Adaptive Wake and Sleep Detection for Wearable Systems

THÈSE Nᵒ 4391 (2009)

PAR

Walter KARLEN

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2009

Acknowledgements

I owe much to my supervisor Prof. Dario Floreano who introduced me to the exciting world of science. His lab, the Laboratory of Intelligent Systems, which assembles a unique group of people with incredibly diversified knowledge and interests, allowed me to feel comfortable and at home for four illuminated years of my life.

I would like to express my gratitude to the jury president Prof. Sabine Süsstrunk, and the members of my jury, Prof. José C. Príncipe, Prof. Robert Riener, and Prof. José del R. Millán, for taking their time to examine my thesis and for their constructive comments on my work.

Dr. med. Werner Karrer, Dr. med. Thomas Rote and Isabelle Arnold of the Luzerner Höhenklinik in Montana (VS) introduced me to clinical sleep monitoring and scoring and awakened my interest for wearable sleep monitoring. I would like to thank them for having given me some of their valuable time.

I would also like to thank the management of the Solar Impulse project and all other Solar Impulse team members and partners, who believed in my work and made their equipment available for my research.

I am very grateful to all volunteers who participated in my sleep studies. Without their contribution, many of my findings and developments would not have been possible.

Dr. Claudio Mattiussi and Dr. Jean-Christoph Zufferey coached me all these years. I would like to thank them for being available so often and for always having a good advice.

The feedback, comments and suggestions from Sara Mitri, Dr. Steffen Wischman and Dr. Markus Waibel enriched largely this thesis. I would like to thank them for their valuable discussions and advice. I would like to thank Sabine Hauert and my brother Michael for having agreed to proof-read many parts of

i

this thesis and their patience. Thanks to Mirko Kovac and Daniel Marbach, who enriched my days with their remarks and jokes when sharing the office with me. Thanks to James Roberts for verifying my electronic designs. Thanks also to all other LIS members for contributing to the friendly, scientific atmosphere in the lab. Special thanks go to Anouk Hein, our dynamic administrative assistant.

I am very grateful to my parents Stephanie and Franz, who always encouraged me to follow an engineering and scientific career and continuously supported me during my studies.

A big kiss to Johanneli, my love, for her passion, understanding and for cheering me up in difficult periods. Without her, the difficult days of this thesis would have been more difficult, the enjoyable days less enjoyable.

Lausanne, April 2009

Abstract

Sleep problems and disorders have a serious impact on human health and well-being. The rising costs for treating sleep-related chronic diseases in industrialized countries demands efficient prevention. Low-cost, wearable sleep / wake detection systems which give feedback on the wearer's "sleep performance" are a promising approach to reduce the risk of developing serious sleep disorders and fatigue.

Not all bio-medical signals that are useful for sleep / wake discrimination can be easily recorded with wearable systems. Sensors often need to be placed in an obtrusive location on the body or cannot be efficiently embedded into a wearable frame. Furthermore, wearable systems have limited computational and energetic resources, which restrict the choice of sensors and algorithms for on-line processing and classification. Since wearable systems are used outside the laboratory, the recorded signals tend to be corrupted with additional noise that influences the precision of classification algorithms.

In this thesis we present the research on a wearable sleep / wake classifier system that relies on cardiorespiratory (ECG and respiratory effort) and activity recordings and that works autonomously with minimal user interaction. This research included the selection of optimal signals and sensors, the development of a custom-tailored hardware demonstrator with embedded classification algorithms, and the realization of experiments in real-world environments for the customization and validation of the system. The processing and classification of the signals were based on Fourier transformations and artificial neural networks that are efficiently implementable into digital signal controllers.

Literature analysis and empiric measurements revealed that cardiorespiratory signals are more promising for a wearable sleep / wake classification than clinically used signals such as brain potentials. The experiments conducted during

this thesis showed that inter-subject differences within the recorded physiological signals make it difficult to design a sleep / wake classification model that can generalize to a group of subjects.

This problem was addressed in two ways:

First by adding features from another signal to the classifier, that is, measuring the behavioral quiescence during sleep using accelerometers. Conducted research on different feature extraction methods from accelerometer data showed that this data generalizes well for distinct subjects in the study group.

In addition, research on user-adaptation methods was conducted. Behavioral sleep and wake measures, notably the measurement of reactivity and activity, were developed to build up a priori knowledge that was used to adapt the classification algorithm automatically to new situations.

This thesis demonstrates the design and development of a low-cost, wearable hardware and embedded software for on-line sleep / wake discrimination. The proposed automatic user-adaptive classifier is advantageous compared to previously suggested classification methods that generalize over multiple subjects, because it can take changes in the wearer's physiology and sleep / wake behavior into account without adjustment from a human expert.

The results of this thesis contribute to the development of smart, wearable, bio-physiological monitoring systems which require a high degree of autonomy and have only low computational resources available. We believe that the proposed sleep / wake classification system is a first promising step toward a context-aware system for sleep management, sleep disorder prevention, and reduction of fatigue.

Keywords: Sleep / wake classification; wearable devices; cardiorespiratory signals; neural networks; adaptive systems; human centered systems; context awareness; sleep management.

Zusammenfassung

Schlafstörungen beeinträchtigen das menschliche Wohlbefinden und die Gesundheit. Die steigenden Kosten zur Behandlung schlafabhängiger chronischer Erkrankungen in den Industrieländern erfordern eine effiziente Prävention. Günstige, tragbare Systeme zur Schlafüberwachung, welche den Benutzer über sein Schlafverhalten aufklären, könnten dem Auftreten von gesundheitsschädigenden Schlaferkrankungen vorbeugen.

Leider ist es noch nicht möglich, alle wertvollen bio-medizinischen Parameter mit tragbaren Systemen aufzuzeichnen und zu klassifizieren. Vielfach sind die Sensoren dem Träger hinderlich und die Prozessorleistung zu eingeschränkt. Ausserhalb des Labors sind die Signale oft sehr verrauscht, was die Qualität von Klassifikationsalgorithmen stark beeinträchtigen kann.

Diese Dissertation beschäftigt sich mit der Entwicklung eines tragbaren, autonomen Schlaf / Wach Detektors basierend auf der Messung von kardiorespiratorischen Signalen und Aktivität. Es werden die Auswahl und Integration von Sensoren und Elektronik, die Entwicklung eines angepassten Algorithmus zur Klassifikation und die nötigen Experimente zur Validierung des neuartigen Systems beschrieben. Zur Klassifikation der EKG-, Atmungs- und Bewegungssignale wurden spektrale Merkmale mittels Fourier Transfomation ermittelt und anschliessend mit Neuronalen Netzwerken klassifiziert.

Individuelle physiologische Differenzen zwischen verschiedenen Personen erschwerten die Entwicklung eines generell anwendbaren Klassifizierungsalgorithmus. Deshalb wurde zusätzlich zu den physiologischen Daten auch das Bewegungsmuster der Probanden verarbeitet.

Um den Algorithmus anpassungsfähiger zu gestalten, wurden zwei zusätzliche Messungen eingeführt, welche zwei typische Schlaf / Wach Verhalten (Reaktion und Aktivität) erfassen. Diese Messungen erlaubten in regelmässigen Ab-

ständen, den Klassifikationsalgorithmus automatisch auf neue Benutzer oder andere Veränderungen anzupassen.

Die Resultate dieser Dissertation tragen zu neuen Entwicklungen im Bereich von intelligenten, tragbaren bio-medizinischen Geräten bei, welche auf einen geringen Stromverbrauch und Rechenleistung angewiesen sind.

Schlüsselwörter: Schlaf / Wach Klassifikation; anpassungsfähige Systeme; Menschzentrierte Systeme; Künstliche Intelligenz; Neuronale Netzwerke; Schlafmanagement

Für Frieda und Klemens,
welche das Erscheinen dieser Arbeit nicht mehr
miterleben durften.

Contents

List of Figures

List of Tables

Acronyms

ACC Accelerometer (spectral feature extraction)

ACT Activity Count (feature extraction)

AGE Analog Genetic Encoding

ANN Artificial Neural Network

BAN Body Area Network

BP Blood Pressure

BT Bluetooth

DSP Digital Signal Processor

ECG Electrocardiogram (spectral feature extraction)

EEG Electroencephalogram

EMG Electromyogram

EOG Electrooculogram

FFT Fast Fourier Transform

HR Heart Rate

HRV Heart Rate Variability

MAC Multiply-accumulate

MIPS Million instructions per second

NREM Non-Rapid Eye Movement Sleep

PAT Peripheral Arterial Tone

PCA Principal Component Analysis

PCB Printed Circuit Board

POS Position (feature extraction)

PSD Power Spectral Density

PSG Polysomnography

PVT Psychomotor Vigilance Test

PW Pulse Wave

RAM Random-access Memory

REM Rapid Eye Movement Sleep

ROC Receiver Operating Characteristic

RSP Respiration (spectral feature extraction)

RTC Real Time Clock

SC Skin Conductance

SEM Sensor Electronics Module

SpO2 Blood Oxygen Saturation

TST Total Sleep Time

VLSI Very-Large-Scale Integration

1   Introduction

"Snoring can endanger your health" (2009) public campaign of the Swiss Pulmonary League.

1.1   Motivation

Although humans show similar sleep behaviors, their sleep habits differ: Humans go to bed at different times, they have preferences for different sleep environments and they need different amounts of sleep to be rested in the morning. But sleep habits do not only change between persons, they also change within a person's lifetime. Newborns sleep 80% of the time. An average young adult

however, sleeps between 7 and 9 hours a day and as he or she gets older, this time is further reduced.

Whether sleep is important and necessary for life is for researchers still an open question (Cirelli and Tononi, 2008 and Siegel, 2005). Sleep plays an important role in quality of life of humans (Edinger and Means, 2005). How relevant sleep really is to humans and their lifestyle becomes obvious after a sleepless night. One feels more fatigued, one has to fight to stay awake at certain times of the following day, the mood is changed[1], the attention is reduced (Dinges, 2004), one lacks motivation, bigger risks and less objective decisions are taken (Venkatraman et al., 2007). Most people will not try to stay awake for multiple nights in a row, but if one does so, hallucination and daydreams start and one encounters the so called 'microsleeps'.

Researchers link the involuntary lack of sleep and sleep disorders to many common diseases in industrialized countries like cardio-vascular diseases, diabetes and obesity (Committee on Sleep Medicine and Research, 2006). With the assumption that 10 - 30 % of the US citizens have insomnia, the estimated direct annual costs (medication and health care) are around 14 billion US Dollars (Walsh and Engelhardt, 1999) and the indirect costs (reduced productivity, increased work and traffic accidents, increase of other medical illnesses) could easily reach 100 billion US Dollars (Committee on Sleep Medicine and Research, 2006). Whereas large efforts were put into discovering the effects of sleep loss and increased fatigue in the field of transportation and into the automatic screening and detection of sleep disorders in hospitals, health care professionals only recently become aware that a higher amount of prevention and surveillance in the general population is required (Committee on Sleep Medicine and Research, 2006).

With the recent success of mobile and wearable devices for training and wellness (e.g. pulse or energy expenditure monitors), it becomes apparent that a wearable, cheap and easy to use sleep monitoring or fatigue prediction device could contribute to public health. One can imagine a multitude of intelligent devices that could help to reach better and healthier sleep. For example a system could measure sleep onset times, sleep durations and sleep quality and informing the user of sleep irregularities and trends. Recommending improvements in sleep habits can lead to a more restful sleep. Such a system has the potential to become

---

[1]The positive effects of sleep deprivation on depressed people are used in psychiatry to treat a multitude of depression types without medication and are the most rapid antidepressant available today (Wirz-Justice et al., 2005).

a lifestyle device. The device could also be a primary sleep disorder prevention system that would be more powerful than only passive prevention methods and less expensive than disorder treatments.

1.2 Hypothesis

Advances in building sensors for measuring bio-medical signals, the reduced power consumption of sensors, the increased processing power of electronics and the advances in artificial intelligence for medical applications make it possible to design an intelligent sleep and wake discrimination system which can be accessible to the general public. By an intelligent system we mean a system that can adapt to its environment and does not simply apply some static rules. For example, a device that automatically detects typical physiological characteristics of its user and can associate unknown patterns without calibration to a characteristic.

We estimate that such a sleep / wake based wellness system can only be successful with a large public if it is low cost and easy to use. It also needs to conform to scientific and engineering standards by being safe, robust, reliable (works consistently in any situation), generalizable (works for everyone), sensitive (detects sleep as sleep) and specific (detects wake as wake). This thesis shows that a low-cost, wearable system for sleep / wake discrimination can be built with currently available sensor technology and well known machine learning techniques. As a design paradigm we apply: *Simpler, but smarter and more understandable*. Such a system could then serve as a framework for a multitude of applications that require a comfortable device that combines artificial intelligence for sleep / wake detection with embedded sensors and electronics in a single, wearable system, such as

- Prevention of fatigue of the wearer by detecting lack of core sleep.

- Early detection of sleep troubles and disorders in groups at risk.

- Coaching the wearer using sleep hygiene rules (Stepanski and Wyatt, 2003) to improve their lifestyle.

- Assistant to sportsmen and sportswomen for planning their sleep and regeneration schedules to increase training efficiency.

1.3   Achievements and Contributions

One of the main challenges in this thesis was to render concepts and technology from the fields of biomedical engineering, artificial intelligence and human physiology into a wearable intelligent system.

The main contributions of this thesis are:

- An analysis of signals that are suitable for wearable biomedical applications, in particular the sleep / wake discrimination.

- The research and development of an algorithm for sleep / wake classification relying on physiological signals, designed to fit a low-power system.

- The research and development of a hardware platform embedding the sensors and on-line algorithms for a wearable device. The device is relatively low cost and shows an autonomy of multiple days.

- The research and development of a method for the automatic adaptation of the above system to different users and user situations.

- A comparison with existing wearable methods that rely on behavioral measures of sleep only (actigraphy).

1.4   Structure of the Thesis

Chapter 2 *Sleep, Wake and Fatigue*: Because this thesis is focused on the detection of sleep and wake, in this chapter we will introduce the definition of the sleep behavior. We describe the physiology of sleep and how physiological and behavioral changes in sleep and wake can be measured with available technology. Fatigue and sleepiness are important consequences of bad sleep and are therefore also discussed. We give the state of the art of wearable fatigue detection and management devices.

Chapter 3 *The Problem of Automated, Wearable Sleep and Wake Discrimination*: In this chapter we elaborate the current problems of a wearable sleep / wake classification device. We empirically answer the first pertinent question "Which signals can be measured and which sensors used for a reliable sleep / wake detection with a wearable system?". For this, a portable physiological recording system is used and a qualitative evaluation of the different recordings is made. The cardiorespiratory signals from the electrocardiogram (ECG) and a respiration effort

sensor were selected for further evaluation.

Chapter 4 *Automatic Sleep / Wake Classification*: In this chapter we present an algorithm for low-power mobile sleep / wake classification. It is based on a Fourier Transformation for preprocessing of the ECG and respiration signals. The proposed classifier is an Artificial Neural Network (ANN) with a single-layer, feedforward topology. The performance of the algorithm is initially tested on real-world recordings from a single subject. Then we evaluated the performance of the algorithm when trained on different subjects.

Chapter 5 *From Portable to Wearable*: Because the data used for the experiments in Chapter 3 and 4 was recorded with a portable, and not a wearable device, we had to validate the method also for a wearable system. However, no wearable systems able to record and classify the required signals currently exist. Therefore we had to develop a wearable prototype (*SleePic*). The development of the SleePic system is presented in Appendix B. We then optimized the algorithm for use on the SleePic system. Experiments were conducted with young healthy subjects in a realistic, real-world setting and the subject-specific and subject-independent performance of the wearable classifier analyzed.

Chapter 6 *Activity*: Because the classifiers based on cardiorespiratory features generalize badly for new users and because accelerometer measurements are very practical for wearable applications, we introduce in this chapter three activity measurements. The first is based on the well established actigraphy method by Cole et al. (1992). The second uses body position for the classification. The third uses the spectral analysis of the accelerometer signal for the feature extraction (analog to the physiological signal preprocessing). We then give an in-depth comparison of these methods with the physiological methods using the data from the recordings obtained in Chapter 5.

Chapter 7 *Adaptation*: The introduction of the accelerometer signal could not completely resolve the problem of generalization of the classifier because the physiological signals exhibited too large inter-subject differences. Therefore we introduce in this chapter a new procedure for the automatic and on-line adaptation to new subjects or physiological changes in the known subjects. The presented methods are based on behavioral measures of sleep and wake (activity and reactivity) which are known to generalize well between subjects.

Chapter 8 *Concluding Remarks*: In this chapter we conclude the findings of this thesis and discuss the limitations. We also give an idea about future research and possible applications of the sleep / wake discrimination system.

Appendix A *Physiological Recording Devices*: In this chapter we give an overview

of existing portable and wearable physiological recording devices that can potentially be useful for sleep / wake discrimination.

Appendix B *SleePic Development*: This chapter describes the hardware and software developments for the wearable sleep / wake classification system called 'SleePic' that was used for the experiments in Chapters 5, 6 and 7.

Results that are described in Chapters 4 and 5 have been accepted for publication in the IEEE Transactions on Biomedical Circuits and Systems (Karlen et al., 2009) and the Journal of Artificial Evolution and Applications (Dürr et al., 2009). The content of Chapter 4 and 6 was also presented at peer-reviewed IEEE conferences (Karlen et al., 2007, 2008).

2 Sleep, Wake and Fatigue
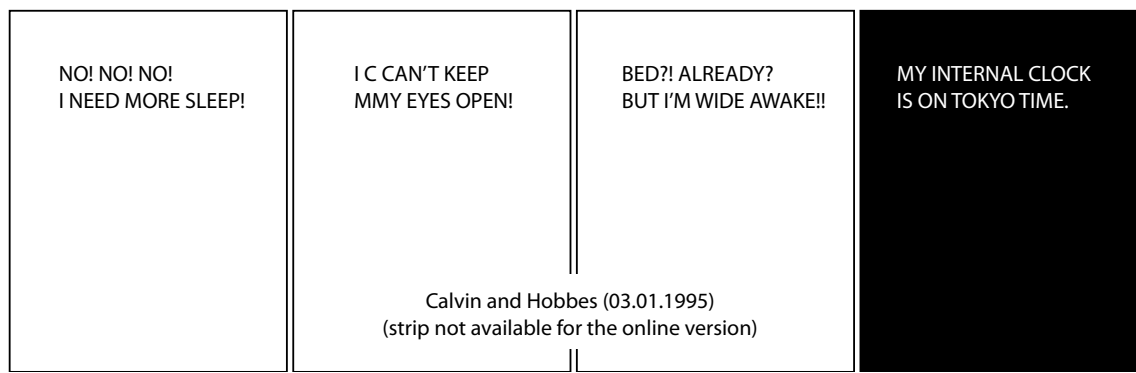
NO! NO! NO! I NEED MORE SLEEP!	I C CAN'T KEEP MMY EYES OPEN!	BED?! ALREADY? BUT I'M WIDE AWAKE!!	MY INTERNAL CLOCK IS ON TOKYO TIME.

Calvin and Hobbes (03.01.1995)
(strip not available for the online version)

2.1 Definition of Sleep

Sleep is primarily a behavior. This behavior is defined by (summarized by Flanigan, 1972):

a) A species-specific body posture;

b) Maintained behavioral quiescence;

c) Elevated arousal threshold; and

d) State reversibility with stimulation.

As Ogilvie (2001) pointed out, this definition is still valid today and can be useful for the detection of sleep. In mammals the measuring of physiological parameters that are correlated and derived from this behavioral differentiation are in

general more convenient. Especially the possibility to continuously measure the electro-physiological correlations make the measurements also more accurate. A description of the physiology of sleep will be given in the next sections.

Although sleep and the transition from wake to sleep were much studied in the last decades, its exact function is not yet clear (Siegel, 2005). The opinions of researchers are split in two whether sleep is essential or not (Cirelli and Tononi, 2008). The main reason for this discrepancy is the lack of having identified a core function of sleep. However, as Cirelli and Tononi (2008) pointed out, sleep is universal in all species and can't be suppressed without consequences in the animal's behavior. Therefore, there might still be a core function of sleep which has not yet been found. Siegel (2005) has summarized the findings about the possible functions of mammalian sleep into energy conservation, neural system recuperation, brain development, emotional regulation and memory actualization. He also mentions that during sleep, brain activity is not much lower than during wakefulness, and therefore sleep should not be seen as a "stand-by" mode of the body solely.

In addition to nutrition, fitness and emotional states, sleep plays an important role for human wellness. Sleep seems to be a behavior complying with the regular need of the human body to rest. This is also expressed by a decrease in the activity of most body parts. Whereas this is true for most body functions, this is not the case for neural activity and should therefore not be generalized.

Achermann and Borbély (1990) and Borbély (1982) suggested that sleep is regulated in the brain by three main processes:

- A circadian rhythm determines the periods of the day with high and low sleep propensity (Figure 2.1a top).

- A homeostatic process, which is defined by the prior amount of sleep, determines the need for sleep (sleep propensity; Figure 2.1a middle).

- An ultradian (cycles of less than 24 hours) process, which regulates the chronology of different sleep stages (Figure 2.1a bottom and Figure 2.2b; see also the explanations in the text below).

The corresponding mathematical models of sleep regulation (Achermann and Borbely, 2003) suggest that sleep pressure increases during waking and is reduced during sleep. The independent circadian process modulates the upper and lower threshold for the homeostatic process (Figure 2.1b). This threshold marks the beginning and ending of sleep pressure accumulation. The three sleep regu-

lating processes are influenced by hormonal, neurological, genetic and environmental factors (Hobson, 2005 and Siegel, 2005). These external factors influence not only the sleep behavior itself, but a multitude of other processes, such as autonomous nerve system (ANS) activity, motor activity, immune system etc. These changes can then be captured by measuring physiological signals.
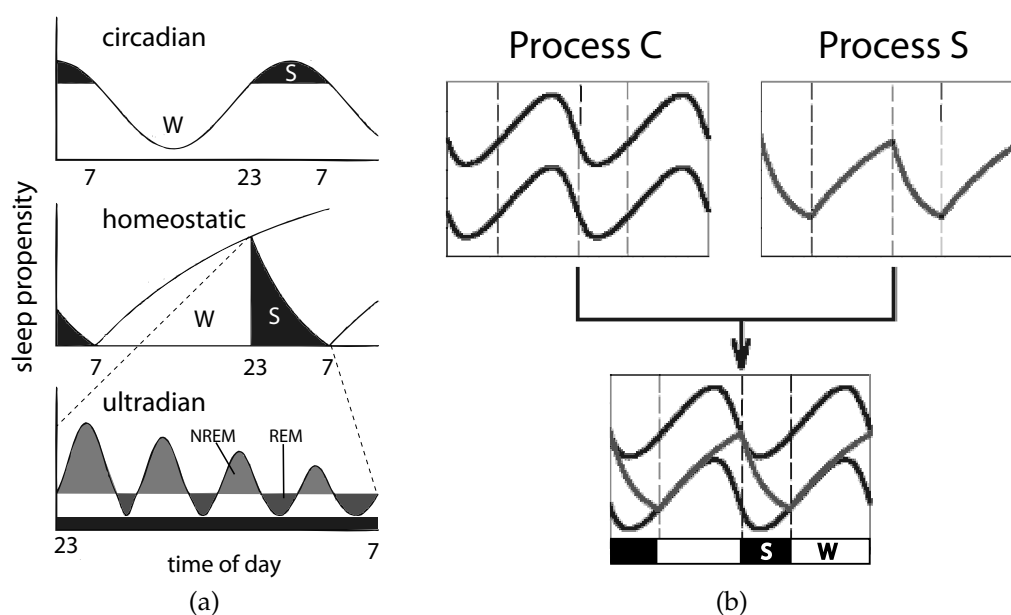
Figure 2.1: a) Three main processes regulate sleep propensity: a circadian (top), a homeostatic (middle) and an ultradian process (bottom). W: wake; S: sleep (black). The ultradian process regulates the duration and intensity of NREM (light gray) and REM (dark gray) sleep. The decrease of NREM correlates with the homeostatic decrease of sleep propensity. b) Model of sleep regulation. The circadian process C, is regulated by the internal clock and modulates the onset and termination of a sleep episode. The homeostatic process S influences the sleep propensity, which as result increases with waking (white bar) and decreases during sleep (black bar). Images adapted from Achermann and Borbely (2003).

Sleep can be separated into two major and clearly separable behaviors (regulated by the ultradian process), the non-rapid eye movement (NREM) and the rapid eye movement (REM). NREM can be subdivided into four more stages (stage 1-4) which are described in the manual of Rechtschaffen et al. (1968). The transitions between the stages are vaguely defined[1] and can be only distinguished by evaluating the proportional content of specific patterns (differing in frequency

---

[1]In the latest sleep scoring definition of Iber et al. (2007) the notion of sleep stage 3 and 4 has been replaced by only one stage 3 because the differences in the EEG were only marginal. Because of the common use of 4 stages this thesis uses the former definition of Rechtschaffen et al. (1968).

and amplitude) in the brain wave measurements (electroencephalogram (EEG); Figure 2.2b left). Stage 1 and 2 are also often referred to "light sleep" and stage 3 and 4 to "deep sleep" or "slow wave sleep" that is characterized by an increased synchronization of brain activity (see also Section 2.2.1). During overnight sleep, as part of the ultradian process, a person undergoes a repeated succession of these states (stages 1-4 and REM) as shown in the example hypnogram in Figure 2.2b. The duration of slow wave sleep is continuously reduced in each cycle and the duration of REM increased (Figure 2.1a and 2.2b). One cycle lasts about 60 minutes in humans. The cycles can be interrupted with awakenings. After wake-up, the subject needs to go again through stages 1-2 to enter slow wave sleep. Frequent awakening (for a multitude of reasons, e.g. noise, obstructive sleep apnea) may limit the amount of deep sleep a person can get and results in poor sleep quality. Depending on the wake-up frequency, this anomaly can become a sleep disorder with severe consequences (e.g. daytime sleepiness).

(a)   (b)
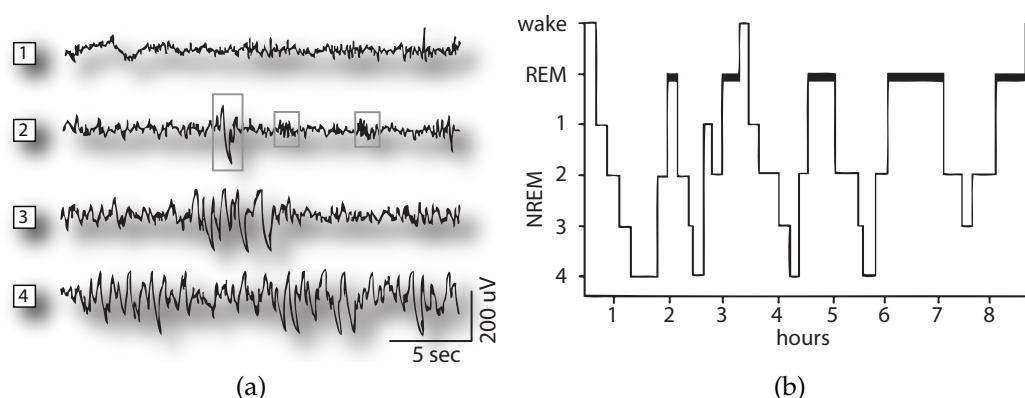
Figure 2.2: a) Brain wave patterns (EEG) associated with the different NREM sleep stages 1-4. The amount of low frequency delta waves (0.5 to 4 Hz) increases with sleep stage (Stage 3 and 4 contain more than 50% of delta waves). A K-complex can be observed in the stage 2 as high amplitude impulse and sleep spindles as two high frequency bursts (squares). b) Hypnogram showing the sleep architecture of a typical night of a healthy subject caused by the ultradian sleep regulating process. An awakening during sleep requires the subject to pass first the light sleep stages (stage 1+2) to fall into deep sleep (stage 3+4). Figures adapted from scholarpedia.org (2009).

2.1.1 Sleep Debt

The human body requires a certain amount of sleep to function effectively. When for whatever reason, the number of sleep hours are reduced, sleep debt is accumulated. Sleep debt is defined as the difference between the hours of sleep a person needs (obligatory core sleep) and the hours of sleep a person actually gets (exceeding the top threshold of the circadian component in Figure 2.1b). It is the cumulative build-up of the sleep pressure (Van Dongen et al., 2003). These lost hours of sleep need to be replaced. Under sleep debt, the sleep propensity is increased. The larger the sleep debt, the stronger the tendency to fall asleep and the more the probability for microsleep (short, involuntary attack of sleep with complete loss of attention) increases (Dinges, 2004). The effects of sleep debt vary periodically with the circadian cycle, but do not disappear by themselves. Sleeping is the only way to reduce sleep debt (Van Dongen et al., 2003).

How much core sleep a person really needs, has not been clearly identified and may vary considerably between subjects (Naitoh, 1992). But it has been hypothesized that it lies around 4-5 hours per day, which can also be taken in short duration sleep segments equally distributed over the day (polyphasic sleep, Stampi (1992)).

The state of being under increased sleep debt is called sleep deprivation. Sleep deprivation is either identified by measuring the time needed to fall asleep or by using questionnaires. The Multiple Sleep Latency Tests (MSLT) are used to measure the time needed to fall asleep in sleep-favorable conditions (in bed, darkness). The Maintenance of Wakefulness Test (MWT) measures the capacity to stay awake when placed in conditions theoretically ideal for falling asleep. Both techniques give indications about the severity of sleep deprivation and the probability for microsleep occurrences.

2.1.2 Sleep Inertia

Sleep inertia is the feeling of sleepiness after awakening and is expressed by a temporary reduction in the ability to perform even simple tasks efficiently. Sleep inertia can last from 1 minute to 4 hours, but typically lasts 5-30 minutes (Tassi and Muzet, 2000). The severity of sleep inertia is dependent on how long the person was asleep and the stage of sleep at awakening (most severe when waking up from deep sleep). Effects can be even more severe if a person is very sleep deprived. However, sleep inertia can usually be reversed within 15 minutes by

activity and noise. Sleep inertia can cause impairment of motor and cognitive functions and can affect a person's abilities, for example to drive safely. Sleep inertia can be very dangerous for people who have to perform risky tasks shortly after waking up from deep sleep or a nap, for example night call service.

Sleep inertia is determined with performance measurements: simple or complex reaction time, grip strength, steadiness and coordination, perception tasks, memory tasks, logical reasoning and a number of cognitive tasks, such as mental arithmetic. It is also hypothesized by Kräuchi et al. (2004) and Tassi and Muzet (2000) that EEG shows a higher synchronization during sleep inertia than in normal wakefulness (reduced alpha (high frequencies) and presence of theta brain waves like in sleep stage 1).

2.2 Differentiation between Sleep and Wake

The transition between the wake and the different sleep stages is fuzzy and its detection not easy (Ogilvie, 2001). Human experts who analyze signal patterns visually can have an inter-rater agreement lower than 80% (Danker-Hopfe et al., 2004). Davis et al. (1937) started to analyze electroencephalograms (EEG) to identify the transitions of sleep. Since then, measurements of (electro-) physiological parameters for sleep discrimination have become more and more popular. Rechtschaffen et al. (1968) published a sleep scoring manual that was used in medical sleep analysis until it was refreshed in Iber et al. (2007). The scoring is based on the measure of brain waves called electroencephalography (EEG), the measure of eye movements called electrooculography (EOG) and the measure of electrical muscle activity called electromyography (EMG) and is widely known as polysomnography (PSG). PSG is the gold standard in medical sleep screening. Most of the time, PSG also includes recordings of other physiological signals to identify sleep disorders.

Signals for sleep and wake discrimination typically have either origin in the central nervous system (EEG, EOG, EMG) or the autonomic nervous system activity (ECG, blood pressure, respiration, skin conductance, temperature). Some observation methods rather rely on behavioral detection of sleep (motor activity, video, pressure systems) which cannot be associated clearly to one of the two nervous system activities. In the following sections we discuss in more detail each sleep recording method and present state-of-the-art discrimination techniques for each method.

2.2.1 Physiological Methods

Electroencephalogram (EEG)

The measurement and frequency analysis of brain waves is so far the most accurate and important sleep / wake discriminative method (Carskadon and Dement, 1989). It has been introduced by Davis et al. (1937) for the human sleep/wake discrimination and allows in combination with EOG and EMG the distinction of wake and sleep stages (1-4 and REM, Rechtschaffen et al. (1968)). The basic difference between wake and sleep stages is observed in the frequency spectrum and amplitude of the brain potentials. These differences are not always clear-cut and it is possible that the behaviorally perceived sleep-onset does not coincide with the changes in the EEG patterns (Carskadon and Dement, 1989). Typical EEG patterns that can be observed during NREM are (see also Figure 2.2a) theta (4 to 7 Hz) and delta waves (high voltage slow waves, 0.5 to 4 Hz), sleep spindles (12 to 16 Hz) and K-complexes (high voltage bursts). Normal wake and REM is mainly characterized by alpha waves (8 to 13 Hz).

However, EEG also displays some crucial difficulties in measuring the electrical activity of the brain. Due to the very low potentials (range of micro Volts) measured on the scalp, it is extremely vulnerable to technical and physiological artifacts. Common sources of artifacts are other and stronger electrical potentials across the body coming from eye movements, heart rate or muscle activity. Motion artifacts and electromagnetic fields are serious noise sources too. Automated analysis of the EEG signals without complex filtering and signal rejection becomes therefore almost impossible. Besides the artifacts and noise, the application of the necessary gel-electrodes (minimum 5) on the scalp is cumbersome and uncomfortable for the subject. As the placement of the electrodes is crucial (location and contact quality), it must be done by a trained professional. The duration of recordings without changing gel electrodes or loosing quality is limited to approximatively 12 hours which makes long-term measurements nearly impossible. All these factors are reasons for recording EEG mainly in controlled environments like hospitals and research laboratories.

EEG (and also EOG and EMG) analysis can be automatized with rule based and machine learning algorithms. An automated analysis is typically composed of (Penzel and Conradt, 2000)

a) a removal of artifacts and noise (filtering step);

b) feature extraction and waveform recognition (preprocessing step); and

c) a classification step.

The filtering step typically consists of removing wave patterns not belonging to the brain activity, such as heart rate, eye movements or electromagnetic noise.

Traditionally, the preprocessing step consists of a spectral analysis of the EEG, either with Fast Fourier Transformation (FFT) algorithms or with more advanced wavelet based transformations (Penzel and Conradt, 2000). Additional features are then extracted to reduce the dimension of data. Typical features can be amplitudes or the power of specific frequencies.

For the classification of EEG signals, a vast number of algorithms have been suggested[2]. A big success has been shown by Artificial Neural Networks (ANN) (review by Robert et al., 2002, 1998) and ANN derived methods, probably because they are relatively easy to implement and theoretically they can solve any nonlinear classification problem.

Electrooculogram (EOG)

EOG is mainly used to detect eye movement patterns associated to the transition of wake to sleep (such as the slow rolling of the eyes) and the REM stage (rapid eye movements). There is no particular work that analyzes EOG signals of different NREM sleep stages. Recordings of EOG show similar problems as EEG, whereas the EOG potentials are larger and therefore easier to measure. However, as the electrodes have to be placed on the forehead and temple (with less precision than EEG), the sensors cannot be hidden and the measurement is more intrusive than EEG for the subject.

Electromyogram (EMG)

In polysomnography, EMG is routinely measured from below the chin to detect the REM sleep phase (low tonic level) and sleep onset (continuous reduction of tonus). However, EMG can theoretically also be measured from other muscles, which is less intrusive for the subject and the changes in the signal pattern may be more pronounced at sleep onset at these locations. The problems of EMG are similar to those of EOG. An automatic analysis of EMG signals uses frequency spectrum and the signal power features.

---

[2]At the time of writing, a search on the PubMed database (PubMed, 2009) gave over 100 publications for the search terms {automated, EEG, sleep, analysis }.

2.2.2 Cardiovascular Measurements

Heart rate (HR) is derived from the electrocardiogram (ECG) and well known to decrease at sleep onset. Pivik and Busby (1996) have shown that HR decreases with adolescents 30 seconds before entering sleep stage 1. The heart rate variability (HRV) is the change of inter-heartbeat intervals and its change during sleep has been intensively studied by many research groups (Bonnet and Arand, 1997 and Burr, 2007 and Scholz et al., 1997 and Shinar et al., 2006 and Telser et al., 2004). HRV is directly linked to the autonomic nervous function, which also varies over different sleep stages. It is common to analyze the power spectrum of a high frequency (0.15 - 0.5 Hz), low frequency (0.04 - 0.15 Hz) and a very low frequency (0.005 - 0.04 Hz) band. Redmond and Heneghan (2006) used similar frequency bands for the automatic sleep stage classification with a probabilistic classifier. However, due to differences in the methods used to calculate HRV, the results are sometimes contradictory (Malik et al., 1996). A wearable application of this technique is difficult, but has been suggested by Bianchi et al. (2006).

Another cardiovascular derived measurement is the peripheral arterial tone (PAT). PAT reflects the sympathetic tone variations of the ANS and can be measured by capturing the change of the pulsatile arterial volume at the finger (Schnall et al., 1999). Like the HRV, the PAT signal has potential to differentiate between REM and NREM sleep (Bresler et al., 2008) and deep and light sleep (Herscovici et al., 2007). The WatchPAT^{TM}-200 (Itamar Medical, Israel; Figure 2.3) is the product integrating the sensor used in the mentioned PAT studies. This watch-like device connecting two finger sensors (PAT and blood oxygen saturation) can also be used to discriminate between sleep and wake (Hedner et al., 2004). However, sleep / wake discrimination is based solely on actigraphy (see Section 2.2.3). The major drawback of this system is that the finger sensors are intrusive and impair the maneuverability of the wearer.

Respiration

An experiment by Naifeh and Kamiya (1981) has shown that the process underlying sleep onset is also closely linked to the regulation of respiration. Trinder et al. (1997) have shown that when respiration was reduced at sleep onset, EEG patterns also shifted from high to low frequency. Redmond and Heneghan (2006) used respiratory-derived features together with a multitude of ECG-derived features for the automatic classification of three sleep states (wake, REM, NREM) in subjects with obstructive sleep apnea. The advantage of using cardiorespira-

Figure 2.3: The WatchPAT^{TM}-200 measures blood oxygen saturation, peripheral arterial tone and activity to discriminate sleep stages and respiratory sleep disorders. It is designed to be used in ambulatory sleep monitoring to detect obstructive sleep apnea, the most frequent sleep disorder. From Itamar (2009).

tory signals lies in the potential use in a portable system, the usefulness of these signals for sleep disorder detection and that the sensors can be applied by inexperienced users. Respiration is either measured by airflow detection or by chest and abdomen efforts. Respiratory measurements are commonly used in PSG to detect sleep disorders.

Temperature

Temperature regulation in humans follows a circadian characteristic. As sleep is also initiated by a circadian process (Section 2.1), a correlated variation is consequently observed. Body core temperature follows a sinusoidal circadian cycle and sleep onset is located where core temperature is decreasing and the wake-up is initiated when the core temperature is increasing (Czeisler et al., 1980 and Gilbert et al., 2004). It is however not clear whether sleep and temperature regulation are two distinctive processes regulated by the same circadian clock or if sleep is controlled by a thermo-regulatory process. Kräuchi et al. (2004) showed that distal skin temperature increased at sleep onset and had a dramatic decrease at wake up (2 degree Celsius).

The problem with these temperature measurement studies is always that rela-

tively low temperature variations are measured that only can be detected when the experiments are conducted under tight laboratory conditions, for example with a constant ambient temperature and a constant activity level of the subject. Outside the laboratory, the temperature variations induced from the environment are similar to the variations detected under strict conditions and can hardly be distinguished. Core temperature is difficult to measure and a reliable measurement is only possible with invasive methods (rectal or intravenous probes) so far.

Skin Conductance (SC)

Hori (1982) showed that the sleep onset time was accompanied by a decrease in skin potential level measured on the palm. In the study of Shiihara et al. (1998) the changes in skin conductance have been associated with the sleep duration and the differentiation of sleep habits in 32 subjects. Despite these findings, it is unclear how specific this measure is to sleep, as similar changes were observed for stress (Picard, 1998). For sleep / wake discrimination, no automated analysis of SC has been proposed for far.

2.2.3 Active and Passive Behavior Based Detection

Questions and Tests (Active)

Active behavioral sleep tests consist of simple tests, randomly distributed in time, which require a response from the subject. Ogilvie and Wilkinson (1984) were asking the question "Are you awake?" and the subject had to respond with a button. Failing to do so was considered as being asleep. It has also been shown that reaction time correlates with the changes in EEG during the process of falling asleep (Hori et al., 1994).

The disadvantage of these active techniques is that they cannot be conducted continuously and the detection resolution corresponds to the (random) sampling interval. Further, the procedure of active behavioral tests may prevent the subject from falling asleep or wake him/her up.

Activity (Passive)

In home environments, where PSG is typically not available, physicians rely on actigraphy for sleep monitoring (Sadeh and Acebo, 2002). In this method, the ac-

celerations of the extremities (typically the wrist) are recorded over several days with a watch-like device using miniature accelerometers and a storage medium. Periods of low activity are later classified as sleep by off-line computer processing. Many different classification algorithms have been suggested (Cole et al., 1992 and de Souza et al., 2003 and Hedner et al., 2004 and Jean-Louis et al., 1996 and Lotjonen et al., 2003 and Sadeh et al., 1994), but often they cannot cope with the problem of misclassifying low activity tasks like reading and watching television or the case where the sensor band is not worn (de Souza et al., 2003 and Sadeh and Acebo, 2002). In particular, sleep time is overestimated for subjects with low daily activity patterns, which may lead to wrong diagnosis. Further, even if actigraphy gives a good estimation of global sleep patterns, it is not sufficiently accurate to detect and diagnose sleep related disorders (Ancoli-Israel et al., 2003 and Sadeh and Acebo, 2002). As activities other than sleeping are associated to low activity, postprocessing of the data is needed (Cole et al., 1992 and Hedner et al., 2004 and Jean-Louis et al., 1996 and Webster et al., 1982). Typically, this induces discrepancies higher than 10 minutes which is not suitable for on-line applications where reactivity of the system is important.

Recently, alarm clocks using accelerometers have been commercialized (Axbo, 2008 and Sleeptracker, 2007). Activity is used to detect the best sleep phase for easy wake-up in a given time window (10 to 30 minutes). However, the devices are only active at night and the clocks cannot calculate sleep duration (see also Section 2.5.3).

