ACTLab

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Dissertation

Methods for monitoring the human circadian rhythm in free-living

Florian Wahl

Passau, April 2019

A dissertation submitted to the faculty of computer science and mathematics in partial fulfillment of the requirements for the degree of doctor of natural sciences.

Supervisor: Prof. Dr. Oliver Amft
Internal Examiner: Prof. Dr. Michael Granitzer
External Examiner: Prof. Dr. Alois Ferscha

Copyright Information

Abbreviations

ADL	Activity of daily living
AE	Average error
ANN	Artificial neural network
ARMAX	Autoregressive–moving-average model with exogenous inputs
Acc	Accelerometer
CAD	Computer aided design
CART	Classification and regression tree
CBT	Core body temperature
CCFL	Cold cathode fluorescent lamp
CNN	Convolutional neural network
CRMA	Constant routine melatonin assessment
CV	Cross validation
DLMO	Dim light melatonin onset
DS	Duration of sleep
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
FAA	Federal Aviation Administration
FPR	False positive rate
GMM	Gaussian mixture model
GPS	Global positioning system
HLM	Head light measurement
HMM	Hidden Markov model
HR	Heart rate
IMU	Inertial measurement unit
KM	Kronauer model
LED	Light-emitting diode
LODO	Leave one day out
LOO	Leave one out
LOPO	Leave one participant out
LSQ	Least squares
LSVM	Linear support vector machine
MAD	Mean absolute deviation
MCTQ	Munich chronotype questionnaire
MCU	Main controller unit
ME	Median error
MEQ	Morningness-eveningness questionnaire
MS	Midsleep timing
MSe	Midsleep error
NN	Neural network
PC	Personal computer
PCA	Principal component analysis
PLEA	Passive light estimation application

PPG	Photoplethysmography
PPV	Positive predictive value
PSG	Polysomnography
RMSe	Root-mean-square error
RNN	Recurrent neural network
ROC AUC	Receiver operating characteristic area under curve
RSI	Repetitive strain injury
SCN	Suprachiasmatic nucleus
SNR	Signal-to-noise ratio
SO	Sleep onset time
SSID	Service set identifier
SUS	System usability scale
SVM	Support vector machine
TPM	Two process model
TPR	True positive rate
TV	Television
WS	Within subject
WU	Wake up time
XGB	Extreme gradient boosting
mMAD	Mean median absolute deviation

Summary

Our internal clock, the circadian clock, determines at which time we have our best cognitive abilities, are physically strongest, and when we are tired. Circadian clock phase is influenced primarily through exposure to light. A direct pathway from the eyes to the suprachiasmatic nucleus, where the circadian clock resides, is used to synchronise the circadian clock to external light-dark cycles.

In modern society, with the ability to work anywhere at anytime and a full social agenda, many struggle to keep internal and external clocks synchronised. Living against our circadian clock makes us less efficient and poses serious health impact, especially when exercised over a long period of time, e.g. in shift workers. Assessing circadian clock phase is a cumbersome and uncomfortable task. A common method, dim light melatonin onset testing, requires a series of eight saliva samples taken in hourly intervals while the subject stays in dim light condition from 5 hours before until 2 hours past their habitual bedtime.

At the same time, sensor-rich smartphones have become widely available[1] and wearable computing is on the rise. The hypothesis of this thesis is that smartphones and wearables can be used to record sensor data to monitor human circadian rhythms in free-living. To test this hypothesis, we conducted research on specialised wearable hardware and smartphones to record relevant data, and developed algorithms to monitor circadian clock phase in free-living. We first introduce our smart eyeglasses concept, which can be personalised to the wearers head and 3D-printed. Furthermore, hardware was integrated into the eyewear to recognise typical activities of daily living (ADLs). A light sensor integrated into the eyeglasses bridge was used to detect screen use. In addition to wearables, we also investigate if sleep-wake patterns can be revealed from smartphone context information. We introduce novel methods to detect sleep opportunity, which incorporate expert knowledge to filter and fuse classifier outputs. Furthermore, we estimate light exposure from smartphone sensor and weather information. We applied the Kronauer model[2] to compare the phase shift resulting from head light measurements, wrist measurements, and smartphone estimations.

We found it was possible to monitor circadian phase shift from light estimation based on smartphone sensor and weather information with a weekly error of 32 ± 17 min, which outperformed wrist measurements in 11 out of 12 participants. Sleep could be detected from smartphone use with an onset error of 40 ± 48 min and wake error of 42 ± 57 min. Screen use could be detected smart eyeglasses with 0.9 ROC AUC for ambient light intensities below 200 lux. Nine clusters of ADLs were distinguished using Gaussian mixture models with an average accuracy of 77%. In conclusion, a combination of the proposed smartphones and smart eyeglasses applications could support users in synchronising their circadian clock to the external clocks, thus living a healthier lifestyle.

[1] J. Poushter. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center*, 22, 2016

[2] M. E. Jewett, D. B. Forger, and R. E. Kronauer. Revised limit cycle oscillator model of human circadian pacemaker. *Journal of Biological Rhythms*, 14(6):493–500, 1999

Acknowledgements

First and foremost, I would like to thank Prof. Dr. Oliver Amft for his patient guidance. Despite your busy schedule, you always found time for fruitful discussions and giving valuable feedback. Oliver, thank you for being a great mentor and providing me with advice and encouragement over the past years.

Furthermore I would like to thank Prof. Dr. Michael Granitzer and Prof. Dr. Alois Ferscha for serving as examiners. I am thankful for your feedback on this thesis and grateful for the time you invested.

This dissertation started at TU of Eindhoven. I would like to thank all former colleagues and students for the valuable discussions and great times in the ACTLab. Especially, I would like to thank Piero and Gabriele for introducing me to the Italian way of life. Thanks to Steffen and Ferdinand for being excellent roommates who turned into friends for life.

After moving to Passau, Adrian shared an office with me for two and a half years. Adrian, thank you for the good discussions, countless laughs, and the good time in Japan. Thanks to my colleague and roommate Luis for the excellent time in our apartment, where we lived in harmony for over two years, and the valuable advice especially towards the end of this thesis. Thank you Martin for your contributions to the WISEglass project, especially for 3D-printing and hardware hacking. I would like to thank all former colleagues and students at the University of Passau. I enjoyed working with all of you.

Within the OnTime project, I had the chance to meet chronobiologists. I would like to thank Roelof Hut, Thomas Kantermann, Martha Merrow, Till Roenneberg, and Giulia Zerbini for the valuable input. Without the financial support of the Technologiestichting STW, this thesis would not have been possible. Thanks to everyone in the OnTime project for the fruitful discussions and interesting talks at the symposiums.

Two years ago, I met a young duo of entrepreneurs, Michael and Sebastian, who had just started a new venture. Thank you for the opportunity to work with you and the Smartricity team and for letting me write parts of this dissertation at your office in the weekends. Sebastian, thank you for good conversations during our evening runs.

Thanks to all my wonderful colleagues at Technologie Campus Grafenau, where I am fortunate to work at. Thank you Diane and Robert for giving me the opportunity to work while finishing this thesis. Benedikt, I am glad we met! It is a pleasure to work with you. Thank you for your companionship, proofreading parts of the thesis, and your valuable feedback. Daniela, thank you for the help with the cover design.

I would like to thank Rebecca, Matthieu, Sophie, and Jürgen for the warm welcome into their family. Thank you for watching Charlotte in the weekends so I could find time to finish this thesis. Sebastian, you are one of the few friends I met long before the beginning of this thesis. I am grateful for your company over the past 22 years.

Especially I would like to thank my brother Philipp, my mother Ingrid, and my father Manfred. Thank you for your support and encouragement throughout my life. This work could not have been done without you. Marion and Helmut, thank you for your support and guidance as my godparents.

Bastian & Charlotte, thank you for being wonderful children, sound sleepers, and bringing love and joy into our lives. Your loving and gentle characters are a pleasure to be around.

Finally and most importantly, I would like to thank Dorothée for her support and patience over the past years. Thank you for loving me as I am and being such an encouraging partner. You are my *Lieblingsmensch*. I love you to the moon and back.

for Dorothée

Contents

1

Introduction

Since the stone age human life was guided by the solar cycle. Living a diurnal routine increased chances of survival for humans, as they could defeat their predators more likely during daytime. During nighttime, humans seeked shelter and recovered by sleeping. With time of day being so important to survival, our organism adapted to keep in sync with the solar cycle over the years. In today's developed world, we have artificial light sources and most of us do not need to hunt their own meals. In this chapter, human circadian rhythms and common measurement techniques are introduced. Next, we discuss how wearable and portable computers can be used in combination with context recognition and expert models to support better synchronisation of circadian and social clocks in free-living users. Finally, we introduce the thesis objectives and give an overview over the remaining chapters.

1.1 Human circadian rhythms

Circadian rhythms have a period of approximately one day (from latin circa diem). Human circadian rhythms follow the solar light/dark cycle and are entrained through timing and intensity of light exposure. Depending on the timing and intensity of light exposure, the circadian phase is compressed or expanded, shifting the circadian clock forwards or backwards, respectively [76]. Circadian clocks are distributed over all cells in the human body. The master clock resides in the suprachiasmatic nucleus (SCN), which receives light/dark information from the eyes through a direct pathway. Timing information is relayed from the SCN to peripheral clocks through autonomic innervation, body temperature, humoral signals, and feeding-related cues [65]. Light information received by the SCN does not originate from classic photoreceptors on the retina (rods and cones) used for vision, but from specialised photosensitive ganglion cells. While human vision is most sensitive at 555 nm wavelength, ganglion cells are most sensitive to blue light, peaking at 464 nm [20].

