitous Computing Applications, in *Proceedings of the 10th International Conference on Ubiquitous Computing 2008*.

S. N. Patel, J. A. Kientz, G. R. Hayes, S. Bhat, and G. D. Abowd (2006). Farther Than You May Think: An Empirical Investigation of the Proximity of Users to Their Mobile Phones, in *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp) 2006*, pp. 123–140, Springer.

D. Patterson, D. Fox, H. Kautz, and M. Philipose (2005). Fine-Grained Activity Recognition by Aggregating Abstract Object Usage, *Proceedings of the 9th IEEE International Symposium on Wearable Computers*, pp. 44–51.

H. Peter, T. Penzel, and J. H. Peter (2007). *Enzyklopädie der Schlafmedizin*, Springer.

M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, and D. Hahnel (2004). Inferring Activities from Interactions with Objects, *IEEE Pervasive Computing*, vol. 3(4), pp. 50–57.

T. Plötz, N. Y. Hammerla, A. Rozga, A. Reavis, N. Call, and G. D. Abowd (2012). Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities, in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing 2012*.

R. Polikar (2006). Ensemble Based Systems in Decision Making, *Circuits and Systems Magazine, IEEE*, vol. 6(3), pp. 21–45.

C. P. Pollak, W. W. Tryon, H. Nagaraja, and R. Dzwonczyk (2001). How Accurately does Wrist Actigraphy Identify the States of Sleep and Wakefulness?, *Sleep*, vol. 24(8), pp. 957–65.

S. J. Preece, J. Y. Goulermas, L. P. Kenney, and D. Howard (2009). A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities from Accelerometer Data, *IEEE Transactions on Biomedical Engineering*, vol. 56(3), pp. 871–879.

N. C. Rattenborg, B. Voirin, A. L. Vyssotski, R. W. Kays, K. Spoelstra, F. Kuemmeth, W. Heidrich, and M. Wikelski (2008). Sleeping Outside the Box: Electroencephalographic Measures of Sleep in Sloths Inhabiting a Rainforest., *Biology Letters*, vol. 4(4), pp. 402–5.

N. Ravi, N. Dandekar, P. Mysore, and M. Littman (2005). Activity Recognition from Accelerometer Data, *Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence (IAAI)*, pp. 1541–1546.

A. Rechtschaffen and A. Kales (1968). A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects.

E. Roze, P. Bounolleau, D. Ducreux, V. Cochen, S. Leu-Semenescu, Y. Beaugendre, M. Lavallard-Rousseau, A. Blancher, F. Bourdain, P. Dupont, *et al.* (2009). Propriospinal Myoclonus Revisited, *Neurology*, vol. 72(15), p. 1301.

A. Sadeh (2011). The Role and Validity of Actigraphy in Sleep Medicine: An Update., vol. 15(4), pp. 259–267.

A. Sadeh, P. Hauri, D. Kripke, and P. Lavie (1995). The Role of Actigraphy in the Evaluation of Sleep Disorders, *Sleep*, vol. 18(4), pp. 288–302.

A. Sadeh, K. Sharkey, and M. Carskadon (1994). Activity-Based Sleep/Wake Identification: An Empirical Test of Methodological Issues, *Sleep*.

A. Sahami Shirazi, N. Henze, T. Dingler, K. Kunze, and A. Schmidt (2013). Upright or Sideways?: Analysis of Smartphone Postures in the Wild, in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services 2013*.

S. S. Sahoo, C. Ogbuji, L. Luo, X. Dong, L. Cui, S. S. Redline, and G.-Q. Zhang (2011). MiDas: Automatic Extraction of a Common Domain of Discourse in Sleep Medicine for Multi-center Data Integration, in *AMIA Annual Symposium Proceedings 2011*.

B. Schilit, N. Adams, and R. Want (1994). Context-aware Computing Applications, in *First Workshop on Mobile Computing Systems and Applications 1994*.

C. Schlenoff, B. Weiss, and M. P. Steves (2010). Lessons Learned in Evaluating DARPA Advanced Military Technologies, in *Proceedings of the 10th Performance Metrics for Intelligent Systems Workshop 2010*.

P. M. Scholl, N. Kücükyildiz, and K. V. Laerhoven (2013). When Do You Light a Fire?: Capturing Tobacco Use with Situated, Wearable Sensors, in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing 2013*.

P. M. Scholl and K. Van Laerhoven (2012). A Feasibility Study of Wrist-Worn Accelerometer Based Detection of Smoking Habits, in *Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 2012*.