The most recent development in actigraphy for sleep detection is the Vivago$^®$ Active multi-functional watch shown in Figure 2.4. The watch detects if the user is active and displays activity histograms on the screen. A newly developed algorithm by Lotjonen et al. (2003) detects sleep on-line. Unfortunately, the algorithm has not yet rigorously been tested and compared to previously presented actigaph algorithms.

Observation (Passive)

Direct observation, or more frequently video analysis, is part of every PSG and complementary to the physiological recordings. Many behavioral effects of sleep can be observed on video like posture, movements, eye movements and muscle relaxation effects (twitching). However, the time resolution of this behavioral observations is limited. The analysis of the video has to be done manually and is less accurate compared to the automated EEG methods (Closs, 1988 and Cole

Figure 2.4: Vivago$^®$ Active. An actigraphy based wellness watch that displays past activity, sleep histograms and statistics on the screen. Image from Vivago (2009).

et al., 1992 and Ogilvie, 2001). Automated video analysis has been suggested by Okada et al. (2008), but the proposed method has only been tested on one child so far.

Pressure Systems (Passive)

The effects of muscle relaxation at sleep onset have been identified in early studies of sleep research (Blake et al., 1939). The inability to maintain body tension during sleep has since been used as a passive indicator of sleep. The effect of relaxation is visible in the EMG and occurs a few seconds (1-25) after the EEG alpha waves have disappeared. The most known application of this sleep behavior detection is the "dead man pedal/switch". Such a pedal has to be maintained during the manipulation of a machine at half level. If the operator is, for whatever reason, loosening control of muscles, the pedal is either released or fully pressed and an alarm goes off. As an example, dead man pedals are installed in every operator stand of Swiss train cabins to detect if the train operator is conscious and alert. In this particular case, the pedal is combined with an active behavioral test, which randomly requires response to an auditory signal.

The main disadvantage of dead man pedal's and similar is the incapability to distinguish sleep from other muscle relaxing/blocking behaviors such as cerebral attacks and show therefore low validity. Further, the continuous or very frequent requirement of holding a pedal or a switch restricts the maneuverability of the subject and limits time resolution.

2.3   Sleepiness, Attention and Fatigue

Sleepiness (or drowsiness) can be defined as a precursor of sleep onset. It can be seen as a process of increasing loss of attention. Fatigue is a phenomenon of physical and psychological tiredness which also includes sleepiness. But fatigue can also have other sources than sleepiness, such as boredom, mental and physical exhaustion, chronic or genetic diseases and others. For a more complete discussion of sleepiness and fatigue, the reader is referred to Lal and Craig (2001) and Shen et al. (2006).

Fatigue and sleepiness have a big effect on people's life style, wellness and health (Schmid et al., 2007 and Spiegel et al., 2005, 2002). They have also been identified as a major source for accidents in transportation and hazardous operator settings (Akerstedt, 2000 and Dinges, 1995) (see also Chapter 1). The reason for fatigue detection technologies were precisely summarized in Dinges and Mallis (1998). The subjective fatigue estimations are known to be unreliable and because the technology advances have made the detection of fatigue possible, the use of technology is an alternative to the regulation of work and rest hours.

Sleepiness is measured with attention test such as the very well established Psychomotor Vigilance Test (PVT) developed by Dinges et al. (1985). Increased sleepiness can also be observed in the presence of slow wave patterns in the EEG (similar to sleep stage 1), which is normally during wake dominated by alpha wave patterns (review by Lal and Craig, 2001). Note that the subjects' motivation has an impact on the subjective estimation of sleepiness and fatigue and it also influences the outcome of sleepiness tests (Limor et al., 2009). Therefore, a sleepiness or fatigue evaluation device should also include sleep history, which could qualify sleepiness more objectively.

2.4   Fatigue Management

As fatigue and the associated increased sleepiness level correlates with a higher accident risk of persons, in many industrialized countries federal laws regulate the work, rest and/or sleep times for workers in risk domains. This can be a protection for them when operating in a hazardous environment (as it is typically the case for shift workers) and also a protection of the persons in the operator's environment. In transportation, they can be other road users or passengers of public transport. However, regulating the hours of sleep and work does not necessarily guarantee that the worker can maintain a high attention for the whole duration

of his work. For example, alcohol intake or a severe sleep disorder can reduce the recovering power of sleep and the worker cannot reach the required attention level for his task.

In the last two decades, federal safety boards, insurance institutions and the involved industries started to search for more effective ways to measure and manage fatigue of workers. This lead to a significant amount of research and product developments in the fatigue management field. Dinges and Mallis (1998) classified the fatigue detection and prediction technologies into four classes:

a) Readiness-to-perform technologies (or fitness-for-duty);

b) Vehicle-based performance technologies;

c) In-vehicle, on-line, operator status monitoring technologies; and

d) Mathematical models of alertness dynamics, joined with ambulatory technologies.

A short summary of these technologies is presented in the next sections. For an extended review of these technologies and existing systems we refer to Barr et al. (2005) and Hartley et al. (2000) and Williamson and Chamberlain (2005).

2.4.1 Readiness-to-perform Technologies

Fitness-for-duty or readiness-to-perform approaches consist of performance or physiological tests, which evaluate the level of alertness and vigilance of a subject at a given time and extrapolate this to an upcoming time frame. The tests seek to detect impairment due to a sleep loss or active substance intake and in case of incapacity alert the subject or the supervisor before the work is performed. The tests can consist of psychomotor tests, eye tracking or combination of these (overview in Hartley et al., 2000).

The tests are relatively easy to conduct and can be done with portable devices. The detection of low attention due to a lack of sleep is quite good. However, they show a major drawback: The tests give only a snapshot of the subjects' alertness at a given time and the predictive capacity of the tests is very limited. They do not sufficiently account for the subjects' increase of fatigue over time, which results from task complexity, a circadian low or increased sleep debt.

2.4.2   Vehicle-based Performance Technologies

These technologies are measuring the interaction behavior of the subject with a machine that he or she is supposed to control. For example in an automobile, the steering behavior, speed changes, lane deviations etc. can be measured. They are hypothesized to be altered when a driver is fatigued as compared with his normal driving behavior.

Such systems are commercially available and are also on the way of being implemented into serial produced cars and trucks. The validity of these systems is not resolved completely because a change of driving behavior can have other sources than an increased fatigue. For example, the vehicle can perceive a different steering activity because it moved from an urban to a rural environment or the user adapted to modified driving conditions like a wet road or the driver of the vehicle changed at the last stop. Even if valid measures can be demonstrated, this technology is not interesting for a wearable life-style device, because it can only measure the fatigue level when the subject is operating a machine or performing a regular task.

2.4.3   In-vehicle, On-line, Operator Status Monitoring Technologies

This category of technologies include all trials to measure fatigue from physiological and bio-behavioral measures. This includes typically the on-line measure of brain activity, heart rate, eye movements, eye blinking, head nodding, and skin conductance. These techniques are often not very sensitive (not all drowsy events can be detected) and like the previously described techniques the measurement is hard-linked to a setup. In particular cameras need to be fixed externally. An interesting wearable approach is the mounting of an infra-red distance sensor on an eyeglass frame to measure eye blink velocity by Johns et al. (2005) (see also the Section 2.5.2), however this technique has not been validated sufficiently yet. One of the most successful and intensively researched techniques is the "Percentage of eyelid closure" (PERCLOS; Dinges and Grace, 1998), which measures the duration of the eyelid closure of subjects (closed more than 80%). PERCLOS is measured with an external camera that observes the operator's face. The "dead man pedal" and other switches described in Section 2.2.3 belong also to this category of devices.

2.4.4 Mathematical Models of Alertness Dynamics Joined with Ambulatory Technologies

This approach involves the application of mathematical models that predict the subject's alertness at regular intervals based on sleep, circadian rhythm and other fatigue related information. The relevant source information is either obtained from devices that automatically monitor sleep or from log books. One can argue that such systems represent a certain danger: A favorable output of such algorithms could mislead the subjects to increase their own work time. Like every model, it is only as good as the input values. Especially if the source information is imprecise and overestimates for example sleep duration, the error propagates through the model equations and the output may overestimate the subject's readiness to perform the given task.

The US Army Sleep Management System

The *US Army Sleep Management System* is a mathematical model based on the effects of sleep deprivation periods and circadian cycles (Belenky et al., 1998). It calculates the degradation of performance by prolonged waking and the restoration of it by sleep. The model has been embedded into a wristwatch that measures activity. A sleep detection algorithm based on the detection of low activity pattern (Cole et al., 1992) calculates sleep times, which are used by the model for the alertness prediction. We discuss the device in Section 2.5.4. A major drawback is the static model. A static model means that the watch does not distinguish between different users and will always indicate the same sleepiness level for a given sleep history, independent of the user's regular sleep durations and its circadian phase shift. A successor of this model is called "Sleep, Activity, Fatigue and Task Effectiveness". It has been developed by the U.S. Department of Defense (Balkin and McBride, 2005) and includes an additional model that allows the prediction of sleep inertia.

The Three-Process Model of Alertness

The *Three-Process Model of Alertness* has been developed by Akerstedt and Folkard (1995) and predicts alertness on a scale from 1 to 16. It is inspired by the model of sleep regulation developed by Borbély and Achermann (1999) (see Section 2.1). A measure of sleep time is used as input to compute the effects of a homeostatic process and is then added up with a circadian process (Figure 2.5). Sleep inertia

is added to the model as a third process. Besides alertness the model can also predict the sleep latency (time to fall asleep). The model has been used by transport companies to analyze and plan shifts schedules for their drivers. With these experiences the model has continuously been updated (Åkerstedt et al., 2008), but so far it has not been used in a mobile device. The interest in this model lies in its proximity to the real underlying processes of sleep and sleepiness. However, it does not take into account task dependent effects of fatigue (like monotony). To calibrate this model, a measure of the circadian component is needed. The component is heavily subject (and timezone) dependent and is difficult to measure non-invasively with current sensor technologies.
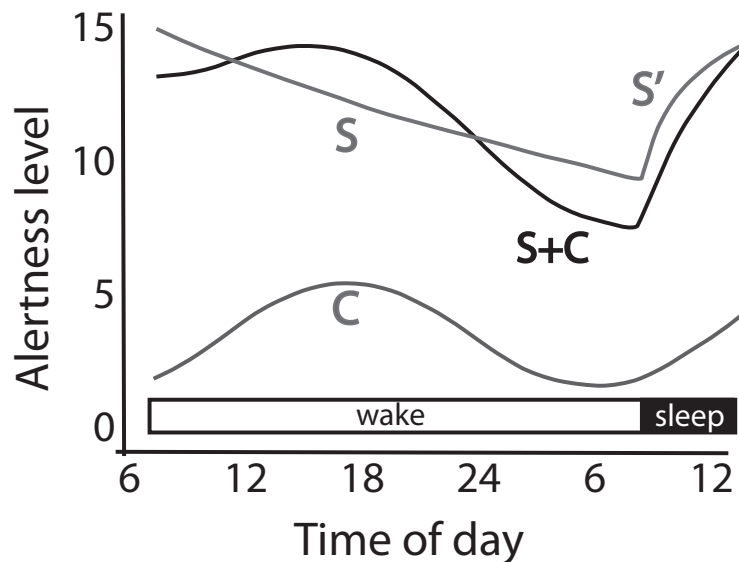
Figure 2.5: Sleep (S, S′) and circadian (C) components of the Three Process Model of Alertness. The third process of sleep inertia is not shown. The arithmetical sum S+C equals to the predicted alertness. Adapted from Åkerstedt et al. (2008).

2.5 Related Projects and Products

Only few existing wearable systems offer the functionality of detecting automatically the wearer's fatigue or sleep over short or long term. In this section we present devices/projects which have a minimum degree of autonomous decision making and are (potentially) wearable. An overview of portable and wearable systems which are intended for recording physiological signals can be found in Appendix A.

2.5.1 Alertness Monitoring Helmet's

A concept suggested by Kaefer et al. (2003) is to measure heart rate, head activity, eyelid closure and reaction time (by means of an auditory reaction test) with sensors mounted on a helmet. An "Alertness Monitoring and Control Unit" would integrate the signals from the sensors, perform the measures and output alertness information. Figure 2.6 shows an idea of a prototype. The project never exceeded the concept phase.

<div align="center">(a)   (b)</div>
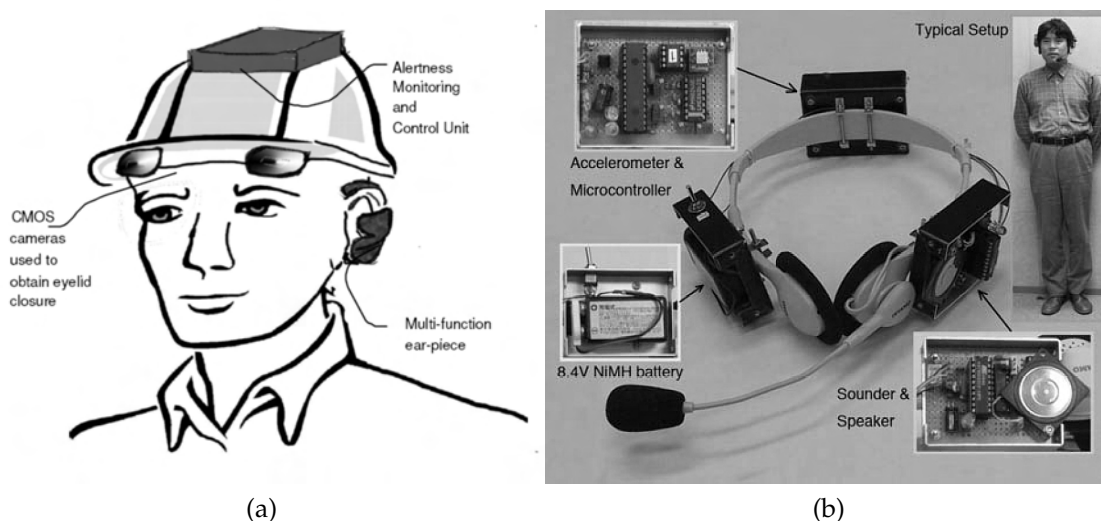
Figure 2.6: Examples of alertness monitoring helmet concepts: a) Alertness monitoring helmet concept suggested by Kaefer et al. (2003); b) Alertness monitoring earphones concept suggested by Matsushita et al. (2006).

A similar approach is the head nodding detector from Matsushita et al. (2006). It is composed of a headset with an on-board 2-axis accelerometer and fatigue is estimated based on typical head nodding patterns when the user stands still for 30 seconds. These regular tests seem to disturb the subject's regular task seriously. An evaluation of the device by giving detection accuracy or other performance parameters is missing.

2.5.2 Optalert

Optalert^{TM} (Optalert, 2009) are glasses on which an infrared proximity sensor is mounted. The sensor measures eye blink amplitude and eye blink velocity. The glasses connect over USB to a board computer that classifies the eye blink features into a custom drowsiness scale. The system alerts the operator with auditory and

visual alarms if the drowsiness increases above a certain threshold. So far, an in-depth validation and scientific analysis of the system is missing (Johns et al., 2005). This device is mentioned here, because the drowsiness classification could be easily embedded in a wearable device, which would make a general fatigue alarm device for a multitude of situations. However, glasses are a limited wearable system because they are not worn during sleep.

Figure 2.7: Optalert™alertness detection glasses. IR proximity sensors are measuring the velocity of the eyelid movements to estimate the alertness of the wearer. From Optalert (2009).

2.5.3   Gentle Wake-Up Devices

Two tiny, wrist worn devices called *Axbo*® and *Sleeptracker*® (Axbo, 2008 and Sleeptracker, 2007) appeared recently on the market, which can be considered as wearable sleep inertia reduction devices (Figure 2.8). They are designed to wake up the wearer at the 'ideal time' in the morning where he/she can get up easily without being sleep drunken (see Section 2.1.2 sleep inertia). These 'gentle wake-up' devices are functionally similar to actigraphs , which consist of a watch-like device that measures the acceleration of the wrist of the subject with tiny accelerometers (see also Section 2.2.3). They have a basic on-line processing capability. In a user defined wake-up time window they check if the wearer is (almost) awake. If a limb movement event is detected with the accelerometer, the wearer is considered as awake. In such an event, the alarm goes off instantly to facilitate the wake-up of the user. The accuracy and usefulness of such life-style devices is controversial and does not work for all users equally. Most importantly, they are not able to detect whether a user is truly asleep.

We tested both devices inside the laboratory and users were unsatisfied. Com-

mon remarks were that the devices wake up too early (the Axbo has a fixed wake up window of 30 minutes) and that the wake-up is not gentle at all.

Independent of the validity of both devices we can conclude that they have only limited interest for a wearable fatigue prediction system because they reduce only a sub-phenomenon of fatigue which is the sleep inertia. Sleep inertia is very variable within persons and lasts under normal conditions between 5 and 30 minutes after wake up.

(a) Axbo® alarm clock   (b) Sleeptracker®

Figure 2.8: 'Gentle wake-up' devices to increase alertness after wake-up based on accelerometer recordings. From Axbo (2008) and Sleeptracker (2007).

2.5.4 US Army SleepWatch

The so far most interesting device that was described and tested in Dinges et al. (2005) is the *SleepWatch®*. The SleepWatch® is a wrist-worn actigraph that embeds an algorithm for sleep management. This "Sleep Management Model" has been developed by the Walter Reed Army Institute of Research (see Section 2.4.4). It estimates the users' sleep in 1-minute intervals based on the actigraphy algorithm from Cole et al. (1992) that is described in more detail in Section 6.1. Based on the sleep / wake patterns collected over several days the Sleep Management Model computed the total-sleep-time and the sleep-amount-per-day to

output a performance index between 0 and 100%. The index is an estimation of the wearer's performance (or performance-readiness) and displayed as a "fuel gauge" on the watch screen.

As this watch uses a model which is based on actigraphy inputs that are known to overestimate the sleep duration (Section 2.2.3) and no other sensors are used to compensate this, it is suspected that the performance prediction might be largely influenced. Further, as the Sleep Management Model is dependent on circadian parameters and these vary largely between subjects, it is not clear how the watch can cope with these variations.

The only scientific study that was analyzing the SleepWatch$^\circledR$ (Dinges et al., 2005) tested the effects of the watch to the sleep behavior of truck drivers. The study did not assess the validity and accuracy of the devices and also did not apply any of these criteria for the selection of the devices to study. The device is not commercially available from the manufacturer.
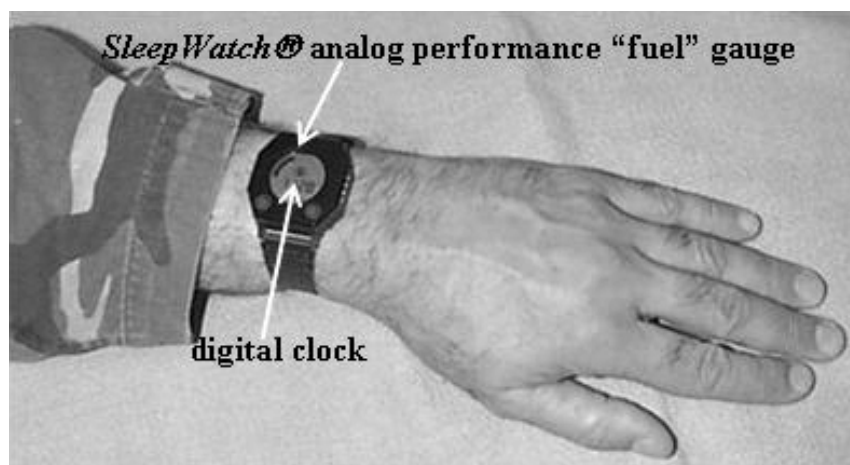
Figure 2.9: SleepWatch$^\circledR$ device for the prediction of alertness indicated with a "fuel" gauge. From Dinges et al. (2005).

2.5.5   European Project SENSATION

SENSATION (Sensation-IST, 2009) is an European project that aimed to develop new sensor technologies for the real-time monitoring and prediction of wakefulness and fatigue. One of the main focuses was the development of nano- and micro-sensors for brain and autonomic function monitoring. The resulting sensors for brain activity monitoring are based on carbon-nanotubes and can be used without gel which is a great step toward wearable EEG electrodes (Ruffini et al.,

2007). A product called ENOBIO$^{\circledR}$ is using this technology and is in its launching phase for commercialization (Starlab, 2009). However, it is far away from being nano or micro sized (Figure 2.10, the price is estimated to 6000 CHF and the duration of use is limited to 6 hours. A rigorous scientific validation is missing so far and the implementation of the complete sleepiness detection system requires a PC and is therefore not wearable at all.

Figure 2.10: ENOBIO$^{\circledR}$ gel-less sensor system for EEG measurement and on-line sleepiness prediction developed during the European project SENSATION. From Starlab (2009).

2.6   Context Awareness and Personal Health

Recently, research groups started to focus on human centered, context aware computer systems. So called pervasive systems are usually embedded into the user's environment and make assumptions about the user's situation in real time and respond appropriate and in an adaptive way to these situations. These concepts were extended to wearable devices to facilitate the all-day life of the user in diverse application fields such as communication (Biever, 2004 and Schmidt et al., 1999), education (Sung et al., 2005), manufacturing (Stiefmeier et al., 2008), navigation (Wilson et al., 2007), nutrition (Amft et al., 2004) and health care (Bonato, 2005 and Lukowicz, 2008) by giving appropriate on-line feedback to the user with a mobile device.

The application of context aware, personalized systems to health care and biomedical applications is particularly interesting because it can potentially solve

one of the main challenges of heath care in industrialized societies: Rapidly increasing costs of screening, prevention of chronic diseases and the assistance to elderly people (Korhonen and Saranummi, 2007 and Martin et al., 2000). Further, because of age, gender and physiological differences in patients, medical analysis and assistance often needs an adapted solution to the user and should not only be given in specialized centers, but everywhere. While an intelligent wearable system can provide a continuous motivation for the individual to recover[3], it is clear that the contact with human experts and caregivers will still be necessary and should not be underestimated. Several European projects (under the activity of e-health) took up the idea of continuous monitoring and personalized diagnosis, therapy and feedback for a multitude of diseases such as cardiovascular (Amft and Habetha, 2007 and MyHeart-IST, 2009), stress and fatigue (Sensation-IST, 2009, see also Section 2.5.5) or just monitoring of vital signs (Wealthy-IST, 2009). Another approach aims to assist caregivers with patient personalized tools (PIPS, 2009). All these projects have in common to rely on some kind of wearable/portable system to sense and process localized information and to transmit it further to centralized databases where human experts can access it and take decisions. This means that the personal profiles are not generated locally on the device, but with the help of professional support externally and are mostly transmitted over the Internet. This raises concerns for privacy and dependability of service providers, which have not been solved yet. A system with locally embedded intelligence may reduce these concerns and can involve the subject to the health process in a higher degree.

Two categories of electronic wearable devices were commercially successful so far (Gerasimov, 2003): Personal electronic assistive devices (watches, photo cameras, etc.) and personal electronic monitoring devices (pacemakers, holter systems, etc.). A third emerging class of devices is in between these two categories and usually has the look of a watch or a mobile phone, but can display and collect some additional information for example pulse rate, geo-location or directions. Extending this category of inexpensive consumer electronic devices to personal health systems is probably the most promising way to reach user acceptance.

Particularly interesting is the detection of sleep and wake states and related fatigue of the user with a personalized wearable system because of the pronounced

---

[3]It is considered that the chronic diseases like cardiovascular, stress, asthma, etc. are largely linked to the person's lifestyle. If lifestyle can be altered with education and coaching, the probability of recovery is increased and the cases of recurrences reduced.

difference of sleep behavior and physiology (see Section 2.1). Hospitals and sleep centers have high costs and have already long waiting lists. It is estimated that 30% of the general population exhibit a sleep disorder but the majority are not aware of having such or do not seek help (Committee on Sleep Medicine and Research, 2006). With the current infrastructure and screening costs, it is not possible to identify and take care of all the potential patients. The lack of sleep represents a serious health problem (Committee on Sleep Medicine and Research, 2006). The increased daytime sleepiness causes fatigue and has severe consequences on the road safety (Akerstedt, 2000). A comfortable and lightweight wearable system may tackle parts of these problems. For example by detecting sleep troubles early (prevention of causes) or by detecting lack of sleep and increased fatigue (prevention of effects).

2.7 Wearable Fatigue Prediction Approach

If we want to continuously assess a fatigue or sleep profile of persons for a healthier lifestyle and risk reduction, a wearable solution which does not influence the persons' comfort is a favorable approach. From the analysis of the previous described fatigue management classes it becomes clear that only a model of sleepiness combined with an accurate measure of the model's input (typically sleep and circadian parameters) can be integrated into an embedded wearable device. Whereas the use of such a technology might be a good solution for the prediction and prevention of fatigue, it is not recommended to use it as alertness system for emergency sleep detection in risky environments. For this, an on-line, machine-dependent operator sleepiness detection system such as the techniques described in Section 2.4.2 and 2.4.3 would be more appropriate. However, it is not excluded that signals measured for the wearable sleep and fatigue prediction can also be used for the alarm generating system.

To conclude, a system for an accurate sleepiness prediction requires accurate information about previous sleep times, sleep durations and sleep quality. Amongst others, in this thesis the problem of measuring accurately sleep duration and time with a wearable system will be addressed.

3 The Problem of Automated, Wearable Sleep and Wake Discrimination
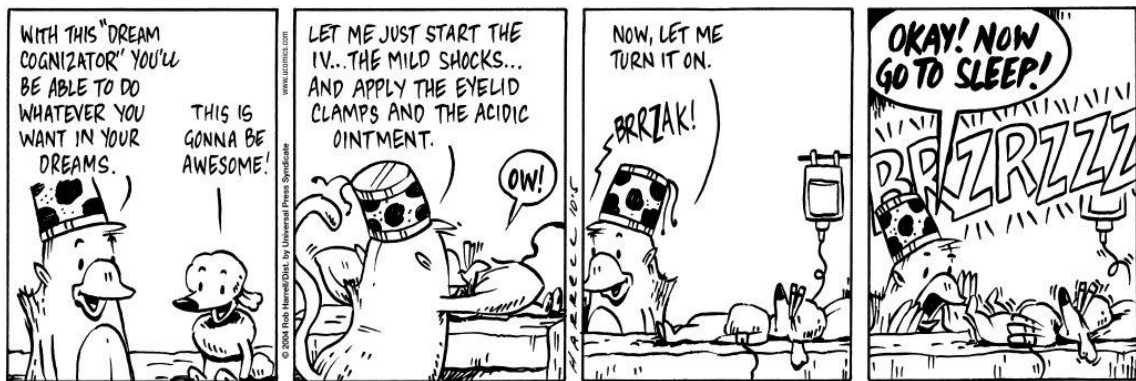
Big Top © 2004 Rob Harrell. Used by permission of Universal Press Syndicate.

The sleep discrimination methods based on physiological signal recording described in Section 2.2.1 are normally carried out in controlled hospital or laboratory environments and need medical assistance for setting up sensors, monitoring and analysis. Although the analysis is typically computer-assisted (Penzel and Conradt, 2000), it still requires a sleep expert and is therefore expensive and time consuming. It is difficult to integrate PSG sensors into a wearable device, as they are rather bulky, power-consuming and highly susceptible to noise. Furthermore, EEG, EOG and EMG recordings require many electrodes to be glued to the scalp and face, which make them very cumbersome and uncomfortable for the user.

Portable and wearable devices and sensors exist that allow the recording of one or more physiological parameters and that are potentially useful for sleep discrimination (Appendix A).

None of the algorithms for automatic sleep discrimination that were presented in Section 2.2 have been tested on data obtained from wearable devices. This and the absence of devices for automatic sleep / wake discrimination based on physiological signals arise the following research questions:

- Which signals can be measured with a wearable device and have enough quality for sleep / wake discrimination?

- Can an algorithm be designed to discriminate on-line sleep / wake states on a wearable device?

- Can the use of physiological signals overcome the limitations of the behavioral measures of actigraphy or dead-man switches/pedals?

- How do inter-individual differences in physiology influence the classification accuracy with respect to the wearable recordings?

- Can such a wearable device be used in general public applications?

These are the main research questions that will be addressed in this thesis. The wearability of the system will be the crucial, but also the limiting factor. If the system is intended to be used in preventive health and wellness applications as has been suggested in the introduction, it has to be low cost, comfortable and comprehensive for the general public. The drawback of wearable systems it that they cannot reach the performance of immobile biomedical clinical systems. They have reduced processing power and the energy resources are limited. More importantly, the quality (signal-to-noise ratio, artifacts, sampling rate, etc.) is dramatically reduced in wearable sensors or some signals cannot be measured at all.

The question if signals can be recorded with wearable systems in an ambulatory setup without having dramatic consequences on the sleep / wake classification is discussed in the next sections.

3.1   Useful Signals for Wearable Sleep and Wake Classification

It is not clear from the literature, which signals do contain enough accurate information about sleep and wake, when they are recorded with a non-clinical device and which signals are consequently suitable for a wearable sleep detection device. To answer this question, we decided to record a multitude of signals during

sleep and wake states of a human subject using a commercially available and certified portable recording system.

In a first step, it was not possible to record all potentially interesting signals with a portable device. Therefore, the most interesting signals were selected according the criteria of wearability (sensor integration), clinical accuracy and power consumption.

A literature and market based pre-study showed that the acceleration measurements are the most efficient solution for a wearable system. Accelerometers are small, highly integrated and low-power and can be worn by virtually everybody. As the use of actigraphy for sleep research is well established, there was no need to investigate it further with our preliminary recordings.

EEG measurements are seen as the signal patterns with the most accurate sleep / wake prediction capabilities. However, the difficulty to render the EEG sensors wearable and low power make them not adapted for a wearable device (see Section 2.5.5). We therefore did not select the EEG for the empirical measurement tests. We also resigned the measurement of blood oxygen saturation (SpO2), because it had little association to sleep, the measurement of peripheral arterial tone (PAT) because it was used to discriminate only between sleep disorders so far and this type of sensor technology is difficult to obtain, and the measurement of blood pressure (BP), because it cannot be measured continuously and non-invasively at the same time.

We decided to empirically determine the signal quality (noise susceptibility) of the ECG, EOG, EMG, respiration effort (RSP), SC, and distal skin temperature signals when measured with a mobile device in a real-world setup. We also analyzed the correlation with the sleep/wake behavior determined by a human expert from simultaneous video, EMG and EOG recordings. The comfort of the sensors for recording the above mentioned signals was also evaluated.

3.1.1   Recording System

The portable recording system that was used is called "Heally" and is produced by Koralewski Industrie Elektronik, Celle, Germany (Koralewski.de, 2008). The Heally system is a medical certified system that uses different recording modules. It has previously been used by the European Space Agency (in a long-term project on the International Space Station (ISS) called Health Lab/Neurolab) and in other stress assessment studies in the field (Johannes et al., 2008). The system that was used for our recordings consisted of three modules (Figure 3.1):

Signal	Abbreviation	Sampling Rate [Hz]	Resolution [bits]
Electrocardiogram	ECG	100	12
Electromyogram	EMG	200	12
Electrooculogram	EOG	200	12
Respiration	RSP	50	12
Skin conductance	SC	10	12
Finger temperature	Temp	2	12
Pulse Wave	PW	100	12

Table 3.1: Sampling frequencies of physiological signals with the Heally recording system.

a) A master module for the configuration of the slaves, storing the data to the internal Flash memory (16 MB) or sending it to a PC via a Bluetooth wireless link.

b) A slave module that records a 2-lead ECG with electrodes, distal temperature, pulse wave, and skin conductivity with a finger probe and that had a marker button.

c) A second slave module that records respiration effort with an inductive belt around the chest, EMG and EOG with electrodes.

The recording frequencies and the sampling resolution for the signals are given in Table 3.1. The system was powered by a nickel-metal hydride (NiMH) 2.4V 2000 mAh cell pack that allowed an autonomy of 16 hours with the selected modules and configurations (Table 3.1). We chose not to use an additional EEG module, as the EEG recordings were judged to be too invasive and too susceptible to noise (see Section 3.1). Also, an additional module would have reduced the autonomy of the system to below 12 hours. All sensors and electrodes were connected to the modules with cables. The electrodes for ECG, EMG and EOG were gel electrodes and had to be glued to the skin. To give a more "wearable" feeling to the subjects a washable shirt was designed, onto which Velcro straps for easy attaching and removal of the cables and modules (Figure 3.1) were sewn.

3.1.2 Recordings

The initial measurements that were conducted were on a single male subject of age 26 (hereafter referred to as subject A). We chose sampling frequencies f ac-
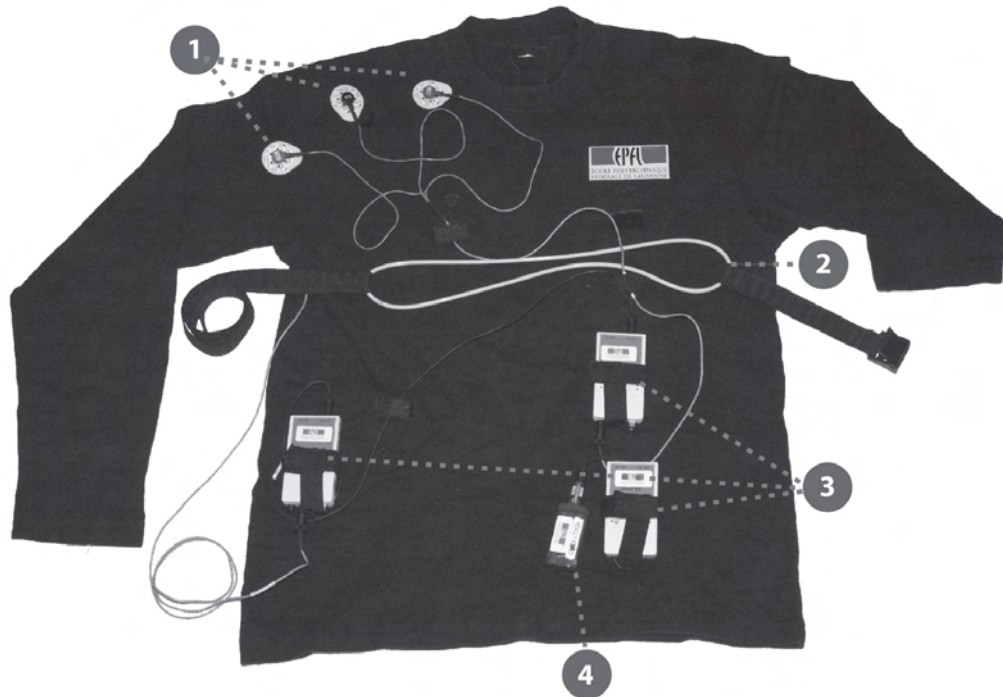
Figure 3.1: Heally recording system mounted on a home-made shirt. 1) ECG gel electrodes; 2) inductive belt sensor; 3) electronics modules; 4) NiMH battery. The EMG and EOG electrodes as well as the finger sensor for SC, temperature and PW are not shown.

cording to the requirements for digitalized PSG (Penzel and Conradt, 2000).

The EMG was recorded from the right shoulder muscle *trapezius* (f_{EMG}=200 Hz). The EOG, which was measured with electrodes glued on forehead and temple, was also sampled at 200 Hz, but it was only measured during the night, in order not to disturb the subject too much during daily activities. A video of the upper part of the body was recorded during night-time.

In order to obtain an equal amount of data for both sleep and wake, the subject wore the recording system for 16 hours per session. This recording time corresponded to the double of the average sleep time for the studied age group (Steptoe et al., 2006). The recording started approximately 4 hours before the regular bedtime of the subject. A total of eight recording sessions were carried out. Each session contained a mean of 8.15 ± 1.26 hours of sleep and 8.06 ± 2.04 hours of wake. A total of 65.23 hours of sleep and 64.51 hours of wake were analyzed. In the case of sensor failure or detachment, the corresponding data segments were discarded.

3.1.3   Analysis

As EEG was not available, the video, EOG and EMG recordings were analyzed by a human expert to determine if the subject was asleep or not. We did not distinguish between light, deep or REM sleep, because the binary discrimination between sleep and wake was sufficient for this experiment. The video was divided into segments of 10 seconds and each segment was labeled using the following criteria derived from Carskadon and Dement (1989) and Closs (1988) and Cole et al. (1992) and Ogilvie (2001):

a) The person was considered to be awake if his/her eyes were open or body movements occurred for more than 10 seconds.

b) If the eyes were closed, the subject was considered to be asleep when muscle tonus was released or slow eye movements were present. If segments of the video analysis were uncertain, the EOG and EMG signals were examined.

c) In doubtful cases, where neither video, EOG nor EMG signals could help to clearly identify sleep, the state was set to awake.

To prevent undetected wake with closed eyes, the subject was asked to open his eyes if he woke up during the night.

We compared the labeling for the mobile experiment based on video, EMG and EOG (visual) with the labeling from standard PSG using the scoring criteria of Rechtschaffen et al. (1968) for a recording of 10 hours. The PSG recording and labeling was performed at the sleep laboratory of the Luzerner Höhenklinik, Montana (VS) and the visual labeling was performed by the human expert who labeled all portable and wearable recordings used in this thesis. The inter-rater agreement was computed using Cohen's Kappa Coefficient (Landis and Koch, 1977). κ varies from 1 for perfect agreement to 0 for no agreement and is a chance adjusted measure.

In our comparison, we obtained a $\kappa = 0.87$, which is considered as almost perfect in inter-rater agreements of PSG (Danker-Hopfe et al., 2004). Note that the inter-rater PSG agreement between two human raters of different sleep laboratories is usually below 82%. For the tested night, Total Sleep Time (TST; see also Section 5.2.3) is overestimated by the mobile analysis by 2.67 minutes and 8 supplementary awakenings were detected (see also Figure 3.2). A closer look on the PSG unveils that the disagreement origins mainly in the detection of arousals or movements as awake, which were considered in PSG as REM sleep. Compared
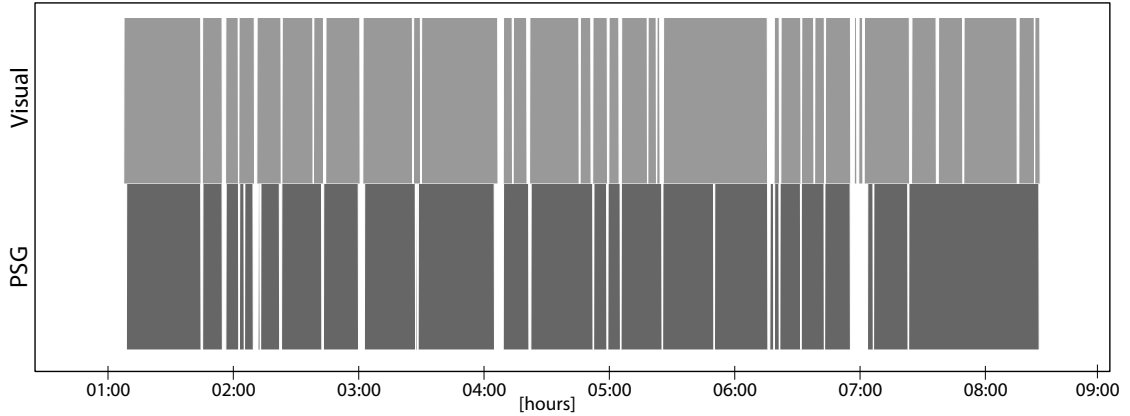
Figure 3.2: Comparison of polysomnography (PSG; dark gray) and video/EMG/EOG (visual; light gray) labeling of a sleep session. The PSG labeling was performed by a professional sleep technician at the Luzerner Höhenklinik, Montana (VS) using the scoring criteria from Rechtschaffen et al. (1968). The visual labeling was done using the criteria for the mobile sleep experiment that was used later for this study. The inter-rater agreement is considered as almost perfect (Cohen's Kappa $\kappa = 0.87$).

to the obtrusiveness of EEG recordings, the differences in agreement seem to be minor and therefore the visual rating was solely used for our experiments.

3.1.4 Results

Table 3.2 shows an evaluation summary of the signals that were evaluated qualitatively for this study. Each signal type was evaluated based on information from the literature and data sheets in a pre-study for theoretical accuracy in sleep discrimination, possibilities to integrate the necessary sensors in a mobile device and the required energy consumption thereof. For each criterion, one of four qualitative measures was given:

- 3 points: best, gold standard (+++).

- 2 points: good, recommended/promising (++).

- 1 point: acceptable (+).

- -1 point: not appropriate (-).

If no information could be retrieved, no measure was given (?). When an alternative measurement technology was available (but not empirically tested), the measure for the alternative was given in rounded brackets. The comfort of the

sensors, the quality of the non-filtered signals (including noise and movement artifacts), and the correlation of features with sleep and wake were evaluated. The wearable comfort was determined from the direct feedback given by subject A. The signal quality was evaluated by comparing the signal to theoretical outcomes and variations and on the content of artifacts. The correlation of features with sleep / wake was determined in the time and frequency domain. For this the time series were compared to the a priori sleep / wake labeling of the technician. To compare the signals in the frequency domain, the power spectral density (PSD) with a sliding window of 30 seconds equal to PSG was calculated and displayed as spectrogram.

Testing	Signal	clinical (theoretical) accuracy	pre-study		experimental study			final rating
			sensor integration	power consumption	signal quality	wearable comfort	features & sleep clues	
empirical tests	**RSP**	+	++	++	++	++	++	++
	ECG	++	+(++)	+	+	+(++)	++	+ (++)
	EMG	+(++)	+	-	+	+(-)	++	+
	EOG	+(++)	+	+	+	-	+	+
	Temp	+	++	++	-	+	-	-
	SC	+	++	++	-	-	-	-
	PW	?(+)	+	-	-	-	-	-
literature only	**Accel**	+	+++	+++	++^{a}	+++	+	
	EEG	+++	-	-	-^{a}	-^{a}	++^{a}	
	SpO2	-	+	-	-(+)^a	+(-)^a	?	
	BP	+	+	-	+^{a}	-^{a}	?	
	PAT	+	+	?	?	-^{a}	?	

^{a} These values have been estimated from the literature.

Table 3.2: Qualitative evaluation of quality and suitability of different physiological and behavioral signals for wearable sleep discrimination. [+++]=best (gold standard); [++]=good [+]=acceptable; [-]=not appropriate; [()]= variation depending on different implementation or sensor technology.

According to Table 3.2, the most promising empirical results were obtained with the RSP and ECG signals. The spectrogram of both signals in Figure 3.3 showed an interesting, visually detectable correlation with the a priori sleep / wake information. The changes of the PSD patterns did not exactly match in time. Whereas the ECG PSD seemed to anticipate the sleep onset and wake-up, the RSP PSD patterns are much more timely overlapping with the a priori knowledge. This suggests that ECG and RSP may show complementary frequency related features for sleep / wake classification. During posture changes, both signals showed movement artifacts. Despite the presence of these artifacts, the respiration patterns were still detectable. To improve the comfort of ECG recordings, dry electrodes would be needed.