Personal circadian timing preference is expressed through chronotype. Chronotypes are naturally distributed ranging from early (larks) to late (owl) types. Chronotype can be estimated, e.g. using the Munich Chronotype Questionnaire [77] from midsleep points on days without alarm clock use (i.e. waking up without an external trigger). Chronotype changes throughout the lifespan of a person. Starting as early types, adolescents shift towards a later type on average. During adulthood, chronotype drifts towards an earlier type again.

Many functions of the body depend on the circadian phase, e.g. physical strength or the ability to focus peak at a certain time of day. The personal circadian phase thus may determine the biologically optimal time for certain activities. If the phase of the circadian clock is known, daily activities could be scheduled such that they are timed optimal according to our circadian rhythm. However, due to many obligations it is difficult to synchronise circadian and social clocks.

Social jet lag is defined as the amount of dissynchrony of internal (circadian) and external (wall clock) time. Living against the internal clock can have serious impact on health, e.g. obesity and cancer [56, 75]. School timing has been found to contribute to social jet lag of school children [90]. Another contributor to social jet lag is artifical light, especially from LED light sources, as LED light is rich in the blue part of the visible light spectrum, to which the ganglion cells are most sensitive to. Thus, LED backlit computer and smartphone screens contribute to social jet lag, as working anywhere at any time has become normal.

Chronobiology is a field of biology that examines circadian rhythms in organisms. Research on chronobiology started in the 18th century, when astronomer Jean Jacques d'Ortous de

Mairan reported on the daily leaf movements of the mimosa plant. In the 1960s the first symposium on chronobiology was held, and in 2017 the Nobel Prize in Physiology or Medicine was awarded to the chronobiologists Jeffrey C. Hall, Michael Rosbash and Michael W. Young for their discoveries of molecular mechanisms controlling the circadian rhythm.

1.2 Common measurement methods

To minimise social jet lag, measurement of circadian phase is key. Once circadian phase is known, lifestyle can be adjusted and circadian phase can be tracked to measure impact of changes made. In chronobiology, a set of established measurement methods exists. In the following, a summary of measurement methods is presented from more intrusive to less intrusive methods.

Polysomnography (PSG) is used in sleep laboratories. PSG monitors multiple body functions, including brain activity (EEG), eye movements (EOG), muscle activations (EMG), and heart activity (ECG). PSG requires a large number of wires to be attached to a subject (typically at least 22 wires) by a technician. Evaluation is performed using an expert or algorithm for scoring. PSG can be used to measure sleep onset latency, sleep efficiency, and distinguish sleep stages.

Core body temperature (CBT) also follows a circadian rhythm. Thus, CBT is a common marker in chronobiology studies. CBT can be obtained using a rectal probe or a temperature sensor with wireless transmitter encapsulated in a digestible pill. Skin temperature can be recorded using small temperature sensors combined with a data logger, which are placed on the skin [59].

A common standard in measuring circadian clock phase is a dim light melatonin onset (DLMO) assessment. Melatonin secretion takes place during internal night and in the absence of blue light. Thus, onset timing of melatonin production in a dim light environment is considered an important circadian phase marker. DLMO is commonly measured over a period of 8 hours, during which subjects need to remain in a dim light environment. Saliva samples are taken during hourly intervals and DLMO is computed using regression.

Actigraphy is commonly used in circadian biology to track subjects activity, especially their sleep/wake rhythm. Wrist-worn actigraphs record motion intensity accumulated over 30 sec to 5 min intervals. Subsequently, sleep/wake patterns are identified from motion patterns either manually by a technician or automatically using algorithms. In addition, many actigraphs feature a light sensor, which records the light intensity the subject is exposed to.

Multiple standardised surveys can be used to assess chronotype and sleep habits. Chronotype can be estimated using the Munich Chronotype Questionnaire (MCTQ) [77]. Similarly, the morningness-eveningness questionnaire (MEQ) [49] estimates timing of the circadian temperature rhythm. Sleep diaries are often used to determine sleep/wake patterns and alarm clock use.

1.3 Wearable and mobile computers

Many of the common measurement methods cannot be applied in free-living. In addition, existing methods do not provide continuous measurements. Advances made in wearable computing over the past years could enable chronobiology researchers and patients to benefit from continuous measurement methods in free-living. Wearable computers have become popular,

due to advances in electronic miniaturisation, wireless communication, battery technology, and energy efficiency of integrated circuits. While smart watches and fitness bands struggle to gain and maintain a larger user base [30], specialised systems can help to answer specific research questions. We believe adoption rates of wearable computers suffer from a number of shortcomings, including the lack of personalisation and social acceptance.

Personalisation has benefits beyond increased adoption: Wearable sensor technology benefits from being fitted to the wearer, as sensor noise can be reduced. Thus, wearable sensor devices should be personalised in order to optimise fit. As most industrial products use injection moulding technology to fabricate device enclosures, personalisation of wearables has hardly been addressed in the past. For injection moulding, an expensive mould is used to produce enclosures. Thus, injection moulded enclosures are typically offered in only one or few sizes. In contrast, 3D-printing technology can be used to print personalised enclosures.

Social acceptance is a key factor for the adoption of wearable computers. If a vis-à-vis feels observed during a conversation, e.g. by a wearable's camera, acceptance drops according to Koelle et al. [58]. Ideally, wearable computers integrate into existing everyday accessories to minimise public stigmatisation for the user. A particular example is the Google Glass project: Glass consisted of a head-mounted display, a camera, wireless connection, and computer in an eyeglasses frame. The omnipresence of a camera pointing at the vis-à-vis of the wearer sparked heated public discussions. As result, the project failed as a consumer device due to stigmatisation of the wearers.

In contrast, smartphones are widely accepted and have become our everyday companions with ownership rates of 72% in the US and a median of 68% in developed countries [70]. Packed with sensors and wireless connections, they are already carried by many, making them the perfect platform for daily monitoring. Another advantage is cost, as study participants can use their own smartphones, with an application built to record relevant data or for answering survey questions.

1.4 Context recognition

Context describes user activity (e.g. walking, standing), user state (e.g. heart and respriation rate), environmental state (e.g. light, temperature), and extended location information (e.g. in the office). The current situation of a user can be estimated using context recognition. Knowledge of a persons daily routine can be used to identify the influence of an activity on the circadian phase. Such knowledge could be used to give personalised feedback to the monitored person, e.g. avoid long working nights on the weekend.

To recognise daily routines and activities of daily living (ADLs), machine learning methods can be used. Head motion and orientation information be used to detect common ADLs, e.g. eating or reading. Detection of ADLs is a well studied problem. Seiter [82] used hierarchical topic models to infer ADLs and daily routines from activity primitives. Often, multiple sensors spread over the body were used for ADL recognition. Bao and Intille [12] used five accelerometers to detect a set of 20 ADLs.

In addition to typical ADLs, such as eating or walking, some activities are especially relevant for circadian research, e.g. computer screen use. Computer screen use at night is known to suppress melatonin production due to the strong blue spectral component of LED backlit computer screens. Knowledge of sleep/wake routines is key for circadian biology and is currently recorded with pen and paper diaries or inferred from wrist activity recorded by actimetry.

While pen and paper diaries are known to be inaccurate, actigraphy requires additional hardware. Using smartphones, which participants already carry with them, sleep/wake routines could be detected using in a ubiquitous and cost-efficient manner.

1.5 Expert models

In chronobiology, expert models are available, which were developed based on data from studies in controlled conditions. Such models are valuable as they provide a mathematical model for circadian processes. In contrast, machine learning methods typically attempt to extract a stochastic model, often resulting in volatile predictions. The Two Process Model (TPM) by Daan et al. [31] models homeostatic sleep pressure and circadian phase to determine the transition points between sleep and wake. The TPM was initially developed to investigate the influence of circadian amplitude and period on sleep/wake rhythms. Expert knowledge embedded into the TPM could be used to fuse and filter estimator outputs in order to improve machine learning estimations of sleep timing.

Expert models can further be used to compute performance estimates in the relevant domain. Light intensity estimates are key for estimating circadian phase. Circadian sensitivity to light exposure changes with circadian phase, thus light intensity estimates should not be evaluated in the light intensity domain. Instead, an evaluation in the time domain, i.e. the amount of circadian phase shift, is desirable. The Kronauer model [55, 61] can be used to compute the resulting circadian phase shift in the time domain from continuous light estimations.

1.6 Thesis objectives

1.6.1 Objectives

This thesis addresses three main objectives: First, we propose a personalised, 3D-printed, multi-modal wearable sensor system to measure light exposure and monitor patterns of typical activities of daily living in free-living conditions. Second, we use smartphones to estimate sleep/wake patterns and light exposure without requiring additional hardware. Third, we introduce novel estimation algorithms for sleep/wake pattern estimation and circadian phase shift estimation using machine learning methods in combination with expert models. In combination, the three objectives can be used to monitor human circadian rhythms in free-living conditions at greater detail and lower cost compared to the state-of-the art.

1.6.2 Motivation

Figure 1.1 shows state of the art methods and novel methods proposed in this thesis with respect to their system cost per participant (i.e. hardware and labour cost per participant) and detail level of the collected data. Sleep laboratories provide the most detailed insights (e.g. PSG recordings), but are cost-intensive starting at 600 USD per participant and night. The thesis proposes to use personalised, 3D-printed smart eyeglasses to collect data with an increased level of detail, compared to actimetry, at similar system cost per participant. Similarly, smartphones can be used to increase information detail over surveys when monitoring larger groups.