J. Scott, A. J. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar (2011). PreHeat: Controlling Home Heating Using Occupancy Prediction, in *Proceedings of the 13th International Conference on Ubiquitous Computing 2011*.

C. Seeger, A. Buchmann, and K. Van Laerhoven (2011). myHealthAssistant: A Phone-based Body Sensor Network that Captures the Wearer's Exercises throughout the Day, in *Proceedings of the 6th International Conference on Body Area Networks 2011*.

K. Shelley (2005). Developing the American Time Use Survey Activity Classification System, *Monthly Labor Review*, vol. 128, pp. 3–15.

S. L. Shuger, V. W. Barry, X. Sui, A. McClain, G. A. Hand, S. Wilcox, R. A. Meriwether, J. W. Hardin, and S. N. Blair (2011). Electronic Feedback in a Diet- and Physical Activity-based Lifestyle Intervention for Weight Loss: A Randomized Controlled Trial, *International Journal of Behavioral Nutrition and Physical Activity*, vol. 8(41), pp. 1–8.

J. M. Siegel (2009). Sleep Viewed as a State of Adaptive Inactivity, *Nature Reviews Neuroscience*, vol. 10(10), pp. 747–753.

V. Stanford (2002). Wearable Computing Goes live in Industry, *Pervasive Computing, IEEE*, vol. 1(4), pp. 14–19.

U. Steinhoff and B. Schiele (2010). Dead Reckoning from the Pocket - An Experimental Study, in *IEEE International Conference on Pervasive Computing and Communications (PerCom) 2010*.

T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster (2008). Wearable Activity Tracking in Car Manufacturing, *IEEE Pervasive Computing*, vol. 7(2), pp. 42–50.

M. Stikic, D. Larlus, and B. Schiele (2009). Multi-graph Based Semi-supervised Learning for Activity Recognition, in *Wearable Computers, 2009. ISWC'09. International Symposium on 2009*.

M. Stikic and K. Van Laerhoven (2007). Recording Housekeeping Activities with Situated Tags and Wrist-Worn Sensors: Experiment Setup and Issues Encountered, in *Proceedings of the 1st International Workshop on Wireless Sensor Networks for Health Care 2007*.

M. Stikic, K. Van Laerhoven, and B. Schiele (2008). Exploring Semi-supervised and Active Learning for Activity Recognition, *12th IEEE International Symposium on Wearable Computers*, pp. 81–88.

Y. Su, S. Shan, X. Chen, and W. Gao (2009). Hierarchical Ensemble of Global and Local Classifiers for Face Recognition, *Image Processing, IEEE Transactions on*, vol. 18(8), pp. 1885–1896.

A. Subramanya, A. Raj, J. Bilmes, and D. Fox (2006). Recognizing Activities and Spatial Context Using Wearable Sensors, in *Proceedings of the Twenty-Second Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-06) 2006*.

L. Sun, D. Zhang, B. Li, B. Guo, and S. Li (2010). Activity Recognition on an Accelerometer Embedded Mobile Phone with Varying Positions and Orientations, in *Ubiquitous intelligence and computing 2010*, pp. 548–562, Springer.

M. Sunseri, C. B. Liden, J. Farringdon, R. Pelletier, M. S. LC, S. Safier, J. Stivoric, A. Teller, and M. S. SureshVishnubhatla (2009). The SenseWear Armband as a Sleep Detection Device, *BodyMedia internal white paper Healthy MDD Sleep duration min Sleep Quality Score Sleep Quality Score*.

M. Takeshima, M. Echizenya, Y. Inomata, K. Shimizu, and T. Shimizu (2014). Comparison of Sleep Estimation Using Wrist Actigraphy and Waist Actigraphy in Healthy Young Adults, *Sleep and Biological Rhythms*, vol. 12(1), pp. 62–68.

D. M. J. Tax, M. Van Breukelen, R. P. W. Duin, and J. Kittler (2000). Combining Multiple Classifiers by Averaging or by Multiplying?, *Pattern recognition*, vol. 33(9), pp. 1475–1485.

H. L. Taylor, D. R. Jacobs Jr, B. Schucker, J. Knudsen, A. S. Leon, and G. Debacker (1978). A Questionnaire for the Assessment of Leisure Time Physical Activities, *Journal of Chronic Diseases*, vol. 31, pp. 741–755.

B. H. W. te Lindert and E. J. W. Van Someren (2013). Sleep Estimates using Microelectromechanical Systems (MEMS)., *Sleep*, vol. 36(5), pp. 781–9.