For the two other signals measured with gel electrodes (EMG and EOG) a correlation with the sleep times and the spectral patterns was also observed, but less pronounced. Like the RSP and ECG signals, they showed the presence of noise, especially during posture changes. The EOG and EMG measures showed a major disadvantage: The muscle activity signals are relatively weak and were often corrupted with ECG signals. Further, they needed to be measured at a high sampling rate (200 Hz) which also increased power consumption. The EOG needed electrodes glued on the face, which was not practical and not accepted by the subject during the day. In fact, the use of gel largely improved the signal quality, but only until the gel started to dry out. Further, the gel reduced the capability of the skin to breath and induced uncomfortable itching and skin irritations after approximatively 12 hours of wearing the electrodes. In textile electrodes, sweat (which is conductive) can be used as natural conductor between skin and electrodes. However, it cannot prevent the electrodes from moving during heavy body motion. This suggests that EOG and EMG are difficult to measure with a wearable system using only textile electrodes and without gel that could improve skin contact.

The empirical measures showed that all the finger sensor data was strongly corrupted with movement artifacts. This was expected, as the subject was supposed to follow regular activities during the day and used the finger often (e.g. type writing). The finger sensor itself was judged to be very uncomfortable, hurting and disturbing during activities. Neither for skin temperature nor for skin conductance a trend or correlation with the sleep times were identified.
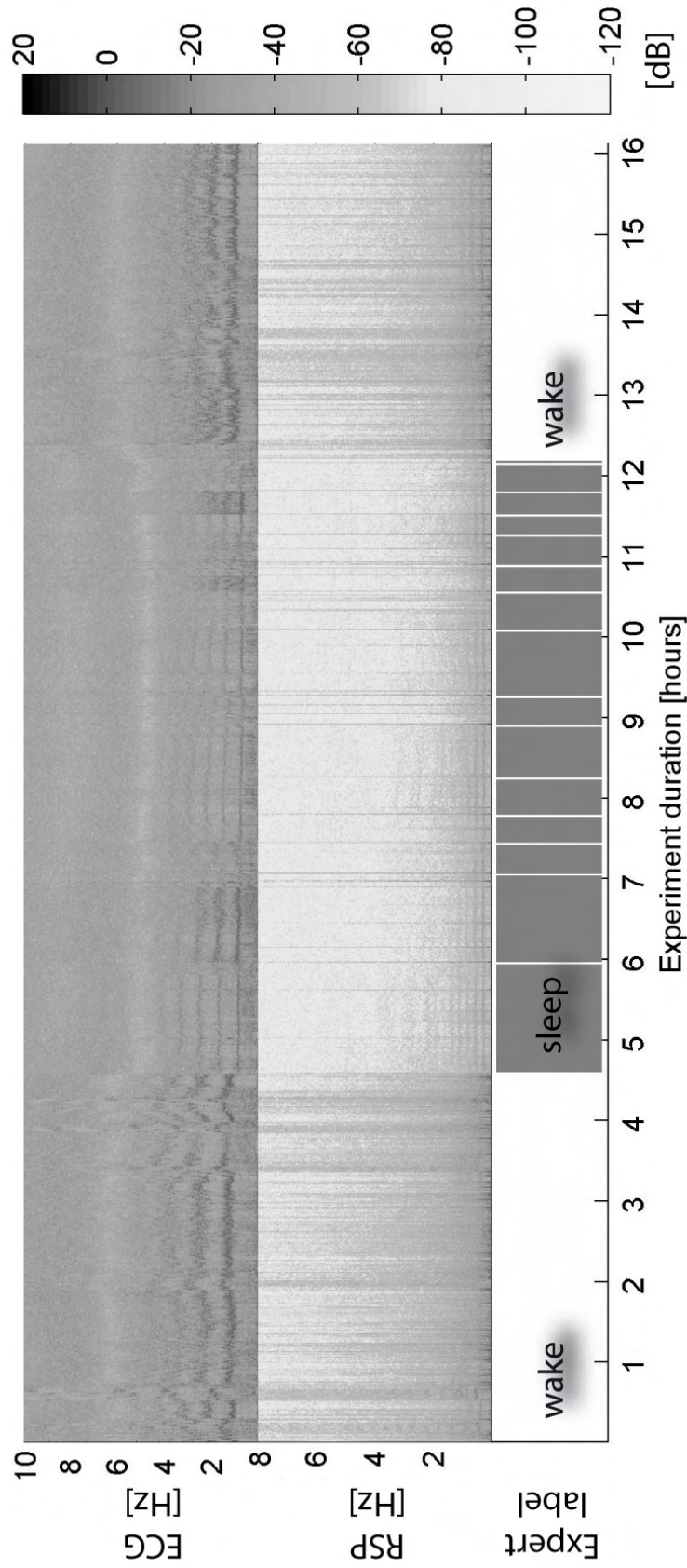
Figure 3.3: Power spectral density (PSD) of the ECG (top) and respiratory signal (middle) of one recording session of 16 hours using the portable Heally recording system. In the middle of the recording session a sleep of 7.5 hours took place (with short wake ups). The bottom graph shows these sleep times (filled blocks) obtained from labeling the video, EMG and EOG by a technician. Changes in the spectral patterns can be observed as well for the ECG as also for the RSP PSD on sleep onset and at wake-up. However, the changes are not exactly located at the same times. The PSD patterns also change during the sleep period and the changes can be associated to the wake-up.

3.1.5   Conclusion

From a wearable perspective, the use of accelerometers mounted on the body to detect movements is the most wearable and comfortable solution existing today. The sensors are low cost, low power and can be integrated into tiny devices of the size of a watch. However, with a wrist actigraph only one of the typical sleep behaviors can be detected (maintained behavioral quiescence) which influences the validity (and specificity) of this technique because quiescence is not unique to sleep. The use of additional accelerometers mounted on the body might help to measure body posture, a further behavioral expression of sleep.

If high sleep detection accuracy is desired, the measure and analysis of EEG signals would be the gold standard approach. EEG frequency analysis shows a high sleep detection rate in a high time resolution. It is the state of the art for detecting different sleep stages that might be valuable for sleep quality estimations, but which is not the primary goal. However, the very weak brain potential signals to measure and the high susceptibility to different noise sources makes it also one of the most difficult signals to measure outside the laboratory. The current sensor technology and the requirement for gluing multiple electrodes on the skin make EEG recordings not yet suitable for wearable applications. Recent advances in sensor development (e.g. Enobio, Figure 2.10) show that in the future, wearable EEG recording might become practical.

Different research groups have shown the link between the changes of the autonomic nervous system at sleep onset, the different sleep stages and the regulation of heart and respiration activity. Cardiorespiratory signals have been successfully used for the automatic screening of sleep / wake and a reduced number of sleep stages. So far, the only study (Redmond and Heneghan, 2006) that compared the automatic screening from cardiorespiratory signals with one using EEG features, obtained a lower accuracy for the cardiorespiratory classifier. However, the classifier had to discriminate between different sleep stages, which is a more difficult task that we aim for. Our qualitative tests with a portable recording system showed that cardiorespiratory signals can be measured easily in an every-day environment and also show features for sleep discrimination outside the laboratory. The tests also showed that the recording of this signal has potential to be comfortable enough and unobtrusive.

Other physiological recordings that might be usable for a sleep discriminative task (skin temperature, skin conductance, EOG and EMG) were shown to be either too obtrusive, delicate to measure and/or are without consistent information

about sleep and wake states.

All these findings suggest that the most promising signals for non-behavioral sleep / wake discrimination with a wearable system are ECG and respiration effort signals using frequency features. Although the respiratory patterns can be derived partially from the ECG measurements and would reduce the number of hardware components, the use of two distinctive sensors makes more sense for a wearable device, because the use of a distinctive measurement technology may offer redundancy, especially when the required dry ECG electrodes recordings are corrupted with noise. Further, the respiratory belt sensor is easy to integrate and does not consume much energy.

Therefore we will focus on cardiorespiratory features for the further development of a wearable sleep / wake classifier.

4 Automatic Sleep and Wake Classification

In the previous chapter we have identified and selected ECG and respiration effort (RSP) as the most interesting physiological signals for automated and wearable sleep/wake discrimination. On an embedded, low-power microcontroller that is used typically in wearable devices the computational resources are limited. In this chapter the design of a classification algorithm for sleep/wake is presented.

The algorithm offers low computational needs for the preprocessing and the classification and makes it suitable for use on a microcontroller. The preprocessing consists of calculating an estimation of the power spectral density (PSD) for the raw ECG and RSP signals with the help of a discrete Fast Fourier Transform (FFT). The classification method was based on Artificial Neural Networks (ANN), a well established method for sleep discrimination. The designed ANN had a single-layer, feed-forward topology with the signal spectral density as input and

one neuron as sleep / wake output. The performance of the classifier was evaluated for subject-specific and subject-independent data.

4.1 Algorithm

The sleep / wake classification algorithm was designed to be fast and low in computational complexity such that it could easily be embedded into a wearable device without stressing the tight power budget of a potential wearable device.

4.1.1 Preprocessing and Feature Extraction

The classification algorithm that was suggested by Redmond and Heneghan (2006) required 27 features from ECG and RSP, 18 of them being spectral functions of either signal. The preliminary analysis in Chapter 2 showed visual differences in the spectrogram of RSP and ECG for sleep and wake. Given this, spectral features seem to be promising for an automatic classification and spectral transformations can be done in many ways efficiently on microcontrollers. We therefore decided to calculate the power spectral density (PSD) of raw ECG and respiratory signals for feature extraction. PSD estimation methods are widely used for the frequency analysis of biomedical signals (Cohen, 2006). The periodogram method that we chose partitioned the original sequence of samples into equally sized segments s. Each segment s was windowed with a Hamming window function w to reduce the effects of spectral leakage (border effects) (Harris, 1978). The Hamming function is defined as follows:

$$w(n) = 0.53836 - 0.46161 cos\left(\frac{2\pi n}{N-1}\right) \qquad n = 0, 1, \ldots, N-1 \qquad (4.1)$$

where N is the number of samples that form a segment s. For each segment a periodogram $\hat{\mathbf{S}}$ was then calculated with an FFT, as follows

$$\hat{\mathbf{S}} = |FFT\{s(n)w(n)\}|^2 \qquad (4.2)$$

The periodogram $\hat{\mathbf{S}}$ gives an estimation of the frequency content in the given time segment. Computationally efficient FFT calculation in a digital signal processing (DSP) microcontroller requires a number of samples $N = 2^l$ where l is a positive integer. For our measurements we used the values of $N_{ECG} = 4096$ and $N_{RSP} = 2048$ according to this constraint and corresponded to a segment

length of 40.96 seconds. This was the next larger possible segment size compared to the traditional segment length of 30 seconds of PSG. We analyzed the effects of decreasing and increasing this segment size for a better frequency or time resolution, which will also be discussed in Section 5.4. Since the performance of the classifier did not improve with larger segment sizes, there appears to be no reason to incur the additional computational cost entailed by the increased segment size. As the input values are real numbers, the PSD output is symmetrical around the DC component and it is sufficient to use half of the output points ($N/2$). The DC component was eliminated, because it mainly contained the offset of the uncalibrated sensors. Further experiments that will be described in Section 5.5 showed that the information in high frequency components of the ECG and respiratory spectrogram are redundant and can be pruned without degrading the performance of the classifier. For this algorithm development, the ECG spectrogram was truncated at 10 Hz and the respiratory spectrogram was truncated at 8 Hz and the lower frequency part was kept. Correspondingly, the input size of the ANN was reduced from $N_{ECG}/2$ and $N_{RSP}/2$ to $m_{ECG} = 409$ and $m_{RSP} = 327$, respectively. The reduced input size did not only reduce the training time, but also reduced the search space that avoided over-fitting on noisy data.

4.1.2 Neural Classifier

A feed-forward ANN with no hidden layers and one single output unit with a tangent-sigmoid transfer function (see Figure 4.1, ANN classifier) was selected. Since the introduction of ANNs for sleep discrimination by Principe and Tome (1989), single and multilayer ANNs have shown to be suitable for sleep analysis. ANNs also display clear advantages over other classification algorithms, as they are easier to implement, have very simple topologies where one hidden neuron layer can theoretically solve all nonlinear problems and many efficient algorithms are available for automated learning. Similar to Principe and Tome (1989), we experimented with ANNs with no, one or more hidden layers, but according to our (Dürr et al. (2009), see Section 5.5) and Principe and Tome (1989) findings, the performance was not better in the multilayer topologies. It seems that a linear separation of the input vector is sufficient and the gradual increase of the hidden layers also increased the risk of over fitting. As the training time increased considerably (up to 10 times for 1 hidden layer), it was reasonable to keep the single-layered structure. For the classification into the classes, a symmetrical threshold was applied to the output neuron. That means that the output of the tangent-

sigmoid transfer function ([-1 1]) was thresholded so that $y(x) \geq 0$ was mapped
to sleep and $y(x) < 0$ was mapped to wake.

Training

To train the ANN and update the synaptic weights we used the Levenberg - Mar-
quardt back-propagation algorithm (Hagan and Menhaj, 1994). Using this algo-
rithm, the weights $\mathbf{x}_k$ are updated as follows[1]:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[\mathbf{J}^T\mathbf{J} + \mu\mathbf{I} \right]^{-1} \mathbf{J}^T\mathbf{e} \qquad \mu \geq 0 \tag{4.3}$$

where $\mathbf{J}$ is the Jacobian matrix with the first derivatives of the network errors e
and $\mathbf{I}$, the identity. The network error was calculated with the mean square dif-
ference between the actual and the desired ANN output. μ is a variable that is in-
creased at each step by $\mu increase$ until the weights changes obtained by Equation
4.3 result in a reduced network error. After this training iteration, the synaptic
weights are updated and μ is decreased by $\mu decrease$. The algorithm begins a
new training iteration by calculating $\mathbf{J}$ for the updated weights and calculating
Eq. 4.3 again.

Initialization of the weights was done with the Nguyen-Widrow method (Ngu-
yen and Widrow, 1990).

To train the networks, the physiological data obtained by the later described
experiments were divided into three sets: training, validation and test. The train-
ing set (TR) contained the data used to update the synaptic weights. The perfor-
mance of the network was evaluated on the validation set (VA) after each training
iteration and the training was stopped if the performance of VA did not increase
for more than 15 training iterations or the minimal gradient (10^{-10}) was reached.
The test set (TE) was used to measure the performance of the network with the
best validation performance after the entire training.

Topology

Three different ANN topologies, which differed in the type of input signal, were
studied. The input vector of the first topology *ECG+RSP* was composed of the
logarithm of the periodograms $\hat{\mathbf{S}}_{ECG}$ and $\hat{\mathbf{S}}_{RSP}$ (Figure 4.1). The other two archi-

---

[1]The parameters were: μ: 0.001; $\mu increase$: 10; $\mu decrease$: 0.1; μ max: 10^{10}; min gradient:
10^{-10}; max validation failures: 15.
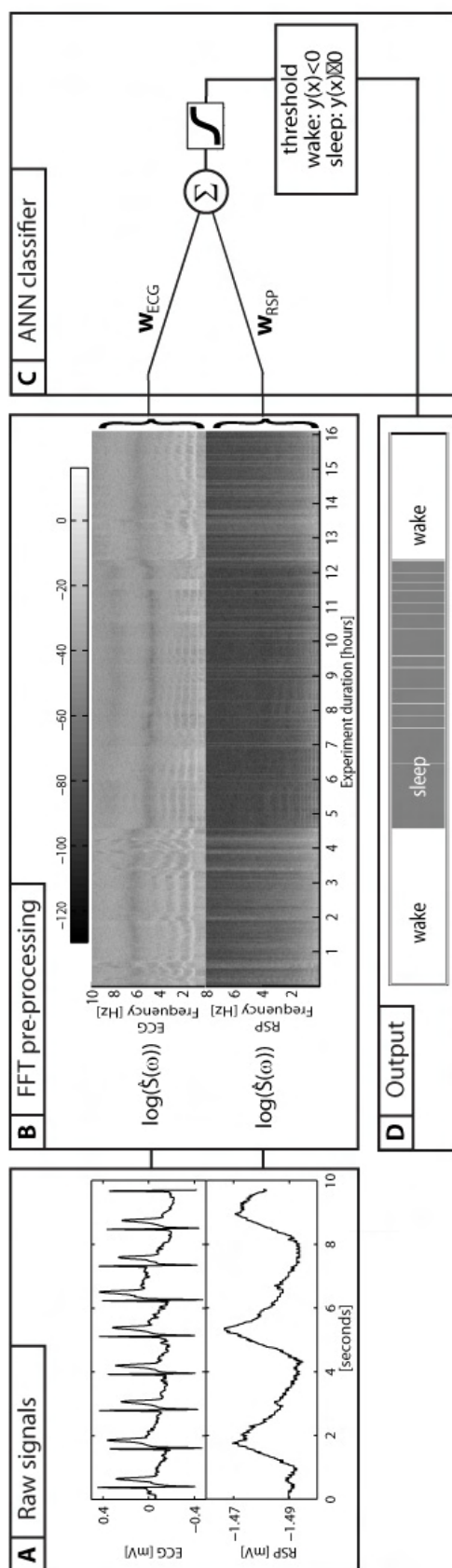
Figure 4.1: Overview of the sleep / wake classification system: Raw ECG and respiratory effort signals (A) are mapped to the frequency space with the help of a Fast Fourier Transform (B). The resulting frequency data (represented here by a spectrogram) are fed to a feed-forward, single-layer Artificial Neural Network (C) with a tangent-sigmoid transfer function and a symmetric classification threshold. The output of the classifier is sleep or wake (D).

tectures *ECG* and *RSP* used only the logarithm of the periodogram of one of the two signals, ECG or RSP, respectively.

4.2   Performance Measures

We introduced a series of measures to determine the performance of our algorithm.

   We calculated the accuracy by computing the percentage of correct classifications of sleep (*true sleep*) and wake (*true wake*) of TE as described by

$$accuracy = \frac{true\ sleep + true\ wake}{all\ sleep + all\ wake} \tag{4.4}$$

   where *all wake* and *all sleep* are the number of classified wake (-1) or sleep (1) points, respectively. Between different methods and topologies, the accuracies were compared using Student's t-test. If not stated differently, the data was first tested on its normal distribution using a Kolmogorov-Smirnov test at the 5% significance level.

   To further quantify the performance of the system, we computed two additional quantities (Figure 4.2): *Total Sleep Time* (TST), the total duration of segments classified as sleep per session, which is an important parameter in the sleepiness estimation models described in Section 2.4, and *Awakenings*, the total number of sleep-to-wake transitions during the period between the first and last segment labeled as sleep in a session, which is an indicator of the sleep quality of the subject. Both sleep parameters are commonly calculated for PSG reports.
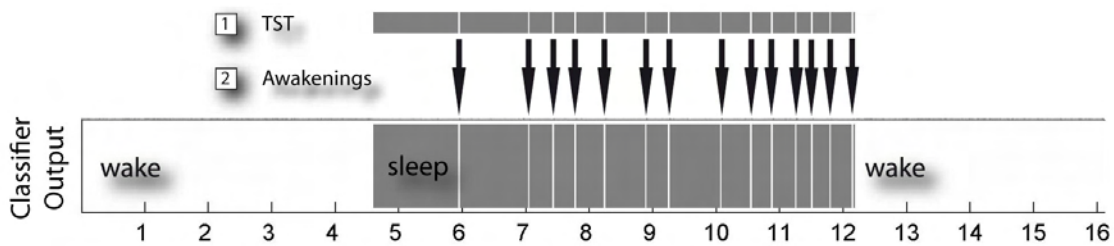
Figure 4.2: Sleep quality indicators; 1) Total Sleep Time (TST) is the duration of the epochs classified as sleep; 2) Awakenings is the sum of transitions from sleep to wake (arrows).

   The comparison of the TST labeled according to the video data and estimated by the classifier does not take into account whether the estimated sleep epochs are

correctly classified. Therefore, we plotted a series of Receiver Operating Characteristic (ROC) curves. ROC curves allow the assessment of the results of classifier data in which the a priori labeled data are not equally distributed (which is the case for the present data). For this reason it is often used in medical decision making and has been introduced in sleep analysis comparisons by Tryon (1991). A ROC curve shows the fraction of correctly classified sleep points called *sensitivity* (Equation 4.5) vs. the fraction of wrongly classified sleep points (1-*specificity*, Equation 4.6), when the classification threshold of the ANN output is varied from -1 to 1.

$$sensitivity = \frac{true\ sleep}{true\ sleep + false\ sleep} \tag{4.5}$$

$$1 - specificity = 1 - \frac{true\ wake}{true\ wake + false\ wake} \tag{4.6}$$

In such a representation the output of an ideal classifier is located at (0,1), which corresponds to perfect classification of all wake and sleep states. An example of such a ROC curve is shown in Figure 4.3.

4.3 Subject-specific Experiments

We investigated the performance of the algorithm when trained and tested on recordings from subject A. For this, the same recordings as in Chapter 3 were used. The 8 available sessions were randomly divided into TR containing 5 sessions, VA containing 1 session, and TE containing 2 sessions (Figure 4.4). 5 independent runs were performed from different initial weight values. In order to prevent performance biases due to the choice of sessions used for training and testing, we repeated the experiment 4 times with different sessions in the training and test set.

4.3.1 Results and Discussion

The classification accuracy of the networks using both ECG and RSP signals was significantly better than that of the networks using only ECG signals (95.4% vs. 91.7%, p<0.001, Student's t-test) and the networks using only respiration signals (95.4% vs. 93.3%, p<0.01, Student's t-test) (Table 4.1). Figure 4.5 shows a 2 hour detailed view of an output of a single-user classifier using respiration and ECG
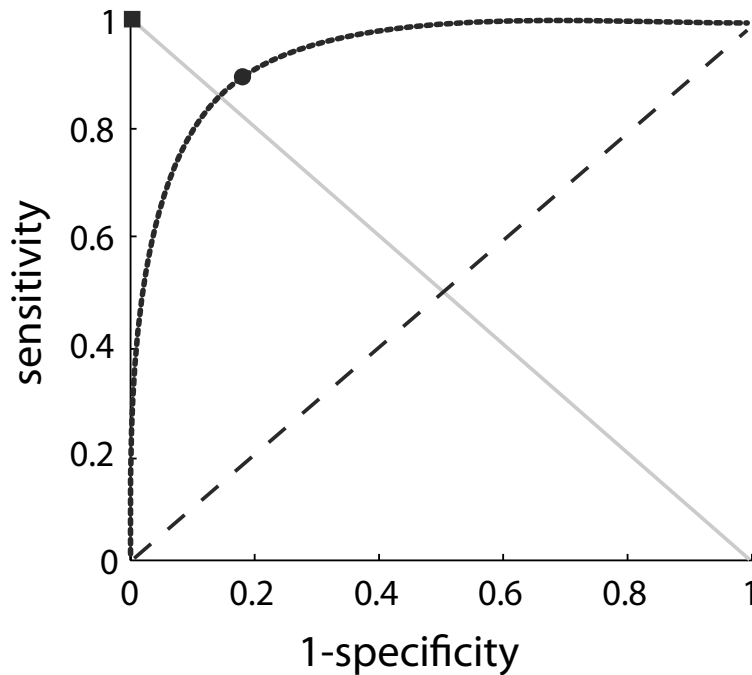
Figure 4.3: Receiver Operating Characteristics (ROC) curve example. The curve is obtained by varying the classification threshold from -1 to 1. The circle is the ROC point of the test set classification with the symmetrical threshold. The square corresponds to an optimal classifier that classifies all sleep as sleep and wake as wake. The dashed line corresponds to the ROC line of a random classifier. The full gray line corresponds to ROC points that have perfectly balanced sensitivity and specificity. The area below the ROC curve (AUC) is an estimation for the quality of the classifier and is highest when equal to one.

signals as input. In this example we can observe that the process of falling asleep is gradually detected by the classifier, although the labeling is binary (black line composed of triangles). From a physiological point of view this makes sense because sleep onset is a gradual, rather than a discrete process. A more detailed probabilistic classification taking into account the uncertainty of the classification would presumably reveal that the uncertainty is high in this transition phase, reducing the reliability of the classification. The same figure also shows that the first short awakening of the subject after 48 minutes could not be correctly classified. This difficulty to detect short awakenings inside long sleep epochs has also been reported by various actigraphy studies (de Souza et al., 2003 and Kushida et al., 2001 and Pollak et al., 2001).

The circles in the top of the ROC graphs for the single-user experiments shown
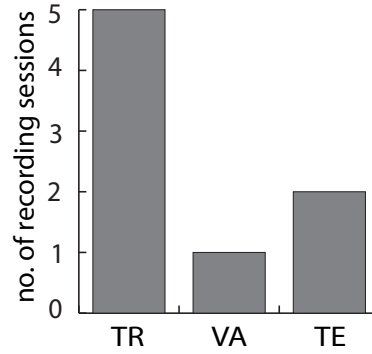
Figure 4.4: Experimental design for training the neural classifier. Four experiments were conducted by altering the sessions in the training (TR), validation (VA) and test (TE) set.

Experiment	ECG	Resp	ECG+RSP
accuracy	91.67 ± 2.74	93.27 ± 1.48	95.42 ± 1.61

Table 4.1: Test classification accuracies ± SD in percent.

in the top of Figure 4.6 correspond to the output of the test set with a classification threshold set to 0. We observe equally balanced sensitivity and specificity in most cases of the RSP and the ECG+RSP topology, which means that the classifier classifies sleep as accurately as wake. For the ECG topology, the solutions are either biased to an increased sensitivity or specificity. For the same topology we observe that a series of curves are shorter, which indicates that ECG may not classify all situations equally good.

In comparison, the only other study in the literature where both ECG and respiration signals from a single subject were combined, reported an accuracy of 81% (Redmond and Heneghan, 2006). However, a direct comparison of the two results is difficult, because

a) the value reported in Redmond and Heneghan (2006) was measured in the more difficult task of classifying wake, NREM sleep and REM sleep;

b) it was obtained with data recorded in a controlled hospital environment using polysomnography equipment;

c) it only used data from night recordings;

d) it used data from patients with obstructive sleep apnea; and

e) it used a computationally more expensive pre-processing algorithm calcu-

Figure 4.5: Unthresholded output of a subject-specific classifier using the ECG+RSP topology. The figure illustrates the transition from wake to sleep. The crosses represent correctly classified segments when a classifier threshold of 0 is applied, the circles represent the wrong classifications. The black triangles are the labeling values from the technician.

lating 27 features and a probabilistic classifier algorithm, which may be difficult to implement in a low-power and wearable device.

Table 4.2 shows that on average the ECG and the ECG+RSP classifiers overestimate the Total Sleep Time, whereas the classifier using only respiratory data slightly underestimates the Total Sleep Time. Similar findings were reported in actigraphy studies (de Souza et al., 2003 and Kushida et al., 2001). A possible explanation of sleep overestimation when the ECG signal is used is that the ECG signal might vary substantially between different sleep stages. This might be the case for REM-sleep, for example, but further analysis with a more detailed labeling of the sleep data would be needed to confirm this statement.

The statistics on Awakenings in Table 4.2 show that the mean number of awakenings is overestimated by all classifiers. This can be explained by the relatively noisy output of the classifier around the classification threshold, where consecutive epochs are classified alternatively as sleep and wake and therefore counted as several awakenings. An example of this phenomenon can be observed in Fig-
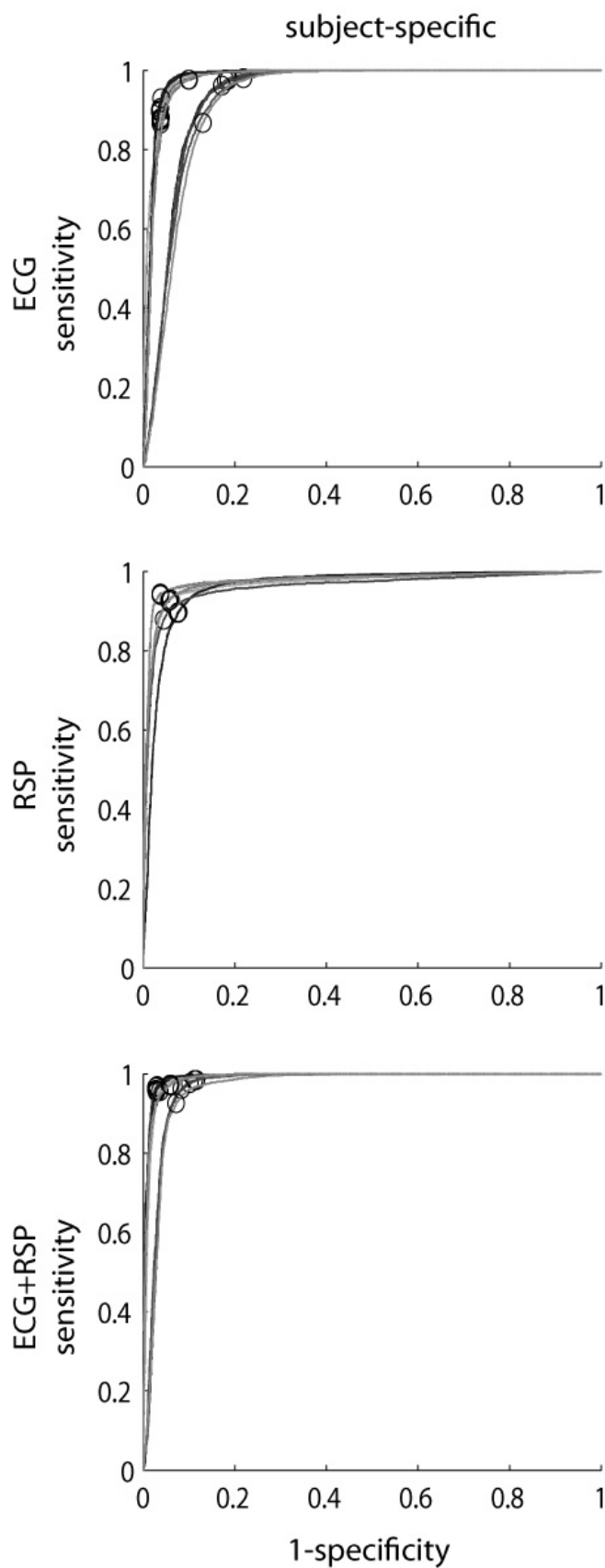
Figure 4.6: Receiver Operating Characteristics (ROC) curves of the three subject-specific neural classifiers topology configurations. Each curve corresponds to a different run obtained by changing the training set configuration or the initial network weights. The circles are the ROC points of the actual classification of the test set with the same classification threshold as in the training.

Sleep parameters[a]	Mean ± SD	Mean difference from Video ± SD
1. Total Sleep Time [hours]		
Video (label)	7.55 ± 1.7	
ECG	7.79 ± 2.11	-0.25 ± 1.22
RSP	7.46 ± 1.54	0.09 ± 0.3
ECG+RSP	7.77 ± 1.74	-0.22 ± 0.44
2. Awakenings [numbers]		
Video (label)	23.75 ± 8.81	
ECG	88.71 ± 86.96	-64.96 ± 85.24
RSP	92.61 ± 39.67	-68.86 ± 36.52
ECG+RSP	49.6 ± 24.37	-25.85 ± 20.61

[a] The mean is calculated for each recording session individually (and not for the TE) and the SD is taken over all sessions in all TE.

Table 4.2: Comparison of calculated sleep parameters for the video labeling and the three ANN topologies. Total Sleep Time (TST) is the total amount of sleep, and awakenings is the number of wake-up events per night.

ure 4.7. The occasional mis-labeling of the video data might also contribute to this overestimation. Another source of error could also be the difference in resolution of the video labeling (10 seconds) and the FFT (40.96 seconds). A short wake-up of 10 seconds might be lost if it is located at the border of a recording segment and therefore attenuated by the windowing function. Note that the combination of ECG and respiration performs relatively better in terms of the estimate of the number of awakenings.

4.3.2   Conclusion

With this set of experiments we have shown that sleep / wake classification can be done with a reasonable accuracy of more than 90% when training was carried out on data recorded from a portable system that was worn by a single user. This can be achieved by using only FFT preprocessing and a single-layer, feed-forward ANN.

4.4   Multiple Users or the Problem of Generalizing

On a wearable device for a general public, it is particularly interesting to have a subject-independent algorithm that is able to generalize for a large number of
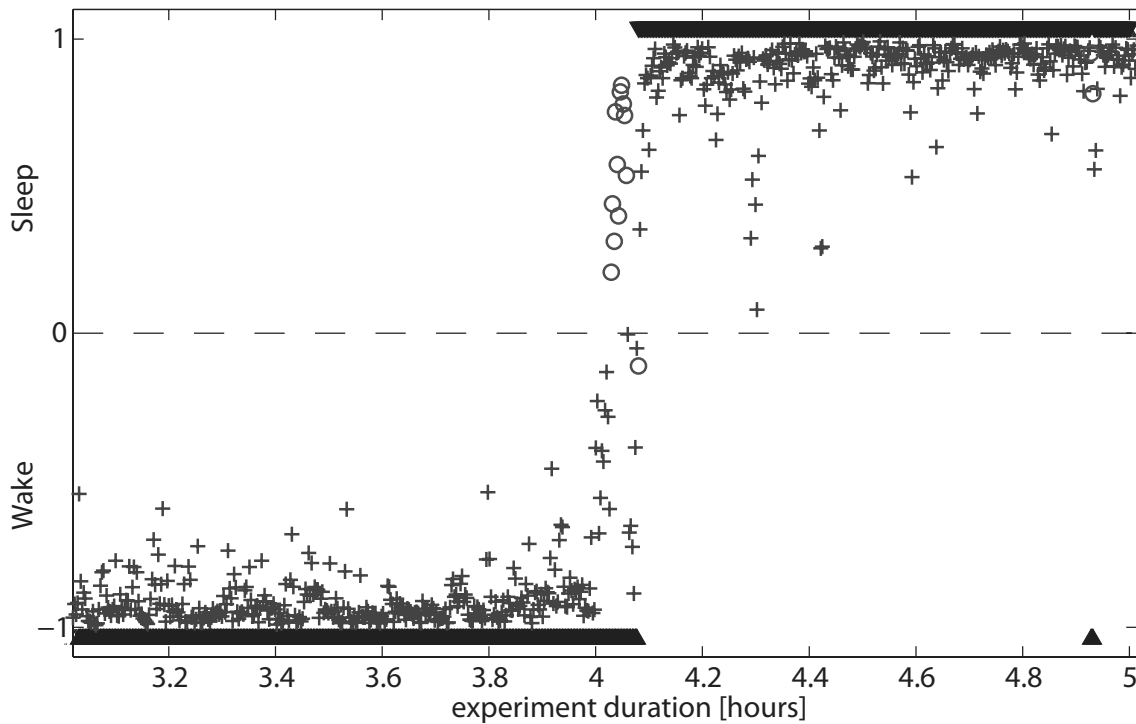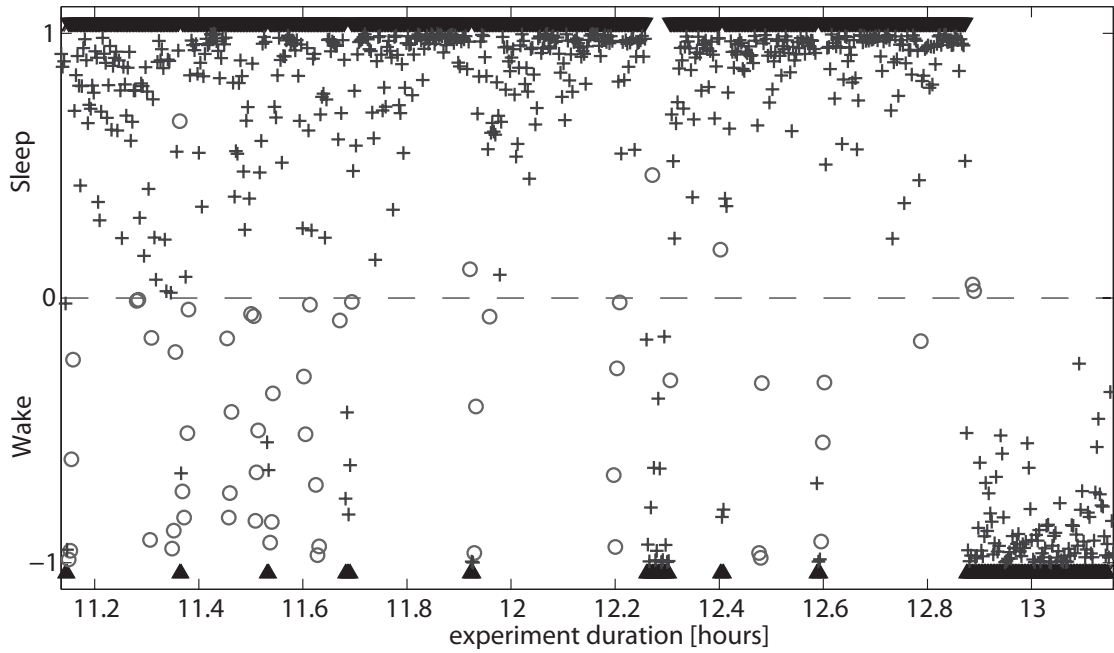
Figure 4.7: Unthresholded output of a subject-specific classifier using the RSP topology. The figure illustrates the transition from sleep to wake. The crosses represent correctly classified segments when a classifier threshold of 0 is applied, the circles represent the wrong classifications. The black triangles are the labeling values from the technician.

users, and can give accurate sleep predictions without the need for calibration. Calibration or retraining would need user interaction that can be disturbing. For example, only few people would accept to undergo the procedure of recording video, EMG and EOG.

The spectral patterns of EEG were found to have significant inter-subject differences during deep sleep (Buckelmüller et al., 2006), whereas the patterns did not show any intra-subject changes within weeks. Redmond and Heneghan (2006) remarked that the accuracy of their cardiorespiratory sleep stage classifier dropped by more than 10% when the classifier was tested on data from a different set of users than the one that contributed to the training. To analyze whether this is also true for the described sleep / wake classifier in Section 4.1, a new set of recordings involving 5 more subjects were recorded. To investigate the capability to generalize to multiple users, all three previously tested topologies were trained and tested using data from multiple users in different combinations.

4.4.1 Data Recordings

We conducted home recordings with 5 healthy male subjects, aged between 23 and 29 years (subjects B to F). As described in Chapter 2 ECG and RSP were again recorded with the Heally system. However the finger sensor was removed, as it was found disturbing and painful by subject A. The measurements of the EMG and EOG were kept, but only as an additional reference measure to the video. Similar to the initial test, a night-time video of the upper part of the body was recorded.

The duration of each recording session was again 16 hours. Each subject presented himself for 2 sessions. Each session (including subject A) contained a mean of 7.19 ± 1.65 hours of sleep and 8.45 ± 3.19 hours of wake. A total of 130.47 hours of sleep and 161.64 hours of wake were analyzed. The subjects reported no major discomfort because of the recording system during sleep. However, at the end of the recording, itching at the electrode sites was reported.

4.4.2 Experiments

The performance of the method when trained on a single person and tested on multiple persons, and when trained on multiple persons and tested on multiple persons was analyzed. Six experiments were carried out, each with an increasing number of persons in the training set (1 to 6) and all remaining persons in the test set (Figure 4.8). In the experiment where all 6 persons were in the training set, we made sure that the two sets contained different recording sessions. In all experiments, the validation set was composed of 2 hours of data from each user, randomly sampled over the available 2 sessions and containing an equal amount of sleep and wake labels. This data was neither used for training nor for testing. Five independent runs of each experiment were performed from different initial weight values.

4.4.3 Results

In all six subject-independent conditions, the accuracies in the test condition dropped compared to the initial subject-specific conditions in Section 4.3.1. The accuracy drop was the largest for the networks using only ECG signals (Table 4.3, first column). Table 4.4 shows that, as expected, in all experiments the training performance of the networks using both ECG and RSP is higher than the training performance of the networks using only one of the two signals. However,
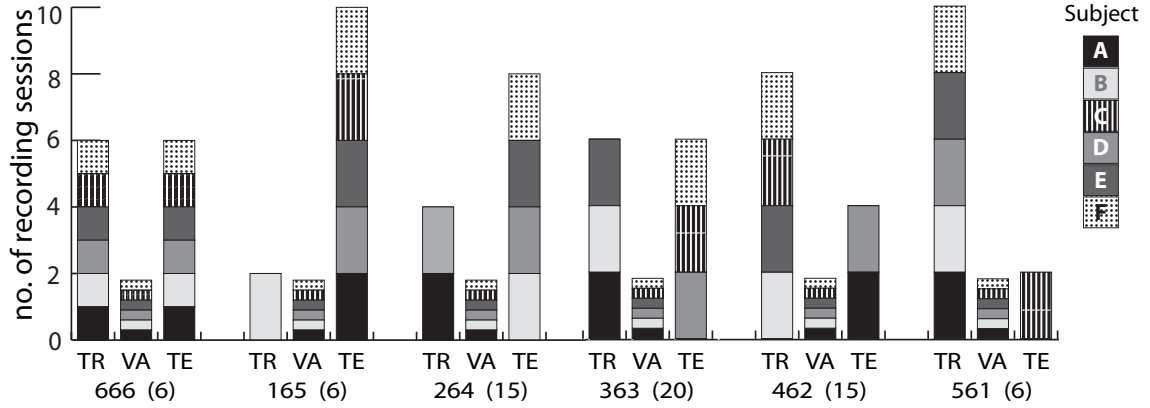
Figure 4.8: Experimental design for training the neural classifier where the classifier was trained and tested on multiple users (subject-independent). Numbers indicate users in the training set (TR), users in the validation set (VA), users in the test set (TE) (no. of repetitions with different combinations of users/sessions in training and test).