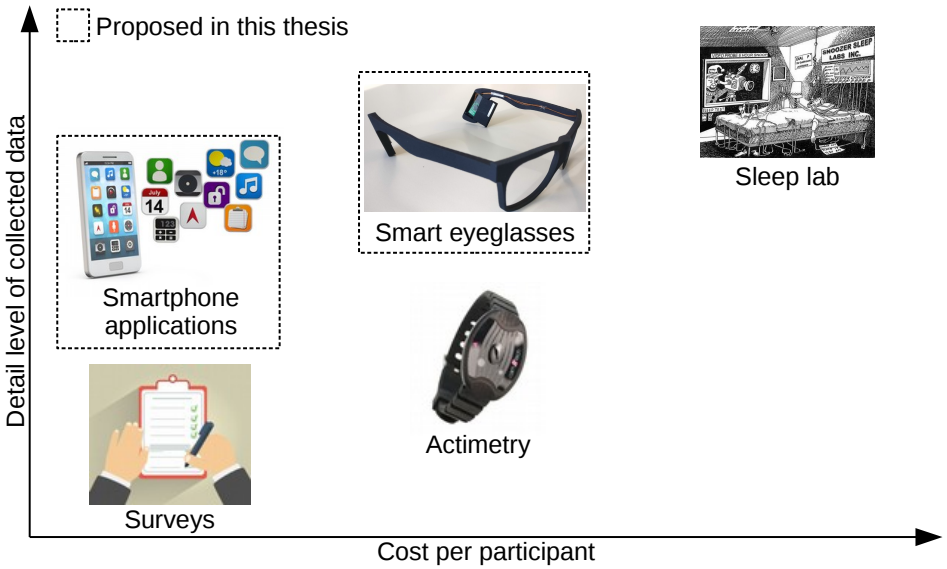


Figure 1.1: State of the art methods and novel methods proposed in this thesis with respect to their system cost per participant and detail level of the collected data. We hypothesise that smartphones and personalised wearables can be used to monitor human circadian rhythms in free-living environments.

1.6.3 Smart eyeglasses

Actimetry is a popular wearable monitoring tool in circadian rhythm research. Our proposed replacement are personalised, 3D-printed, multi-modal smart eyeglasses. While actimetry devices only record motion intensity and light exposure information, smart eyeglasses are equipped with more sensor modalities. In addition, wearers of eyeglasses are unlikely to forget their eyeglasses at home, as they require them to improve their vision. In comparison, wrist wearables are commonly neglected after some use [30].

1.6.4 Smartphones

We propose the use of smartphone applications instead of surveys. Smartphones are widely available [70] and equipped with a variety of sensors. In addition, they are connected to the internet, thus energy intensive data processing can be offloaded to the cloud, in order to preserve smartphone battery runtime. While surveys are often inaccurate and subject to self-reporting bias, smartphone sensors can be used in combination with digital surveys to adjust for such inaccuracies. Furthermore, smartphone sensors can monitor users continuously, thus providing continuous insight into user behaviour.

1.6.5 Novel algorithms

Our proposed methods for estimation of sleep/wake patterns and circadian phase shift extend machine learning methods by integrating domain expert models. Existing domain expert models were used as filter and fusion for machine learning estimator outputs, resulting in more accurate estimation of sleep/wake patterns and allowing us to estimate circadian phase shift from light exposure estimations.

1.7 Thesis outline and related publications

The thesis is organised in three parts as depicted in Figure 1.2: (1) Personalisation of wearables, (2) Context recognition, and (3) Algorithms for circadian monitoring.

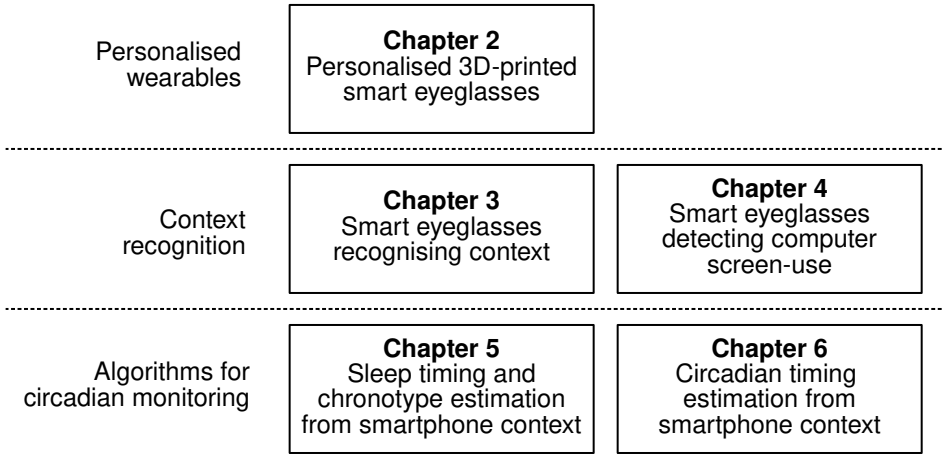


Figure 1.2: Outline of remaining chapters of this thesis in three groups: Personalised wearables, context recognition, and algorithms for circadian monitoring.

Chapter 2 introduces the personalisation of 3D-printed smart eyeglasses to record physiological parameters. Eyeglasses are personalised to the wearers head using few simple head measurements. We outline three applications for smart eyeglasses. Subsequently, data from smart eyeglasses is used in Chapter 3 to recognise groups of common daily activities. Daily routines could be reorganised such that they take place when the user is at their corresponding circadian peak, e.g. going to the gym at the physically strongest time of day. Smart eyeglasses are used to detect computer screen use in Chapter 4. Computer screen use at night is one of the key contributors to circadian misalignment. In Chapter 5 smartphone context information is used in combination with the Two Process Model [31] to estimate sleep timing. Sleep/wake routines are required to compute social jet lag and key in understanding an individuals circadian rhythm under free-living conditions. Smartphone context information and cloud-sourced weather reports were used in combination with the Kronauer model [55, 61] to estimate circadian phase shift in Chapter 6. Continuous phase shift estimations could reveal the circadian impact of performing an activity and thus support users in reducing their social jet lag. Conclusions over the whole thesis are presented in Chapter 7. Table 1.1 lists previously published chapters.

Table 1.1: Published chapters.

Chapter	Publication
2	F. Wahl, R. Zhang, M. Freund, and O. Amft. Personalizing 3D-printed smart eyeglasses to augment daily life. *IEEE Computer*, 50(2):26–35, 2017
3	F. Wahl, M. Freund, and O. Amft. WISEglass: Smart eyeglasses recognising context. *EAI Endorsed Transactions on Pervasive Health and Technology*, 16(5), 2016
4	F. Wahl, J. Kasbauer, and O. Amft. Computer Screen Use Detection Using Smart Eyeglasses. *Front. ICT*, 4, 2017
5	F. Wahl and O. Amft. Data and Expert Models for Sleep Timing and Chronotype Estimation from Smartphone Context Data and Simulations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3):139:1–139:28, Sept. 2018

1.8 Additional publications

- F. Wahl, M. Freund, and O. Amft. Using smart eyeglasses as a wearable game controller. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 377–380. ACM, 2015

- F. Wahl, M. Freund, and O. Amft. WISEglass: Multi-purpose context-aware smart eyeglasses. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 159–160. ACM, 2015

- O. Amft, F. Wahl, S. Ishimaru, and K. Kunze. Making Regular Eyeglasses Smart. *Pervasive Computing, IEEE*, 14(3):32–43, 2015

- F. Wahl, T. Kantermann, and O. Amft. How much light do you get? Estimating daily light exposure using smartphones. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, pages 43–46. ACM, 2014

- F. Wahl and O. Amft. Personalised phone placement recognition in daily life using RFID tagging. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference On*, pages 19–26. IEEE, 2014

- F. Wahl and O. Amft. Using RFID tags as reference for phone location and orientation in daily life. In *Proceedings of the 4th Augmented Human International Conference*, pages 194–197. ACM, 2013

2

Personalised 3D-printed smart eyeglasses

Chapter originally published as: F. Wahl, R. Zhang, M. Freund, and O. Amft. Personalizing 3D-printed smart eyeglasses to augment daily life. *IEEE Computer*, 50(2):26–35, 2017.

2.1 Introduction

Eyeglasses are widely accepted as wearable accessories to augment sight. We believe that embedded technology will allow us to use eyeglasses for many more things: Since eyeglasses sit at a key site for continuous monitoring of physical, physiological, and environmental parameters, they may become the primary choice of health monitoring and assistance tool for many. Future smart eyeglasses could support applications from tracking daily activity of elderly users, alerting computer workers due to extended screen use, or assisting patients in diet interventions, to name a few. All of the functions may be used via software apps on the same smart eyeglasses. In contrast to smartphones however, the smart eyeglasses adpotion will be driven by a rapid digital development process. Essential design considerations for any wearable accessories are wearer comfort, everyday usability, and fashionable perception, where needs and preferences vary on a personal level. For example, eyeglasses wearing comfort is related to the frame that remains in constant skin contact. Skin irritations may occur if the fitting is too tight, or the frame may loosen from the head if the fitting is too wide. Thus, form and shape must be adjusted to individual needs and several anthropometric parameters of the head and face affect eyeglasses fitting. For classic spectacles, opticians perform 'personalisation' by bending and tweaking various frame components by heating plastic frame parts. For smart eyeglasses containing electronics, mechanical deformation may not be a viable option. Instead, smart eyeglasses should be personalised during the manufacturing process already. Device personalisation does however not only relate to ergonomics and perception. Fitting eyeglasses to individual head anatomy can further contribute to maintain skin contact at specific anatomical landmarks that are essential for recording physiological parameters, such as heart rate.