E. Tews, R. Weinmann, and A. Pyshkin (2007). Breaking 104 bit WEP in less than 60 seconds, Cryptology ePrint Archive 2007/120, Cryptography and Computeralgebra group, Darmstadt University of Technology.

J. Tilmanne, J. Urbain, M. V. Kothare, A. V. Wouwer, and S. V. Kothare (2009). Algorithms for Sleep-Wake Identification Using Actigraphy: A Comparative Study and New Results, *Journal of Sleep Research*, vol. 18(1), pp. 85–98.

S. Tominaga, M. Shimosaka, R. Fukui, and T. Sato (2012). A Unified Framework for Modeling and Predicting Going-out Behavior, in *Proceedings of the 10th International Conference on Pervasive Computing 2012*, pp. 73–90, Springer.

H. P. Van Dongen, G. Maislin, J. M. Mullington, and D. F. Dinges (2003). The Cumulative Cost of Additional Wakefulness: Dose-response Effects on Neurobehavioral Functions and Sleep Physiology from Chronic Sleep Restriction and Total Sleep Deprivation, *Sleep*, vol. 26(2), pp. 117–129.

V. T. van Hees, M. Pias, S. Taherian, U. Ekelund, and S. Brage (2010). A Method to Compare New and Traditional Accelerometry Data in Physical Activity Monitoring, *2010 IEEE International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pp. 1–6.

T. Van Kasteren, A. Noulas, G. Englebienne, and B. Kröse (2008). Accurate Activity Recognition in a Home Setting, in *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp) 2008*.

K. Van Laerhoven, M. Borazio, D. Kilian, and B. Schiele (2008). Sustained Logging and Discrimination of Sleep Postures with Low-Level, Wrist-Worn Sensors, in *12th IEEE International Symposium on Wearable Computers (ISWC), 2008. 2008*.

K. Van Laerhoven, N. Kern, H. Gellersen, and B. Schiele (2003). Towards a Wearable Inertial Sensor Network, *In Proceedings of the IEEE Eurowearable*, pp. 125–130.

K. Van Laerhoven, D. Kilian, and B. Schiele (2008). Using Rhythm Awareness in Long-Term Activity Recognition, in *Proceedings of the 12th IEEE International Symposium on Wearable Computers (ISWC) 2008*.

J. Virkkala (2012). Using Accelerometers as Actigraphs, *Poster in ESRS'12*.

H. M. Wallach (2004). Conditional Random Fields: An Introduction, *Technical Reports (CIS)*, p. 22.

L. Walsh, E. Moloney, and S. McLoone (2011). Identification of Nocturnal Movements during Sleep Using the Non-contact Under Mattress Bed Sensor, in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2011 2011*.

J. Ward, P. Lukowicz, and G. Tröster (2005). Gesture Spotting Using Wrist Worn Microphone and 3-axis Accelerometer, *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence*, pp. 99–104.

J. Ward, P. Lukowicz, G. Tröster, and T. Starner (2006). Activity Recognition of Assembly Tasks Using Body-worn Microphones and Accelerometers, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28(10), pp. 1553–1567.

A. R. Weiss, N. L. Johnson, N. Berger, and S. Redline (2010). Validity of Activity-Based Devices to Estimate Sleep, *Journal of Clinical Sleep Medicine (JCSM)*, vol. 6(4), pp. 336–42.

E. Weitzman *et al.* (1981). Delayed Sleep Phase Syndrome: A Chronobiological Disorder with Sleep-Onset Insomnia., *Archives of General Psychiatry*, vol. 38(7):737-46.

G. Welk, J. Schaben, and J. Morrow (2004). Reliability of Accelerometry-Based Activity Monitors: A Generalizability Study., in *Medicine & Science in Sports & Exercise Vol.36 No.9 2004*.

W. H. Wu, A. A. Bui, M. A. Batalin, L. K. Au, J. D. Binney, and W. J. Kaiser (2008). MEDIC: Medical Embedded Device for Individualized Care, *Artificial Intelligence in Medicine*, vol. 42(2), pp. 137–152.

P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster (2008). Activity Recognition from On-body Sensors: Accuracy-power Trade-off by Dynamic Sensor Selection, in *Wireless Sensor Networks 2008*, pp. 17–33, Springer.

Z. Zhu, U. Blanke, A. Calatroni, and G. Tröster (2013). Prior Knowledge of Human Activities from Social Data, in *Proceedings of the 17th Annual International Symposium on Wearable Computers (ISWC) 2013*.

A. Zien, U. Brefeld, and T. Scheffer (2007). Transductive Support Vector Machines for Structured Variables, in *Proceedings of the 24th international conference on Machine learning 2007*.