Experiments	Topologies		
	ECG	RSP	ECG+RSP
165	65.54 ± 7.76	83.09 ± 2.78	76.53 ± 5.39
264	66.74 ± 5.88	85.69 ± 2.05	81.65 ± 3.82
363	68.17 ± 5.98	86.64 ± 2.33	83.59 ± 3.21
462	68.63 ± 8.99	87.34 ± 2.91	84.96 ± 3.52
561	69.51 ± 14.96	87.31 ± 4.68	85.25 ± 5.25
666	78.72 ± 1.66	89.77 ± 1.37	88.34 ± 2.30

Table 4.3: Subject-independent test classification accuracies $\pm$ SD in percent.

the networks using both ECG and RSP (Table 4.3, third column) displayed lower accuracy in the test condition than the networks using only respiration signals (Table 4.3, second column). Also, the variability of the accuracy across multiple replicates was higher when respiration was used in combination with ECG as compared to the condition when respiration alone was used. These results may suggest that ECG signals have unique features that are specific to each individual, so that networks trained on the data obtained from a small number of subjects do not generalize well to other individuals for the purpose of discriminating sleep and wake states. This points to the necessity of increasing the number of subjects contributing to the formation of the training and validation datasets, when using both ECG and RSP signals, in order to prevent the classifier from over-fitting the peculiarities of the training data obtained from a restricted sample of subjects.

subject-independent 561
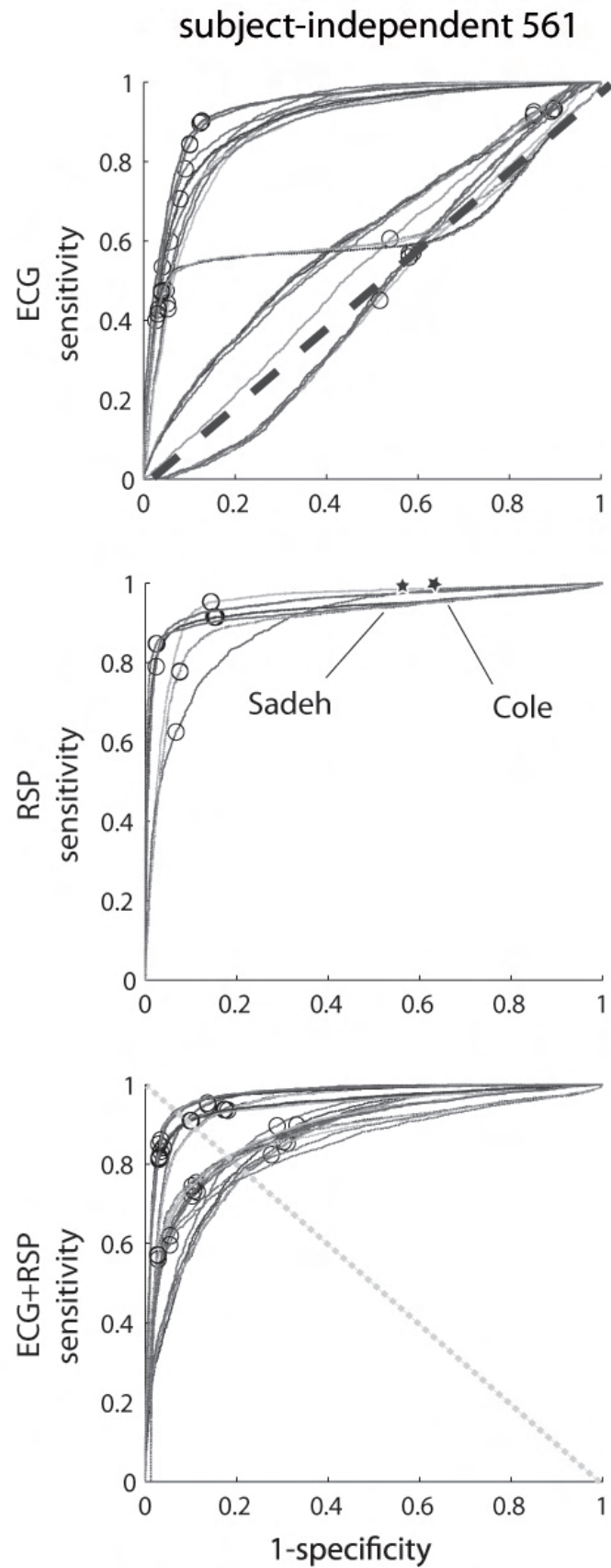
Figure 4.9: Receiver Operating Characteristics (ROC) curves of the three neural classifiers topologies (ECG, RSP, ECG+RSP) for the subject-independent 561 configuration. Each curve corresponds to a different run obtained by changing the training set configuration or the initial network weights. The circles are the ROC points of the actual classification of the test set. Top: The dashed line corresponds to the ROC line of a random classifier. Middle: The stars correspond to the ROC points obtained by the two actigraphy algorithms proposed by Cole et al. (1992) and Sadeh et al. (1994) that were used in de Souza et al. (2003).

Experiments	Topologies		
	ECG	RSP	ECG+RSP
165	96.63 ± 1.81	95.81 ± 1.45	99.53 ± 0.34
264	92.62 ± 1.99	93.83 ± 1.25	97.62 ± 1.12
363	90.31 ± 2.04	93.06 ± 0.84	96.02 ± 0.89
462	88.46 ± 1.98	92.64 ± 0.60	95.05 ± 0.67
561	86.92 ± 1.63	92.34 ± 0.34	94.63 ± 0.56
666	88.35 ± 3.31	92.72 ± 1.07	95.84 ± 1.02

Table 4.4: Subject-independent mean training accuracies $\pm$ SD in percent.

Sleep parameters[a]	Mean $\pm$ SD	Mean difference from Video $\pm$ SD
1. Total Sleep Time [hours]		
Video (label)	4.27 ± 1.14	
ECG	4.44 ± 3.07	-0.17 ± 2.6
RSP	3.88 ± 1.12	0.38 ± 0.9
ECG+RSP	4.17 ± 1.6	0.09 ± 0.9
2. Awakenings [numbers]		
Video (label)	31.38 ± 13.27	
ECG	118.5 ± 68.7	-86.79 ± 67.52
RSP	106.08 ± 50.99	-74.7 ± 50.49
ECG+RSP	98.82 ± 52.59	-67.82 ± 49.84

[a] The mean is calculated on each session individually (and not per user or TE) and the SD is calculated over all means from each TE.

Table 4.5: Comparison of the sleep parameters *Total Sleep Time* and *Awakenings* for the subject-independent 561 experiments.

Although one may notice a positive correlation between accuracy and number of users in the training set, our data sets are not sufficiently large to draw conclusions on the observed differences among the six subject-independent conditions.

The subject-independent experiment with a single subject in the TE and all other subjects in TR (561) is the most realistic setup for a general public use and similar setups to this were also used by Redmond and Heneghan (2006) and in actigraphy studies (de Souza et al., 2003 and Kushida et al., 2001). Therefore, we chose this experiment for further analysis. Table 4.5 shows that in the 561 experiment, the ECG and the ECG+RSP classifiers produce a good mean estimate of the TST, whereas all classifiers overestimate the number of Awakenings, which has also been observed in the subject-specific experiments (Table 4.2). This means

that the quality of sleep is constantly underestimated. If the measure of sleep quality is required, a post-processing to filter out unrealistic oscillations in the sleep / wake output is recommended.

The ROC data for the 561 experiment also reveals, as previously indicated, that the ECG signal alone does not generalize well to other users, and for some cases, does not find a solution that is better than random (Figure 4.9, circles close to dashed line, top).

Although the accuracy of our subject-independent classifier with respiration signals is comparable to the 91% accuracies obtained by de Souza et al. (2003), which used the algorithms of Cole et al. (1992) and Sadeh et al. (1994) using only accelerometer signals, it produces a smaller fraction of wrongly classified wake points. As we mentioned in Chapter 6, actigraphy may often mis-classify wake periods of low activity (reading, watching television, lying in bed) as sleep periods. Indeed, the specificity of the actigraphy methods in (de Souza et al., 2003) was only 44% (Sadeh) and 34% (Cole) during the wake periods (stars, Figure 4.9, middle), whereas in our case it was 88.12% $\pm$ 9.6 when only the respiration signal was used and 91.87% $\pm$ 5.23 when both ECG and respiratory effort were used. In general, we can say that our solution using the respiration signal as input achieves a better balance between sensitivity and specificity than those obtained using actigraphy. The findings of Kushida et al. (2001), who compared actigraphy with PSG data from a large group of sleep disordered patients, support this statement (accuracy = 78%, sensitivity = 92%, specificity= 48%). A comparison between the cardiorespiratory and the actigraphy sleep / wake classification will be presented in more detail in Chapter 6.

4.4.4   Conclusion

As expected and already described in similar cases in literature (Redmond and Heneghan, 2006), the cardiorespiratory sleep / wake classifier performed worse when generalizing to different users than it was trained for. Interestingly, the performance drop was more pronounced on the ECG topology (-20% of the mean) than the RSP (-6%) and ECG+RSP (-10%) topologies. The obtained generalized accuracies are still comparable to the existing actigraphy methods. Although, the specificity of the classifier is higher than the actigraphy-based sleep detection in literature. It is therefore reasonable to transfer the method to a more wearable technology, which is described in the next chapter.

5   From Portable to Wearable

Non Sequitur © 1998 Rick Stromoski. Used by permission of Universal Press Syndicate.
All rights reserved.

In the previous chapter, we showed that with frequency-domain features of cardiorespiratory signals obtained from a portable medical recording device and an ANN, sleep / wake classification is possible and accurate. These findings are not necessary valid for a wearable device, because the comfort of wearability comes with constraints in sensors, processing power, size, energy consumption and signal quality.

In wearable systems, typically the less power hungry microcontrollers (<0.1 Watts) are preferred to the high-performance 32-bit application processors (1-2 Watts) used in mobile devices. Even if more advanced wearable computing systems are available[1], the processing load for the classification should be kept low,

---

[1]Highly integrated wearable computing devices using mobile application processors have been developed with success. The showcase is the QBIC wearable belt described in Amft et al. (2004). The energy autonomy for these devices is still relatively short (3 hours in continuous work mode).

so that as much computing resources as possible are free for high level computing (e.g. context awareness algorithms or fatigue models). Microcontrollers cannot run an operating system and the data format (8 or 16 bit) as well as the computational resources are restricted (if the computing costs have to stay low). On the other hand, microcontrollers can offer specialized digital signal processing (DSP) architectures that reduces computational load on signal processing. Therefore, the design of wearable systems requires the reduction of the algorithm complexity to the absolute minimum and an adjustment to the capabilities of the specific microcontroller. In our case, beside the algorithm complexity, the power consumption of the system can be further reduced by lowering the sampling rate of sensors and the reduction of the number of features for the classification.

In this chapter we use a wearable hardware system called *SleePic*[2] that was developed for this thesis (Appendix B). The SleePic system is based on a commercially available, wearable sensor module and two custom electronic modules for processing and user interaction. We seek methods to minimize the workload of the classification algorithm on the embedded microcontroller and therefore the power consumption of the wearable system. For that, data from the portable experiments (Chapter 4) and newly recorded data from the SleePic system are used. The segment size for the preprocessing and the ANN search space are reduced with evolutionary and statistical selection methods. With the data obtained from the wearable SleePic system, the findings of the wearable methods are then validated and discussed.

5.1   Requirements for a Wearable Sleep and Wake Discrimination System

For the development of the wearable sleep / wake discrimination system, we set up a series of design criteria that were mainly based on the findings of the previous chapters and the general requirements for a bio-medical wearable device (Martin et al., 2000 and Scheffler and Hirt, 2004). The system should:

- Be *wearable*: From sensors to user feedback, all elements of the device should be integrated into a wearable system. This implies low size and low weight. A medical device is considered wearable if it is not perceived as a medical device by an observer. The materials that are in contact with skin should be

---

[2]The name SleePic is derived from *Sleep* discrimination with a *Programmable Interface Controller*.

washable and bio-compatible. The system should be easy to put on and off by the user himself. A wearable system should also be comfortable and not display any cables.

- Have enough *autonomy*: The system makes only sense if it can run for at least 48 hours. Further, it should be easy to recharge.

- Be *low cost* and *realizable*: To build a prototype no specific hardware developments should be needed. The components should be low cost. This is also in respect to a possible general public application.

- Record *ECG* and *RSP*: The device should record ECG and RSP signals so that the algorithm from Chapter 4 can be applied.

- Be able to *execute the classification algorithm on-line*: The processing resources on the system need to be able to cope with the FFT and ANN computation in a reasonable time without excessive energy consumption.

- Give *feedback to the user*: The system needs an interface to communicate with the wearer. The feedback method has to adhere to the energy restrictions.

There exist no such systems on the market today. Appendix A gives an overview of the existing portable and wearable devices, as well comparative tables with respect to the criteria stated above.

5.2 Wearable Experiments

A series of recordings were conducted to obtain sleep / wake data in real-life situations. The data was then used for modifying the sleep / wake algorithm presented in Chapter 4 to fit on the SleePic system.

5.2.1 SleePic System

As no commercially available device was able to fulfill all our needs, we decided to develop a prototype ourselves. To keep the costs of the prototype low, we decided to rely on commercially available sensors and components for the prototyping of the wearable system. Missing functionalities were added by custom designed extension modules featuring DSP microcontrollers. Custom made chips that implement the FFT and ANN on silicon would have been a more power efficient alternative to microcontrollers. For example, Aziz et al. (2009) have sug-

gested a very-large-scale integration (VLSI) chip integrating a neural recording interface and a wavelet based ANN on a single chip for an automated epileptic seizure detection. However, these types of chips are very task specific and their development expensive. Therefore we preferred the more flexible and cheaper microcontroller solution for the prototyping.

The SleePic sleep / wake discrimination system (Figure 5.1) that was developed during this thesis, is composed of three modules:

- The *SleePic Sensor Module* which is based on the commercially available wearable sensor system Equivital™ (Hidalgo Ltd, UK) and is worn around the chest.

- The *SleePic Core* processing module that incorporates a digital signal microcontroller (dsPic33™ from Microchip™ Technology Inc, USA). It is responsible for all signal preprocessing and classification. The module connects directly to the SleePic Sensor Module.

- The *SleePic Watch* that is worn on the wrist. The SleePic Watch is responsible for remote sensing and user feedback with LEDs and a button. It communicates wirelessly with the SleePic Core.

The design of the modules as well as the operating modes and power consumption of the SleePic system are described in Appendix B.

5.2.2   Recording Sessions and Subjects

The SleePic system was worn by nine (two female and seven male) volunteers in the age between 24 and 30 years. The subjects were in good health and reported no cardio-respiratory disease or any sleep disorders.

The subjects came to the laboratory in the evening after work, were then instructed about the experiment procedure and how to wear the device. The subjects were asked to wear the SleePic device for 36 hours. They were allowed to remove the belt when doing heavy sport or when they wanted to take a shower. During the whole experiment, they had to fill out a logbook by indicating their sleep times, system-off times and particular events related to the system that may happen during the experiment. During the two nights of the experiment, the subjects were asked to sleep at home. For both nights an infrared video camera was installed in their bedroom to record the sleep behavior during bedtime. During the whole experiment the subjects had to perform a reaction task using the button
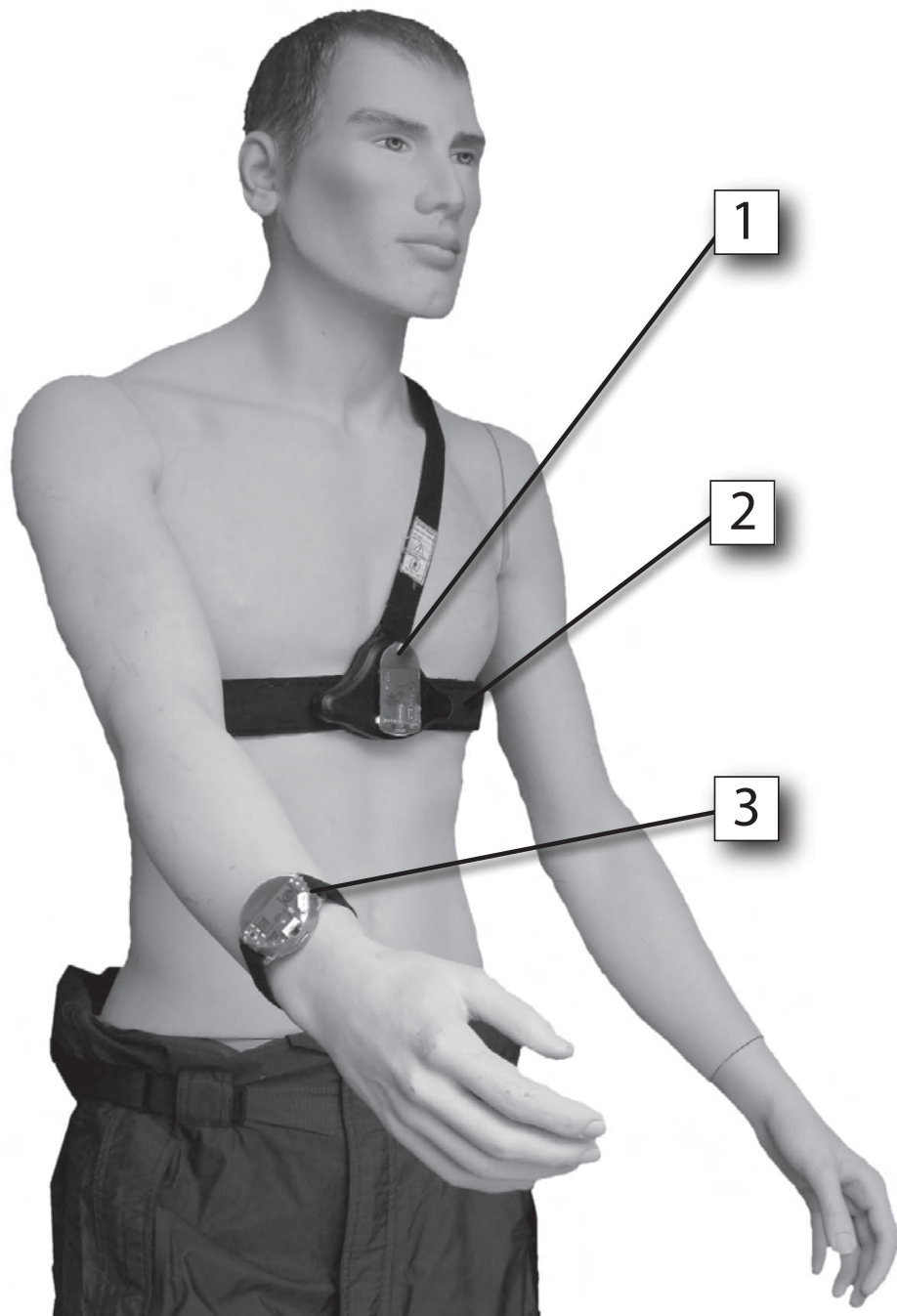
Figure 5.1: A subject wearing the complete SleePic sleep / wake detection system. 1) SleePic Core extension module; 2) Equivital™ Sensor Belt; and 3) SleePic Watch for user interaction.

on the Sleepic Watch. A detailed description of this task is given in Chapter 7. After 36 hours of recording, the participants brought back the complete recording system to the lab for a debriefing and filled out a questionnaire with questions
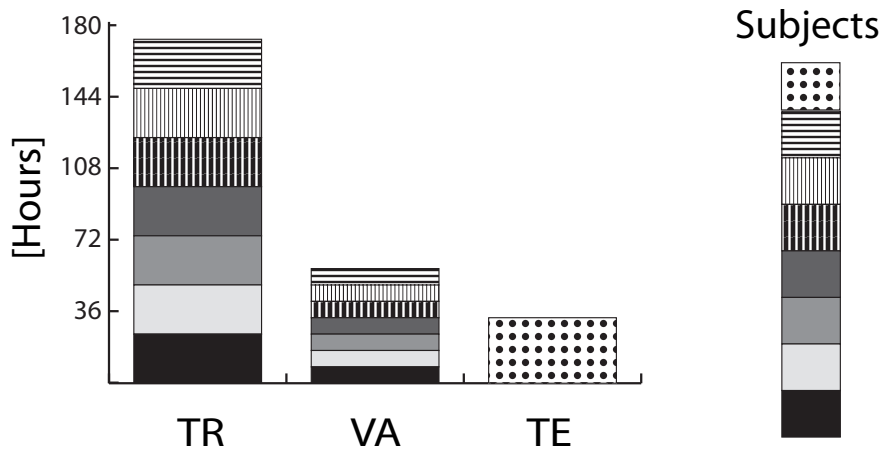
Figure 5.2: Repartition of the sessions for the subject-independent experiment. One user contributes completely to the test data set (TE). The data from the other users are divided into 75% for the train data set (TR) and 25% for the validation data set (VA).

about the usability and comfort of the system.

5.2.3   Data Analysis

A technician analyzed the video recordings and labeled the wake / sleep periods in 10-second intervals in the same way as it was done in the previous study for the portable recordings (Section 3.1.3). The data from the periods where the SleePic was not worn were discarded. When the recording device failed to give any data, the missing data epochs were discarded. However, signals with movement artifacts or other task-dependent disturbances were not discarded, since they might contain useful information for the classification task. With one female subject, the sensor belt became too loose, which was not detected instantly. Therefore an additional 3.5 hours with bad data needed to be discarded. During the recording of a male subject, a battery connection failed and we were not able to collect enough data from this subject. The data from this subject was dropped from the analysis. From the remaining eight subjects, a total of 250 hours of valid recordings were obtained, thereof were 37% labeled with sleep.

With the obtained wearable data we conducted subject-specific and subject-independent experiments like in Chapter 4.

For the subject-specific experiments, the first 18 hours of recordings (that included the 1st night) were randomly divided into a train data set (75%) and validation data set (25%) the rest of the recordings from the same user were used as

test set.

For the subject-independent experiments, all recording sessions but one were combined and randomly divided into a train data set (75%) and validation data set (25%). The recording session from the remaining subject was used as test data set (Figure 5.2). This was repeated until each subject was once in the test data set (8 times).

The subject-specific data sets were also used for the adjustment of the computational load of the preprocessing (Section 5.4) and the ANN processing (Section 5.5).

5.3   User Acceptance

From the post-study questionnaire, we were able to evaluate the subjects' acceptance of the SleePic prototype. All subjects had no problem with the study. Interestingly, all female subjects felt the wearable device comfortable enough, whereas the male subjects either could not decide (50%) or were not comfortable with it (50%). The men argued that they could not get used to the belt because of it was too tight or too big. The SleePic system was disturbing the subjects when they were lying face down. All subjects found the light vibration that accompanied the reaction task helpful and not as disturbing. The subjects also agreed with the frequency of the reaction time tasks. Wearing the device was influencing the daily activities of the subjects only slightly (60%) or not at all (40%). However, only 50% would wear the device again for a private purpose. A possible reason for this could be that the users did not get any direct feedback from the device and did not know the real purpose of the experiments during this study. Therefore, compared to the perceived decrease in comfort, no visible advantage for the user existed. Additional experiments with a fully working device are required to clarify this assumption. A point that was raised several times was the uneasiness of the subjects with the video recordings during the night. This has to be taken into account for further developments.

5.4   Reducing Computational Load of Preprocessing

Although the dsPic33 microcontroller with the highest amount of RAM memory available was embedded the SleePic Core, the RAM was not sufficient to calculate the FFT for ECG and RSP using the recording parameters used in Chapter

Topology	RSP	ECG	
Sampling Resolution [bits]			
portable	12	12	
wearable	10	10	
Sampling Rate [Hz]			*RAM*
portable	50	100	*utilization*
wearable	25.6 (25.6)	51.2 (256)	*ratio*
FFT Input Size [number]			*[%]*
portable 40 seconds	2048	4096	150
wearable 40 seconds	1024	2048	95
wearable 20 seconds	512	1024	37

Table 5.1: Sampling configuration and preprocessing for the portable and the wearable algorithm (original sampling rate of Equivital^TM).

4 (Table 5.1). Further, because the Equivital^TM recorder did not allow the same sampling rates, we had to change the recording configuration (Table 5.1). As the high frequency features were not needed for the classification, it was reasonable to reduce the sampling rate accordingly. We therefore subsampled the ECG from 256 Hz to 51.2 Hz (51.2% of the original size) and used the RSP at the recording sampling rate of the Equivital^TM (25.6 Hz, 51.2% of the original size). Table 5.1 shows the reduction effects. Because of the FFT calculation requirements, the segment size had to be slightly adjusted from 40.96 to 40 seconds to keep a power of two number of sampling points in one segment. The new segment sizes of 1024 (RSP) and 2048 (ECG) for a 40-second window fitted the available RAM, but left only little space for other variables (Table 5.1).

The processing load can be further reduced by decreasing the segment size for the FFT preprocessing. However, by the nature of the Fourier Transform, with an increase in time resolution, the resolution of the frequency output decreases. The reduction of the segment from 40 to the next possible segment size of 20 seconds resulted in a frequency resolution shift from 0.0122 Hz to 0.025 Hz. We conducted an experiment where the classification accuracies were compared depending of the size of the segments. Figure 5.3 shows the test results when the features are calculated on a segment size of 5, 10, 20 and 40 seconds respectively. It made from a physiological point of view no sense to take smaller segments than 5 seconds, because all frequencies contained in the original physiological signal would lie in one frequency bin. Larger segments cannot be computed with the given dsPic33 and were not considered. Although the accuracies for the different segment sizes
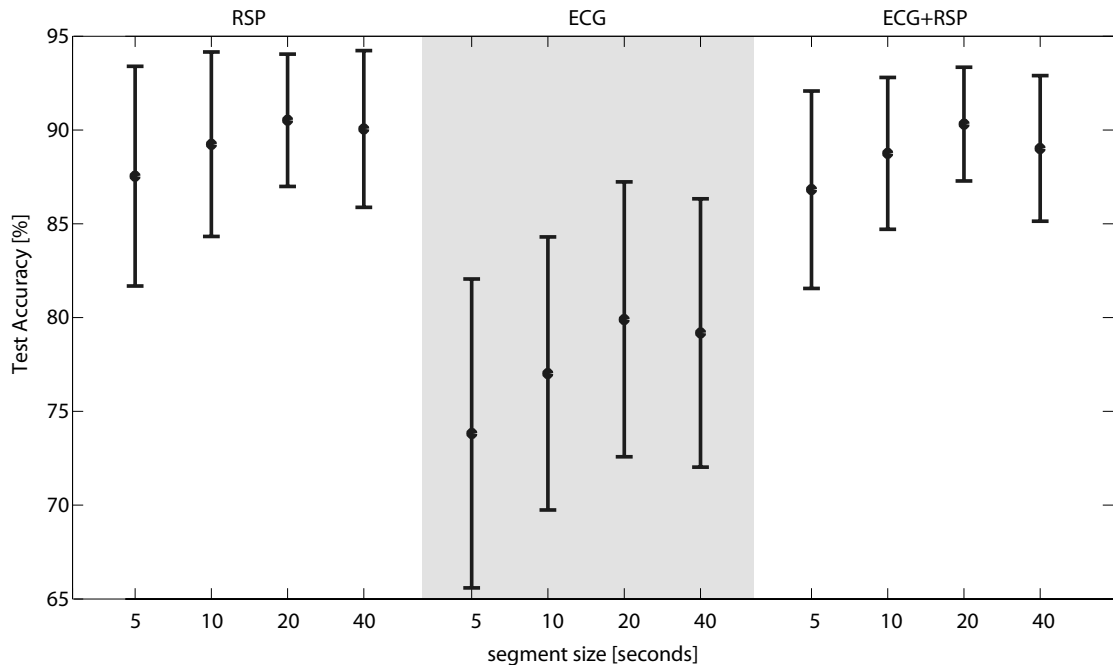
Figure 5.3: Performance analysis for different segment sizes (5, 10, 20 and 40 seconds) used in the preprocessing. The dots are the mean over all trials and subjects, the whiskers the standard deviation. The means of the 20-second segments are better for all three analyzed topologies (ECG, RSP and ECG+RSP).

are not significantly different, we observe a positive trend toward the 20-second segments. This is expressed with the highest mean performance, minimal standard deviation and the highest total accuracy in all three topologies (ECG, RSP, ECG+RSP).

We decided to use 20-second sized segments for the computation of the FFT preprocessing on the wearable device because it was the closest possible segment size to the standard 30-second segment used in polysomnography that freed precious RAM space (Table 5.1).

5.5 Improving Network Topologies

The topology of the ANN networks used in Chapter 4 was hand-coded and the chosen solution was selected for its simplicity and functionality. The single-layer feed-forward neural network that was used gave a functional solution, but no information was obtained whether this topology is the optimal solution. Eventually, a network topology with at least equal performance and less computational needs can be found in an automatic way.

For this optimization problem we used a method which is able to automatically synthesize ANN topologies and the search for the network weights in the corresponding search space (Dürr et al., 2009). The selected network encoding method called Analog Genetic Encoding (AGE) uses evolutionary algorithms for exploring the search space. AGE was developed at the Laboratory of Intelligent Systems (LIS) by Mattiussi and Floreano (2007) and was improved by Peter Dürr for the use with ANNs. AGE encodes artificial neurons and the interaction between them in one single artificial genome.

For our optimization problem the AGE representation was combined with a standard genetic algorithm. The evolved network could connect to an arbitrary subset of the 409 inputs from the ECG data, the 327 inputs from the RSP data and a constant bias unit (Figure 5.4). The deletion of a weight connection in the artificial genome resulted in the removal of the corresponding input from the network. Additionally, the evolutionary process was able to insert hidden neurons into the network in order to generate more complex network structures.

The measure of quality of the evolved classifier was the same as for the hand-coded, single-layer networks (mean square error between the actual and the desired classifier output, Section 4.1.2). The subject-independent 561 ECG+RSP experiment was repeated using the same data as for the hand-designed, fixed topology.

5.5.1   Network Layers

We compared the performance of the ANNs that were evolved with AGE with the networks trained in Chapter 4. Figure 5.5 shows that the evolved networks did not display a significantly different classification accuracy (Wilcoxon rank sum test p=0.48).

An analysis of the topology of the evolved networks showed that out of the 30 evolved networks, 19 featured no hidden neurons, 7 featured one hidden neuron, and 4 featured two hidden neurons. There was no correlation between the number of hidden neurons and the classification accuracy (Spearman's rank correlation coefficient ρ = -0.06, p = 0.74).

These findings showed that the minimalistic approach with no hidden layer can be used without performance drop and supports the previous choice (Chapter 4) of a less computational intensive single-layer ANN topology. A similar finding of using a single layer network for the discrimination of wake and a combination of sleep stages (stage1/REM, stage 2, stage 3/4) based on EEG frequency
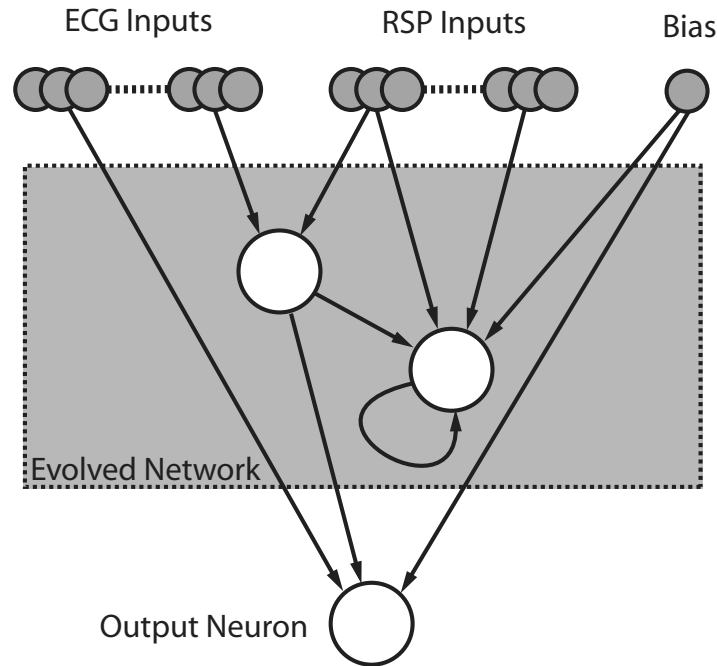
Figure 5.4: Experimental setup for the automatic synthetization of ANN using Analog Genetic Encoding (AGE). An evolutionary algorithm was able to add hidden neurons and to strengthen/weaken the synaptic weights and to add hidden neurons (gray box). Theoretically, the synthetization was able to create recurrent connections between the neurons, but this feature was not used for our experiments. Adapted from Dürr et al. (2009).

features was found by Principe and Tome (1989), where they compared zero, one and two hidden layer networks.

5.5.2 Network Inputs

While the hand-designed fixed topology networks, as used in Chapter 4, employed all of the 736 input features (ECG+RSP), many of the evolved networks used a drastically reduced set of inputs (see Figure 5.6, the median of the number of inputs used is 244.5). There was no correlation between the number of inputs used by the evolved networks and their performance (Spearman's rank correlation coefficient $\rho = 0.02$, p = 0.94). This indicates that many input features are indeed redundant and that it is possible to synthesize networks with a very small number of inputs which perform as well as the hand-designed network using all inputs.
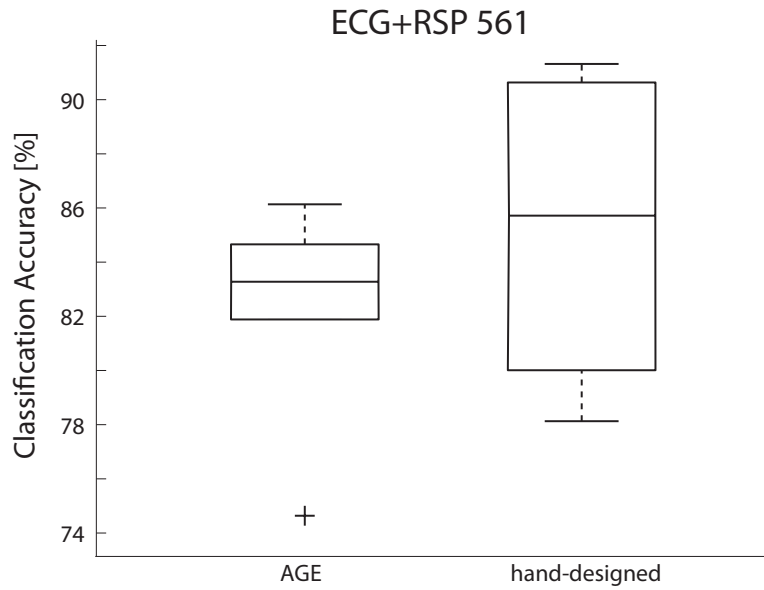
Figure 5.5: Average classification accuracy of the evolved ANNs (AGE) and the single-layer, fixed ECG+RSP topology from Chapter 4 (hand-designed). The midline in each box is the median, the borders of the box represent the upper and the lower quartile. The whiskers outside the box represent the minimum and maximum values obtained, except when there are outliers which are shown as small crosses. The outliers are data points which differ more than 1.5 times the interquartile range from the border of the box. Adapted from Dürr et al. (2009).

Because artificial evolution always selected features from both ECG and RSP (Figure 5.7), the use of the combined ECG+RSP topology is reasonable.

The AGE experiments did not show a clear preference for a particular set of features. A possible reason for not consistently choosing smaller sets is that the genetic algorithm was only evaluated on classification error but not network size.

We analyzed the input network weights of the single layer ANN obtained from the SleePic experiments where the entire frequency spectrum served as input. We observed that the weights of the neural inputs with the higher frequencies of the ECG and RSP spectrum did show a very weak activation and variation compared to the weights of the low frequencies. Based on physiological reasons and because the simple structure of the ANN, we hypothesized that the low activation and variation in the domain above 3 Hz is linked to a reduced importance of these features for the classification. Therefore we designed a different network topology that only uses the most relevant input weights. In our particular case (single-layered network), the input features were considered as relevant when the mean weight over all training runs was larger than the median standard de-
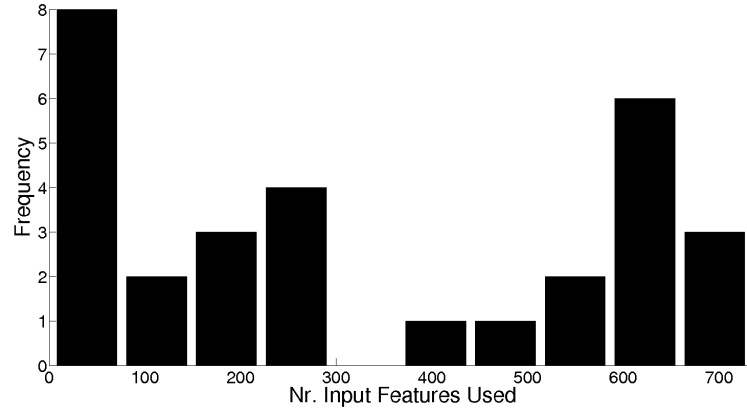
Figure 5.6: Distribution of the number of input feature of the ANN obtained with the 30 artificial evolution runs with AGE. With permission from Dürr et al. (2009).

viation of all layer weights, as follows

$$w_x \equiv \begin{cases} select & \forall \quad mean(w_x) > median(std(w_{1,\dots,N})) \\ discard & \text{otherwise} \end{cases} \tag{5.1}$$

where w_x are the weights of all training runs at position $x \in 1, \dots, N$. The resulting pruned network input numbers and the pruning frequencies for each signal are shown in Table 5.2. The input size of the pruned network was reduced to 8.3% of its original size.

We compared the performance of the reduced networks with the networks using all frequency inputs. Figure 5.8 shows the accuracy for each topology and the original and pruned network inputs. We observed that the accuracies show no statistical difference, but revealed a tendency for a higher median in all topologies with the pruned input. The positive effects in accuracy of the pruned ANNs can be attributed to the reduction of the search space and therefore to the reduction of over-fitting due to the increased noise on the higher-frequency inputs.

We also investigated the use of Principal Component Analysis (PCA) for the reduction of the ANN input vector. The analysis showed that many input features are correlated (redundant). For ECG, 92% of the components contributed to less that 5% to the variation of the data. For RSP, 89% of the components contributed to less that 5% to the variation of the of data. However, the computation of PCA introduced an additional computing step and offered less flexibility for changes of the ANN weights later on. Further, the PCA transformation did not allow us to understand the origin of the principal components completely but
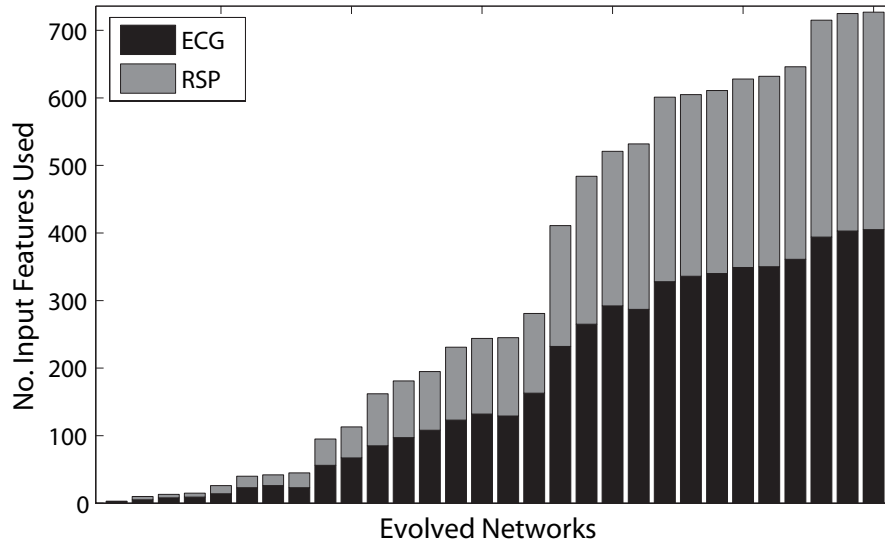
Figure 5.7: The evolved networks with AGE sorted by network input size. Artificial evolution selected independent of input size always features from the ECG and RSP spectrum. With permission from Dürr et al. (2009).

showed that the low frequencies that were selected with Equation 5.1 were also contributing heavily to the principal components.

A smaller input feature set did accelerate the training and the network computation and also reduced the required amount of training input vectors (curse of dimensionality). It was therefore reasonable to use the pruned network for further studies.

Topology	RSP	ECG	ECG+RSP
Frequencies [Hz]			
original	0-12.8	0-25.6	
pruned	0-1.4	0-2.25	
ANN Input Size [number]			
original	257	513	770
pruned	28	46	74

Table 5.2: Properties of the input feature space for the three ANN topologies RSP, ECG and ECG+RSP. The original features are all the features that are available from the preprocessing. The pruned features correspond to the features obtained by applying Equation 5.1.
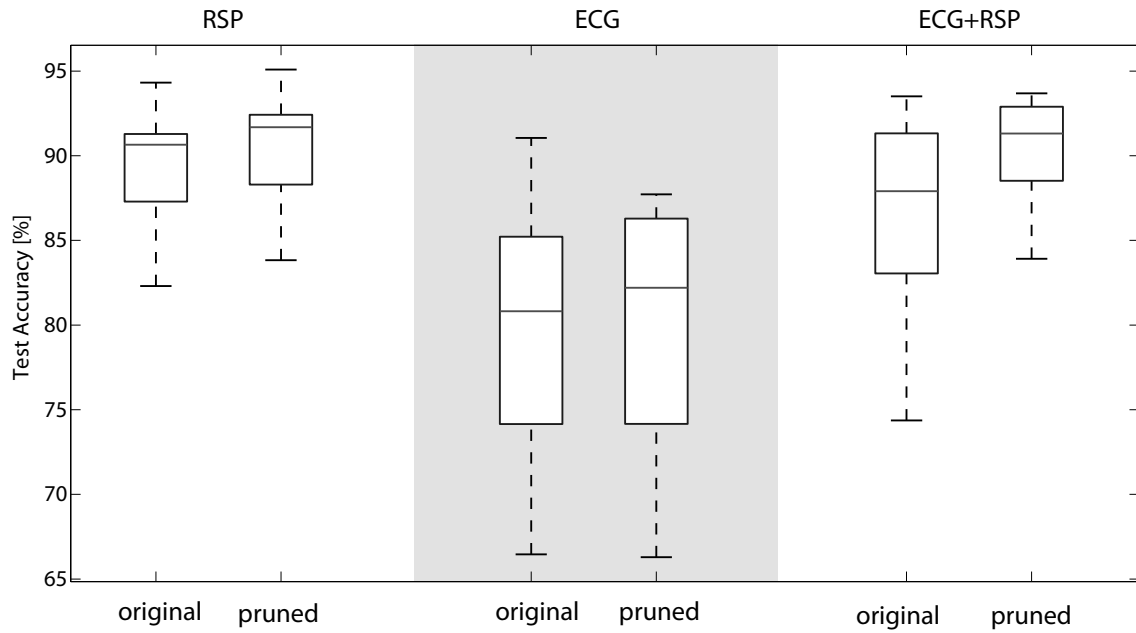
Figure 5.8: Performance comparison when the entire feature set (original) or the reduced feature set (pruned) is used as input of the ANN (subject-specific experiments). Whereas a positive effects of the pruning can be observed on the combined topology (ECG+RSP), no clear advantage for the single signal inputs can be found (ECG, RSP). The horizontal lines of each box are the lower quartile, median, and upper quartile values (from bottom to top). The whiskers represent the most extreme values within 1.5 times the interquartile range from the quartile. The outliers are data with values beyond the ends of the whiskers and are represented with crosses.