Novel 3D-printing techniques and materials could provide the basis for personalised wearable accessories without the initial production setup costs such as when creating moulds. With printing, the development and production process of devices becomes mostly digital: 3D volume models are developed using CAD software and adapted for a particular printing technique by subsequent software filters, including thermoplastic deposition and selective laser sintering. So far, digital development processes to design and manufacture personalised wearable accessories are lacking. The human head and face are especially complex surfaces, reflected in a wide set of anthropometric paramters and thus results in a challenging modelling problem.

Here we show how novel 3D-printing techonology can be used to personalise regular-looking smart eyeglasses to match relevant anthropometric data of the head by elaborating a CAD-based digital development process. We approach eyeglasses personalisation from anatomically correct human models of face and head and fit a smart eyeglasses frame model with few free model parameters only. Our smart eyeglasses are hybrid systems, where the frame and some functionality is printed and some is embedded using integrated electronics. We show examples of smart eyeglasses being used in monitoring tasks related to physical activity, circadian phase alignment, and dietary monitoring. The examples reflect the application space in augmenting human health using sensor data acquired at the head.

2.2 Trends in smart eyeglasses

Eyeglasses are established, as a vision aid and fashion accessory. Adding sensors and computing expands the eyeglasses' augmentation features, without critically altering device appear-

ance and handling. Functional versatility is vital for wearable systems to address individual wearer expectations. Smart eyeglasses can serve various functions while the personalised fit and design makes them convenient to wear.

Considering the head as monitoring location, Amft et al. [5] presented an overview on the technical and application opportunities for smart eyeglasses. Ishimaru et al. [52] used J!NS MEME, which is an electrooculography eyeglasses prototype, to perform activity recognition based on eye movement. Wahl et al. [97] built a smart eyeglasses prototype onto a regular eyeglasses frame, and evaluated inertial and other sensors for recognising activity types. Zhang et al. [107] investigated smart eyeglasses for dietary monitoring using Electromyography (EMG) electrodes. The variety of functions confirms that smart eyeglasses could become a central monitoring component in human augmentation. However, acquiring reliable sensor data requires accurate fitting to the head. The personalisation process investigated in this work is essential to realise smart eyeglasses.

Researchers demonstrated further applications of head-worn wearable devices, e.g. using the eyeglasses add-on Google Glass. Rahman et al. [72] used Google Glass to detect eating episodes in daily life and Hernandez et al. [46] derived heart rate from the Google Glass gyroscope while wearers remained still. Unobtrusive design however determines successful adoption: Koelle et al. [58] found that bystanders felt less comfortable around people wearing head-mounted displays because the wearers intent is not clear. The integrated camera in Google Glass started many privacy discussions and slowed adoption rates. Smart eyeglasses require careful choices of design, sensors, and interaction, to remain unobtrusive and clarify the wearer-focused monitoring intent.

2.3 Designing personalised smart eyeglasses

We analysed head and face characteristics that influence the eyeglasses shape and derived a set of free frame parameters. We consider a regular eyeglasses design that was modified to embed sensors and processing electronics in the frame and temple ends. The free frame parameters are subsequently fitted to head configurations.

2.3.1 Head modelling

Natural head and face shapes vary widely. For a viable personalisation process, we identified key head characteristics affecting frame design by analysing parameters of the open-source human modelling software MakeHuman. Out of 146 available parameters to define head and face shape, we derived 26 that affect frame fitting. We reduced head model configuration to three parameters by removing parameters having a redundant effect on frame fitting.

2.3.2 Parametric frame model

An example CAD frame model was constructed that closely resembles regular eyeglasses. We aimed at creating a parametric frame model with a minimal set of measurements that could be taken from the head. Initially, the 26 head and face shape parameters were mapped to six free parameters controlling the frame model. In further refinements, five free frame parameters were kept and nose pad symmetry was omitted due to its low practical relevance. Figure 2.1c lists head and face shape parameters and the final free frame parameters used. Frame temples were modelled by parameters for length, angle, and bend to ensure appropriate fit. The

eyeglasses front has parameters for adjusting to nose width and height, lens height, and lens width. Lens and nose width determine the overall width of the eyeglasses. Lens height and nose height control the vertical eyeglasses position. Lens height is critical to prevent the frame from resting on the wearers cheekbone.

2.3.3 Frame fitting simulation

To verify the five CAD frame model parameters are sufficient to fit a wide range of head shapes, we used anthropometric head and face shape data from the FAA Human Factors design guide [37]. From the design guide, we derived measurements to generate nine head model configurations using MakeHuman. We selected the head length and width to match the 1, 50, and 99 percentile human.

After importing a generated head model into the CAD software, an unfitted eyeglasses model was aligned to a fixed point on the head model's noses. Subsequently, head measurements were taken to fit the frame starting with the temples length. Next, head width was measured to adjust lens and nose width. Nose height offset was adjusted to ensure ergonomic viewing. Finally, the angle of the temples was used to ensure that temple ends fit by touching the head behind ears.

We performed an interference analysis to ensure that no overlaps occur between the models of fitted eyeglasses and head. Figure 2.1b shows an example head model configuration with the fitted frame and measurements for the head model.

2.3.4 Smart eyeglasses hardware

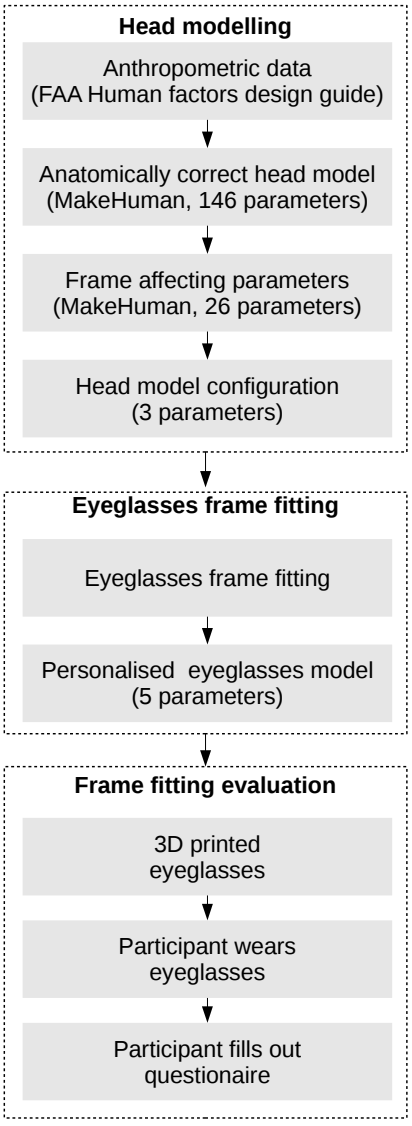
The final personalised smart eyeglasses are based on a 3D-printed frame and integrated electronics. The main difference to regular eyeglasses are pockets integrated into temple ends, housing battery and main processing unit with MCU, flash memory, and wireless unit. To balance component weight, the components are distributed onto both temple ends. The pockets sit behind the wearers' ears when the eyeglasses are worn. Wiring inside the frame interconnects the different components.

2.3.5 Physical frame fitting evaluation

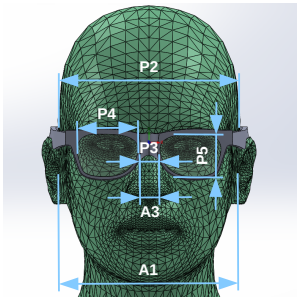
In addition to simulations, we fitted the eyeglasses CAD model to four users (2f/2m) using the same personalisation procedure. Participants were asked to wear the eyeglasses for one day and rate wearing comfort. All participants were regular eyeglasses wearers and switched to contact lenses for the study duration. After receiving detailed information about the evaluation process, participants gave written consent to the use of their data and images.

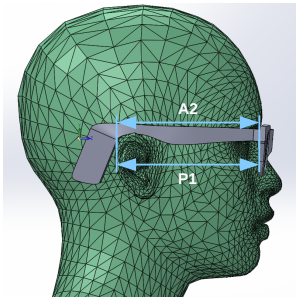
We measured temple length, ear to nose distance, and nose width. Subsequently, we adjusted the CAD frame model parameters to personalise the frame to each participant's head. We printed the personalised eyeglasses on a 3D printer (Witbox with diamond hotend) to evaluate frame fitting. As the focus was to evaluate frame personalisation, we did not integrate lenses nor electronic components. The frames were printed as a single piece. After removing printer support material, frames were sanded. Due to the printing process, frame weight was 28 g, similar to actual eyeglasses. Figure 2.1d shows a participant wearing a 3D-printed, personalised eyeglasses frame.

After the evaluation day, participants completed a questionnaire on fit and wear comfort with six items, rated on a 5 point Likert scale. For comparison, participants completed the



a)





b)

	Head and face anthropometrics			Smart eyeglasses model parameters				
	A1 Head width [cm]	A2 Nose-ear distance [cm]	A3 Nose width [cm]	P1 Temple length [cm]	P2 Temple ends width [cm]	P3 Nose width [cm]	P4 Lens width [cm]	P5 Lens height [cm]
Model 1	12.5	18.8	1.3	8.7	11.5	1.5	4.3	3.5
Model 2	12.5	10.3	1.6	10.1	12.1	1.6	3.8	3.0
Model 3	12.5	11.5	1.9	11.5	12.2	1.9	3.8	3.0
Model 4	14.2	9.2	1.5	9.1	13.4	1.5	5.0	3.8
Model 5	14.2	10.6	1.9	10.6	13.4	1.6	4.7	3.3
Model 6	14.2	11.8	2.1	11.7	14.1	1.8	4.6	3.0
Model 7	15.9	9.2	1.7	9.1	15.1	1.6	5.4	3.4
Model 8	15.9	10.6	2.1	10.5	15.3	1.8	5.2	3.2
Model 9	15.9	11.8	2.4	11.8	15.7	1.8	5.1	3.0
Participant 1	15.0	11.0	1.4	10.9	14.3	1.4	5.9	3.1
Participant 2	13.1	10.6	1.8	10.5	10.8	1.8	5.3	3.8
Participant 3	14.5	11.0	1.1	10.9	13.5	1.4	5.2	3.3
Participant 4	14.1	11.1	1.6	11.0	12.2	1.6	4.2	3.0

c)

d)

Figure 2.1: (a) Digital development process to personalise smart eyelgasses frames. (b) Example head model configuration according to data from the FAA Human Factors design guide and fitted eyeglasses frame. Anthropometric and free frame parameters are indicated. (c) Parameters of head model configurations, participants, and fitted CAD frame model parameters. (d) Study participant wearing fitted, 3D-printed eyeglasses.

same questionnaire to rate their usual eyeglasses. Participants rated the printed eyeglasses frame comfortable to wear overall (4.25 of 5). The fit at nose, ears, and frame width were rated comfortable. Participants rated the design slightly futuristic (3.5 of 4). When compared to their own eyeglasses frames, participants found similar wear comfort in the printed eyeglasses. All result data are available in the electronic appendix.