CURRICULUM VITAE

Marko Antonio Matteo Borazio

Education:	2010–2014	**TU Darmstadt, Germany**
		PhD student at the Embedded Sensing Systems Group of Dr. Kristof Van Laerhoven
	2000–2008	**TU Darmstadt, Germany**
		Studies of computer science, graduation with degree *Dipl.-Inform.*
		Major subjects: Robotics, computer graphics, communication networks and cryptography
		Minor subjects: Bionic
		Diploma thesis: *Toward Sustained Logging in Sleep Studies through Energy-Efficient Actigraphy*, supervised by Dr. Kristof Van Laerhoven and Prof. Bernt Schiele
Experience:	2010–2014	Research assistant, Embedded Sensing Systems Group, Darmstadt, Germany.
	2008–2009	Junior consultant, software development, adept consult AG, Frankfurt/M, Germany.
	2002–2007	Student trainee, system administration, adept consult AG, Frankfurt/M, Germany.
	2002	Intern, system administration, adept consult AG, Frankfurt/M, Germany.
	1997–2002	Gastronome and IT-assistant, Ristorante Brighella, Frankfurt/M, Germany.

PUBLICATIONS

[11] *"Wear is Your Mobile? An Empirical Comparison between Wearable and Mobile User Monitoring"*
Marko Borazio and Kristof Van Laerhoven.
In submission.

[10] *"Towards Benchmarked Sleep Detection with Inertial Wrist-worn Sensing Units"*
Marko Borazio, Eugen Berlin, Nagihan Kücükyildiz, Philipp M. Scholl and Kristof Van Laerhoven.
IEEE International Conference on Healthcare Informatics (**ICHI**), Sept. 2014, Verona, Italy.

[9] *"Diary-Like Long-term Activity Recognition: Touch or Voice Interaction?"*
Philipp M. Scholl, Marko Borazio, Martin Jänsch and Kristof Van Laerhoven.
IEEE Workshop on Glass and Eyewear Computers (**GEC**), June 2014, Zurich, Switzerland.

[8] *"Already up? Using Mobile Phones to Track and Share Sleep Behavior"*
Alireza Sahami Shirazi, James Clawson, Yashar Hassanpour, Mohammad J. Tourian, Ed Chi, Marko Borazio, Albrecht Schmidt and Kristof Van Laerhoven.
International Journal of Human Computer Studies: Elsevier, 2013, vol. 71, issue 9, pp. 878-888.

[7] *"Improving Activity-Recognition without Sensor Data: A Comparison Study of Time Use Surveys"*
Marko Borazio and Kristof Van Laerhoven.
4th ACM International Augmented Human Conference (**AH)**, Mar. 2013, Stuttgart, Germany.

[6] *"Enhancing Accelerometer-based Activity Recognition with Capacitive Proximity Sensing"*
Tobias Grosse-Puppendahl, Eugen Berlin and Marko Borazio.
International Joint Conference on Ambient Intelligence (**AmI**), Nov. 2012, Pisa, Italy.

[5] *"Combining Wearable and Environmental Sensing into an Unobtrusive Tool for Long-Term Studies"*
Marko Borazio and Kristof Van Laerhoven.
2nd ACM SIGHIT International Health Informatics Symposium (**IHI**), Jan. 2012, Miami, Florida, USA.

[4] *"Predicting Sleeping Behaviors in Long-Term Studies with Wrist-Worn Sensor Data"*
Marko Borazio and Kristof Van Laerhoven.
International Joint Conference on Ambient Intelligence (**AmI**), Nov. 2011, Amsterdam, Netherlands.

[3] *"How to Log Sleeping Trends? A Case Study on the Long-Term Capturing of User Data"*
Holger Becker, Marko Borazio, and Kristof Van Laerhoven.
The 5th European Conference on Smart Sensing and Context (**EuroSSC**), Nov. 2010, Passau, Germany.

[2] *"Visualizing Sleeping Trends from Postures"*
Marko Borazio, Ulf Blanke and Kristof Van Laerhoven.
In Proceedings of the 14th International IEEE Symposium on Wearable Computers (**ISWC**), Oct. 2010, Seoul, Korea.

[1] *"Sustained Logging and Discrimination of Sleep Postures with Low-Level, Wrist-Worn Sensors"*
Kristof Van Laerhoven, Marko Borazio, David Kilian and Bernt Schiele.
In Proceedings of the 12th International IEEE Symposium on Wearable Computers (**ISWC**), Oct. 2008, Pittsburgh, PA, USA.