5.6 Classification Results

According the findings of the previous sections, a final preprocessing and classification configuration was selected for the wearable sleep / wake classifier: single-layer feed forward networks, FFT window size 20 seconds, pruning of the spectrogram at 1.4 Hz (RSP) and 2.25 Hz (ECG). With this configuration we trained and tested the classifier and compared the results with the results from the portable system presented in Chapter 4. As described in Section 5.2 we tested a subject-specific and a subject-independent classifier.

Subject-Specific

Figure 5.9 compares the accuracy of the wearable classifier with the results of the portable classifier obtained in Chapter 4 for the subject-specific experiments.
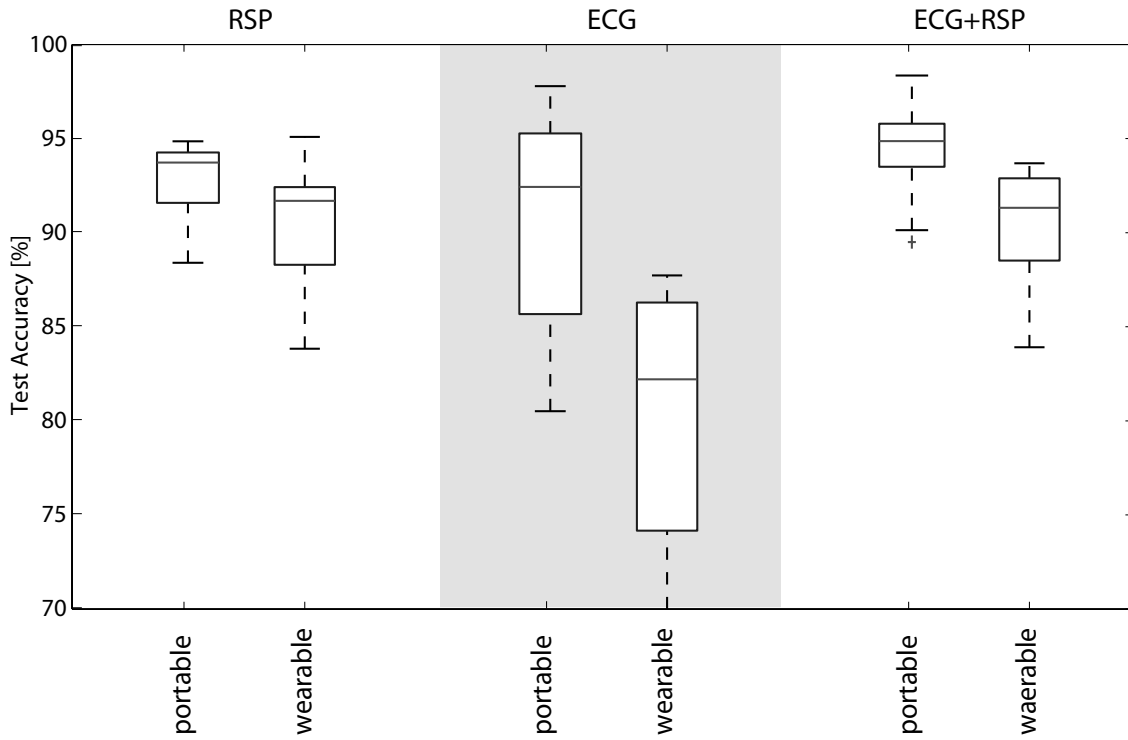
Figure 5.9: Accuracy comparison between the wearable and portable (Chapter 4) classifiers for a subject-specific classifier and the network topologies RSP, ECG, ECG+RSP.

We can observe a drop in the median accuracy for all wearable classifiers and topologies (Student's t-test, $p<0.01$). The most dramatical drop is for the ECG experiments where the median accuracy drops by 10%. The median accuracies of the two other experiments are above 93%. Parts of these drops in accuracy can be due to the reduction of the sampling resolution of the sensors and the reduction of the precision of the FFT calculation. Inter-individual differences might also contribute to this drop, as the wearable experiments were conducted on a higher subject number.

Additionally, the striking drop of performance when using the ECG topology might suggest that the use of dry electrodes instead of gel electrodes has also a negative effect on the classification. If dry electrodes are used, the ECG signal quality depends on the humidification of the electrodes with sweat, which can vary largely also within a subjects recording. As the electrodes are not fixed to the body with glue, movement artifacts become also more pronounced. Unfortunately it was not possible to quantify the individual contributions of the performed changes to the accuracy decrease with the SleePic system configuration and the conducted experiments.
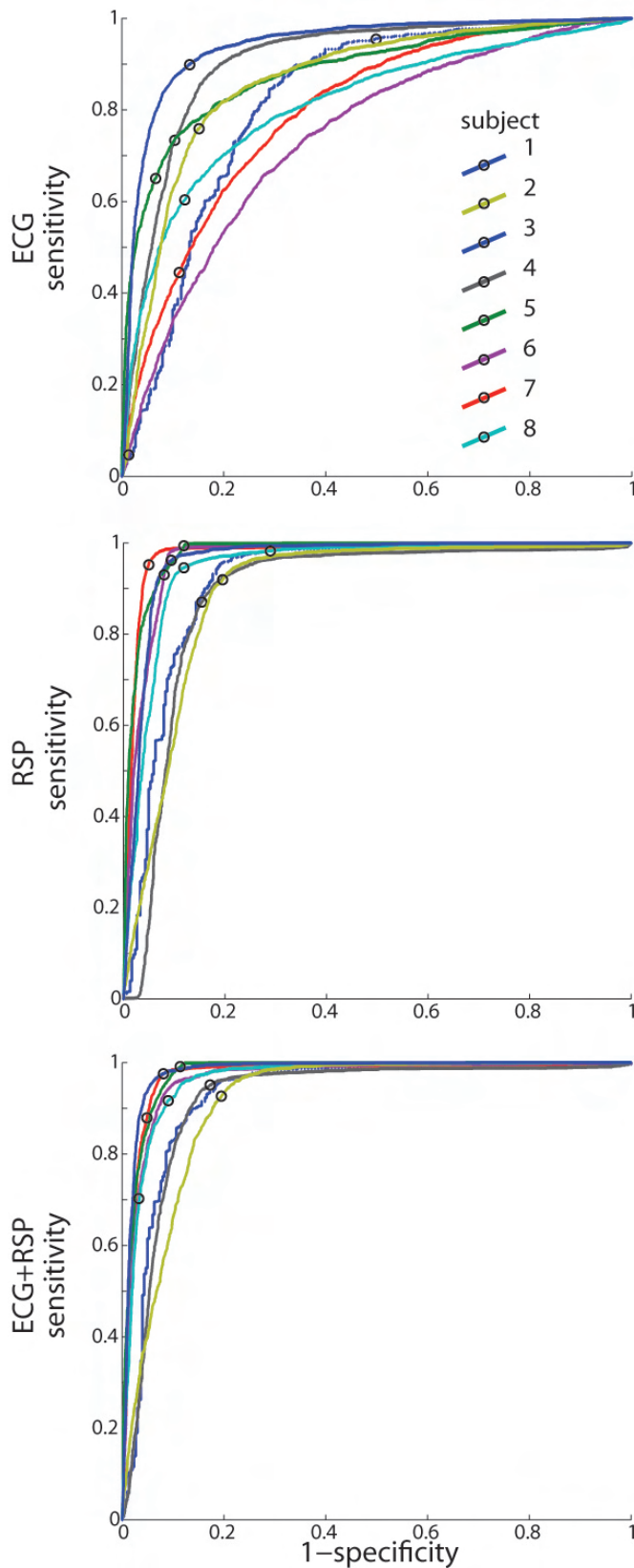
Figure 5.10: Receiver Operating Characteristics (ROC) curves of the three neural classifiers topologies for the subject-specific wearable experiment. Each curve corresponds to the best out of 10 runs from each user. The circles are the ROC points of the actual classification of the test set. Top: The classification with the ECG topology resulted in relative low sensitivities. Middle and bottom: The classification resulted in a relative good balanced sensitivity and specificity for most users.

The ROC curves in Figure 5.10 show that the radius of the curvature for the ECG topology classifiers (top) are much larger than for the two other topologies whose curves show a clear trend toward the (0,1) corner. The majority of the ROC points (circles) in the ECG experiment also show that the actual classifier tends for a low sensitivity while specificity stays in the same range (except subject 1). Compared to the ROC plots of the portable experiments (Figure 4.6) we observe a higher dispersion of the ROC points for the wearable experiments.

Subject-Independent

The classification accuracy of the subject-independent experiments where the test data comes from a new subject who was not present for the training is shown in Figure 5.11. As it was already observed in the portable experiments, the median accuracy for the ECG topology drops significantly compared to the two other topologies (Student's t-test, p<0.01) and also compared to the subject-specific experiment Student's t-test, p<0.01). The increase of the curvature radius of the ROC curves (Figure 5.12 top) and the low sensitivity (47.54% ± 14.2) is even more pronounced than in the subject-specific experiment (Student's t-test, p<0.01). Surprisingly, also the accuracy of the RSP and the ECG+RSP topologies dropped significantly from the subject-independent (Student's t-test, p<0.01) and portable experiments (Figure 5.11. The median accuracy drops for all topologies below 75%, which cannot be considered as very accurate.

An analysis of the ROC curves of the RSP experiment in the middle of Figure 5.12 unveils that subject 3 does show a complete different respiration pattern during sleep (S shaped ROC curve with very low sensitivity ROC point) which cannot be separated from the wake class of the training data set. As the data of this user is used in the training of all other training runs, this has also a negative effect on the classification of the other subject data. This can be either observed in the shift of the ROC points away from the center for all other subjects on the same ROC figure or by looking at the mean square error (mse) of the training. Whereas the mse of the generalization training without the critical subject 3 reaches 48.06 ± 0.01 %, the other trainings including the subject perform worse 53.66 ± 1.61 % (Student's t-test, p<0.01; note that the mse is calculated before the classification thresholding). As this subject showed high accuracies and well balanced sensitivity and specificity in the subject-specific experiments (Figure 5.10), the origin for this performance drop is rather from physiological inter-subject differences than from issues with the recording quality. One can argue similarly for the ECG+RSP

topology experiments where both data sources were combined (Figure 5.12 bottom).
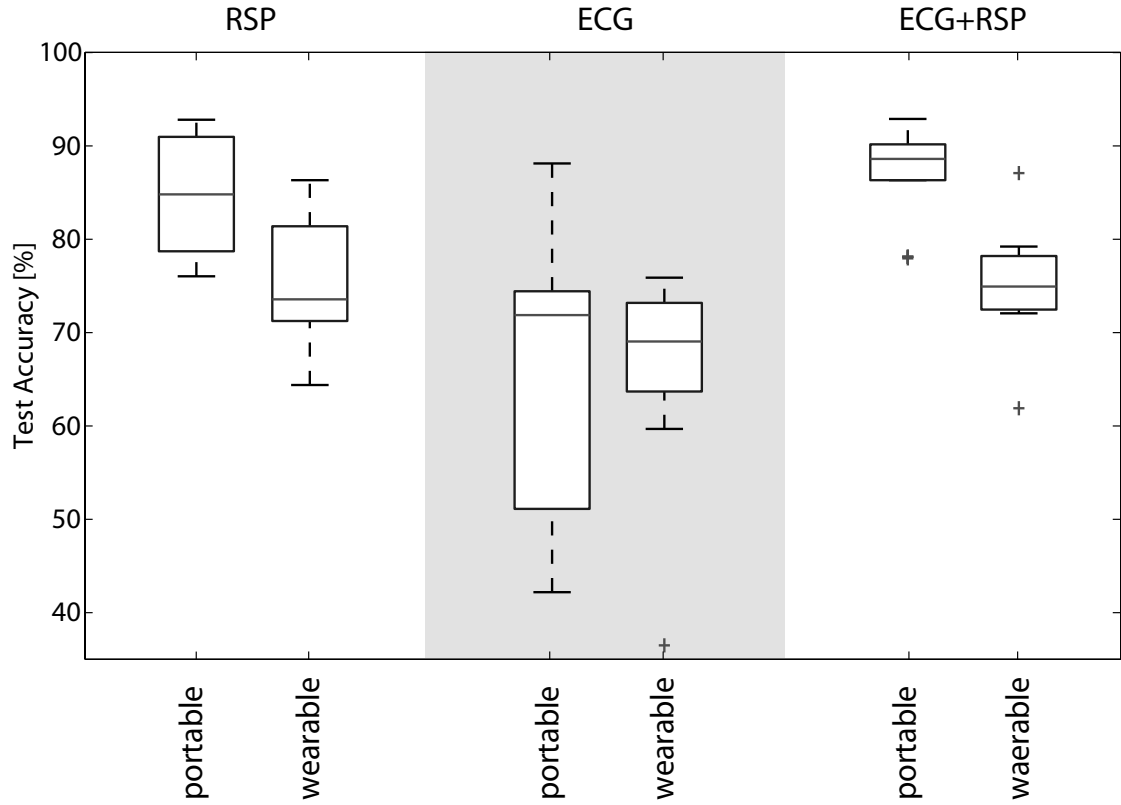
Figure 5.11: Accuracy comparison between the wearable and portable classifiers for a subject-independent classifier.

5.7   Conclusion

The low median classification accuracies obtained from the subject-independent wearable classifier design (<75%) suggest not to use a generalized ANN classifier for the sleep / wake classification with the physiological signals ECG and RSP. The use of such a classifier in a subject-specific setup seems to be more reasonable where the median accuracy for the RSP and ECG+RSP topologies are above 91%. However, certain individuals show accuracy drops of up to 8% also for the subject-specific experiments.

   In general, a performance drop was observed with the transition from a portable to a wearable system. Possible reasons might be the lower recording resolution given by the recording devices (12-bit to 10-bit) or the quality decrease of

Figure 5.12: Receiver Operating Characteristics (ROC) curves of the three neural classifiers topologies for the subject-independent experiment configuration. Each curve corresponds to the best run obtained by changing the training set configuration or the initial network weights and with a different subject in the test set. The circles are the ROC points of the actual classification of the test set. Top: The flattened ROC curves show the degradation of performance. Middle: The S-shaped curve of subject 3 shows that the subject had a different respiration pattern during sleep which was not present in the training patterns. All classifiers show an unbalance of specificity and sensitivity opposite the subject-independent case in Figure 5.10. Bottom : The combination of both signals in the input does not improve the classification compared to the RSP topology.
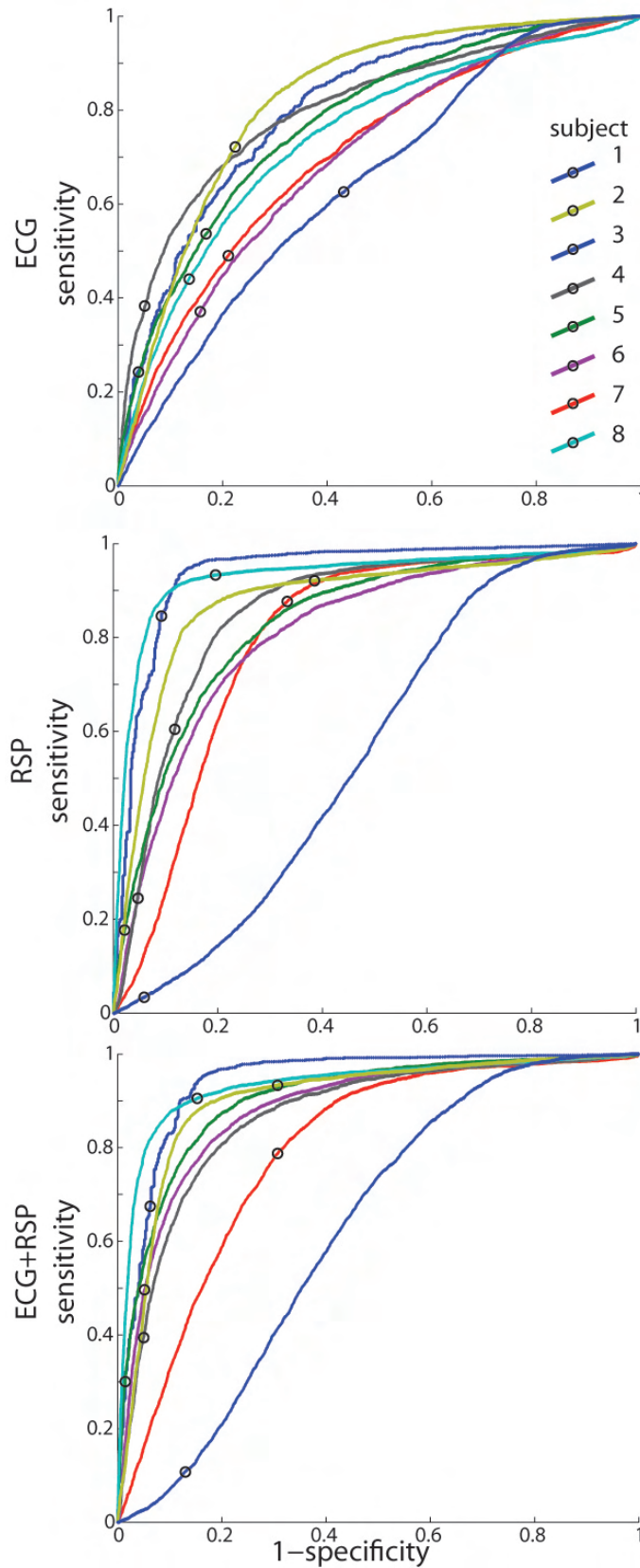
the ECG recordings because of the use of dry electrodes instead of glued gel electrodes. Additional filtering and signal reconstruction might improve the signal quality. The exact quantification of these influences was not possible and requires further investigation.

We observed a performance drop due to physiological differences between the subjects within the tested population. The sample size is not large enough to draw a direct conclusion on this, but with further increasing the population size, more distinct differences between users have to be expected.

These findings indicate that a physiological wearable sleep / wake classification system for an accurate classification requires a priori knowledge about the user. The way how this knowledge was obtained so far (Video, EMG and/or EOG analysis) is not adequate for a wearable system. These procedures only few people would accept to undergo. The labeling of the physiological data also required a human expert. Further, the analysis was time consuming and therefore costly.

To overcome these limitations of generalizability and obtrusiveness, we suggest exploring two paths:

a) The passive detection of sleep behavior patterns with accelerometers has been identified in Chapter 3 as an interesting option for a wearable sleep / wake detection system. In Chapter 6 we are going to evaluate this approach.

b) Eventually, as subject-specific systems clearly performed better, an adaptive algorithm that automatically adjusts the classifier to the user can bring the desired accuracy improvement to a general public device. This aspect will be discussed in Chapter 7.

6

Activity

The subject-independent classification accuracies of sleep and wake that were obtained with the ECG and RSP features in Chapter 5 were relatively low. This raises the question if the behavioral based actigraphy sleep and wake detection methods (Section 2.2.3) would be a good alternative. The advantages of activity based methods lies in its simplicity of recording and analysis. However, as actigraphy is a passive method of measuring a single sleep behavior (Section 2.1), it does not only measure low activity periods related to sleep. To minimize classification errors, actigraphy algorithms have to detect movements in large segments (e.g. Cole et al. (1992)'s algorithm measures the activity on seven 1-minute segments) and require some postprocessing. Still, activity methods do overestimate total sleep duration because of wake patterns with low activity periods that are easily misclassified as sleep (de Souza et al., 2003). In this chapter, a commonly used actigraphy sleep / wake algorithm from Cole et al. (1992) is tested on the data from the SleePic device. Two other algorithms based on body position and

spectral analysis of the accelerometer data are proposed. A fourth algorithm combines the actigraphic with the physiological signals ECG and RSP seen in Chapter 5. The classification performances of the activity based topologies are then compared to the results obtained in Chapter 5.

6.1   Accelerometer Preprocessing and Classification Algorithms

6.1.1   Activity Counts (ACT)

Activity count is the most common and simplest way to obtain activity related features (Gorny and Spiro, 2001). Three main methods exist to get an activity count measure (Figure 6.1):

a) *Time-above-threshold*: calculation of the time the activity is above a certain threshold in a segment.

b) *Zero-crossing*: counting the instances the accelerometer signal crosses a given threshold in a segment.

c) *Digital integration*: The surface below the activity pattern is integrated.

One of these activity counts can then be used for the estimation of how much the subject moved in a given time segment and eventually was asleep.

The model described by Cole et al. (1992) is commonly used for the sleep / wake estimation. This model is based on a weighted sum of a bias b, the activity of the current (A_0), the four previous (A_{-x}) and two future (A_{+x}) 1-minute activity segments as follows

$$D = b + w_{-4}A_{-4} + w_{-3}A_{-3} + w_{-2}A_{-2} +$$
$$w_{-1}A_{-1} + w_0A_0 + w_{+1}A_{+1} + w_{+2}A_{+2}$$

(6.1)

This algorithm can be optimized with logistic regression or discriminant analysis by updating the weights $w_{-4...+2}$ and using a constant threshold for D (Cole et al., 1992 and Edward et al., 2004). It is designed for off-line use, as it makes use of future data. Other algorithms like Artificial Neural Networks (ANN) (Edward et al., 2004) or statistical methods (Jean-Louis et al., 1996) have also been suggested.
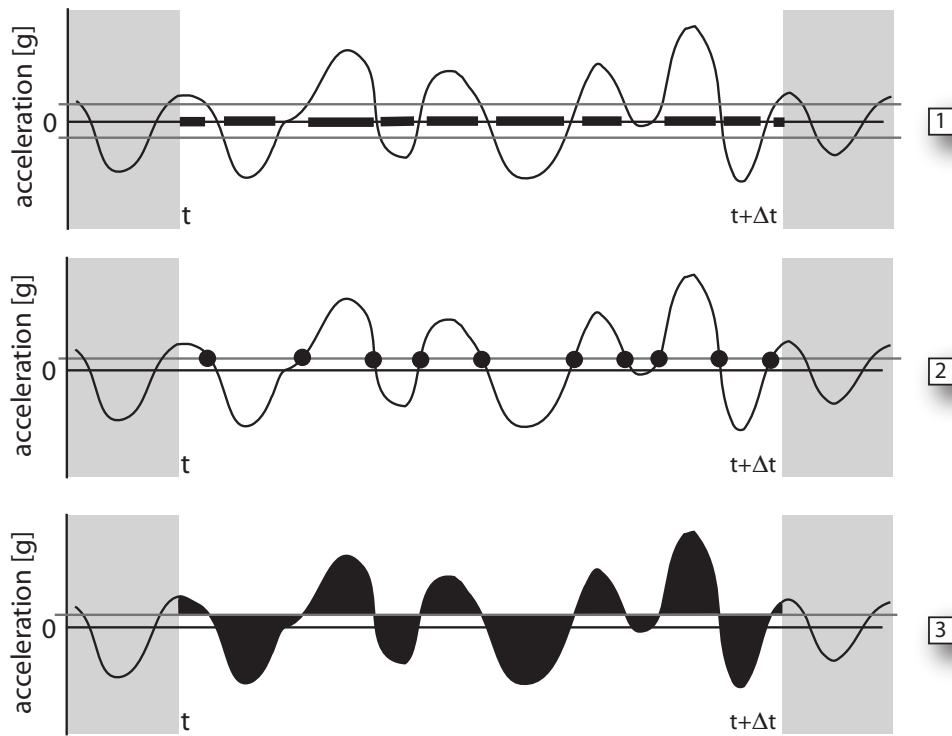
Figure 6.1: Three most commonly used algorithms to calculate activity: 1) Time-above-threshold; 2) Zero-crossing; and 3) Digital integration. The sum of the black units gives the activity count for a segment of length Δt. The horizontal gray line is a threshold to which the signal is compared.

For our implementation of the activity count algorithm, we applied a 0.5-11 Hz bandpass filter on the raw accelerometer signal to remove slow (gravitational) movements and excessive accelerations that are not related to human movements. On the filtered signal we applied the zero-crossing algorithm (Figure 6.1). Equation 6.1 can also be represented by a single-layer ANN having A_n as inputs and one output neuron. This allowed us to keep the previously used ANN classifier design and to change only the input. However, we kept also the non-linear tangent-sigmoid transfer function, which was more adapted for a classification problem, but therefore had to alter slightly Cole's model. We did not use any postprocessing as it has been suggested by Cole et al. (1992) and Webster et al. (1982), because the suggested postprocessing rules (like "10 minutes or less scored as sleep surrounded by at least 20 minutes (before and after) scored wake are re-scored wake" Webster et al. (1982)) added an additional delay of up to 20 minutes to the classifier. Further, in Cole et al. (1992) the postprocessing was designed to improve only the bad specificity performance and had only little effect on the accuracy (+ 1%) because the tested data contained little wake. If

such empirically designed postprocessing rules are desired for a particular application, we suggest to use recurrent connections in the ANN network and to learn the time related dependence based on the train data set. We did not investigate this option in this thesis because the increase of complexity due to the recurrent ANN. Recurrent networks in a microcontroller lead beside other problems to difficulties in presenting the time-dependent data to the network for training and to a significant increase in training time.

6.1.2   Body Position (POS)

Sleep behavior is linked to typical positions (Section 2.1). Sleep should therefore be detectable by recognizing typical sleep positions. Body position measurements are often used in context aware applications (e.g. Stiefmeier et al. (2008)), but usually accurate posture information is needed which requires the measurement of position from different body parts. In a wearable system this requires multiple, distributed sensors that form a body area network (BAN). Because BANs have limited energy autonomy, we focus here only on the detection of the main body orientation. As the SleePic device measures the acceleration on all three body axes, the body orientation can easily be determined with the DC values of the accelerometer signals.

The proposed position algorithm (POS) calculates the mean acceleration for 20-second segments for each axis as follows:

$$POS_{X,Y,Z}(s) = \frac{\sum_{n=0}^{N-1} a_{X,Y,Z}(n)}{N} \tag{6.2}$$

where $a_{X,Y,Z}$ is the raw accelerometer output and N is the number of samples that form a segment s. Figure 6.2 shows the outcome of this algorithm for one recording session. A single-layer ANN was designed with the three $POS_{X,Y,Z}$ values as inputs and one output neuron.

6.1.3   Spectral Analysis (ACC)

Analog to the ECG and RSP algorithms in Chapter 4, we introduced the power spectral density calculation for the accelerometer signal (Equation 4.2). As the accelerometer had the same recording configuration as the respiration signal (10 bit, 25.6 Hz), the FFT preprocessing was calculated using the same procedure
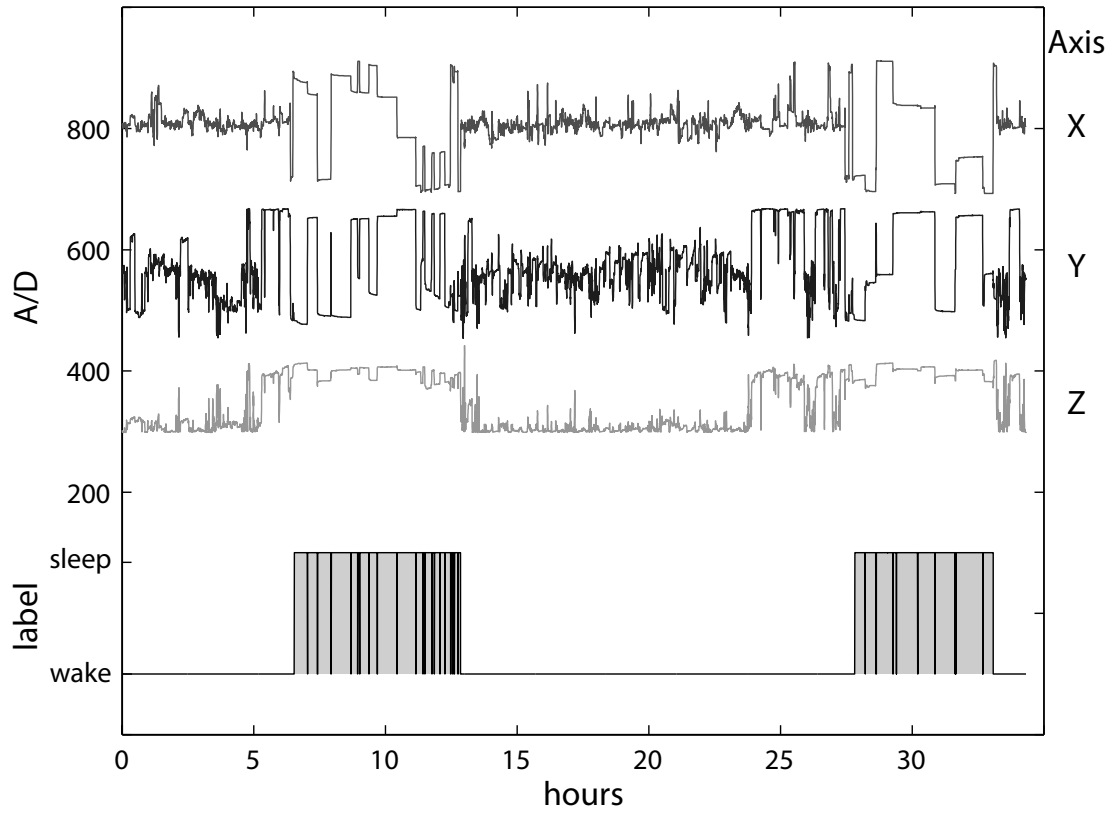
Figure 6.2: Position features for the three accelerometer axes X,Y,Z for one subject recording session. The feature is calculated by taking the mean of each of the three accelerometer outputs over a 20 second segment. Bottom: Corresponding video sleep / wake labeling of the recording session. The amplitude change of the Z axis shows that the subject went to a lying position for 4 hours before falling asleep the second night (24 hours after the session started and with interrupts at 26 hours).

($N = 512$). The input feature vector was also pruned using the same statistical method as described in Section 5.5.2. The resulting configuration (ACC) is shown in Table 6.1. The same single-layer ANN topology was used as in the RSP experiments in Section 5.6.

Combination of Physiological and Actigraphic Signals

To observe the possible effects of combining features of two distinctive type of data sources (physiological and behavioral), we combined the RSP, ECG and ACC spectral features to one set of ANN inputs (ECG+RSP+ACC).

Topology	ACT	POS	ACC	ECG+RSP+ACC
Frequencies [Hz]				
original			0-12.8	
pruned			0-1.4	
Segment size [s]				
	7*60	20	20	20
ANN Input Size [number]				
original	7	3	257	-
pruned	-	-	11	85

Table 6.1: Properties of the input feature space for the four ANN topologies ACT, POS, ACC and ECG+RSP+ACC. The original features are all the features that are available from the preprocessing. The pruned features correspond to the features obtained by applying Equation 5.1.

6.2   Experiments

We repeated the subject-specific and subject-independent experiments using the data recorded with the SleePic system as it was described in Section 5.2.3.

6.3   Results and Discussion

Figure 6.3 shows no significant difference between the subject-specific and the subject-independent accuracies for each accelerometer topology. This shows, contrary to the RSP and ECG features used in Chapter 5, that the sleep behavior measured with the accelerometer signals, independent of preprocessing, generalize well between users. Interestingly, the subject-independent ACT, POS and ACC experiments show a narrower accuracy range (lower SD). This improvement of the minimal accuracy and decrease in maximal accuracy also suggest that these experiments tend toward a generalized solution.

The activity count algorithm topology (ACT) shows a median accuracy of 80% for the subject-independent experiments (Figure 6.3). The accuracy is better than the topologies based on cardiorespiratory features in Section 5.6 (Figure 5.11; Student's t-test, $p < 0.01$). However, the ACT topology shows a lower accuracy compared to the other actigraphic classification methods (Figure 6.3; Student's t-test, $p < 0.01$). The corresponding ROC curves (Figure 6.4 top) show the limitation in specificity which were also reported in literature (de Souza et al., 2003). This is manifested by the large steps in the ROC curves, in which no change in the speci-

ficity appear. Note that the steps are initiated for each subject at the same speci-
ficities in the subject-independent and the subject-specific experiments. For these
experiments we observe a lower sensitivity than in literature (>90%, de Souza
et al. (2003) and Edward et al. (2004) and Sadeh and Acebo (2002)). A possible
reason might be the placement of the accelerometer on the chest, which differs
from most studies where the placement on the wrist is preferred. Hilten et al.
(1993) and Middelkoop et al. (1997) showed that the different placements of acti-
graphs have an effect on the classifier accuracy. However, Edward et al. (2004)
made a successful use of actigraphs on the chest of infants, arguing the advan-
tage of being able to measure additionally the body position with the same sen-
sor. Another reason might be the fact that the previous study results were based
on clinical data where the subjects' movements did not correspond to real world
settings.

Surprisingly, a high accuracy was obtained by only using the minimalistic
position information (POS) for the classification. This can be explained by that
the subjects were sleeping mainly in a horizontal position. Only one tested sub-
ject showed sleep in another position (and was not correctly classified), which is
manifested by the single ROC point with lower sensitivity in the middle left of
Figure 6.4. Interestingly, for the POS topology, many ROC points lie at the Sec-
tion where the ROC curves start to leave the 100% sensitivity. The specificities of
these points do correlate with the sleep efficiency (Table 6.2) of the corresponding
subjects (Spearman's rank correlation coefficient ρ = 0.93, p = 0.007). This, the
high sleep efficiency, and the very low number of awakenings in Table 6.2 sug-
gest that the POS topologies are rather trained for detecting 'lying-in-bed' than
'sleep'. On the contrary, the other topologies show a lower correlation between
sleep efficiency and specificities (Spearman's rank correlation coefficient ρ < 0.80,
p>0.05) and a higher amount of awakenings which suggests that these classifiers
do not over-fit to a body position.

When the frequency spectrum of the accelerometer (ACC) is used, the increase
of ANN input features smoothen the ROC curves (Figure 6.4, bottom). Whereas
the accuracy does not decrease significantly from the POS topology experiments,
this method allows obtaining classification results with higher specificity (with
costs in sensitivity). Comparing the ACC ROC curves with the ones from the
ACT and POS topologies, one could hypothesize that the ACC combines differ-
ent features of both topologies. This is supported by the fact that the zero-crossing
feature extraction method used for the ACT preprocessing can be seen as a fre-
quency measure which was also included in the spectral representation. Further,
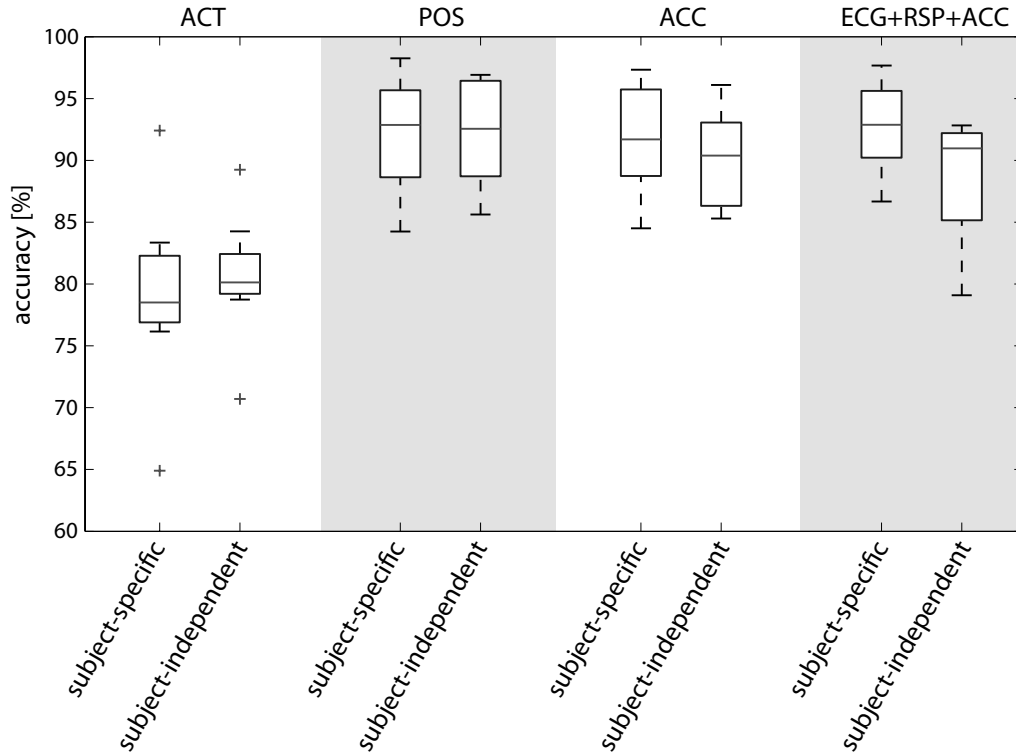
Figure 6.3: Accuracies for the topologies using different accelerometer features (ACT, ACC, POS, ECG+RSP+ACC). The boxes in the left are the results from the subject-specific experiments, the ones in the left from the subject-independent. The horizontal lines of each box are the lower quartile, median, and upper quartile values (from bottom to top). The whiskers represent the most extreme values within 1.5 times the interquartile range from the quartile. The outliers are data with values beyond the ends of the whiskers and are represented with crosses.

the mean position value from the Y axis used in the POS topology (POS_Y) is part of the 0-0.025 Hz frequency band used as an input feature for the ACC topology.
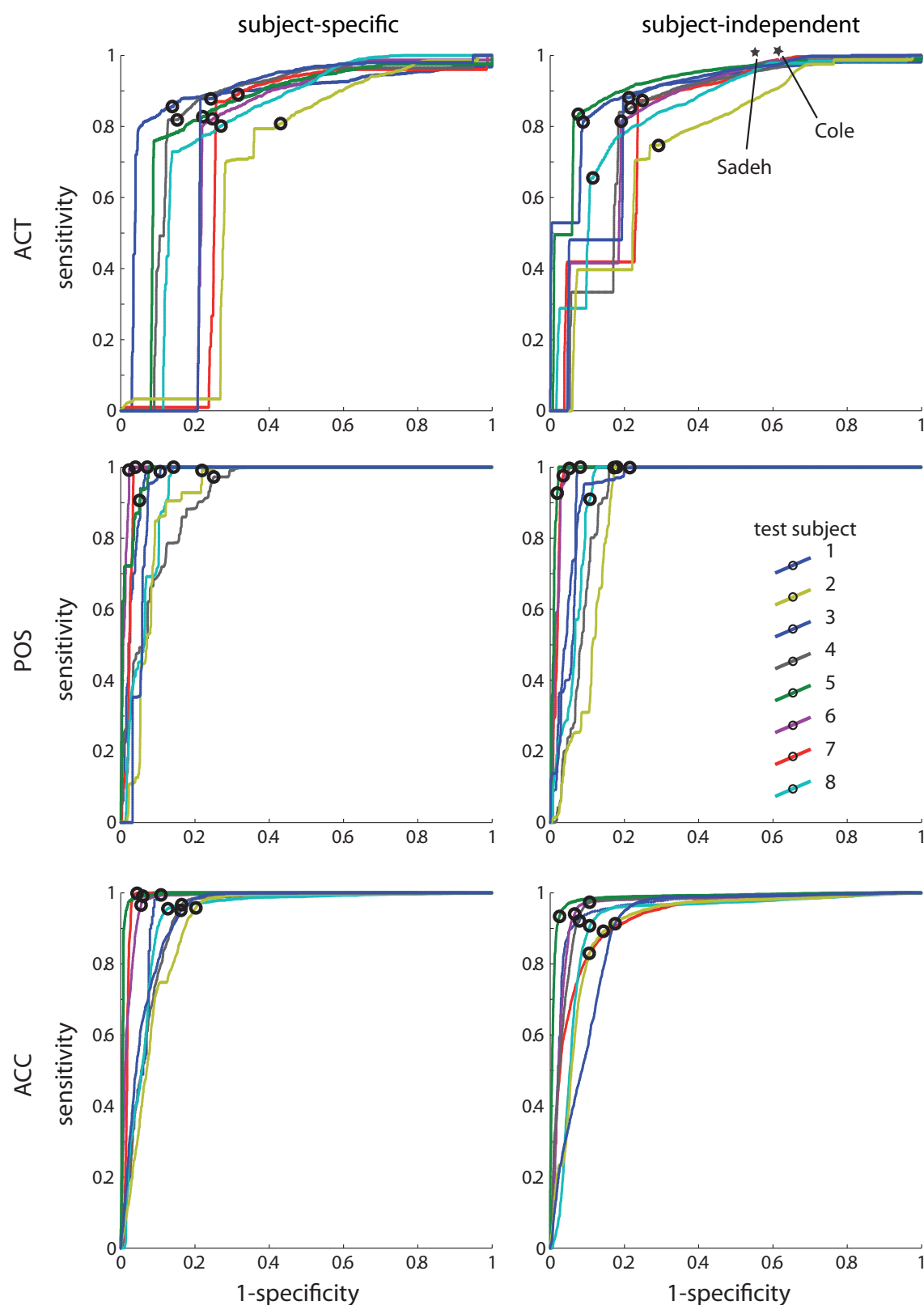
Figure 6.4: ROC curves for the accelerometer preprocessing (ACT, POS and ACC) experiments. Top right: The stars correspond to the ROC points obtained by de Souza et al. (2003) using the two actigraphy algorithms proposed by Cole et al. (1992) and Sadeh et al. (1994).
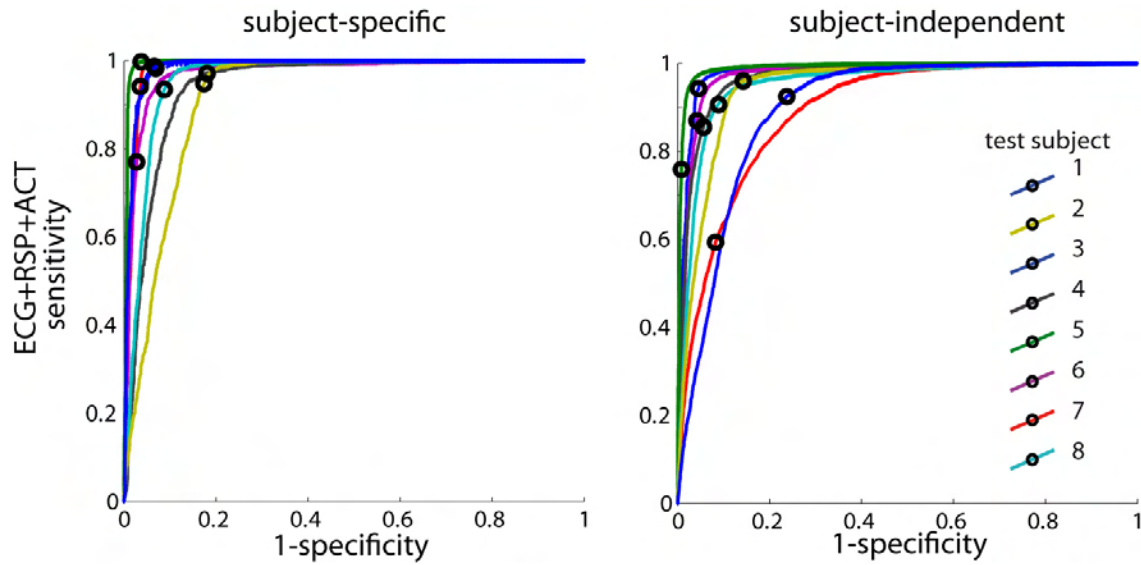
Figure 6.5: ROC curves for the combined architecture ECG+RSP+ACC. Because of large inter-subject variations in the ECG and RSP features, the ROC curves are flatter and the ROC points are more scattered in the subject-independent experiments.