2.4 Applications for smart eyeglasses

We describe three application clusters out of the variety of possible activity and context recognition functions.

2.4.1 General health and fitness

Wrist-worn wearable devices to track general health and fitness have become popular along with the quantified self movement, e.g. the Fitbit activity tracker, but suffer from low long-term compliance [30]. Smart eyeglasses could replace regular eyeglasses and activity trackers for many wearers.

Inertial sensors, including accelerometers and gyroscopes, can help to recognise various activities of daily living (ADLs) in smart eyeglasses. The inertial sensors measure head motion and posture that reflects different repetitive movements, such as walking, running, and cycling, as well as static activities, such as reading. Inertial sensors do not depend on a particular position at the head, thus can be placed anywhere in the eyeglasses frame. Moreover, heart activity can be monitored using a pulse oximeter (PPG) sensor at the eyeglasses temples. While smart eyeglasses acquire physical and cardiac activity, feedback of aggregated behaviour patterns and trends could be provided daily or weekly through a smartphone or website. Eyeglasses frame personalisation helps to maintain skin contact of the pulse oximeter and recognise diverse activities, thus maximising coverage throught the day.

A smart eyeglasses prototype with inertial sensors (Invensense, MPU-9250) placed at the eyeglasses temple ends and pulse oximeter at temples are shown in Figure 2.2a. Figure 2.2c shows example data from the sensors. Different movement and posture patterns depending on the activity are clearly distinguishable.

In our earlier work, we investigated the recognition of ADLs using smart eyeglasses including nine activity clusters, such as walking, eating, and reading [98]. We recorded data from nine participants, who either did not require prescription glasses or wore contact lenses during recordings. An observer labelled the recordings using a smartphone [85], while participants followed a scripted protocol to maximise the ADLs performed. In a total of 66 hours of data, 25 time-domain features of acceleration and gyroscope axes were derived using a 30 s sliding window with 1 s step size. Principal Component Analysis (PCA) was applied to reduce features from 175 to 78 but still explain 99.9% of the variance. Gaussian mixture models (GMMs) were used to classify activity clusters using Leave-One-Participant-Out (LOPO) cross-validation. LOPO ensured that the classifier was tested independently of its training data. Using a GMM with three Gaussian components and diagonal covariance matrix, all activity clusters, except cycling, were detected with an accuracy of at least 80%. Using a class-reject design, overall classifier accuracy was 77% on average, varying from 70% to 84% per participant.

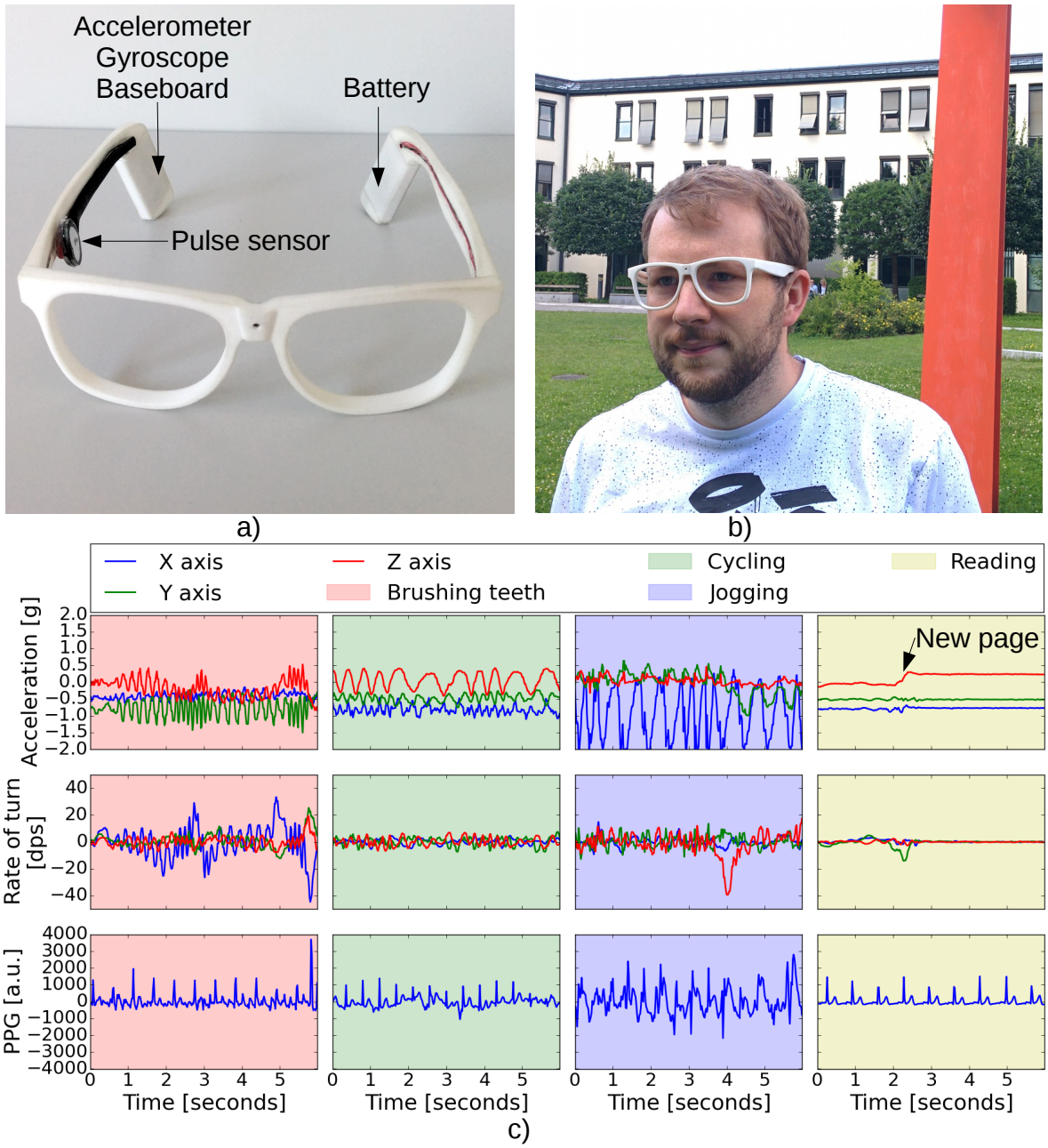


Figure 2.2: Activity recognition results. (a) Smart eyeglasses components used for detection of ADLs. (b) Person wearing smart eyeglasses during ADLs. (c) Example accelerometer, gyroscope, and PPG sensor data during typical ADLs. The data indicate, for example, the head movement when begin reading a new page, as well as repetitive patterns of different frequencies during brushing teeth, jogging, and cycling.

2.4.2 Circadian phase and screen use

The human body has an internal clock, the circadian phase, controlled by intensity and timing of light received though the eyes. Light in the morning advances the circadian phase while light in the evening delays it. Shifts of -2 to +3 hours per day are possible [74, 35]. Misalignment of circadian phase and external time can lead to impaired performance, alertness, and upset gastrointestinal functions. In today's built environments, staying up until late is common, leaving insufficient rest time before the next daylight phase. Circadian misalignment and regular sleep deprivation are thus a known issue for a large part of the population. In addition, the relatively high energy in blue light spectrum emitted by popular LED backlight screens affects the circadian phase. Blue light acts as inhibitor of the sleep-stimulating hormone Melatonin [20]. Extended screen use during evening and night may thus lead to circadian phase delays and reduced sleep quality. Smart eyeglasses could monitor light exposure and blue light content continuously, in close proximity and viewing direction of the eyes.

Our smart eyeglasses feature a colour light sensor (ams, TCS34725), which can measure light exposure throughout the day. The sensor is built into the eyeglasses bridge to align it closely with the eyes' viewing direction. Light exposure profiles and detected screen use could be used to estimate circadian phase shifts. Based on accumulated light and its timing across a day, general behavioural recommendations could be provided as well as specific advice on a day, based on current and expected remaining light exposure. Real-time information, in particular during evenings and nights, could support wearers in adjusting their behaviour and screen use to achieve or maintain a certain circadian phase. For example, the accumulated light exposure information combined with the blue light detected in nightly screen use could be used to recommend blue light software filters, or enable such filters for the computer screen automatically. Eyeglasses personalisation helps here to maximise wearing comfort in daily use.

We analysed screen use detection with smart eyeglasses and data from the colour sensor. In particular, ratios between spectral band regions contain relevant information, e.g. the ratio of blue light vs. overall light intensity. Figure 2.3c shows an example of the light intensities for red, green, and blue colour channels during screen use and no screen use activities.