The combination of the physiological features ECG and RSP with the ACC (ECG+RSP+ACC) shows no significant difference in accuracy (Student's t-test, p=0.7), although mean accuracy is slightly higher and the SD lower compared to the ACC topology in the subject-specific experiments. For the subject-independent experiments, the over-fitting due to the physiological signals shown in Section 5.6 is still present (Student's t-test, p=0.09). However, the use of the ACC make the classification significantly better than when using only physiological signals (Student's t-test, p<0.01). This improvement is also observable on the ROC plots in Figure 6.5.

6.4   Conclusion

Accelerometer based features have shown good classification results for subject-specific and subject-independent experiments. Because accelerometers measure sleep and wake behaviors, the sleep / wake classifier generalize well between the studied users. However, because the measured behaviors are not unique to sleep, the classifiers show a tendency for low specificity, especially for subjects which have low activity during wake or show low sleep efficiency. These observations have also been shown previously by many other research groups (de Souza et al.,

Sleep parameters	Mean ± SD	Mean difference from Video ± SD
1. Total Sleep Time (TST) [hours]		
Video (label)	6.06 ± 0.94	
ACT	6.99 ± 1.48	-0.93 ± 1.39
POS	7.11 ± 1.46	-1.05 ± 1.21
ACC	6.69 ± 1.09	-0.63 ± 0.85
ECG+RSP+ACC	6.13 ± 1.62	-0.07 ± 1.62
2. Awakenings [numbers]		
Video (label)	28.07 ± 11.7	
ACT	30.07 ± 13.57	-2 ± 10.67
POS	1.64 ± 1.70	26.42 ± 11.46
ACC	114.61 ± 55.40	-86.53 ± 53.48
ECG+RSP+ACC	176.71 ± 89.83	-148.64 ± 92.38
3. Sleep Efficiency [%]		
Video (label)	82.81 ± 10.01	
ACT	94.76 ± 16.31	
POS	94.9 ± 4.32	
ACC	89.38 ± 3.76	
ECG+RSP+ACC	82.16 ± 12.01	

Table 6.2: Comparison of sleep parameters for the subject-independent experiments of the ACT, POS, ACC and the ECG+RSP+ACC topologies with the labeling from the human expert (Video). The sleep efficiency is calculated by *Time In Bed* over *Total Sleep Time*.

2003 and Edward et al., 2004 and Sadeh and Acebo, 2002).

As the state of the art highlighted in Section 2.2.3, classifiers using only activity related features do not have the potential to extract information about sleep disorders (Sadeh and Acebo, 2002) and overestimate Total Sleep Time. To make advantage of the generalization properties of the actigraphy features and the specificity of the physiological features we introduced a combined ANN architecture that uses frequency features from ECG, RSP and ACC. Whereas this architecture shows very promising results for the subject-specific experiments, due to the poor generalization capabilities of the physiological signals, the subject-independent experiments do not perform as good as when using accelerometer features only. However, the combined architecture still shows a significant improvement compared to the only physiological topologies seen in Chapter 5.

7
Adaptation

In Chapter 5 we have seen that when using RSP and ECG signals for a wearable sleep / wake classifier, no accurate subject-independent ANN could be trained. Further, when using current labeling techniques (video and electrophysiological recordings), the supervised training of a subject-specific network was not practical and not welcomed by the users. To overcome these limitations and to design a wearable system which can potentially classify sleep and wake in anybody without intrusive or inconvenient calibration methods, we present in this chapter two possible adaptation strategies. The core innovation of these strategies concerns the methods how to label the sleep / wake behavior of the subjects automatically and without being intrusive. We found two simple automatic labeling methods using a priori knowledge from sleep behaviors, which need little or no user interaction, respectively. At the end of this chapter we will discuss the performance of

the adapted networks.

7.1   User Adaptation

In automated medical diagnosis, the physiological differences of human subjects are normally taken into consideration by preliminarily classifying the subjects into different groups, e.g. groups for gender, age or pathologies. For each group, a different diagnosis or prediction model is then developed. Often the separation of these groups is diffuse and individuals may belong physiologically to another group than the data used for the grouping might suggest. For example, a well trained, healthy 60 year old man may, from a physiological point of view, rather belong to the age class of 35-45 instead 55-65. Therefore, as the physiological differences occur gradually and are subject-specific, the discrete separation into groups leads to errors in the model selection and eventually also in the model outcome.

In mobile applications, user adaptation is limited to systems that adjust the high-level response to the user. For example, the training model of a personal assistant fitness trainer is adjusted according the average amount of training the user performs (Buttussi and Chittaro, 2008) or the time and location for medication intake reminders are carefully selected according to the user's preferences (Kaushik et al., 2008). In principle, these approaches are not real adaptation processes, as they only select the predefined output of the applied model to the changed context of the user, but not the static rules to obtain the desired output. These types of implementations have difficulties to give accurate output to previously unseen situations.

In automated sleep scoring and sleep disorder classification from polysomnography recordings, a common approach is the use of large databases containing signal samples for different groups (Penzel et al., 2001). Then complex expert systems classify the bio-medical signals by clustering them into specific groups already present in the database (Anderer et al., 2005 and Penzel and Conradt, 2000). This approach requires a huge amount of processing and storage resources. Also, although the database is extended with each new scoring, it is never complete. Further, the need for large data sets with accurate pre-labeled data requires considerable time investments and human intervention. A database look-up approach is therefore not conceivable for an autonomous wearable system.

Hedner et al. (2004) suggested an adaptation method for off-line actigraphy

analysis of sleep and wake. The density of background movements of the subject are used to adjust the activity count threshold on the digital integration (see Figure 6.1) and a second threshold that discriminates the activity count between sleep and wake for a 5-minute segment. However, there are two fundamental problems. First, the density of background movements is calculated for the entire night. This makes it impossible to determine the two thresholds before the recording ended. Second, the background movements need to be calculated over the duration where the subject sleeps. But because this duration is defined by the classifier whose threshold is not yet determined, an alternative measure must be used. Hedner et al. (2004) do not state explicitly over which duration the background movements were calculated ("throughout the night"). We suspect, as the recording system was only worn for one night per subject, that the background movements were either computed for the whole recording duration or for the sleep time determined from the polysomnography reference recording. Both ways of gathering this a priori knowledge are not practical or available in wearable applications.

An inspiring approach comes from the field of brain-computer interaction (BCI) where the typical EEG patterns reveal important intra-subject and inter-subject differences. The presence of error-related potentials (ErrP) in the EEG has been used to improve the performance (bit rate and accuracy) of the brain-machine interface (Schalk et al., 2000). ErrP are negative potentials that can be detected in the fronto-central EEG within half a second after the occurrence of an error. The ErrP can be observed at the occurrence of either human related errors (response and feedback ErrP described by Blankertz et al., 2003 and Holroyd and Coles, 2002 and Schalk et al., 2000) or machine related errors (interaction ErrP described by Ferrez and Millán, 2005). In a BCI problem, where intra-subject EEG differences within a recording session are detectable (Buttfield et al., 2006), the to be performed action of a computer can be adapted according to the presence or absence of the ErrP. For example, in Ferrez and Millán (2005) a brain-controlled virtual robot is executing a previously classified direction command only if the visual user-feedback did not generate an interaction ErrP in the users EEG. Another interesting way to use ErrP in BCI is to use the presence of ErrP as reinforcement feedback for adapting the classifier on-line (Chavarriaga et al., 2007).

Ideally, a change in the physiological signal similar to ErrP, but with respect to the sleep / wake detection, would lead to an automatic association of new patterns to sleep and wake of the user. Such an approach would be advantageous because the user would not need to enter personal data to select an appropri-

ate model. However, the sleep and wake classification operates in a longer time frame than the delays of ErrP. Further, due to the difficulty of acquiring EEG signals in a wearable setup, this approach does not seem to be realistic.

Therefore, a new approach comprising a dynamically adapted classification is required, which changes according to the user's context. For example, the system should adapt to changes in lifestyle of the user, to a new health condition, but also to a new user.

7.2 Adaptation Strategies

Two basic ANN adaptation strategies are possible in our microcontroller environment. We describe them in the next sections.

7.2.1 Modifying Classification Threshold

A possible strategy is the adjustment of the classification threshold, which corresponds of moving the classification point on a ROC curve. The optimal threshold is determined by selecting the point with the highest accuracy for the new data. This very simple method, where only one parameter in the classification model needs to be adjusted, is however very limited and does not allow to change major classification criteria. In the off-line tuning that was described in Karlen et al. (2009), the mean accuracy could only be increased by 1.43% for the given ANN topology. With an on-line method, where in general less data points are available, we cannot expect to increase this gain in accuracy.

7.2.2 Updating Neural Weights

The supervised modification of the synaptic weights to reduce the classification error is the technique that has already been used for the training of the ANN by using a back-propagation learning algorithm in Chapter 4 and 5. It is also possible to do this with the new automatically labeled data. The difference is that the adaptation data set is smaller than the initially data set used for the training. Further, because of the RAM limitations in the SleePic microcontroller, the adaptation of the classifier cannot be done in batch mode by presenting all data simultaneously, but the labeled data is presented sequentially or semi-sequentially (Randall Wilson and Martinez, 2000). The use of sequential adaptation changes the way how the search space is explored and requires substantial fine-tuning of

the training parameters for the back-propagation algorithm. The unsupervised update of neural weights (e.g. with clustering) is also possible. However, clustering techniques are in general computational expensive and require a large data set. Both conditions are not fulfilled with our wearable system.

Since it has been shown in Chapter 4 that the intra-subject differences of the sleep physiology are less important than the inter-subject differences (for inter- and intra-subject EEG pattern stability in deep sleep over several weeks see Buckelmüller et al., 2006), it is reasonable to update the neural weights not continuously on-line, but at regular intervals using a semi-sequential method. In order to compare the results of the adaptation experiments with the results from the subject-independent experiments obtained in Chapter 5, we selected the adaptation interval duration equal to the recording duration (approximatively 36 h). Experiments not described in this thesis showed that the accuracy of the ANNs that were adapted on-line for the duration of the recording was not different compared to the accuracy of the ANNs trained with the suggested semi-sequential method.

7.3 Automatic Sleep and Wake Labeling

To adjust the classification performance of an ANN in a supervised manner, prior information about the users' sleep and wake states is needed. The video analysis used until now or any other known sleep / wake discrimination methods described in Chapter 2 (EEG, PSG, actigraphy) would need some off-line analysis by a human or machine expert and they are not suitable for an on-line, wearable system. Equally, unsupervised clustering methods like the one used in Huynh and Schiele (2006) and Krause et al. (2006) are too computationally intensive. Further, it is very unlikely that unsupervised training can find a more accurate classifier than a supervised, subject-specific training as described in Chapter 4.

So far, nobody has found a unique change in physiological signals that can accurately and automatically predict a difference between the sleep and wake behavior. Therefore, we suggest creating a new way to gather on-line information, which can automatically label the recorded physiological data to adapt the classifier. This new way of labeling the data should comply with the following criteria:

- High accuracy of the labeling: The trained classifier can only be as good as the a priori knowledge used for training. Imprecise labeling will cause

training on wrong data patterns.

- Low power: The information gathering should be possible with equal computational and sensing resources than what are used for classification.

- Low user interaction: The less the subject is disturbed in her/his daily activities, the more the device is going to be accepted.

We present in the following two sections two new algorithms for the automated, on-line labeling that can be used together with the previously presented SleePic device (Chapter 5):

a) *Button Feedback*, which requires minimal user interaction with the system; and

b) *Activity Feedback*, which automatically labels the data based on activity data of the user.

7.3.1 Button Feedback

The basic idea of this approach is to exploit a behavioral phenomenon of sleep (see Chapter 2), which is the elevated arousal threshold. Under the effect of sleep, the user does not respond to external stimuli of weak or moderate intensity (Carskadon and Dement, 1989). The response to a stimulus is therefore evidence for being awake. Possible stimuli can be auditory, visual or tactile. The difficulty lies in finding the right stimuli and the corresponding intensity for not waking up the person, especially if he or she is in the beginning of light sleep (sleep stage 1, Section 2.1) where the arousal threshold is lower (Carskadon and Dement, 1989). If the stimulus is meaningful to the subject, the arousal threshold is also expected to be lower than normal (Oswald et al., 1960 and Williams et al., 1966).

We suggest stimulating the subject with a blinking LED on the SleePic Watch and simultaneously with a single, light vibration on the chest, neither of which can be perceived during sleep under normal circumstances. This stimulation is randomly generated by the SleePic Core module every 15 to 60 minutes. If the subject reacts to this stimulation by pressing a button on the SleePic Watch within one minute of blinking ($button = 1$), he/she is considered to be awake ($l = -1$). If a response is missing ($button = 0$), the subject is either asleep or he missed the stimulus. In that case, the activity (act) within the stimulus period, as it was introduced in Section 6.1 is analyzed. If it is below a threshold of 0.1, the user is considered to be asleep ($l = 1$) and otherwise, no labeling is carried out ($l = NaN$)

because the subject probably missed the stimulus due to her/his activity. This labeling method can be summarized as:

$$l(button) = \begin{cases} -1 & \text{if } button = 1 \\ 1 & \text{if } button = 0 \text{ and } act \leq 0.1 \\ NaN & \text{otherwise} \end{cases} \tag{7.1}$$

The physiological data from the 20 seconds segment previous to the stimulus were labeled to avoid training on the button pressing movement patterns which may eventually arise in the signal patterns.

7.3.2 Activity Feedback

From our studies on actigraphy in Chapter 6 we know that certain movement patterns of a user can reliably be associated to his/her sleep and wake states. We therefore propose to use typical movement patterns from actigraphy to obtain a number of labeled physiological data segments without any user interaction. Inspired from the actigraphy algorithm of Cole et al. (1992), we decided to examine four prior and two posterior activity data segments of 1-minute size (Equation 6.1). If the activity of each of the segments in the resulting 7-minute window is very low ($\leq$0.1) or high ($\geq$1), we consider the central 20 seconds as sleep or wake, respectively. Otherwise we do not label the data ($l = NaN$). This can be formulated by:

$$l(act_n) = \begin{cases} 1 & \text{if } act_{n-4,\dots,n+2} \leq 0.1 \\ -1 & \text{if } act_{n-4,\dots,n+2} \geq 1 \\ NaN & \text{otherwise} \end{cases} \tag{7.2}$$

This labeling is more conservative than the algorithm of Cole et al. (1992) because it requires all seven segments to have low activity (and not only the sum of them) and it does not label uncertain segments.

7.4 Recordings

To generate an adaptation data set with the feedback methods presented in Section 7.3 the SleePic recordings described in Section 5.2 were used. To record Button Feedback data, we conducted a series of reaction tests during the recordings.

The subjects were instructed to press the button as soon as they saw the blinking LEDs on the SleePic Watch. A light vibration on the chest coming from the SleePic Sensor Module reminded them to look at their watch, but the vibration was not perceived when the subject was very active or when she/he was sleeping. To motivate the button pressing, an award for the study participant with the least missed blinking was promised.

For the Activity Feedback we reused the accelerometer recordings obtained from the SleePic experiments directly.

The SleePic Core recorded the reaction test response times and the corresponding physiological signal segments to the Flash card. It was found that the SleePic Core encountered a writing failure for two subjects during the first night. The missing "no answer" messages for these nights were substituted. The Activity Feedback data and the physiological data were still available for the concerned nights because it was recorded to another Flash card.

On average, 1.83 labeled segments per hour were obtained from the Button Feedback. Thereof contained 36% sleep labels. This corresponds to the same sleep / wake proportion as for the entire recording. Using the labeling rule in Equation 7.1, neither false positive nor false negative labels were generated. The labeling from the Activity Feedback contained in average 14 labeled segments per sleep hour and 8.9 per wake hour. Using the labeling rule from Equation 7.2, we generated a total of 27 false sleep and 6 false wake labels, which corresponds to an error rate of 1.3%. The wrong labels were not discarded for the adaptation process.

7.5   Experiments

To evaluate the adaptation strategies presented in Section 7.2 and to qualify the obtained feedback data described in Section 7.4 we conducted a series of experiments using the new data to adapt five different network topologies that were designed in Chapter 5 and 6. The topologies differed in the input signal vector. The networks topologies consisted either of each individual signal (ECG, RSP, ACC), or a combination thereof (ECG+RSP, ECG+RSP+ACC). The labeled data from the Button and Activity Feedback were randomly split into an adaptation data set (80%) and a validation data set (20%). For each user 10 different adaptation and validation data sets were generated.

7.5.1   Threshold Experiments

In this set of experiments, we modified the classifiers obtained with the subject-independent experiments in Chapter 5 and 6 by changing the classification threshold. The adaptation data set composed of the Button Feedback and the Activity Feedback data of the user that was not present for the training, was used to find the optimal threshold with an iterative process. The classification threshold was incremented in steps of 0.05 from -1 to 1 and the classification performance on the adaptation set calculated. The threshold giving the highest accuracy was selected. As test data set, the same set as in the subject-independent experiments (Section 5.2.3) was used. For each generated adaptation and validation data set one run was performed. This was repeated for each user (8 times) and each topology, making a total of 400 runs.

7.5.2   Neural Weight Experiments

In a second set of experiments, the adaptation, the validation and the test data sets remained the same as in the Threshold experiments (Section 7.5.1). The best network obtained from the subject-independent experiments in Chapter 5 and 6 for each topology was used as a start network for the adaptation procedure. 10 independent runs for each adaptation and validation data set were generated. Again, this was repeated for each user (8 times) and topology. The classifiers were tested on the same test data sets as used in the subject-independent experiments and the threshold experiments.

We conducted two additional set of experiments, which were based on the Neural Weight experiments:

Feedback Experiments

To analyze the contributions from the individual feedback strategies to the performance of the networks obtained from the Neural Weight experiments, these experiments were repeated with only the labeled data from the Button Feedback or the Activity Feedback in the adaptation and validation data set, respectively.

Training Experiments

To compare the Neural Weight adaptation experiments, where the neural weights of a generalized ANN were adapted, to a subject-specific ANN (as it was intro-

duced in Section 4.3), we repeated the Neural Weight experiments with initializing the ANN weights randomly at each run. This type of training corresponds to the subject-specific training in Section 5.6, but by using the limited-size adaptation data set instead of the training data set. Again, the test data set stayed the same and the experiments were repeated 10 times for each subject and topology.

7.6   Results and Discussion

In Figure 7.1, the accuracy obtained from the subject-independent experiments in Chapter 5 (generalized) and from the adaptation experiments described in Section 7.5.2 (adapted) are shown. As mentioned before in Chapter 6, the results of the generalization experiments show that topologies containing the frequency features of the accelerometer data (ACC) as input perform statistically better than the topologies without the ACC (Student's t-test, p<0.05, for all cases). Whereas the adaptation did not significantly increase the accuracy of the ACC topology, the adaptation improved the accuracy of all topologies containing cardiorespiratory signals as inputs (Student's t-test, p<0.01). Although the adaptation methods improved the accuracy of the ECG features based topology significantly, the accuracy did not reach the ones of the RSP or ECG+RSP topologies (Student's t-test, p<0.01). This suggests that ECG features have a too important intra-subject variation and are not recommended to be used as only features in a wearable sleep / wake classifier.

7.6.1   Threshold versus Adaptation

To evaluate the different adaptation strategies we conducted experiments using the automatically labeled adaptation data set. Figure 7.2 shows the accuracies of the classifiers for the adaptation experiments described in Section 7.5.1 and 7.5.2. The threshold modification strategy (Threshold; boxes in the left columns) cannot improve the generalized network and results in a lower accuracy range (median accuracy $\leq 81\%$ for the physiological signal based topologies) than the other two methods. The accuracies of the training experiments (Training; boxes in the center columns) and adaptation experiments (Adaptation; boxes in the right columns) are significantly better (Student's t-test, p<0.01) than the threshold accuracies (except for the ACC topology), but there was no significant difference between them for all topologies (Student's t-test, p>0.95). This may suggest that there is no advantage in initially training a general classifier. However, several aspects are in
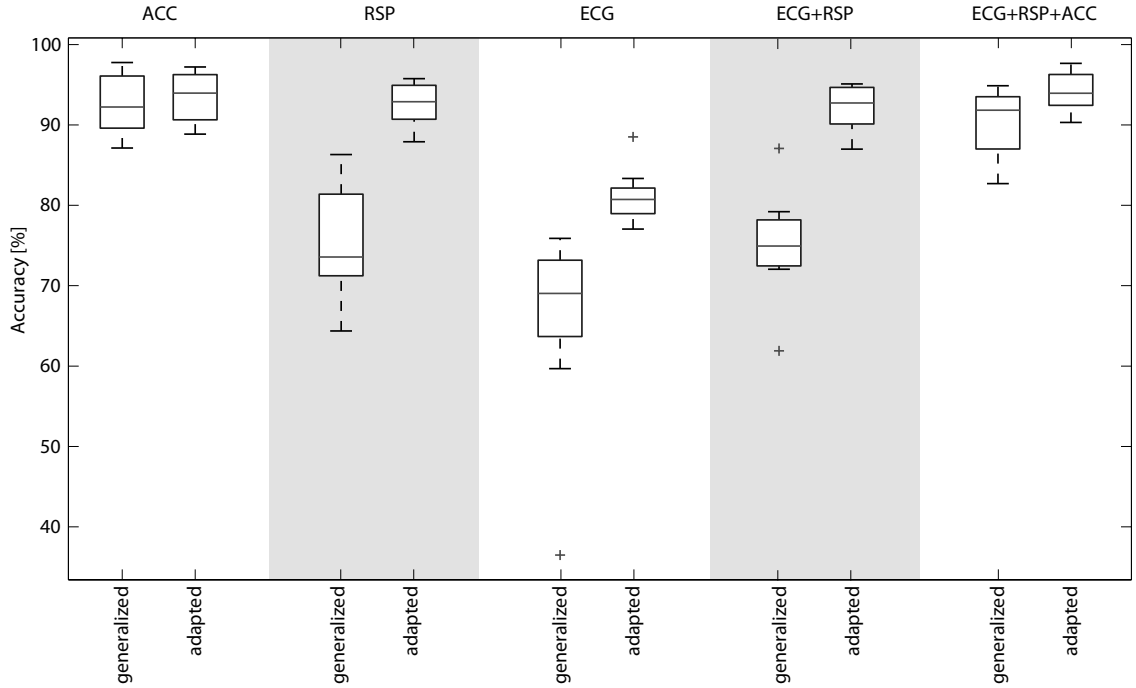
Figure 7.1: Accuracy of the five different ANN input topologies (ACC, RSP, ECG, ECG+RSP, ECG+RSP+ACC). Left boxes: Results of subject-independent networks obtained in Chapter 5 (generalized). Right boxes: Results when adapting the generalized networks with the adaptation data sets. The horizontal lines of each box are the lower quartile, median, and upper quartile values (from bottom to top). The whiskers represent the most extreme values within 1.5 times the interquartile range from the quartile. The outliers are data with values beyond the ends of the whiskers and are represented with crosses.

favor of adapting a generalized network to a new user. As can be seen in Table 7.1, the adaptation experiments need fewer training cycles before the adaptation process can be stopped. This is because the exploration of the search space was limited by local minima. This is not a disadvantage because in our case,

a) the adapted networks do generalize statistically better (Student's t-test, p<0.05) for other users (Table 7.1, 2nd row);

b) the obtained accuracy is not lower for the adapt experiments;

c) the adaptation method escapes potential local minima when the input features display elementary differences for the new subject (outliers of RSP or ECG topologies in Figure 7.1); and

d) when training an initialized networks, it can neither be excluded that the
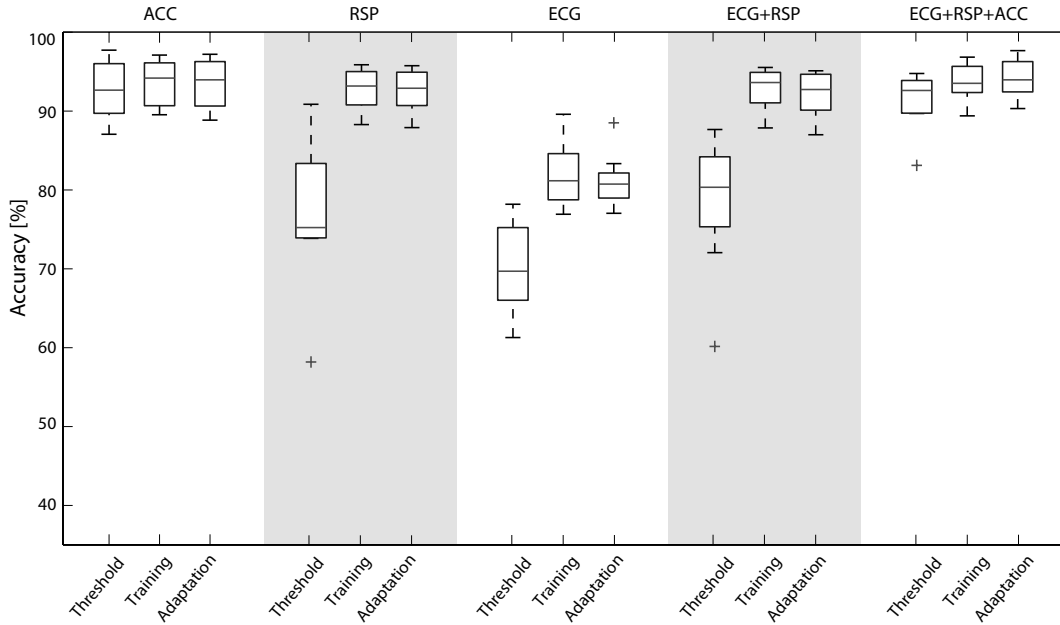
Figure 7.2: Accuracy for the two different adaptation strategies and the subject-specific training of the five different ANN input topologies (ACC, RSP, ECG,ECG+RSP, ECG+RSP+ACC). Left: Result using the Threshold adaptation strategy (Threshold). Middle: Result from the Training strategy (initializing the networks and training with the adaptation data). Right: Result when using a generalized network and adapting the neural weights (Adaptation).

training becomes stuck in a local minimum.

Finally, during the adaptation recording process, the generalized networks will already give reasonable classification results, whereas when using a freshly initialized network, the system will not be able to output a meaningful sleep / wake classification.

7.6.2   Button versus Activity Feedback

To qualify the data obtained from the different feedback methods for the automatic labeling we repeated the adaptation experiment with each labeling source or a combination thereof. Figure 7.3 shows a boxplot for the accuracy when using the Button Feedback, Activity Feedback or the combination of both (used previously for the analysis) in the adaptation data set and the neural weight adaptation strategy. A positive trend in accuracy toward the larger data sets (Button < Activity < Button+Activity) can be observed. It seems that the Button Feedback

	Training	Adaptation
Training cycles [*number*]		
Mean	6.85	4.9
Standard error	0.14	0.10
Maximal cycles	24	8
Accuracy of generalization training set [%]		
Mean	74.76	86.13
Standard error	0.61	1.04
Highest accuracy	86.57	94.99

Table 7.1: Comparison of the Training and Adaptation of Neural Weight Strategy for ECG+RSP+ACC. The training cycles correspond to the number of cycles the algorithm has to modify the synaptic weights of the networks until a stop condition (Section 4.1.2) of the training is reached. The accuracy of generalization is measured by presenting the data set of the subjects not present in the adaptation data set to the adapted networks.

did not provide enough data points for efficient training. This is supported by the fact that the adaptation with the Button data only was not able to reach the accuracy obtained in the subject-specific experiments shown in Figure 5.9. For the combined topology ECG+RSP+ACC, neither of the three feedback methods influenced accuracy, sensitivity or specificity (Student's t-test, p>0.75; Table 7.2).

The sleep and wake behaviors of subjects that displayed a low sleep efficiency (Chapter 6) were particularly difficult to detect with only accelerometer data because the subjects did not move much and were almost in an sleeping position. One subject watched TV before going to sleep in his bed for a total of 135 minutes and was therefore the most difficult subject in our data set. Figure 7.4 shows the ROC curves of the three best adaptation experiments for this particular user and each feedback method. Interestingly, the specificity of the combined case (diamond) was higher than the one of the single Button Feedback (5-point star) and of the Activity Feedback (6-point star). This is also the case for the accuracy (Combined Feedback 91.01%, Button Feedback 86.42%, Activity Feedback 87.64%). This finding indicates that not only does the increasing number of segments in the adaptation data set improve the performance of the classification, but that the different sources for the adaptation data sets are complementary. This can be explained by the nature of the adaptation data. Whereas data from the Activity Feedback came only from clearly classifiable segments of sleep and wake using accelerometer data, the randomly sampled Button Feedback data contained

more segments from periods that were more difficult to classify. However, as the Button Feedback required some attention of the user, it would not be wise increasing this feedback frequency since the users rated the actual frequency as acceptable (Section 5.3).

For an improvement of the method of gathering the adaptation data we suggest to use a combined solution using the Equation 7.2 and a modified Equation 7.1, where the button pressing task is not activated randomly, but by using prior knowledge. A possible source of prior knowledge could be for example the use of the unthresholded output of the ANN. If the output is close to the classification threshold where the classification uncertainty is increased, an additional button pressing request could be useful.

Feedback	Button (B)	Activity (A)	B&A
accuracy [%]	91.59 ± 4.33	92.67 ± 2.83	92.94 ± 3.37
sensitivity	92.64 ± 6.92	96.25 ± 3.04	96.09 ± 3.63
specificity	90.57 ± 4.52	89.62 ± 4.72	90.42 ± 4.72

Table 7.2: Mean $\pm$ SD of accuracy, sensitivity and specificity for different adaptation data sets using the ECG+RSP+ACC topology.

7.6.3 Subject-Specific versus Subject-Independent Systems

As it can be observed in Figure 7.1, 7.2 and 7.3, the accuracy of any adaptation method shows no significant difference to the subject-independent experiments for the simple ACC topology. Table 7.3 shows the mean and standard deviation for the accuracy, sensitivity and specificity for the ACC topology. We observe no statistical difference in accuracy between the three experiments (Student's t-test, p>0.40). Indeed, the training and adaptation to a particular user even has a slight negative effect on the median specificity. This results in an overestimation of the actual sleep time shown in Table 7.4 (Student's t-test, p<0.1). These two observations suggest that the ACC data is able to generalize well between different users and adaptation or retraining for a new user is not necessary. However, as observed in various previous studies, only accelerometer data does not accurately classify wake states with low activity which can be observed in the lower specificity for all topologies in Table 7.3 (89 $\pm$ 6.68 % for the generalization experiment).

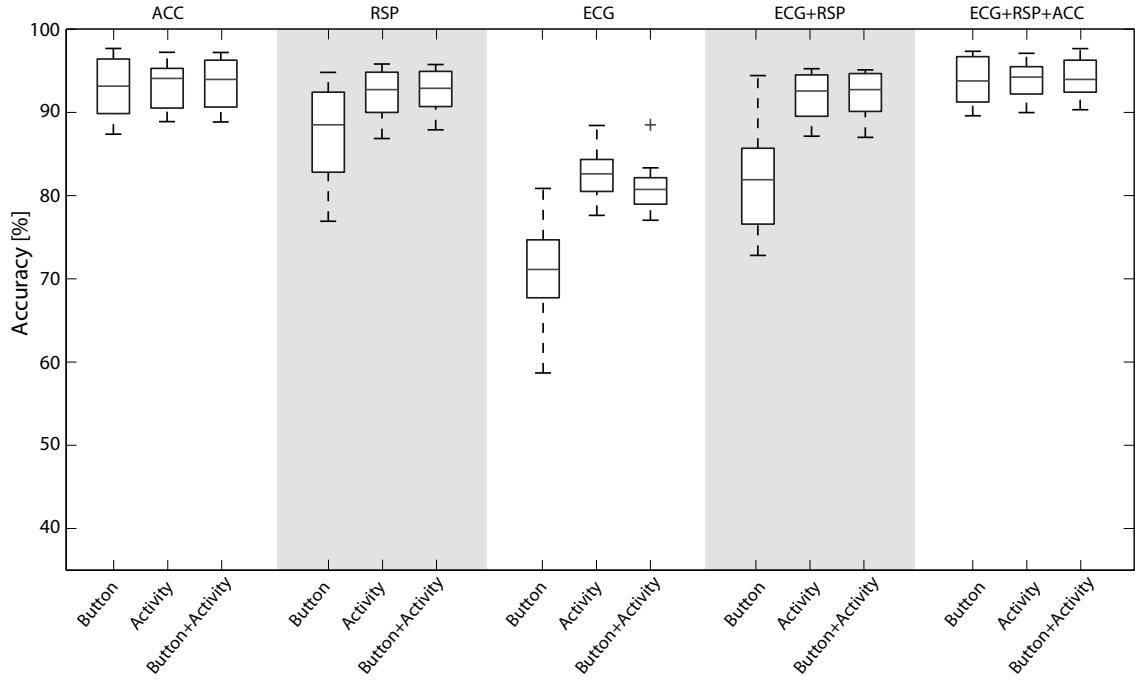In Figure 7.5 we show the ROC curve of the ACC and the ECG+RSP+ACC

Figure 7.3: Accuracy for the adapted ANN when using different methods for generating the adaptation data set for five different input topologies (ACC, RSP, ECG, ECG+RSP, ECG+RSP+ACC) . Left boxes: Results obtained when using only data collected with the Button Feedback method. Middle: Results obtained when using data from the Activity Feedback method. Right: Results when using combined data.

topologies for the generalized and the adapted networks obtained from the same subject as previously described in Section 7.6.2. The generalized ACC network performance (square) shows very high sensitivity and low specificity. Adapting the network improves the specificity at the cost of sensitivity (circle). The generalized ECG+RSP+ACC network has a low (but symmetric) performance (triangle). The adaptation process also improves sensitivity and specificity (diamond) considerably and leads to the highest accuracy of all cases (91.01%) .

The ECG+RSP+ACC topology is interesting because it was the network with the highest adapted accuracy (96.93%) and it had the highest mean accuracy of all experiments (93.18%). Table 7.5 shows the mean and standard deviation for the accuracy, sensitivity and specificity for this topology. We observe a statistical improvement of accuracy (Student's t-test, $p < 0.05$) and sensitivity (Student's t-test, $p < 0.03$) for both the adapted and the trained networks.

The output of the adapted ECG+RSP+ACC classifier is illustrated for two subjects in Figure 7.6. The classification pattern in Figure 7.6a originated from the
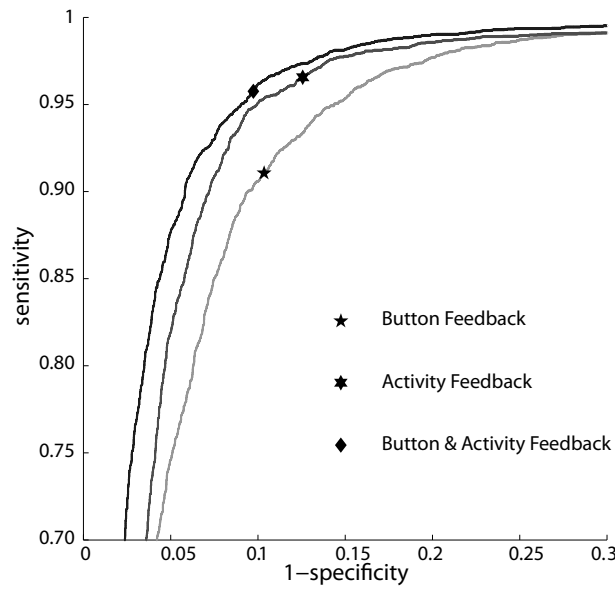
Figure 7.4: Receiver Operating Characteristic (ROC) curves for one particular subject that had many low activity wake events during the study. These events are difficult to classify with traditional actigraphy. Note that the ROC figure is a zoom-in.

[%]	generalized	trained	adapted
accuracy	92.54 ± 3.90	92.78 ± 3.00	92.76 ± 2.82
sensitivity	96.06 ± 3.00	97.17 ± 2.65	97.08 ± 2.38
specificity	89.30 ± 6.68	88.99 ± 4.95	88.29 ± 6.47

Table 7.3: Mean $\pm$ SD of accuracy, sensitivity and specificity for different training methods using the ACC topology.

subject with the highest classification accuracy obtained. The errors were mainly located before sleep onset and during sleep. The classifier tends to anticipate sleep onset as it already has been observed in Chapter 4. Many of the errors during sleep were due the non-overlapping of detected and labeled awakenings, which generated a false positive and a false negative error each. The sleep / wake patterns in Figure 7.6b belong to the same subject that was already presented in Figure 7.4 and 7.5 and which was the most difficult case for the classifier. Most errors were still generated in the period where the subject was lying calmly in bed. However, the classifier did not respond uniquely to the position of the subject (due to the use of the cardiorespiratory features). Further, in the misclassified wake regions, due to the calm lying, the density of correctly classified wake was very high. This suggests that the specificity of the presented classifier could be increased with some post-processing filtering technique.

Sleep parameters	Mean ± SD	Mean difference from Video ± SD
1. Total Sleep Time (TST) [hours]		
Video (label)	6.06 ± 0.94	
ACC	6.79 ± 0.99	-0.73 ± 0.67
ECG+RSP+ACC	6.73 ± 0.92	-0.67 ± 0.53
2. Awakenings [numbers]		
Video (label)	28.07 ± 11.70	
ACC	61.85 ± 37.34	-33.79 ± 33.17
ECG+RSP+ACC	51.17 ± 26.95	-23.10 ± 24.15
3. Sleep Efficiency [%]		
Video (label)	82.81 ± 10.01	
ACC	90.64 ± 6.44	
ECG+RSP+ACC	90.09 ± 7.05	

Table 7.4: Comparison of sleep parameters for the adaptation experiments of the ACC and the ECG+RSP+ACC topologies.
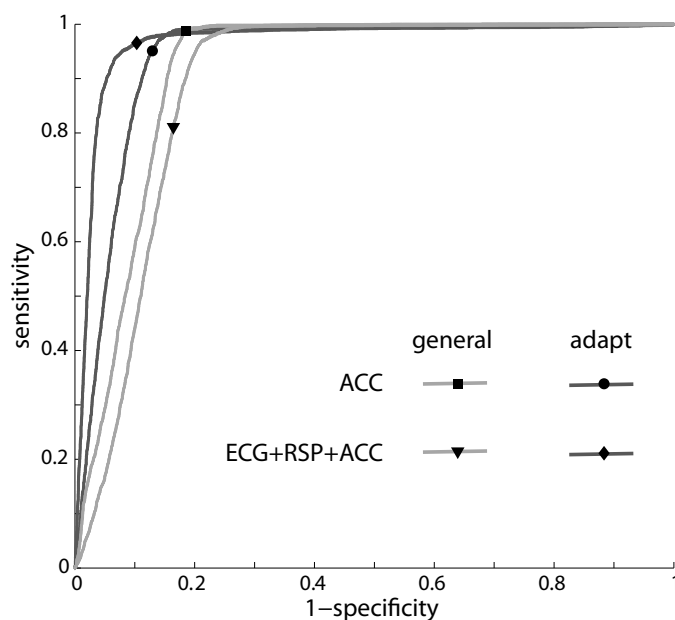
Figure 7.5: Receiver Operating Characteristic (ROC) curve for one particular subject (same as in Figure 7.4) that had many low activity wake events during the study.
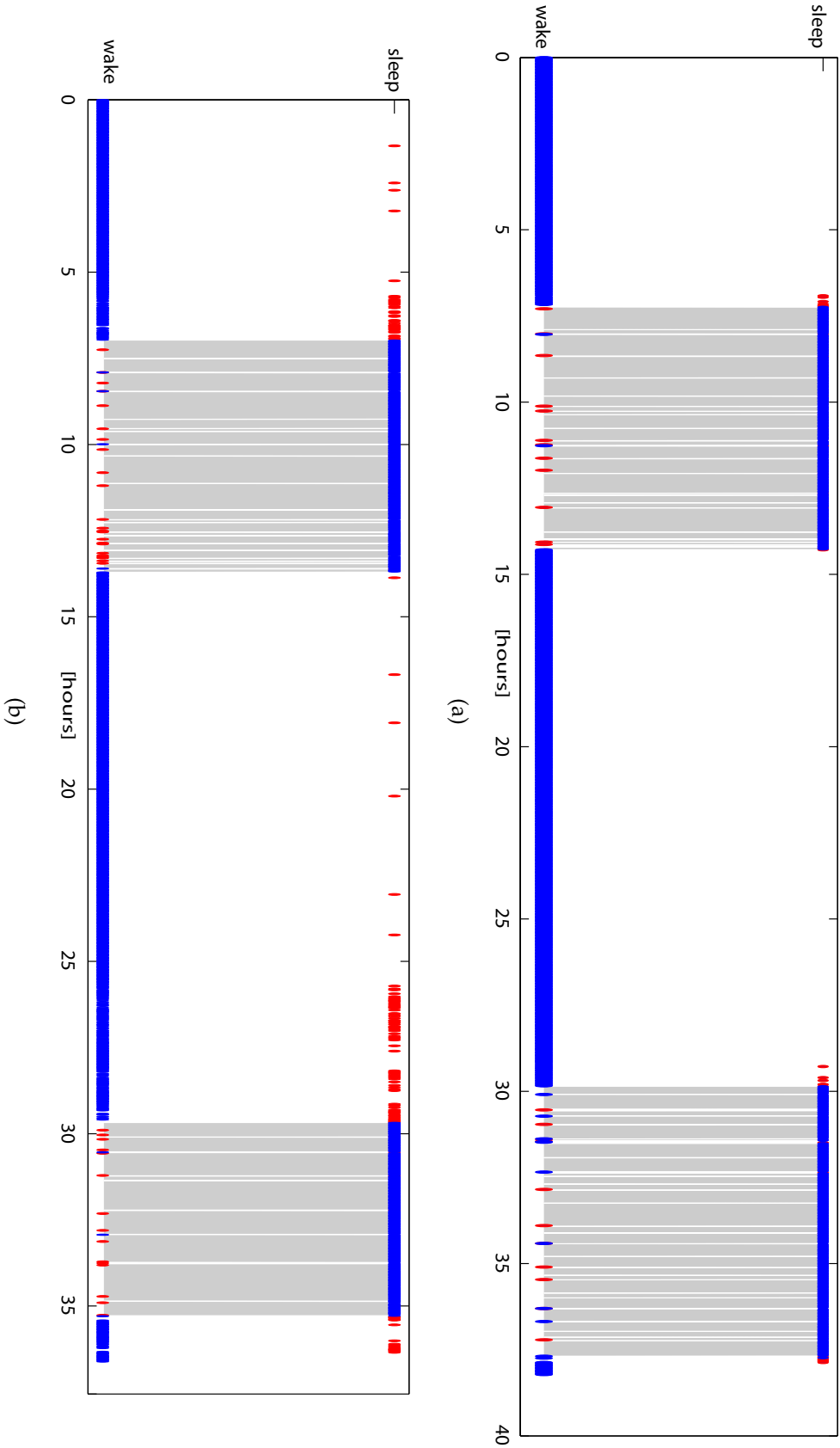
Figure 7.6: Adapted ECG+RSP+ACC classifier output for a) a subject that displayed the highest accuracy; and b) the subject that displayed the lowest accuracy. The blue (red) marks correspond to a correctly (wrongly) classified 20-second segment. The gray boxes correspond to the segments labeled as sleep by the human expert.