We performed a study with 14 participants (2 female, 12 male). Each participant was asked to read a print article and browse the Web on a HP EliteDisplay E241i 24 inch screen at 70 cm distance for 20 minutes per activity. Participants wore the smart eyeglasses prototype to record red, green, blue, and clear colour channels with a sampling rate of 6.5 Hz. In addition, ambient light intensity was recorded at the beginning of each recording using a standard lux meter (AMPROBE, LM-120). In a total of 15.2 hours of data, we derived 18 features of time-domain light channel ratios using a sliding window of 5 seconds and 50% window overlap. With a linear support vector machine, screen use episodes were detected using LOPO cross-validation. The ratios of the light intensities between colour channels clearly distinguish screen use from other activities. Detection accuracies for screen use averaged at 90%. When the ambient light intensity was 500 lux or less, as in a typical evening scenario, average accuracy peaked at 95%. For ambient light intensity above 500 lux, which is the recommended indoor light intensity in office spaces [32], screen use detection accuracy still reached 85%.

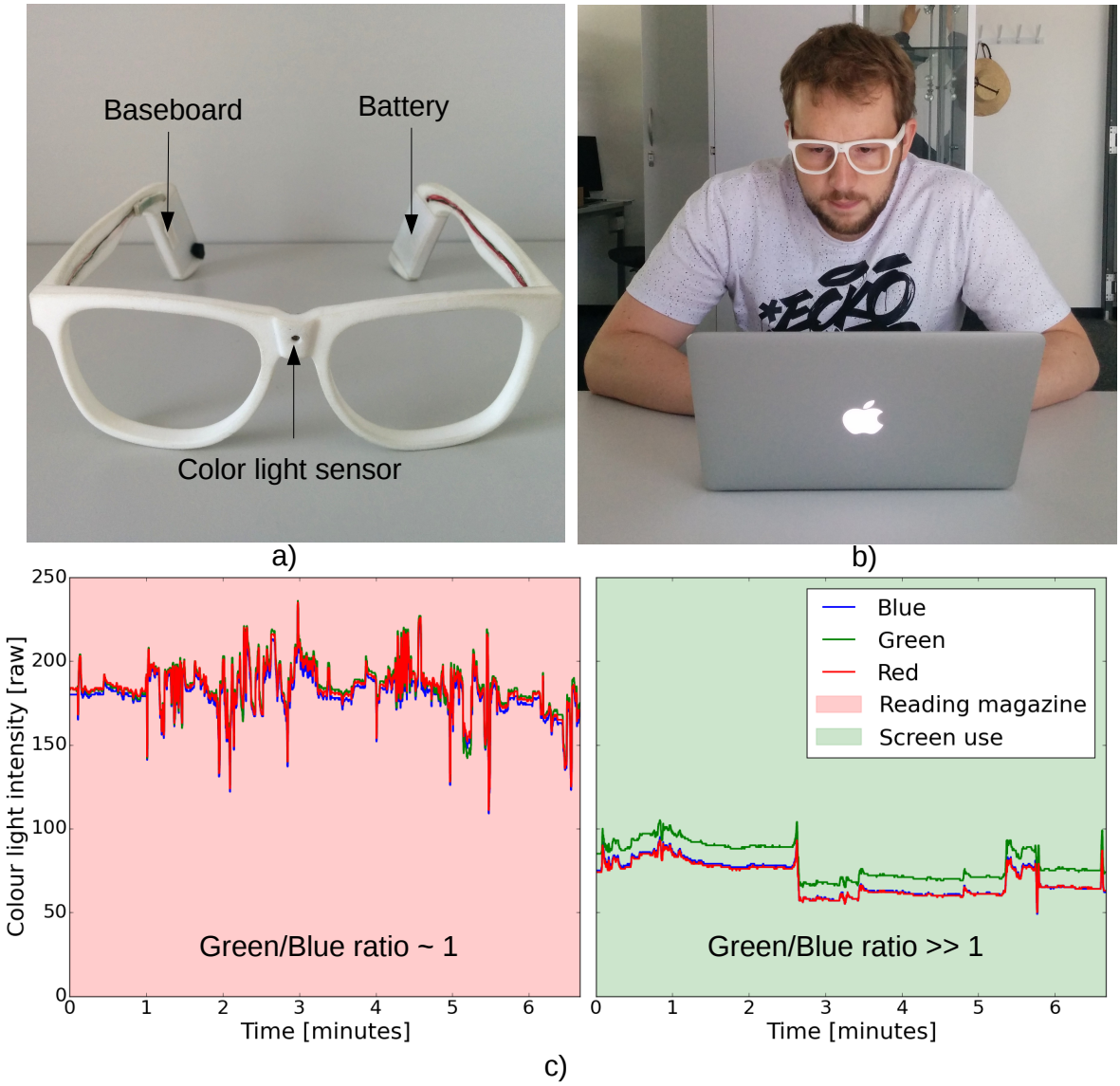


Figure 2.3: Screen use detection. (a) Smart eyeglasses components used for screen use detection. (b) User wearing smart eyeglasses while working at screen. (c) Example light intensity data for red, green, and blue colour channels during screen use and reading a print magazine. Ratios of colour channels clearly differ between activities.

2.4.3 Nutrition

Food intake is a regular routine where food choices and many activities happen unconsciously. Monitoring nutrition can thus help patients suffering from the various diet-related diseases to maintain a healthy diet. Besides the wide-spread cardiovascular diseases and obesity, dietary coaching based on actual intake patterns is of interest for individuals concerned about personal health. Conventional nutrition monitoring requires users to manually log their dietary routines by filling questionnaires on paper, PC, or smartphone. However, the tedious logging is known to result in low adherence of recording every intake detail, rendering self-reporting highly inaccurate. Besides forgetting to log details, users could become subjective to filter their reports, e.g. omitting the high-calorie evening snack. Since a healthy diet may involve lifestyle changes, it frequently requires months or years of coaching and thus a technology that can accompany users without unnecessarily inferring in everyday activities or embarrassing wearers of a monitoring device in the public. Smart eyeglasses may fill in this dietary monitoring challenge and continuously provide relevant food consumption details from different sensors integrated into the eyeglasses frame.

For a long time, facial surface Electromyography (EMG) has been studied in labs and shown to be effective in monitoring chewing. For wearable nutrition monitoring, surface EMG has been disregarded so far due to the prominent placement of electrodes on the head. Eyeglasses however offer a way to integrate EMG-based chewing monitoring in a wearable accessory as the eyeglasses frame is in skin contact at several key landmarks [107]. The bilateral temporalis muscles are of primary interest for chewing monitoring, as they span over a large skull area from temples to ears and are invoked in elevating the jawbone for each chewing cycle. Most of the muscle area is covered by hair, which is inconvenient for surface EMG measurements. Nevertheless, there are small spots around the ears that offer sufficient EMG signal quality. Besides muscle contractions, another relevant information source is the skull vibration generated during chewing. Breaking food pieces into particles generates mechanical vibrations that spread from teeth throughout the skull [4]. The vibrations can be detected at the mastoid bone and other skull regions, such as behind the ears. For nutrition monitoring, eyeglasses personalisation is key to match EMG and vibration sensors to recording landmarks and maintain skin contact. Recorded data is processed at the smart eyeglasses and can either serve for instant interactions with the wearer, e.g. to ask for confirmation of the detected food consumption, or serve as daily or weekly feedback on dietary patterns via a smartphone or website.

We investigated different head locations along the eyeglasses frame to find optimal EMG electrode locations [107]. The temporalis muscle regions just above the ear showed consistently highest signal-to-noise ratios for differential EMG measurements. We used textile electrodes that were convenient to embed in eyeglasses temples at the ear bends. A reference electrode was integrated at the eyeglasses nose pad. While EMG data provide a segmentation of chewing cycles, vibration measurements serve to discriminate different food categories from their material textures.

The eyeglasses prototype integration and example signals from EMG and vibration during chewing and other activities are shown in Figure 2.4. We conducted a study with eight participants (4 female, 4 male, 20 to 56 years old) and recorded chewing of five different foods types, representing different textures: carrot, toast, jelly baby, banana, and biscuits. To analyse a realistic condition for detecting chewing cycles, other activities were performed by the participants too, including speaking, coughing, and head moving. Chewing was performed