[%]	generalized	trained	adapted
accuracy	90.22 ± 4.28	92.99 ± 3.00	92.94 ± 3.37
sensitivity	89.57 ± 8.54	96.78 ± 2.47	96.09 ± 3.63
specificity	90.48 ± 5.56	89.38 ± 5.25	90.42 ± 4.72

Table 7.5: Mean $\pm$ SD of accuracy, sensitivity and specificity for different training methods using the ECG+RSP+ACC topology.

7.7 Conclusion
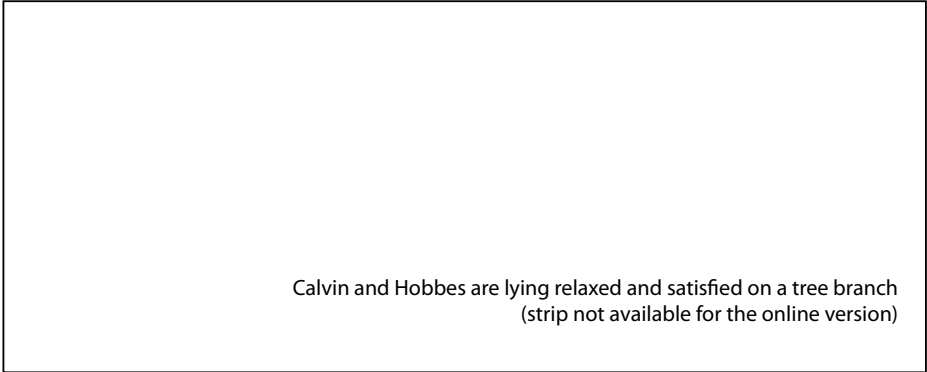
We have shown that the maximal accuracy for the ACC topology can be reached when it is trained for a subject-independent application and an adaptation to a new user has only minimal effects on the performance of such a classifier. In the same way, we have shown that the ANN classifiers that were based on a physiological signal topology can be improved significantly when adapting neural weights. Although the accuracy of the adapted networks did not significantly exceed the ones from the ACC networks, the use of cardiorespiratory signals for the classification displayed an advantage in the particular case when subjects showed low activities during wake, which was impossible to classify when relying on accelerometer data only.

Despite the improvements given by the adaptation methods, the sleep parameters (Total Sleep Time and Awakenings) were overestimated.

However, the main achievement was the description and evaluation of two methods for power and computational efficient gathering of labeled data about the subjects' sleep / wake states in an automatic way. Both methods made use of measurements of typical behaviors that are associated with normal sleep and wake, notably the measuring of an increased arousal threshold and of a maintained behavioral quiescence during sleep. The suggested methodology is based only on occasional button pressing of the subject and the measurement of the user's activity. Because of the simplicity and the low requirements of this method, it is not limited to the cardiorespiratory sleep / wake classification presented here, but could also be used for automatic adaptation of other sleep discrimination algorithms.

8 Concluding Remarks

Calvin and Hobbes are lying relaxed and satisfied on a tree branch
(strip not available for the online version)

Calvin and Hobbes © 1995 Watterson. Universal Press Syndicate. All rights reserved.

8.1   Main Achievements

This thesis aimed at developing methods and technologies toward personalized, context-aware health information systems. The integration of signal processing and classification methods into low-power microcontrollers, as well as the analysis and selection of relevant physiological and non-physiological signals for sleep / wake classification carried out in this thesis have made the development of a self-contained wearable hardware platform possible. The automatic and simple user-adaptation proposed here renders the system user friendly and affordable for a large public.

The challenges of wearable devices that were identified in Chapter 3, such as limited energy, low computational power, increased susceptibility to noise and artifacts, and the need for comfort and unobtrusiveness, led to the use of spectral

119

features of cardiorespiratory (ECG and respiration effort) and activity signals. Single-layer, feed-forward Artificial Neural Networks have proved to be an efficient classifier solution for the classification of these features on a microcontroller (Chapter 4). The sleep and wake classification accuracy of the proposed method was above 90% when data from the same subject was used for training, validation and testing.

Our experiments described in Chapter 4 and 5 have shown that the physiological differences between subjects make it difficult to build a single classifier based only on cardiorespiratory features. However, in order for the wearable system to be used by the general public without expensive calibration or intervention from a human expert, it would be desirable to build a generalized classifier system. A generalized classifier could accurately discriminate sleep and wake for a general population.

Additional features obtained from activity measurements developed in Chapter 6 showed to be more robust to inter-subject variations. When training the artificial neural network (ANN) classifiers with these kind of features the classification accuracy improved for subject-specific, as well as for subject-independent experimental setups. Activity measures have previously been used for long-term ambulatory observations of sleep and wake behavior (Ancoli-Israel et al., 2003). Because low activity is a behavior that is not unique to sleep, the classifiers that used only activity based features showed a higher error rate in detecting wake correctly (low specificity; Chapter 6). However, when computing the newly suggested frequency features of the accelerometer data we have shown an improvement of the wake detection rate compared to the known actigraphy methods. Our new method had a higher specificity because it combines body position and activity information. With the same series of experiments we showed that the ANNs based on accelerometer features generalize well for the studied age group. Furthermore, the inability of traditional actigraphy algorithms to distinguish low-activity wake from sleep was partially eliminated by the combination of the spectral features of activity with cardiorespiratory features. This topology setup showed a better generalization to new users than when only cardiorespiratory topologies were used. However, due to an over-fitting on the physiological signals during the training procedure of the ANNs, this network topology still had lower generalization properties than when only an activity topology was used.

Additional research conducted on adaptation methods in Chapter 7 revealed new methods to automatically adapt the classifiers to individual subjects and to

improve the performance. These methods are based on the principle of actively and passively measuring sleep and wake behaviors such as the measurement of reactivity and activity that show only low variation for distinct subjects in the study group. This a priori information is used to automatically label the corresponding physiological data and to adapt the subject-independent classifiers accordingly. This led to performances of the ANN classifier comparable to the subject-specific experiments. Using such an adaptation approach offers the possibility to distribute the same wearable system to different customers which incorporates the same generalized sleep / wake classifier. After a few days of wearing, the system will then automatically specialize to the wearer without need for cumbersome off-line data collection or classifier training. Such a methodology for user-adaptation may be suitable not only for sleep / wake applications, but also for other domains where subject-specific physiological signals are used, such as evaluation of stress and emotion (Picard, 1998), vegetative state control (Awa-Con, 2009) or assessing the depth of anesthesia (Sartori et al., 2006). Additional development might be necessary to include closed loop control of the adaptation process to ensure long term stability and validity of the classifier.

Although the presented methods are conceptual and the test population of the experiments was limited to young and healthy subjects, the proposed adaptation method has potential to adapt the ANN classifier on-line to any user.

The conducted experiments were focused on real-world, long-term ambulatory measurements. Using realistic data has several advantages. For example the ANNs could be trained on data from environments where the networks actually will be used. This reduced the risk of over-fitting to irrelevant data. Further, the influence of noisy signals and the persons' behavior were integrated into the classifier training, which reduced the required filtering of the signals to a minimum. However, the use of ambulatory data made the validation of the method more difficult. It was not possible to obtain simultaneous, qualitatively good electroencephalographic (EEG) recordings that is used in clinical sleep research and would have allowed for more accurate a priori sleep / wake knowledge for labeling of data. The video, eye movement, and muscle activity recordings that were used in this thesis for the labeling of data correlates with clinical polysomnography labeling (Section 3.1.3), but is not considered as the gold standard for sleep analysis.

The studied group, which was limited in sample size, showed a high sleep efficiency (sleep duration divided by time in bed) and almost no observed sleep in an upright position. Although sleep efficiency does not correlate with the specificity of the physiological classifiers, we could not show that the proposed method can

detect sleep independently of body position. To demonstrate this, a series of experiments would be needed where subjects cannot leave a fixed position different from lying for the whole duration of the experiments.

Another limitation of this work is that we have only shown sleep and wake discrimination in 'normal' situations where the subjects were healthy and not sleep-deprived. For example, our experiments give no indication of the detection of microsleep. Therefore, without further analysis, the method should not be used on an alarm system where the detection of short, involuntary sleep episodes must be detected.

The SleePic hardware prototype developed for this thesis (Appendix B) can be considered as a proof-of-concept for a new generation of health and wellness devices. It shows the feasibility of a wearable, on-line sleep/wake classifier using low-cost components. The minimal processing power requirements of the presented system opens new doors for the use of high-level algorithms like fatigue prediction models. Although cardiorespiratory signals can be measured non-invasively with a wearable belt, it is still more cumbersome to use than a wrist worn actigraphy device. It is therefore application-dependent whether a single actigraph device or a combined physiological system as presented in this thesis should be adopted. If, for example, the application is oriented toward sleep disorder diagnosis, the system with additional cardiorespiratory recordings offer a clear advantage, because current sleep disorder diagnosis requires these signals (Patel and Davidson, 2007).

8.2   Outlook

The applications of a wearable sleep / wake discrimination system are manifold. Although we have shown the feasibility of a wearable sleep / wake discrimination device, in order to become commercially successful, some aspects, such as the level of integration, still need to be improved.

One aspect already in investigation by material scientists is the integration of sensors and electronics into garments (Amft and Habetha, 2007 and Wealthy-IST, 2009). This will lead to higher wearability and therefore acceptance by users. The improvement of wearable sensors and especially the increase of the signal to noise ratio may lead to new features becoming available for classification, which could revolutionize the performance of physiology-based health systems. The main interest certainly lies in the wearable EEG recording, which could eventu-

ally extend the reliable detection of binary sleep / wake to more detailed sleep or sleepiness stages. The demand for wearable and robust EEG has also increased in different other emerging research fields such as non-invasive brain-controlled robots or prosthetics (Millán, 2008 and Sanchez et al., 2008). Beside EEG, the use of more advanced cardiovascular features (which could be extracted from heart rate variability or peripheral arterial tone) may also contribute to the classification of light, deep and REM sleep, which has not been covered in this thesis. The processing, classification and adaptation methods developed and presented in this thesis may be applied to these new approaches. How much the performance can be improved with a more diverse set of features remains to be seen.

With the increase of the number of features from different signal sources, more advanced classification methods might be considered. For example a mixture of ANN experts where behavioral and physiological features are used in individual networks and which are evaluated by a superior classifier could be used. With the continuous improvement of embedded microcontrollers, the implementation of such more advanced classification methods is becoming realistic.

The embedding of fatigue prediction models into wearable devices, as was described in Chapter 2, requires further research. Although preliminary results from different research groups are promising (Akerstedt et al., 2004 and Balkin and McBride, 2005), the models have not yet been evaluated for their robustness. As this thesis has also shown (Chapter 7), imprecision in the calculation of sleep parameters such as sleep onset time, total sleep time and number of awakenings, might not be removed completely with a wearable classifier. These sleep parameters are used by most fatigue prediction models. The impact of these imprecise input values on the models' output has not yet been analyzed. Future pilots of the Solar Impulse plane (SolarImpulse, 2008) have to withstand duties of 48 hours and more. They will operate in a narrow cockpit giving space to only a single pilot and in extreme environments such as altitudes up to 12'000 meters or high temperature variations (-40° to 20° C). To stay alert over the full duty-time they will need to rest regularly, but the mission will allow only short breaks. To get the maximum rest in a minimal time it will be important to plan ahead and coordinate the resting periods with the actual fatigue of the pilot. For this, a polyphasic sleep schedule (Section 2.1.1) combined with an alertness model (Section 2.4.4) can be useful. In a preliminary study (data not shown), we have conducted a series of tests in a flight-simulator using our wearable sleep / wake detection device in combination with a modified alertness model of Akerstedt et al. (2004) (described in Section 2.4.4). In the 48-hour tests, our model planned on-line the

polyphasic sleep schedule of the pilot and adapted the schedule to the increased fatigue over night. The calculated sleep schedules are promising. However, additional experiments are required to obtain conclusive results.

Considering the large impact of low quality sleep on human health and the high ratio of undiagnosed sleep disorders, a deeper research focus on the detection of sleep disorders with low-cost wearable devices is desirable. This would require the screening of larger populations, including people with disorders, using wearable devices. The direct and indirect reduction of cost in health care with automated prevention and screening methods justifies the large amount of investment into research and development that is still needed. We believe that the studied sleep / wake classification system proposed in this thesis is a first, promising step toward a context-aware system for wearable devices that could prevent sleep-related disorders.
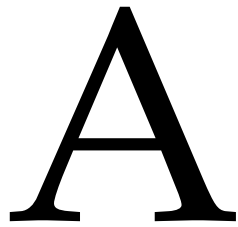
A Wearable Physiological Recorders

In the last decade the recording of physiological signals outside the laboratory became possible because of the miniaturization and power consumption reduction of electronics, the parallel increase of computing power of microcontrollers, increased storage media capacity, advances in embedding sensors, and the improvements in battery technology. With this, a multitude of portable bio-medical and wellness devices appeared on the market or are still in development. In this chapter, the most important developments are presented. Two principal classes are distinguished: Portable and wearable systems.

A.1 Portable Recording Systems

Portable physiological recording systems are so called "holter systems" that are designed for medical applications, where practitioners rely on high quality physiological signals with high sampling rates and low noise levels. They are used where no stationary instrumentation can be used. The basic design is very similar in all holter systems. They are

   a) battery powered

   b) multi-functional

   c) can record a multitude of signals

   d) saving the recorded signals to some storage media

   e) expensive (because of the medical application)

   f) have no or only limited user interface

Figure A.1: Portable physiological recording system used for ambulatory sleep screening. The sensors have to be applied by a technician and the cables and sensors need to be glued to the skin. Picture from Somnomedics (2008).

   g) rather bulky

   h) connect to the sensors with cables.

Because of the later two issues the holter systems are considered as portable and not wearable. An exception is the group of systems that forms a wireless sensor network on the body (so called "body area networks", BANs) as shown in Figure A.2f. Because of the general lack of cables in this group, they are often considered as wearable.

   Figure A.1 shows a holter system used in sleep screening and illustrates nicely the invasiveness of such systems. Patel and Davidson (2007) have reviewed portable home sleep monitoring devices for the detection of sleep disorders that are functionally similar to the device shown in Figure A.1 and concluded that portable recorders are useful for early sleep disorder diagnosis but cannot replace the sleep laboratory procedures.

   Table A.1 shows an evaluation of portable systems that are potentially useful for sleep / wake classification. The list is not exhaustive, many more holter systems exist, but their technology does not substantially differ from the presented systems. Figure A.2 show examples of holter systems.

| Device | Developer | Type | Signals | | | | | Power Autonomy | | Availability |
			ECG	RSP	Acc	EEG	Other	[mAh]	[Hours]	[CHF]
Heally	Koralewski	holter modular	o	o	o	o	o	630	16	x 10'000
Lifeshirt	Vivometrics	holter shirt	o	x	o	o	o	?	?	x 8'000
Somnowatch	Somnomedic	holter watch	o	o	x	o	o	630	18	x ?
Procomp	Thought Tec	holter	o	o	-	o	o	500	-	x 2'000
Lifeguard	NASA	holter	x	x	x	-	o	?	?	- ?
Embletta	Embla	holter	o	o	x	-	o	500	12	x 7'000

| Device	Developer	Type	Data Access	Processing Power Access	User Feedback Display	Other	Medical Certification	Figure
Heally	Koralewski	holter modular	e	-	-	LED	x	A.2c
Lifeshirt	Vivometrics	holter shirt	e	-	x		x	A.2b
Somnowatch	Somnomedic	holter watch	e	-	-	LED	?	A.2e
Procomp	Thought Tec	holter	e	-	-	LED	?	A.2d
Lifeguard	NASA	holter	e	-	x		?	A.2a
Embletta	Embla	holter	e	-	-	LED	x	

Table A.1: Comparison of a selection of portable physiological recording devices. (o) option, (x) yes, (-) no, (e)only PC export, (?) not known.
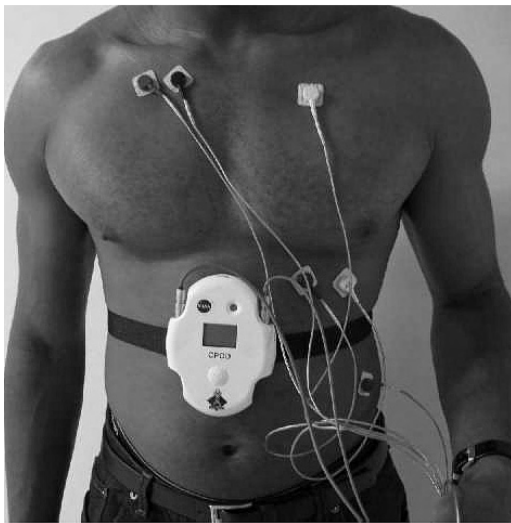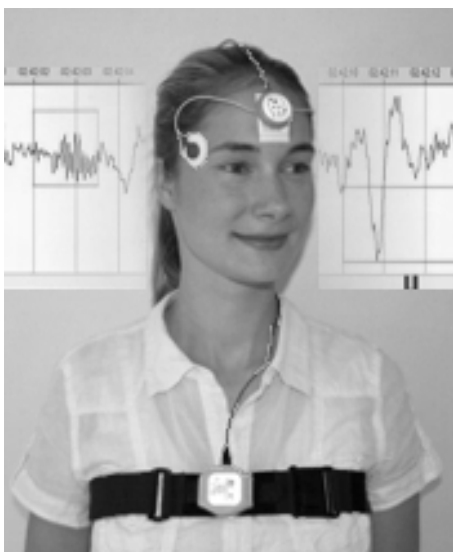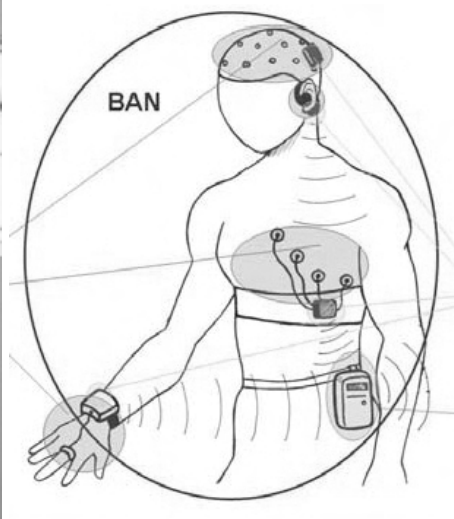
(a) Lifeguard
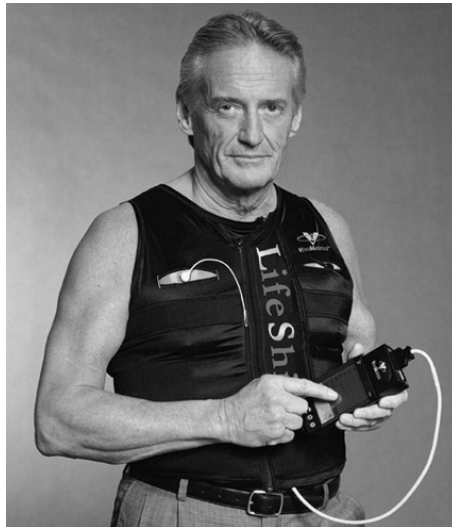
(b) Lifeshirt
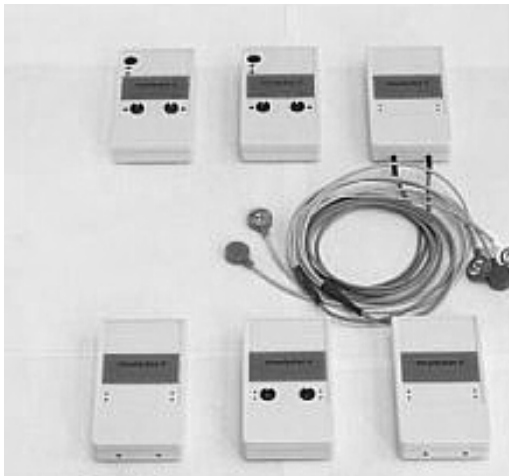
(c) Heally

(d) Procomp

(e) Somnowatch

(f) Sensation BAN Concept

Figure A.2: Examples of portable physiological recording (Holter) systems.

A.2   Wearable Recording Systems

We can classify the existing wearable physiological recording systems into two further categories. First, the category that integrates sensors and electrodes into the garment which is worn on the body (Figure A.3). Second, the category of systems which are watch-like and worn on the wrist (Figure A.4.

Compared to the portable systems, the wearable systems are less expensive and less intrusive. The garment-based systems are mainly still in developmental stage. The already existing products are either wellness driven heart beat monitors (Figure A.3a) or chest belts for first responders (Figure A.3e and A.3f with textile ECG electrodes and integrated respiratory effort sensors. Both product types have detachable electronics modules. The non-textile systems are more autonomous as they combine sensors and electronics on the same device. Often they also feature a display for feedback to the user. An exception is the Vitasense™ device which has independent sensors (temperature pill, ECG patch) which report wirelessly to a master module (Figure A.4e). It can be considered as one of the first BAN in medical use.

Table A.2 compares the different types of wearable systems. For this category, no devices exist that completely fit our needs for a wearable, physiological signal based sleep / wake classification device (Section 5.1). However, we selected the Equivital™ as sensor module for the development of the SleePic device which is described in Appendix B.

Device	Developer	Type	Signals ECG	RSP	Acc	EEG	Other	Data Access	Processing Power Access	User Feedback Display	Other	Power Autonomy [mAh]	[Hours]	Medical Certification	Availability [CHF]	Figure
Equivital	Hidalgo	belt	x	x	x	-	x	x	-	-	vibration	630	48	x	2'000	A.3e
Healthy	EU-IST	shirt	x	x	o	-	o	e	-	-	-	?	?	-	5'000	A.3c
V-tam	Tam-télésanté	vest	x	x	?	-	?	e	-	-	-	?	?	-	?	A.3d
Amon	EU-IST	watch	v	x	x	-	x	e	x	x	-	1000	24	-	?	A.4c
Smartshirt	EU-IST	shirt	-	-	-	-	x	?	-	?	-	?	?	-	?	A.3b
Vivago	VTT	watch	-	-	x	-	-	e	-	x	-	3000	?	x	400	A.4a
Actiwatch	Respironics	watch	-	-	x	-	x	e	-	o	-	2500	?	x	1'000	A.4b
Bodysense	Bodymedia	armband	o	x	x	-	x	e	-	-	-	?	?	x	700	A.4d
WatchPAT-200	Itamar	watch+finger	v	-	x	-	x	-	-	x	-	?	?	x	5'000	A.4f
Polar	Polar	watch+belt	x	v	-	-	-	x	-	x	-	1500	?	x	100-400	A.3a

Table A.2: Comparison of a selection of wearable physiological recording devices. (o) optional, (v) variation, (x) yes, (-) no, (e) only PC export, (?) not known.

(a) Addidas / Polar
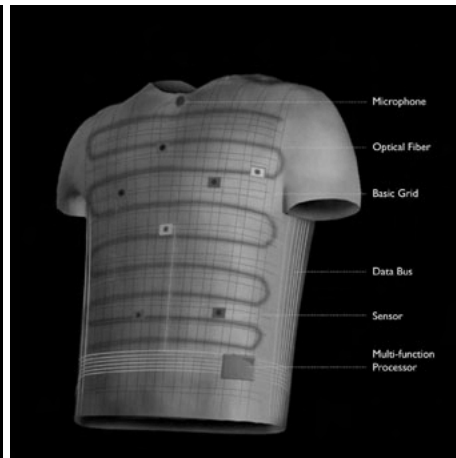
(b) Smartshirt

(c) Wealthy

(d) Vitaljacket

(e) Equivital

(f) Vivometrics

Figure A.3: Garment-based, wearable physiological recording systems.

(a) Vivago

(b) Actiwatch

(c) AMON

(d) BodySense

(e) Vitasense

(f) WatchPaT

Figure A.4: Wearable, non-textile sensor systems.

B SleePic: A Wearable Sleep and Wake Discrimination System

There exist no systems on the market (Appendix B) that comply with all the requirements stated in Section 5.1. To embed the classification algorithm developed in Chapter 4 into a wearable demonstrator, we have to build our own prototype. If costs of the new wearable system should be kept small, the wearable sensors cannot be built from scratch. We therefore decided to rely on commercially available sensors and components for prototyping the wearable system. Missing functionalities were added by custom designed extension modules. A higher integrated product with custom build components and dedicated chips might be more power efficient for our application. However, the development of dedicated chips causes also an increase in costs and consequently also in the final device. Further, it does not allow an as high flexibility as microcontrollers do.

B.1   The SleePic System

The SleePic sleep / wake discrimination system that was developed during this thesis, is composed of three modules:

- The *SleePic Core* processing module (Figure B.1 B) that connects to the

- *SleePic Sensor Module* which is the commercially available wearable sensor system Equivital^{TM} (Figure B.1 C).

- The Sleepic Core also communicates wirelessly with a wrist worn sensor and display / feedback module, called *SleePic Watch* (Figure B.1 A).
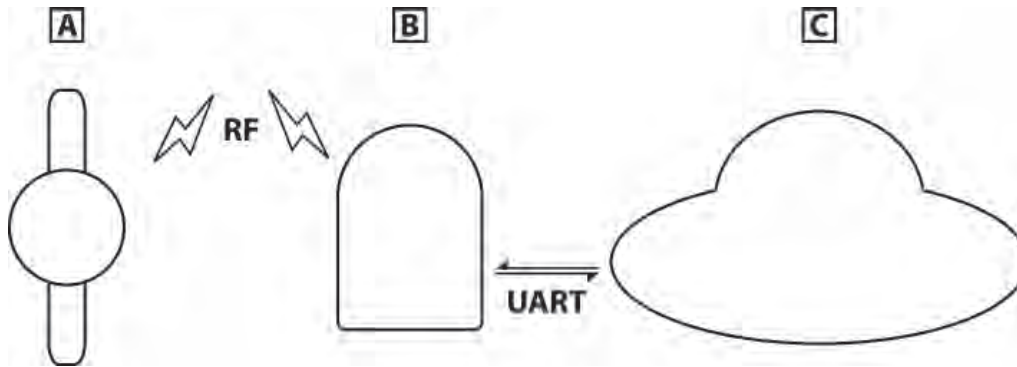
Figure B.1: Communication scheme for the three principal components of the SleePic system: A) *SleePic Watch* with the user interface communicates wirelessly (RF) with the *SleePic Core* processing module (B). The SleePic Core connects physically to the Equivital™ Sensor Electronics Module (C) and receives sensor data over UART.

B.1.1   Wearable Sensor Module

Sensors integrated into textiles appeared only in the last years on the market and the technology is not yet established and still expensive. The Equivital™ physiological monitoring system (from Hidalgo Ltd, UK, Equivital (2008)) is by far the most advanced wearable recoding device with the highest sensor integration (Appendix B). It complies with five of our seven design criteria (Section 5.1). The lack of user feedback possibilities and the missing access to the processing unit of the Equivital™ that prevents to execute the classification algorithm on-line were compensated by the development of two custom made extension modules. The design and construction of the two SleePic extension modules will be described in Section B.1.2 and 2.5.4. The Equivital™ device has been chosen as sensor module for this system development because of the high integration of sensors and textile electrodes in a single belt worn across the upper chest area, which facilitates the correct wearing and use of the system by inexperienced users. It is also light weight, and can continuously work up to 48 hours with the integrated, rechargeable Lithium-Ion battery (760 mAh).

The Equivital™ is composed of two units (Figure B.2a):

a) a washable sensor belt which integrates 3 textile dry electrodes for 2-lead ECG recordings and a piezoresistive strain gauge to measure RSP; and

b) a Sensor Electronics Module (SEM) that can be attached to the sensor belt with five connective clips.

In addition to the ECG and the RSP, the SEM measures 3-axis acceleration (ACC), temperature and also features a real time clock (RTC). To communicate with a PC, the SEM uses a Bluetooth (BT) chip, which can be disabled to save energy. The SEM has a universal serial connector on the top side (Figure B.2a) and offers the possibility to record the sensor data to a removable Flash micro-SD card (up to 2 GB). The SEM also features a vibrating actuator.

The ECG channels are sampled at 256 Hz in a 10-bit resolution and the respiration and acceleration channels at 25.6 Hz in 10-bit resolution. The system uses an open serial UART (57.6 kbaud) communication protocol that facilitates data handling and the interface design. The binary sensor data stream is broadcasted over all communication channels (BT and UART) simultaneously. The drawback of this communication scheme is that the data is only available in the order it is sampled. This form of communication is relatively slow and not energy efficient. The SEM weights 75 grams including the battery. The sensor belt weights 94 grams (size M).

B.1.2   SleePic Core Processing Module

The 'SleePic Core' (Figure B.3) is a printed circuit board (PCB) that plugs to the universal serial connector of the SEM. It was designed for receiving the raw sensor data from the SEM, the processing of the sensor signals and the classification task and the coordination of the tasks between all three modules.

The SleePic Core is composed of a processing unit, a power unit, storage unit and a communication unit. The processing unit is the central element on the SleePic Core and contains the 40 MIPS dsPIC33256GP710 microcontroller from Microchip™ Technology Inc., USA. This programmable interface controller has a 16-bit architecture, 32 KB RAM, 256 KB program memory and digital signal processing (DSP) functionalities. It is the largest available microcontroller from the dsPIC33 series and it has been chosen because of the presence of multiply-and-accumulate (MAC) and a barrel shifter functionality which simplifies the FFT and ANN calculation (see Table B.1 for details on the gained processing time), the high amount of RAM that is also required for the FFT preprocessing and the presence of a multitude of hardware interfaces (2x UART, 2x SPI). Calculus on the dsPic33 can be done in the fractional format (1.15 format, 1 sign-bit and 15 mantissa-bits). The power unit is composed of a step-up/down 3.3 V DC-DC converter which allows to use one single-cell Lithium battery over its full capacity range (2.9 - 4.2 V) or the 3 V power supply from the SEM. Additionally, the
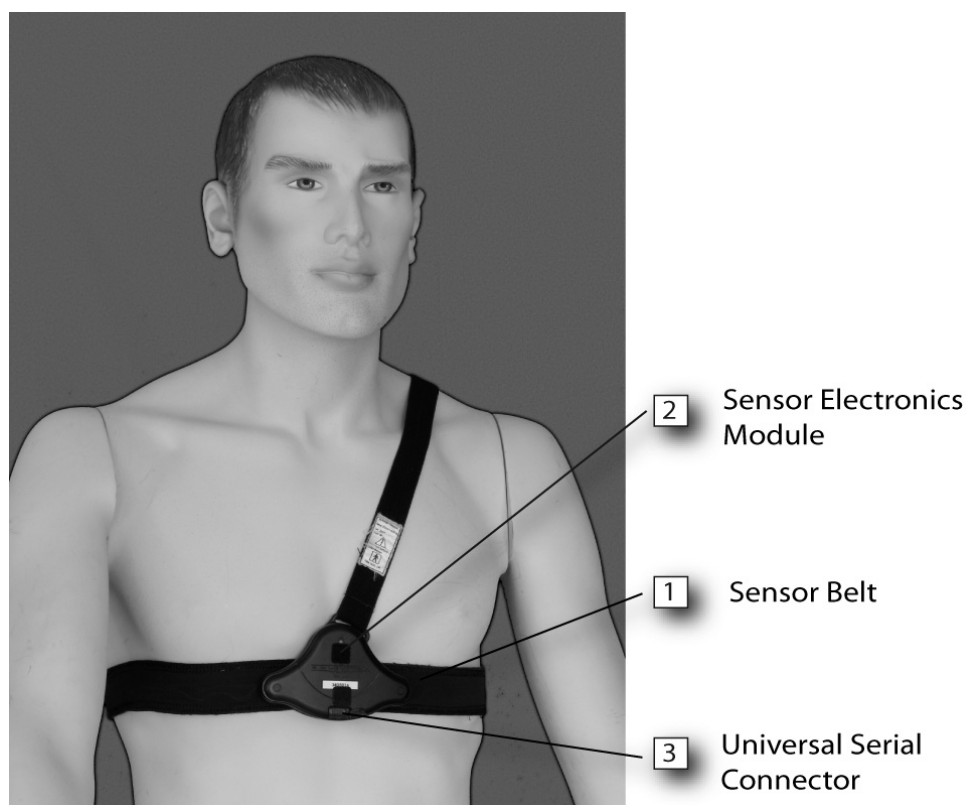
(a)
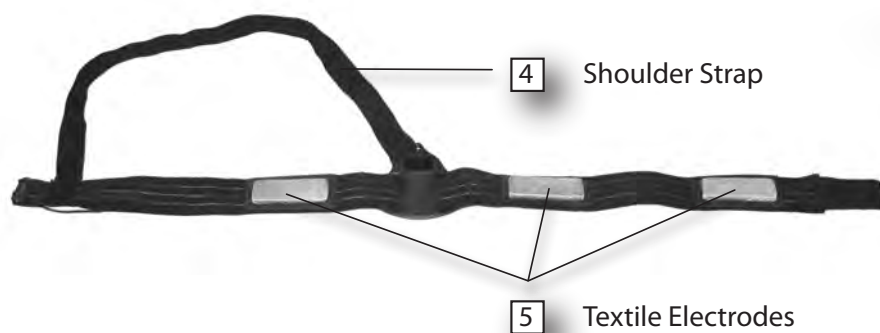
(b)

Figure B.2: (a) A subject wearing the Equivital™ device. (1) washable belt that integrates textile electrodes for ECG (5) and a strain gauge for respiration effort measurements. (2) Sensor Electronics Module (SEM) which does the recording, processing and broadcasting of the signals through the Universal serial and recharging connector of the SEM (3). (b) Inside view of the sensor belt showing the three textile electrodes (5).
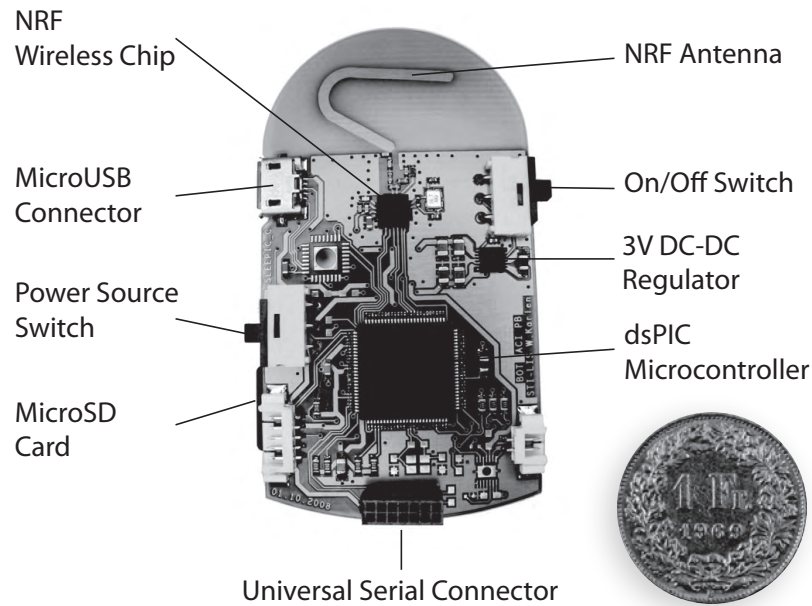
Figure B.3: Bottom view of the *SleepPic Core* processing module. The battery management and the USB converter circuits are not mounted on this version.

unit contains a battery protection circuit that prevent the battery from discharging below 2.9 V and a battery charging circuit. The battery charging circuit is powered over the USB port. The USB port can also power the charging circuit for the battery on the SEM via the universal serial connector.

The communication unit contains a UART port for the communication with the SEM, another UART port to communicate with a PC over a USB converter and an SPI port for the NRF02L wireless chip from Nordic^{TM} to communicate with the SleePic Watch module.

For saving energy, each of the units can be disabled when not needed by the microcontroller, including itself. The dimensions of the SleePic Core module are 65 mm$\times$35 mm$\times$8 mm. Its weight is 8 grams without battery.

The SleePic Core tasks are 1) requesting and receiving over UART RTC timestamps and sensor data from the SEM and the SleePic Watch; 2) computing the pre-processing and classification algorithm; 3) saving the results using the SPI protocol to the storage unit containing a removable Flash 1 GB micro-SD card; and 4) sending user interaction task commands to the SleePic Watch.

Operation Modes

The SleePic Core operates in three main modes (Figure B.4):
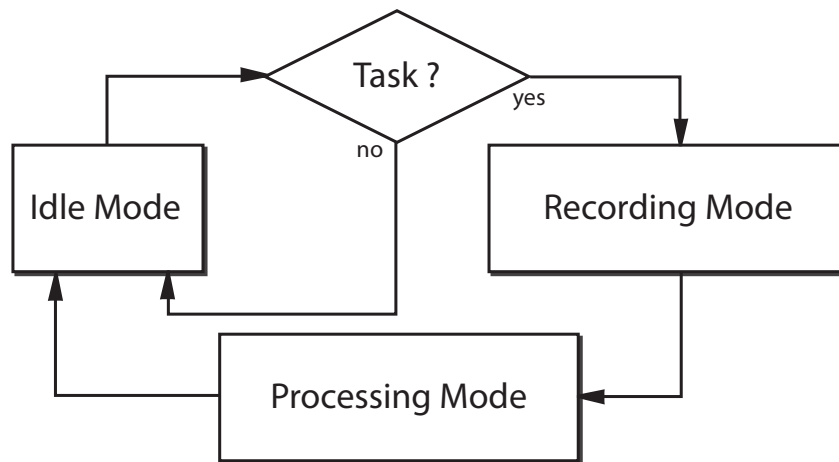
Figure B.4: Operating modes of the SleePic Core module. If no computing task is to be performed, the device stays in a power-saving Idle Mode. If a physiological measure for sleep / wake prediction is required, the module enters into a 20 second Recording Mode. Has all the required data been obtained, the module exchanges data with the SleePic Watch, enters the Processing Mode and computes all algorithms for the sleep / wake classification.

a) The *Idle Mode* is used when no processing or action is needed and all units except the power unit are put into a power-saving mode. A watchdog timer wakes up the microcontroller every 20 seconds. The microcontroller reenters the Idle Mode if no task is on hold. Otherwise it enters the Recording Mode.

b) In the *Recording Mode,* the microcontroller enables the communication unit and starts receiving sensor values from the SEM and stores them in a buffer in RAM. If the buffer is full, the system activates the wireless chip and requests the sensor data from the SleePic Watch. At this point, it can also initiate a user interaction task on the SleePic Watch.

c) In the *Processing Mode* the ECG, RSP and ACC signals are pre-processed and classified according the algorithm described in Section 4.1. The raw data and the results are then stored on the micro-SD card. Afterward, the system returns to the Idle Mode.

Topology	RSP	ECG	ECG+RSP+ACC	ECG+RSP+ACC C Code
prepare[a]	10763	21515	43041	43041
window	2085	4133	8303	55320
FFT	51285	112223	214790	40607313
power	148	220	368	3786
ANN	159	213	330	2358
total	64439	138303	266912	40711818

[a] The preparing steps can be computed during the Recording .

Table B.1: This table illustrates the instruction cycles measurements necessary for the sleep / wake classification algorithm.

B.1.3 SleePic Watch Module

The SleePic Watch module is used for the SleePic user interaction and optionally for sensing. The reason to place this interaction module on the wrist is that it becomes highly visible for the user. The SleePic Watch is build out of 7 PCB's (Figure B.5). The PCB's are on top of each other and 5 PCB rings (Figure B.6B build a cavity to place a Lithium-Polymer battery (145mAh) and electronic components inside it. Only the top and the bottom PCB (incorporate electronic components. The bottom PCB Figure B.6A contains the power unit and the USB-UART converter from the communication unit.

The power unit is composed of the Lithium-Polymer battery, a battery charging circuit and a step-up/down switching 3.3V DC-DC regulator. The recharging of the battery is done over the micro USB connection to a PC, which can deliver up to 100 mA recharge current.

The top PCB contains the control unit, the interaction unit, and the communication unit. The unit is composed of an 8-bit microcontroller from Microchip running at 4 MHz. It features power saving functions, Analog-to-Digital (A/D) conversion and a multitude of communication ports.

The user interaction unit is composed of 5 LED's in the colors of green, orange, red and a button. For this prototype we used no liquid crystal display. To render the feedback more meaningful to the user, it might be advantageous to integrate a liquid crystal display in a next step into the SleePic Watch.