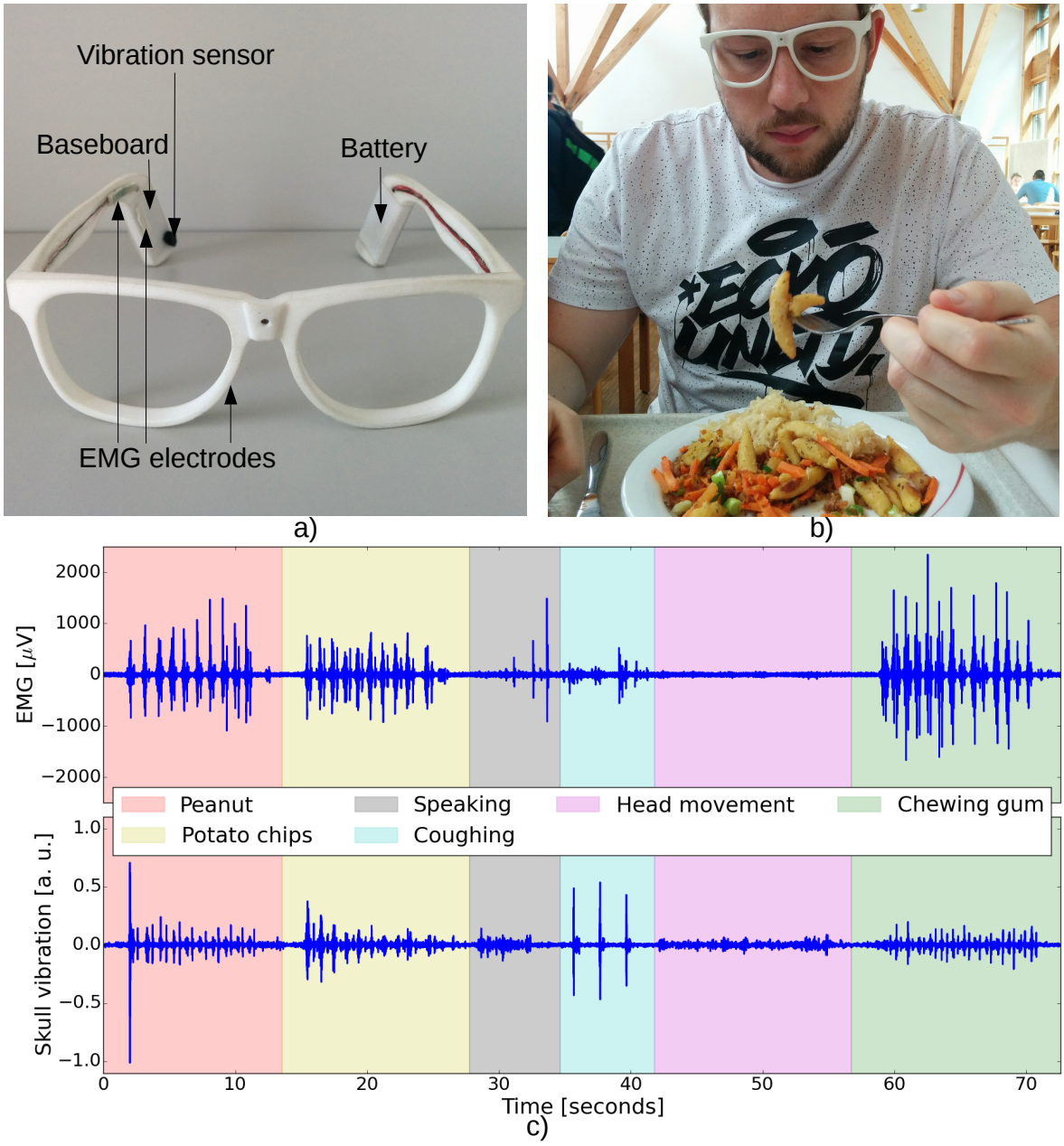


Figure 2.4: Nutrition monitoring. (a) Smart eyeglasses components used for monitoring chewing. (b) User wearing smart eyeglasses in daily life. (c) Example surface EMG and skull vibration data during a series of chewing/non-chewing activities. Signal patterns differ between different activities.

on alternating jaw sides and with different speeds. With ∼38.5 min. of data per participant and totally 5435 chews, chewing cycle detection was performed. Here we employed the mean rectified and filtered EMG signal value within a sliding window of 200 ms, $\bar{x}$, and compared it to a threshold, $\theta = \mu + n \cdot \sigma$, where n is an adjustable scalar, and (μ, σ) are a Gaussian model of the baseline noise. When $\bar{x} > \theta$, the window was regarded as part of a chewing cycle, and a series of continuous windows as a chewing cycle candidate. By majority voting on the windows and retaining chewing cycle candidates within a typical duration of ∼400 ms to 1 s, the finally detected chewing cycles were determined. While the detection could certainly be refined, we already obtained a precision and recall of 80%, confirming that the EMG activation pattern during chewing are a robust chewing cycle indicator.

2.5 Lessons learnt

Our head modelling step used only three input parameters to minimise the measurement effort when creating personalised eyeglasses. Alternative approaches, such as light-based shape scanners could be used, however measurements will be affected by occlusion due to ears and hair.

The frame did not include wearer interaction modalities to emphasise unobtrusive design. We consider that e.g. smartphones pair with smart eyeglasses to support interaction and to offload processing.

While technically all of the described functions could be integrated into one smart eyeglasses model, we built separate prototypes so far. However, there may be applications requiring dedicated eyeglasses, e.g. due to sensor placement. Moreover, the frame design influences how functional components can be embedded and how frame model parameters may be chosen. We expect that for further applications and eyeglasses frame designs the digital development process could be repeated. Although smart eyeglasses can offer exceptional functions, they may suit everyone. The digital development and physical personalisation offer directions for personalising wearable accessories in the future.

2.6 Towards personalised wearable accessories

A new generation of personalised wearable accessories is obtained by customising the device design directly during digital development. Personalising wearable accessories could address a multitude of open challenges in wearable computing: First, personal items may be used more frequently as they are designed and manufactured considering individual preferences regarding fashion and usability. Second is comfort. Accessories that are fitted to individual anatomical requirements align with the body. They could be used continuously and potentially elevate monitoring compliance. Third, with continuously worn smart eyeglasses, multi-day trends in sensor data can be analysed, e.g. to understand behaviour change, rather than relying on a single sample that could be affected by random noise. Fourth, to obtain 'on average' correct samples while being worn in daily life, personalisation to individual anatomy will be essential. Matching anatomical landmarks provides the best possible sensor data that could be obtained with wearable devices, in some situations potentially suitable at the level of medical-grade diagnosis and treatment.

2.7 Future of smart eyeglasses

Smart eyeglasses are a particular example of personalised wearable accessories as eyeglasses are prominently located at the head. With proper personalisation and fashionable design, smart eyeglasses may replace their classic counterparts and the additional functions just makes the glasses more interesting to wear. Eyeglasses choice is particularly related to design preference. Printing offers abundant options for shape and colour choice. Among current top 20 eyeglasses picks [1], 18 models would support integrated electronics and frame personalisation. As a hybrid system, personalisation becomes an essential new feature of eyeglasses that provides the basis for augmentation. For regular wearers, eyeglasses are hardly left off the head, potentially providing many augmentation options based on continuous data. The share of eyeglasses wearers is large: In the US, 64% of the population wears eyeglasses [88]. Even beyond the regular wearers, sunglasses, sports glasses, and safety glasses may feature sensors for human augmentation. Current eyeglasses with music players and for relaxation are first indicators, see box on smart eyeglasses trends.

As diverse as the applications of wearable accessories are their designs and functionality. While each such application will find its audience, their economic impact may remain too small to underwrite the initial costs of classic production processes. Printing may fill in this spot. For example, a commercial printing service charges about 25 Euros for printing a pair of eyeglasses from a CAD model. Printing enables developers to quickly address niche applications. However, integrating electronic and printed components in mass production is an open challenge. As printing and material development progresses, additional functionality may be printed directly, including sensors, wiring, and basic electric components. Integrated electronics are however unlikely to be fully replaced. As electronic components, including sensors and processing units, can be embedded across different smart eyeglasses models and other wearable accessories, they remain volume-produced and thus available at low cost. We believe that hybrid systems of electronic and printed components will open an entirely new, and serious wave of wearable systems.

Acknowledgements

This research was supported by the Dutch Technology Foundation STW under grant number 12184.

3

Smart eyeglasses recognising context

Chapter originally published as: F. Wahl, M. Freund, and O. Amft. WISEglass: Smart eyeglasses recognising context. *EAI Endorsed Transactions on Pervasive Health and Technology*, 16(5), 2016.

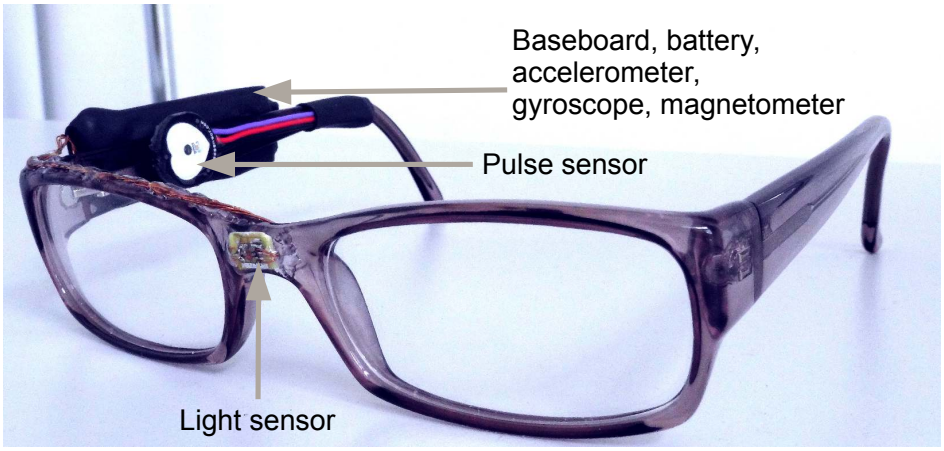


Figure 3.1: WISEglass prototype. Baseboard, battery, accelerometer, gyroscope, and magnetometer were mounted on the outside of a eyeglasses temple. The pulse sensor was mounted on the inside of a temple. The light sensor was integrated onto the bridge.

3.1 Introduction

Context awareness has opened a vast spectrum of applications that benefit from momentary information on user activity, environment, physiology, and similar. Wearable devices are often key to provide context information, as sensors could be placed comfortably at body locations such as wrist, leg, chest, or the ear. While hearing aids and Bluetooth headsets got broadly accepted as daily accessories, head-worn wearables were rarely proposed for context awareness due to potential obtrusiveness of head-attached devices. Some niche applications include ear-worn computers to manage information [86], or the ear-worn e-AR device to monitor activity and physiology [54]. There is nevertheless substantial context information available around the head [5]. Eyeglasses are regularly worn accessories that have a unique opportunity to carry sensors and process data at the head, thus fill a gap for many context-awareness applications. Moreover, eyeglasses are being worn by millions of people, for improved sight, but also as sunglasses, sports glasses, etc.