The sensor unit contains a 3-axis accelerometer from Freescale[TM] and a photodiode to measure ambient light. The data from these sensors were not used for this thesis. However, the measure of the ambient light could form an interesting
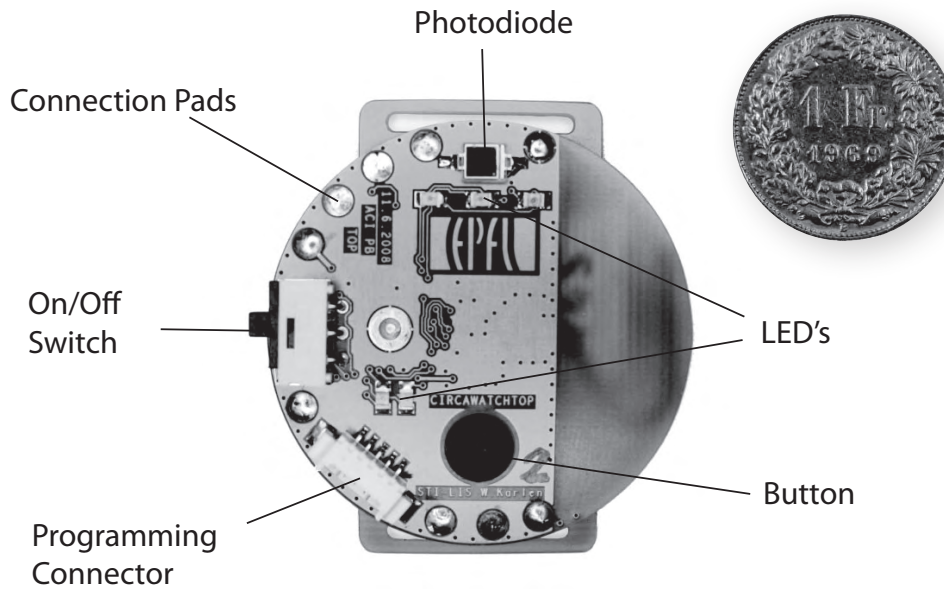
Figure B.5: Top view of the *SleePic Watch* module prototype.

feature for calibrating the circadian components of alertness models in humans such as described in Section 2.4.4.

The communication unit is composed of a USB-UART converter which allows data exchange between the devices via a microUSB connected PC. A Nordic™ NRF02L wireless chip and PCB antenna are used for the data exchange and synchronization with the SleePic Core module.

The diameter of the Watch module is 41 mm and the height 10 mm. This size can be compared to a smaller outdoor watch. A fully assembled SleePic Watch, as shown in Figure B.5, weights 28g.

Operation mode

The SleePic Watch mainly executes the commands received from the SleePic Core. Figure B.7 illustrates the operating flow of the SleePic Watch:

a) Normally the device stays in *Idle Mode*, with all units in battery saving mode and it is woken up regularly by the watchdog timer.

b) After leaving the Idle Mode, the microcontroller enters the *Measure Mode* and powers on the sensor and communication units, samples the A/D ports of the photodiode and accelerometer sensors and saves the values into a buffer in RAM. Then it checks if a message from the SleePic Core has been
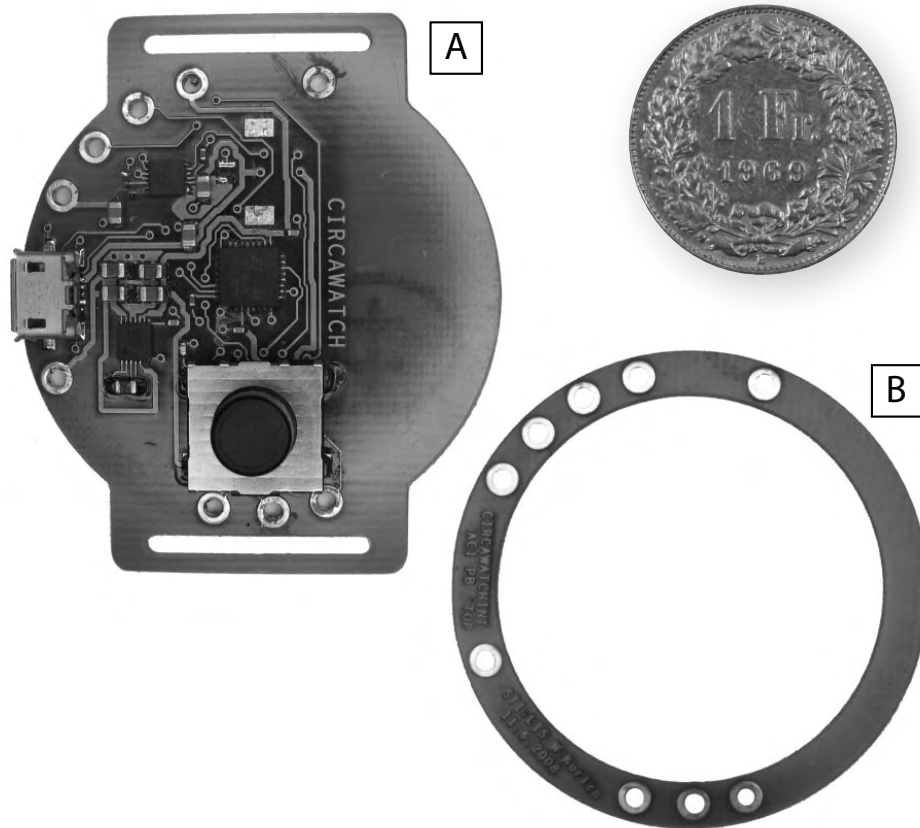
Figure B.6: Top view of the *SleePic Watch* bottom PCB (left) and intermediate PCB layer ring (right). Five of such rings form a cavity for the insertion of the battery. The PCB layers are connected by nine connection soldering pads.

received. Was no message received, the system goes back to Idle Mode.

c) If a "send sensor values" message was received, the SleePic Watch enters the *Answer Mode* where it sends back to the SleePic Core the sensor values in RAM, empties the buffer and goes back to Idle mode.

d) If a "reaction time" command was received, the watch enters *Reaction Mode* and starts to blink 30 seconds with the green LED, then another 30 seconds with the green and orange LED's. If no user action has been taken (no button pressing), it sends back a "no answer" message. Did the user react, the reaction time that was measured with a timer is send back to the SleePic Core and the SleePic Watch system goes back to Idle Mode. The reasons for performing these reaction time measurements are explained later in Chapter 7. In principle, they can be replaced by any other interaction operations, e.g. displaying fatigue level or generation of warnings.
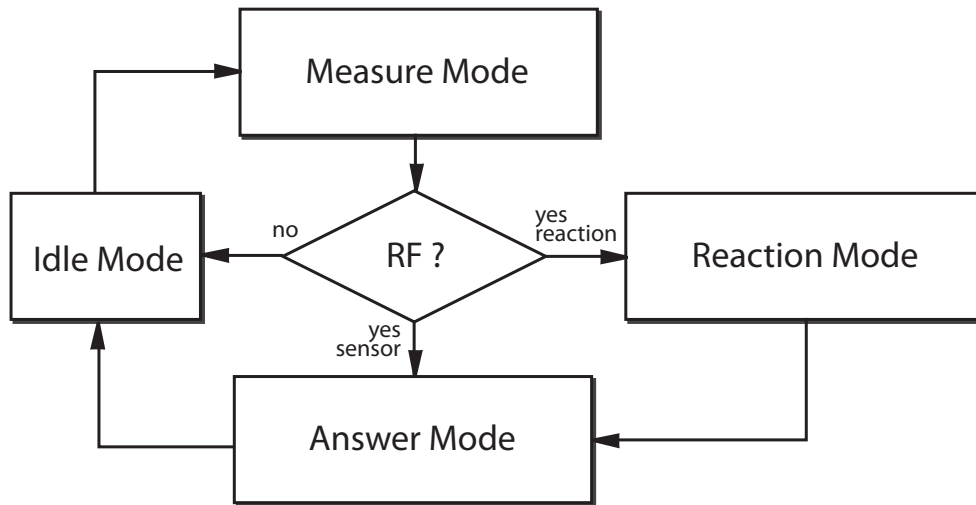
Figure B.7: Operating flow modes of the SleePic Watch module. The module periodically leaves the Idle Mode to perform sensor measures. If the module detects in this time a message from the SleePic Core (RF?), it either enters the Reaction Mode or it sends directly back the requested sensor values (Answer Mode).

B.1.4   Energy Considerations

For a wearable device it is important to keep energy consumption minimal. Both energetically independent systems on chest and wrist achieve an energetic autonomy of 36 hours in the presented configuration, tough they have relative short duty cycles. Both systems can also be easily recharged when connected to a PC via USB.

SleePic Core and Sensor Module

The current consumption of the different modes of the SleePic Core is given in Table B.2. The indicated durations and power values are average values measured at room temperature in a laboratory environment. The total average power consumption of 18.06 mW (6.02 mA @ 3 V) leads to a total autonomy of 36 hours when the 760 mAh battery is shared with the SEM (reduction of 12 hours).

We observe that the major energy resources go to the dsPic and the wireless chip (NRF) in rx (receiving) mode. Compared to a standard microcontroller, the dsPic is relatively power hungry, because of the high clocking and the large RAM size. The power consumption was reduced by putting the dsPic into a very effi-

cient idle mode (0.1 mA) at 93 % of the time. A significant decrease in energy consumption (50%) could be achieved when the Recording Mode can be substituted by a single request for the 20-second data packet instead of the continuous recording of the data stream. This can only be realized when the Equivital^{TM} stores this data internally and the dsPic can communicate with the Equivital^{TM} using a faster protocol (e.g. SPI), which was not possible without changing the Hardware of the SEM. The power dissipation of the dsPic can be further decreased by slowing down the system clock.

The total consumption can be decreased by 20% if the consumption of the wireless chip is decreased by reducing the time the chip has to wait from an answer from the SleePic Watch. This can be achieved by implementing a communication protocol that does not require a continuous powering of the wireless chip while waiting for a response.

A more radical, but surely very energy efficient and power saving option would be the replacement of the dsPic with a dedicated silicon chip for the recording, FFT and ANN calculation. For example, a time-frequency wavelet transform combined with an ANN implemented on a very-large-scale integration (VLSI) chip implementation uses less than 100 μW (Aziz et al., 2009), which would correspond to a power saving factor of 1000 when applied to our problem. However, these types of chips are very task dependent and their development expensive. Therefore we preferred the more flexible microcontroller solution for the prototyping.

SleePic Watch

The power consumption of the different modes of the SleePic Watch is given in Table B.3. The indicated durations and power values are average values measured at room temperature in the laboratory environment. The total average power consumption of 10.36 mW (2.8 mA @ 3.7 V) leads to an autonomy of 51 hours when using the 145 mAh Lithium-Polymer battery. The energy autonomy of the SleePic Watch can be increased in a next step by selecting a smaller and more energy efficient microcontroller that also offers a lower stand-by power consumption and by replacing the power hungry LED's with a low power LCD.

Mode *Duration [s]*	Idle 200 (80%)		Recording 20 (8%)		Processing 30[a] (12%)		AVG
@3 V	mA	%	mA	%	mA	%	mA
dsPIC	0.1	100	54	70	54	7	
NRF rx	-	-	18	1	18	95	
NRF tx	-	-	10	11	-	-	
NRF sleep	0.1	100	0.1	88	0.1	5	
micro-SD	0.01	100	0.01	100	25	3	
Other[b]	0.1	100	0.2	100	0.2	100	
Total	0.31		39.39		21.83		6.02

[a] 30 seconds is a relative conservative assumption. It means the user does not respond to a blinking half of the time.
[b] E.g. battery and power management IC's.

Table B.2: The empirically measured current consumption of the SleePic Core module at 3V which corresponds to the supply voltage from the Sensor module.

Mode *Duration [s]*	Idle 825 (91.5%)		Measure & Answer 15 (2%)		Reaction 60 (6.5%)		AVG
@3.7 V	mA	%	mA	%	mA	%	mA
uC	2	100	4	100	4	100	
NRF rx	-	-	18	7	-	-	
NRF tx	-	-	11	11	-	-	
NRF sleep	0.1	100	0.1	82	0.1	100	
LED1	0	-	-	-	8	50	
LED2	0	-	-	-	8	25	
Other	0.1	100	0.1	100	0.1	100	
Total	2.2		6.6		10.2		2.8

Table B.3: The empirically measured current consumption of the SleePic Watch at 3.7 V.

Bibliography

Achermann, P. and Borbely, A. A. (2003). Mathematical models of sleep regulation. *Frontiers in Bioscience*, 8:683–93.

Achermann, P. and Borbély, A. A. (1990). Simulation of human sleep: ultradian dynamics of electroencephalographic slow-wave activity. *Journal of Biological Rhythms*, 5(2):141–157.

Akerstedt, T. (2000). Consensus statement: fatigue and accidents in transport operations. *Journal of Sleep Research*, 9(4):395.

Akerstedt, T. and Folkard, S. (1995). Validation of the s and c components of the three-process model of alertness regulation. *Sleep*, 18(1):1–6.

Akerstedt, T., Folkard, S., and Portin, C. (2004). Predictions from the three-process model of alertness. *Aviation, Space and Environmental Medicine*, 75(3 Suppl):A75–83. Journal Article.

Amft, O. and Habetha, J. (2007). The myheart project. In van Langenhove, L., editor, *Book chapter in: Smart textiles for medicine and healthcare*, chapter 14, pages 275–297. Woodhead Publishing Ltd, Cambridge, England.

Amft, O., Lauffer, M., Ossevoort, S., Macaluso, F., Lukowicz, P., and Troster, G. (2004). Design of the qbic wearable computing platform. In *Proc. 15th IEEE International Conference on Application-Specific Systems, Architectures and Processors*, pages 398–410.

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., and Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26(3):342–392.

Anderer, P., Gruber, G., Parapatics, S., Woertz, M., Miazhynskaia, T., Klosch, G., Saletu, B., Zeitlhofer, J., Barbanoj, M. J., Danker-Hopfe, H., Himanen, S.-L., Kemp, B., Penzel, T., Grozinger, M., Kunz, D., Rappelsberger, P., Schlogl, A., and Dorffner, G. (2005). An e-health solution for automatic sleep classification according to rechtschaffen and kales: validation study of the somnolyzer 24 x 7 utilizing the siesta database. *Neuropsychobiology*, 51(3):115–133.

AwaCon (2009). http://www.sms.mavt.ethz.ch/research/projects/awacon. (last access: 12.1.2009).

Axbo (2008). http://www.axbo.com. (last access: 26.12.2008).

Aziz, J. N. Y., Karakiewicz, R., Genov, R., Chiu, A. W. L., Bardakjian, B. L., Derchansky, M., and Carlen, P. L. (2009). On-silicon neural activity monitoring and time-frequency analysis for early detection of epileptic seizures. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. (in press).

Balkin, T. and McBride, S. (2005). Managing sleep and alertness to sustain performance in the operational environment. Strategies to Maintain Combat Readiness during Extended Deployments Ű A Human Systems Approach 29-1 Ű 29-10, RTO-MP-HFM-124, Neuilly-sur-Seine, France: RTO. Meeting Proceedings.

Barr, L., Howarth, H., Popkin, S., and Carroll, R. J. (2005). A review and evaluation of emerging driver fatigue detection measures and technologies. In *International Conference on Fatigue Management in Transportation, Seattle, USA*.

Belenky, G., Balkin, T., Redmond, D., Sing, H., Thomas, M., Thorne, D., and Wesensten, N. (1998). Sustained performance during continuous operations: The us army s sleep management system. In Hartley, L., editor, *Managing Fatigue in Transportation. Proceedings of the Third International Conference on Fatigue and Transportation, Fremantle, Western Australia.* Elsevier Science Ltd., Oxford UK.

Bianchi, A., Villantieri, O., Mendez, M., and Cerutti, S. (2006). Signal processing and feature extraction for sleep evaluation in wearable devices. In *Proc. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS '06*, pages 3517–3520.

Biever, C. (2004). Cellphones turn into smart personal assistants. *New Scientist Magazine*, 2475:21.

Blake, H., Gerard, R. W., and Kleitman, N. (1939). Factors influencing brain potentials during sleep. *Journal of Neurophysiology*, 2(1):48–60.

Blankertz, B., Dornhege, G., Schafer, C., Krepki, R., Kohlmorgen, J., Muller, K. R., Kunzmann, V., Losch, F., and Curio, G. (2003). Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial eeg analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):127–131.

Bonato, P. (2005). Advances in wearable technology and applications in physical medicine and rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 2(1):2.

Bonnet, M. H. and Arand, D. L. (1997). Heart rate variability: sleep stage, time of night, and arousal influences. *Electroencephalography and Clinical Neurophysiology*, 102(5):390–396.

Borbély, A. A. (1982). A two process model of sleep regulation. *Human Neurobiology*, 1(3):195–204.

Borbély, A. A. and Achermann, P. (1999). Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms*, 14(6):557–568.

Bresler, M., Sheffy, K., Pillar, G., Preiszler, M., and Herscovici, S. (2008). Differentiating between light and deep sleep stages using an ambulatory device based on peripheral arterial tonometry. *Physiological Measurement*, 29(5):571–584.

Buckelmüller, J., Landolt, H.-P., Stassen, H. H., and Achermann, P. (2006). Trait-like individual differences in the human sleep electroencephalogram. *Neuroscience*, 138(1):351–356.

Burr, R. L. (2007). Interpretation of normalized spectral heart rate variability indices in sleep research: a critical review. *Sleep*, 30(7):913–919.

Buttfield, A., Ferrez, P. W., and Millán, J. d. R. (2006). Towards a robust BCI: Error potentials and online learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):164–168.

Buttussi, F. and Chittaro, L. (2008). Mopet: A context-aware and user-adaptive wearable system for fitness training. *Artificial Intelligence in Medcine*, 42(2):153–163.

Carskadon, M. and Dement, W. (1989). *Principles and Practice of Sleep Medicine*, chapter Normal human sleep: an overview, pages 3–13. W.B. Saunders, Philadelphia.

Chavarriaga, R., Ferrez, P. W., and Millán, J. d. R. (2007). To err is human: Learning from error potentials in brain-computer interfaces. In Shen, R. W. E. and Gu, F., editors, *Advances in Cognitive Neurodynamics ICCN 2007 Proceedings of the International Conference on Cognitive Neurodynamics.*, number IV, pages 777–782. Springer Netherlands.

Cirelli, C. and Tononi, G. (2008). Is sleep essential? *PLoS Biology*, 6(8):e216.

Closs, S. J. (1988). Assessment of sleep in hospital patients: a review of methods. *Journal of Advanced Nursing*, 13(4):501–510.

Cohen, A. (2006). *Biomedical Engineering Handbook*, volume 2 of *The electrical engineering handbook series*, chapter Biomedical Signal Analysis, pages I1–I22. Boca Raton : CRC Taylor & Francis, 3rd edition.

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., and Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5):461–9.

Committee on Sleep Medicine and Research, B. o. H. S. P. (2006). *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. The National Academies press, Washington, DC.

Czeisler, C. A., Zimmerman, J. C., Ronda, J. M., Moore-Ede, M. C., and Weitzman, E. D. (1980). Timing of rem sleep is coupled to the circadian rhythm of body temperature in man. *Sleep*, 2(3):329–46.

Danker-Hopfe, H., Kunz, D., Gruber, G., Klösch, G., Lorenzo, J. L., Himanen, S. L., Kemp, B., Penzel, T., Röschke, J., Dorn, H., Schlögl, A., Trenker, E., and Dorffner, G. (2004). Interrater reliability between scorers from eight european sleep laboratories in subjects with different sleep disorders. *Journal of Sleep Research*, 13(1):63–69.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., and Hobart, G. (1937). Changes in human brain potentials during the onset of sleep. *Science*, 86(2237):448–450.

de Souza, L., Benedito-Silva, A. A., Pires, M. L., Poyares, D., Tufik, S., and Calil, H. M. (2003). Further validation of actigraphy for sleep studies. *Sleep*, 26(1):81–5.

Dinges, D. F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*, 4(S2):4–14.

Dinges, D. F. (2004). Sleep debt and scientific evidence. *Sleep*, 27(6):1050–1052.

Dinges, D. F. and Grace, R. (1998). Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. Technical Report FHWA-MCRT-98-006, Federal Highway Administration, Office of Motor Carrier Research and Standards, Washington, DC.

Dinges, D. F., Maislin, G., Krueger, G. P., Brewster, R., and Carroll, R. J. (2005). Pilot test of fatigue management technologies. Technical Report FMCSA-RT-05-002, American Transportation Research Institute, Smyrna, GA.

Dinges, D. F. and Mallis, M. M. (1998). *Managing fatigue in transportation*, chapter Managing Fatigue by Drowsiness Detection: Can Technological Promises be Realized?, pages 209–223. Oxford: Elsevier Science Ltd.

Dinges, D. F., Orne, M. T., and Orne, E. C. (1985). Assessing performance upon abrupt awakening from naps during quasi-continuous operations. *Behavior Research Methods, Instruments, & Computers*, 17:37–45.

Dürr, P., Karlen, W., Guignard, J., Mattiussi, C., and Floreano, D. (2009). Evolutionary Selection of Features for Sleep/Wake Discrimination. *Journal of Artificial Evolution and Applications*. (in press).

Edinger, J. and Means, M. (2005). *Principles and Practice of Sleep Medicine*, chapter Overview of insomnia: Definitions, epidemiology, differential diagnosis, and assessment, pages 702 – 713. Philadelphia: Elsevier/Saunders, 4th edition.

Edward, S., Nadezhda, S., Stephanie, S., Michael, N., and Group, C. S. (2004). Activity-based sleep/wake identification in infants. *Physiological Measurement*, 25(5):1291. 0967-3334.

Equivital (2008). http://www.equivital.co.uk. (last access: 10.4.2008).

Ferrez, P. and Millán, J. (2005). You are wrong!—automatic detection of interaction errors from brain waves. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, Edinburgh, UK.

Flanigan, W. F. (1972). *The Sleeping Brain: Perspectives in the Brain Sciences*, volume 1 of *Perspectives in the Brain Sciences*, chapter Behavioral states and electroencephalograms of reptiles, pages 14–18. Brain Information Service/Brain Research Institute (UCLA), Los Angeles, CA.

Gerasimov, V. (2003). *EVERY SIGN OF LIFE*. PhD thesis, Program in Media Arts and Sciences, School of Architecture and Planning, Massachusetts Institute of Technology.

Gilbert, S. S., van den Heuvel, C. J., Ferguson, S. A., and Dawson, D. (2004). Thermoregulation as a sleep signalling system. *Sleep Medicine Reviews*, 8(2):81–93.

Gorny, S. and Spiro, J. (2001). Comparing different methodologies used in wrist actigraphy. *Sleep Review*, Summer:40–42.

Hagan, M. and Menhaj, M. (1994). Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6):989–993.

Harris, F. (1978). On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83.

Hartley, L., Horberry, T., Mabbott, N., and Krueger, G. (2000). Review of fatigue detection and prediction technologies. Working paper, National Road Transport Commission, Level 5/326 William Street Melbourne VIC 3000.

Hedner, J., Pillar, G., Pittman, S. D., Zou, D., Grote, L., and White, D. P. (2004). A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep*, 27(8):1560–1566.

Herscovici, S., Pe'er, A., Papyan, S., and Lavie, P. (2007). Detecting rem sleep from the finger: an automatic rem sleep algorithm based on peripheral arterial tone (pat) and actigraphy. *Physiological Measurement*, 28(2):129–140.

Hilten, J. J. V., Middelkoop, H. A., Kuiper, S. I., Kramer, C. G., and Roos, R. A. (1993). Where to record motor activity: an evaluation of commonly used sites of placement for activity monitors. *Electroencephalography and Clinical Neurophysiology*, 89(5):359–362.

Hobson, J. A. (2005). Sleep is of the brain, by the brain and for the brain. *Nature*, 437(7063):1254–1256.

Holroyd, C. B. and Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychology Revue*, 109(4):679–709.

Hori, T. (1982). Electrodermal and electro-oculographic activity in a hypnagogic state. *Psychophysiology*, 19(6):668–672.

Hori, T., Hayashi, M., and Morikawa, T. (1994). *Sleep onset: Normal and abnormal processes*, chapter Topographical EEG changes and the hypnagogic experience, pages 237–253. American Psychological Association,Washington, DC.

Huynh, T. and Schiele, B. (2006). Unsupervised discovery of structure in activity data using multiple eigenspaces. In *Location- and Context-Awareness*, volume 3987/2006 of *Lecture Notes in Computer Science*, pages 151–167. Springer Berlin / Heidelberg.

Iber, C., Chesson, A., and Quan, S. (2007). *The AASM manual for the scoring of sleep and associated events : rules, terminology and technical specifications*. American Academy of Sleep Medicine, c, 2nd edition.

Itamar (2009). http://www.itamar-medical.com/product.asp?pid=3006. (last access: 26.12.2008).

Jean-Louis, G., von Gizycki, H., Zizi, F., Fookson, J., Spielman, A., Nunes, J., Fullilove, R., and Taub, H. (1996). Determination of sleep and wakefulness with the actigraph data analysis software (adas). *Sleep*, 19(9):739–43.

Johannes, B., Salnitski, V., Soll, H., Rauch, M., and Hoermann, H.-J. (2008). De-individualized psychophysiological strain assessment during a flight simulation test–validation of a space methodology. *Acta Astronautica*, 63(7-10):791 – 799. From Dream to Reality: Living, Working and Creating for Humans in Space - A selection of papers presented at the 16th IAA Humans in Space Symposium, Beijing, China, 2007.

Johns, M., Tucker, A., and Chapman, R. (2005). Monitoring the drowsiness of drivers: A new method based on the velocity of eyelid movements. In *International ITS Conference, San Francisco*.

Kaefer, G., Prochart, G., and Weiss, R. (2003). Wearable alertness monitoring for industrial applications. In *Proc. Seventh IEEE International Symposium on Wearable Computers*, pages 254–255.

Karlen, W., Mattiussi, C., and Floreano, D. (2007). Adaptive sleep/wake classification based on cardiorespiratory signals for wearable devices. In *Proc. IEEE Biomedical Circuits and Systems Conference BIOCAS 2007*, pages 203–206, Montreal, Quebec.

Karlen, W., Mattiussi, C., and Floreano, D. (2008). Improving actigraph sleep/wake classification with cardio-respiratory signals. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5262–5265.

Karlen, W., Mattiussi, C., and Floreano, D. (2009). Sleep and wake classification with ecg and respiratory effort signals. *IEEE Transactions on Biomedical Circuits and Systems.* (in press).

Kaushik, P., Intille, S. S., and Larson, K. (2008). User-adaptive reminders for home-based medical tasks. a case study. *Methods of Information in Medicine*, 47(3):203–207.

Åkerstedt, T., Connor, J., Gray, A., and Kecklund, G. (2008). Predicting road crashes from a mathematical model of alertness regulation–the sleep/wake predictor. *Accident Analysis & Prevention*, 40(4):1480 – 1485.

Koralewski.de (2008). http://www.koralewski.de/html/entwicklung _health-lab.html. (last access: 12.2.2009).

Korhonen, I. and Saranummi, N. (2007). Personal health systems - need, market place and challenges to their wide scale adoption. In *Proc. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2007*, pages 6152–6153.

Krause, A., Smailagic, A., and Siewiorek, D. P. (2006). Context-aware mobile computing: learning context- dependent personal preferences from a wearable sensor array. *Mobile Computing, IEEE Transactions on*, 5(2):113–127.

Kräuchi, K., Cajochen, C., and Wirz-Justice, A. (2004). Waking up properly: is there a role of thermoregulation in sleep inertia? *Journal of Sleep Research*, 13(2):121–127.

Kushida, C. A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O., and Dement, W. C. (2001). Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Medicine*, 2(5):389–396.

Lal, S. K. L. and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, 55(3):173–194.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Limor, A., Roni, S., Baruch, E.-A., Ron, P., and Giora, P. (2009). Forty- versus 20-minute trials of the maintenance of wakefulness test regimen for licensing of drivers. *Journal of Clinical Sleep Medicine*, 5(1):57–62.

Lotjonen, J., Korhonen, I., Hirvonen, K., Eskelinen, S., Myllymaki, M., and Partinen, M. (2003). Automatic sleep-wake and nap analysis with a new wrist worn online activity monitoring device vivago wristcare. *Sleep*, 26(1):86–90.

Lukowicz, P. (2008). Wearable computing and artificial intellegence for healthcare applications. *Artificial Intelligence in Medcine*, 42(2):95–98.

Malik, M., of the European Society of Cardiology, T. F., and the North American Society of Pacing Electrophysiology (1996). Heart rate variability : Standards of measurement, physiological interpretation, and clinical use. *Circulation*, 93(5):1043–1065.

Martin, T., Jovanov, E., and Raskovic, D. (2000). Issues in wearable computing for medical monitoring applications: a case study of a wearable ecg monitoring device. In *Proc. Fourth International Symposium on Wearable Computers*, pages 43–49.

Matsushita, S., Shiba, A., and Nagashima, K. (2006). A wearable fatigue monitoring system: application of human-computer interaction evaluation. In *AUIC '06: Proceedings of the 7th Australasian User interface conference*, pages 161–164, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Mattiussi, C. and Floreano, D. (2007). Analog Genetic Encoding for the Evolution of Circuits and Networks. *IEEE Transactions on Evolutionary Computation*, 11(5):596–607.

Middelkoop, H. A. M., Dam, E. M., Smilde-Van den Doel, D. A., and Dijk, G. (1997). 45-hour continuous quintuple-site actimetry: Relations between trunk and limb movements and effects of circadian sleep-wake rhythmicity. *Psychophysiology*, 34(2):199–203.

Millán, J. d. R. (2008). Brain-controlled robots. *IEEE Intelligent Systems*, 23(3):74–76.

MyHeart-IST (2009). http://www.hitech-projects.com/euprojects /myheart/objectives.html. (last access: 26.12.2008).

Naifeh, K. H. and Kamiya, J. (1981). The nature of respiratory changes associated with sleep onset. *Sleep*, 4(1):49–59.

Naitoh, P. (1992). *Why We Nap: Evolution, Chronobiology, and Functions of Polyphasic and Ultrashort Sleep*, chapter Minimal Sleep to Maintain Performance: The Search fo Sleep Quantum in Sustained Operations, pages 199–217. Birkhäuser, Boston.

Nguyen, D. and Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Proceedings of the International Joint Conference on Neural Networks IJCNN*, volume 3, pages 21–26.

Ogilvie, R. D. (2001). The process of falling asleep. *Sleep Medicine Reviews*, 5(3):247–270.

Ogilvie, R. D. and Wilkinson, R. T. (1984). The detection of sleep onset: behavioral and physiological convergence. *Psychophysiology*, 21(5):510–520.

Okada, S., Ohno, Y., Goyahan, Kato-Nishimura, K., Mohri, I., and Taniike, M. (2008). Examination of non-restrictive and non-invasive sleep evaluation technique for children using difference images. In *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, pages 3483–3487.

Optalert (2009). http://www.optalert.com/. (last access: 13.1.2009).

Oswald, I., Taylor, A. M., and Treisman, M. (1960). Discriminative responses to stimulation during human sleep. *Brain*, 83:440–453.

Patel, M. R. and Davidson, T. M. (2007). Home sleep testing in the diagnosis and treatment of sleep disordered breathing. *Otolaryngologic Clinics of North America*, 40(4):761–784.

Penzel, T. and Conradt, R. (2000). Computer based sleep recording and analysis. *Sleep Med Rev*, 4(2):131–148.

Penzel, T., Kemp, B., Klosch, G., Schlogl, A., Hasan, J., Varri, A., and Korhonen, I. (2001). Acquisition of biomedical signals databases. *IEEE Magazine on Engineering in Medicine and Biology*, 20(3):25–32.

Picard, R. W. (1998). *Affective computing*. The MIT Press, Cambridge, MA, 2nd printing edition.

PIPS (2009). http://www.pips.eu.org/. (last access: 13.1.2009).

Pivik, R. T. and Busby, K. (1996). Heart rate associated with sleep onset in preadolescents. *Journal of Sleep Research*, 5(1):33–36.

Pollak, C. P., Tryon, W. W., Nagaraja, H., and Dzwonczyk, R. (2001). How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep*, 24(8):957–965.

Principe, J. and Tome, A. (1989). Performance and training strategies in feedforward neural networks: an application to sleep scoring. *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 341–346 vol.1.

PubMed (2009). http://www.ncbi.nlm.nih.gov/pubmed/. (last access: 20.2.2009).

Randall Wilson, D. and Martinez, T. R. (2000). The inefficiency of batch training for large training sets. In *Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks IJCNN 2000*, volume 2, pages 113–117.

Rechtschaffen, A., Kales, A., Berger, R., and Dement, W. (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Public Health Service, US Government Printing Office*.

Redmond, S. and Heneghan, C. (2006). Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. *IEEE Transactions on Biomedical Engineering*, 53(3):485–496.

Robert, C., Gaudy, J.-F., and Limoge, A. (2002). Electroencephalogram processing using neural networks. *Clinical Neurophysiology*, 113(5):694–701.

Robert, C., Guilpin, C., and Limoge, A. (1998). Review of neural network applications in sleep research. *Journal of Neuroscience Methods*, 79(2):187–193.

Ruffini, G., Dunne, S., Farres, E., Cester, I., Watts, P. C. P., Ravi, S., Silva, P., Grau, C., Fuentemilla, L., Marco-Pallares, J., and Vandecasteele, B. (2007). Enobio dry electrophysiology electrode; first human trial plus wireless electrode system. In *Proc. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2007*, pages 6689–6693.

Sadeh, A. and Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Medicine Reviews*, 6(2):113–124.

Sadeh, A., Sharkey, K. M., and Carskadon, M. A. (1994). Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep*, 17(3):201–207.

Sanchez, J. C., Principe, J. C., Nishida, T., Bashirullah, R., Harris, J. G., and Fortes, J. A. B. (2008). Technology and signal processing for brain-machine interfaces. *IEEE Signal Processing Magazine*, 25(1):29–40.

Sartori, V., Schumacher, P. M., Bouillon, T., Luginbuehl, M., and Morari, M. (2006). On-line estimation of propofol pharmacodynamic parameters. In *Proc. 27th Annual International Conference of the Engineering in Medicine and Biology Society IEEE-EMBS 2005*, pages 74–77.

Schalk, G., Wolpaw, J. R., McFarland, D. J., and Pfurtscheller, G. (2000). Eeg-based communication: presence of an error potential. *Clinical Neurophysiology*, 111(12):2138–2144.

Scheffler, M. and Hirt, E. (2004). Wearable devices for emerging healthcare applications. In *Proc. 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEMBS '04*, volume 2, pages 3301–3304.

Schmid, S. M., Hallschmid, M., Jauch-Chara, K., Bandorf, N., Born, J., and Schultes, B. (2007). Sleep loss alters basal metabolic hormone secretion and modulates the dynamic counterregulatory response to hypoglycemia. *Journal of Clinical Endocrinology & Metabolism*, 92(8):3044–3051.

Schmidt, A., Aidoo, K., Takaluoma, A., Tuomela, U., Van Laerhoven, K., and Van de Velde, W. (1999). Advanced interaction in context. In *Handheld and Ubiquitous Computing*, number 1707/1999 in Lecture Notes in Computer Science, pages 89–101. Springer Berlin / Heidelberg.

Schnall, R. P., Shlitner, A., Sheffy, J., Kedar, R., and Lavie, P. (1999). Periodic, profound peripheral vasoconstriction–a new marker of obstructive sleep apnea. *Sleep*, 22(7):939–946.

scholarpedia.org (2009). http://www.scholarpedia.org/article /neurobiology_-of_sleep_and_wakefulness. (last access: 26.12.2008).

Scholz, U. J., Bianchi, A. M., Cerutti, S., and Kubicki, S. (1997). Vegetative background of sleep: Spectral analysis of the heart rate variability. *Physiology & Behavior*, 62(5):1037 – 1043.

Sensation-IST (2009). http://www.sensation-eu.org/. (last access: 26.12.2008).

Shen, J., Barbera, J., and Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Medicine Reviews*, 10(1):63–76.

Shiihara, Y., Nakajima, M., Miyazaki, Y., Nakamiya, T., Miyamoto, M., Takahashi, S., and Kodama, M. (1998). Evaluation of sleep using ambulatory skin potential recording: differences between morning and evening type. *Psychiatry and Clinical Neurosciences*, 52(2):167–8.

Shinar, Z., Akselrod, S., Dagan, Y., and Baharav, A. (2006). Autonomic changes during wake-sleep transition: A heart rate variability based approach. *Autonomic Neuroscience*, 130(1-2):17–23.

Siegel, J. M. (2005). Clues to the functions of mammalian sleep. *Nature*, 437(7063):1264–1271.

Sleeptracker (2007). http://www.sleeptracker.com. (last access: 26.12.2008).

SolarImpulse (2008). http://www.solarimpulse.com. (last access: 26.12.2008).

Somnomedics (2008). http://www.somnomedics.de/international/products /somnowatch/long-term-eeg.htm. (last access: 26.12.2008).

Spiegel, K., Knutson, K., Leproult, R., Tasali, E., and Cauter, E. V. (2005). Sleep loss: a novel risk factor for insulin resistance and type 2 diabetes. *Journal of Applied Physiology*, 99(5):2008–2019.

Spiegel, K., Sheridan, J. F., and Cauter, E. V. (2002). Effect of sleep deprivation on response to immunization. *JAMA*, 288(12):1471–1472.

Stampi, C. (1992). *Why We Nap: Evolution, Chronobiology, and Functions of Polyphasic and Ultrashort Sleep*, chapter The Effects of Polyphasic und Ultrashort Sleep Schedules, pages 137–180. Birkhäuser, Boston.

Starlab, E. (2009). http://starlab.es/products/enobio. (last access: 13.1.2009).

Stepanski, E. J. and Wyatt, J. K. (2003). Use of sleep hygiene in the treatment of insomnia. *Sleep Medicine Reviews*, 7(3):215 – 225.

Steptoe, A., Peacey, V., and Wardle, J. (2006). Sleep duration and health in young adults. *Archives of Internal Medicine*, 166(16):1689–1692.

Stiefmeier, T., Roggen, D., Troster, G., Ogris, G., and Lukowicz, P. (2008). Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, 7(2):42–50.

Sung, M., Gips, J., Eagle, N., Madan, A., Caneel, R., DeVaul, R., Bonsen, J., and Pentland, A. (2005). Mobile-it education (mit. edu): m-learning applications for classroom settings. *Journal of Computer Assisted Learning*, 21(3):229–237.

Tassi, P. and Muzet, A. (2000). Sleep inertia. *Sleep Medicine Reviews*, 4(4):341–353.

Telser, S., Staudacher, M., Ploner, Y., Amann, A., Hinterhuber, H., and Ritsch-Marte, M. (2004). Can one detect sleep stage transitions for on-line sleep scoring by monitoring the heart rate variability? *Somnologie*, 8(2):33–41.

Trinder, J., Beveren, J. A. V., Smith, P., Kleiman, J., and Kay, A. (1997). Correlation between ventilation and eeg-defined arousal during sleep onset in young subjects. *Journal of Applied Physiology*, 83(6):2005–2011.

Tryon, W. W. (1991). *Activity Measurement in Psychology and Medicine*, chapter 6. Activity and sleep, pages 149–195. Applied Clinical Psychology. Plenum Press, New York.

Van Dongen, H. P. A., Rogers, N. L., and Dinges, D. F. (2003). Sleep debt: Theoretical and empirical issues. *Sleep and Biological Rhythms*, 1(1):5–13.

Venkatraman, V., Chuah, Y. M. L., Huettel, S. A., and Chee, M. W. L. (2007). Sleep deprivation elevates expectation of gains and attenuates response to losses following risky decisions. *Sleep*, 30(5):603–609.

Vivago (2009). http://www.vivago.fi/. (last access: 10.2.2009).

Walsh, J. K. and Engelhardt, C. L. (1999). The direct economic costs of insomnia in the united states for 1995. *Sleep*, 22 Suppl 2:S386–S393.

Wealthy-IST (2009). http://www.wealthy-ist.com/. (last access: 1.1.2009).

Webster, J. B., Kripke, D. F., Messin, S., Mullaney, D. J., and Wyborney, G. (1982). An activity-based sleep monitor system for ambulatory use. *Sleep*, 5(4):389–399.

Williams, H. L., Morlock, H. C., and Morlock, J. V. (1966). Instrumental behavior during sleep. *Psychophysiology*, 2(3):208–216.

Williamson, A. and Chamberlain, T. (2005). Review of on-road driver fatigue monitoring devices. Technical report, NSW Injury Risk Management Research Centre University of New South Wales, Australia.

Wilson, J., Walker, B., Lindsay, J., Cambias, C., and Dellaert, F. (2007). Swan: System for wearable audio navigation. In *Proc. 11th IEEE International Symposium on Wearable Computers*, pages 91–98.

Wirz-Justice, A., Benedetti, F., Berger, M., Lam, R. W., Martiny, K., Terman, M., and Wu, J. C. (2005). Chronotherapeutics (light and wake therapy) in affective disorders. *Psychological Medicine*, 35(7):939–944.

Curriculum vitæ

WALTER CH. KARLEN

I was born in St-Gallen (SG), Switzerland on November 8th 1979 as the son of Stephanie and Franz Karlen-Blatter. I grew up in Arbon (SG), Uhwiesen (ZH), Le-Mont-sur-Lausanne (VD) and Brig-Glis (VS). In Brig-Glis I graduated from the Kollegium Spiritus Sanctus (Type C) in 1999.

In the same year I started my studies in micro-engineering at the Ecole Polytechnique Fédérale de Lausanne (EPFL) from which I received a Master of Science in Micro-engineering in April 2005. I wrote my master thesis at the Université Libre de Bruxelles (ULB), Belgium under the supervision of Prof. Jean-Louis Deneubourg at the Unité d'Ecologie Sociale (ULB) and Prof. Roland Siegwart at the Autonomous Systems Lab (EPFL). The master thesis was focused on the perception of miniaturized robots for the interaction with insects.

Since April 2005, I am a Ph.D. student under the supervision of Prof. Dario Floreano at the Laboratory of Intelligent Systems (LIS) at EPFL. In collaboration with Prof. Floreano I launched within the LIS the research branch of body sensing for the assessment of human wellness. The focus lied in the detection of fatigue and sleep with wearable and intelligent systems. Amongst the work presented in this thesis I also developed a software tool to predict sleepiness of humans performing in extreme environments. The software was used and tested in the Race Across America 2007 by the medical supervisor of the Swiss extreme cyclist Daniel Wyss. The software helped to reduce the cycler's total sleep time to 12

hours while he was en-route for 9 days, 4 hours and 29 minutes and left 3'040 miles behind him.

During my time at the LIS, I proposed and supervised several student projects related to sleep, fatigue and wellness detection. I also prepared and instructed practical exercises for the lectures "Mobile Robots" taught by Dr. Jean-Christoph Zufferey, and "Bio-inspired Adaptive Machines" taught by Prof. Dario Floreano.

From 2005 to 2009 I was a scientific adviser for pilot health monitoring in the Solar Impulse project. During this work I co-supervised several bio-medical experiments for the performance evaluation of pilots in extreme environments. I designed and developed a low-power automatic monitoring system for the oxygen reserve in the cockpit and the surveillance of the blood oxygen saturation (SpO2) of the pilot at high altitudes.

My current research interests lie in the fields of bio-medical signal processing, human physiology, embedded and wearable systems, bio-medical and robotic sensors and actuators, and human-machine interaction.

I am a member of the IEEE Engineering in Medicine and Biology Society, the Société Académique Lémania and the student associations Brigensis and D'OBRU.