Substantial work has been dedicated to developing smart eyewear already. However, functionality was often centred around displaying information in front of the wearers eyes, e.g. for augmented or virtual reality applications [9, 109]. GoogleGlass and others established microinteractions as key feature of smart eyewear and focused on interaction and displaying information using glass-attachable electronics. We believe that instead of displays and direct interaction, smart eyeglasses can be built for context-aware applications, where the focus is on sensing and processing, rather than interaction. Smart eyeglasses would hence focus on acquiring and processing context information relevant to the wearer, but minimise the risk of stigmatising or raising privacy concerns.

Besides the integration of sensing and processing functions, the benefit of smart eyeglasses for different context recognition applications needs confirmation. Previous research on head-attached devices and smart eyeglasses focused on detecting selected activities, such as walking and reading behaviour using inertial sensors and the Electrooculogram (EOG) [54, 52]. A

set of daily living activities was not investigated (see related work for details). Moreover, eyeglasses could be used to acquire environmental and physiological information continuously, if a suitable smart eyeglasses implementation is found.

In this paper, we propose smart eyeglasses as a platform for a variety of context-aware applications. Our aim was to confirm that smart eyeglasses could be built and used instead of various sensors placed at different body positions. We implemented WISEglass with various sensors, including inertial motion, environmental light, and optical heart rate and investigated three application scenarios: daily activity recognition, screen-use detection, and heart rate estimation. This paper provides the following contributions:

1. We present our WISEglass eyeglasses architecture, integrating multi-modal sensing functions corresponding to frequent context types, including user activity, environmental state, and user physiology. We detail the embedded design in regular eyeglasses and data processing.

2. We evaluate WISEglass prototypes in a daily life study with nine individuals that did not wear eyeglasses regularly, including a programme of 20 daily living activities. Based on the study data, we confirm that the sensor placement and wearing across a day is suitable to implement the context recognition tasks for each application scenario.

Towards smart, regular eyeglasses there are many challenges to resolve. Our aim in this work is to establish that context recognition - and thus assistance to wearers - benefits from using smart eyeglasses due to their position at the head. Earlier work on smart eyewear primarily considered interaction and conveying visual information to the wearer. Devices often clipped onto eyeglasses only. Our approach is different, as we focus on single-point, multi-modal sensing and processing, integrated into a typical accessory, and serving different applications. We expect that in a subsequent step, continuous miniaturisation will allow us to even further integrate functionality into unobtrusive eyeglasses.

3.2 Related work

Head mounted sensors were considered for context awareness before. Aziz et al. [7] first used e-AR, an ear-worn sensor, for monitoring patients after abdominal surgery. In a study with 20 users they investigated motion patterns in data obtained from two dual-axis accelerometers. They used a pulse oximeter clipped to the users earlobe to monitor heart rate and oxygen saturation. No performance analysis was provided. Atallah et al. [6] compared seven different on-body motion sensor locations while grouping the activities by physical intensity level. At the head, they used e-AR and found that the ear location delivers good results for 4 out of 5 activity levels. In contrast to e-AR, our sensors are worn as eyeglasses. WISEglass features a 3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer, a RGB light sensor, and a pulse sensor, used to detect a wide range of activities of daily living.

Ishimaru et al. [51] used built-in proximity sensor and accelerometer of Google Glass to classify 5 different activities in a laboratory setting. For user dependent models they achieved an average accuracy of 67 % using blink frequency alone, and 82 % in combination with the accelerometer. Ishimaru et al. [52] used EOG and acceleration signals to classify reading, typing, eating, and talking activities, recorded from two users. They achieved an average of 70 % accuracy using user independent models. In our work, we investigate a full set of

20 activities of daily living grouped into 9 clusters in a field study rather than a laboratory setting. Moreover, our approach yielded comparable performance using user-independent models. Hernandez et. al [46] used the accelerometer, gyroscope, and camera to obtain heart and respiration rate from Google Glass, while participants were asked to remain still. They also attempted to detect emotion of the wearer in combination with an arm-worn sensor [47]. In contrast to Google Glass, which was focussed on micro-interactions and displaying information, we aim at integrating sensing and processing capabilities into regular eyeglasses. We avoid the use of a camera in order to preserve privacy. Furthermore, we gathered measurements in the field instead of a controlled a environment, e.g. heart rate is estimated during different activities.

Previous works often used multiple on-body sensing locations. Bao and Intille [12] used bi-axial accelerometers in five different body locations. They detected a set of 20 different daily living activities and achieved accuracies of up to 84 % in 20 participants. Measuring in multiple sensing locations complicates the setup process. In contrast, we propose a sensor system integrated into an accessory, thus minimising the required setup to putting on a pair of eyeglasses.

Exposure to light entrains our circadian clock, but requires light measurements close to the eye. Figueiro et al.[40] compared three different wearable sensor systems for light exposure measurement. Besides a wrist-worn device and a button worn at the collar, they found that a head-worn solution performs best as the sensor is close to the eyes. For WISEglass, we integrated a light sensor into the bridge of the glasses, thus optimising placement for light exposure measurement. As blue light is dominant in the entrainment of the circadian clock, it is of interest to know the time and duration of screen use that emits elevated blue light levels.

In the past, a variety of sensors have been placed around the head for different applications. WISEglass integrates existing sensors into *one* wearable accessory that allows wearers to benefit from smart eyeglasses in a wide range of applications, which previously required separate sensing systems.

3.3 WISEglass architecture

WISEglass is an approach to retrofit regular eyeglasses with a multi-modal sensor system for different context-aware applications. WISEglass could be used as a everyday accessory, just as regular eyeglasses are used today, or as a special-application device. For our design investigations, we specifically chose standard of-the-shelf eyeglasses and avoided designing the device from scratch, as to ensure that the typical eyeglasses form factor was maintained.

Processor, flash memory, communications interfaces, power controller, as well as inertial sensors were integrated onto a baseboard, the central unit of WISEglass. In our first prototypes, battery and baseboard were bound together using shrink tubing, however we consider that the units could be embedded in opposite ends of the eyeglasses frame in future versions.

With the light sensor, we investigated the embedding of components into the glasses' bridge. The light sensor magnet wiring was routed through miniature holes in the bridge into a milled channel along the top rim over the end pieces ending at the baseboard. We picked the bridge location as it is closest to the eyes and allows us to obtain most accurate light exposure measurements. Environmental light entrains our circadian clock and measuring light exposure is relevant for circadian phase guidance [100]. Another use for light sensor data could be in indoor/outdoor detection.

The pulse sensor was linked to a cable tie to enable users to customise the wearing position. The end of the cable tie was fixed to the temple tip using shrink tubing. A small block of foam was placed between the pulse sensor and the temple of the frame to ensure enough fixation to obtain a relevant signal while maintaining wearing comfort. In further versions, the implementation could be embedded directly into the temple, thus avoiding manual adjustments. Our prototype of WISEglass is depicted in Figure 4.1a. Figure 3.2 depicts the WISEglass hardware architecture.

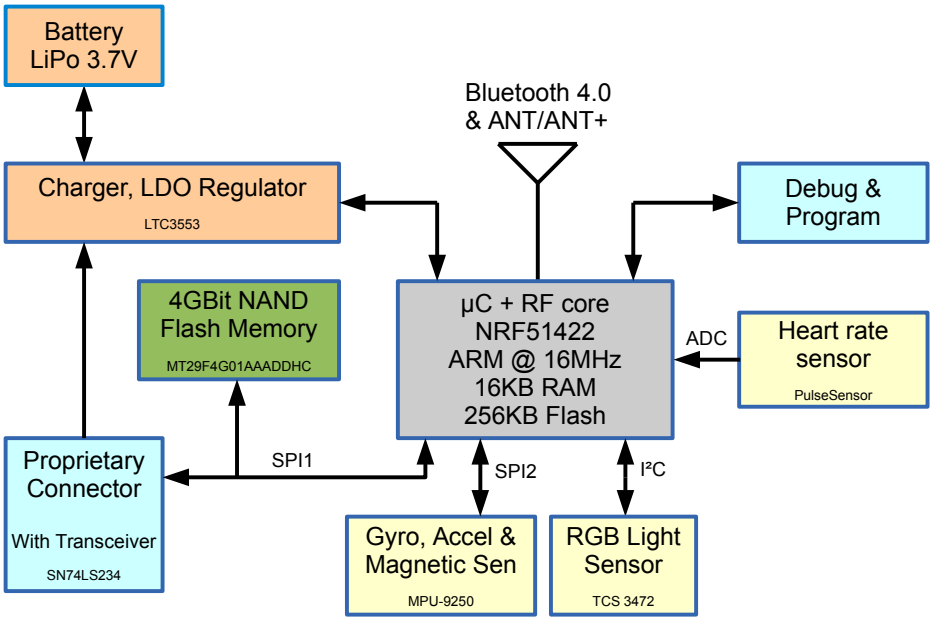


Figure 3.2: WISEglass hardware architecture. Separate SPI busses are used for communication of the main controller unit (MCU) with the flash memory and the inertial measurement unit to optimise data throughput.

The main controller unit (MCU) was a nRF51422 from Nordic Semiconductor that provides a 32 bit ARM Cortex-M0 core running at 16 MHz with a Bluetooth Low Energy and ANT+ wireless module as a System on Chip. We added a 512 MByte NAND flash to complement on-chip memory when storing sensor data. WISEglass is powered by a 3.7 V, 330 mAh Lithium Polymer battery.

For motion and orientation estimates, a MPU-9250 from InvenSense was used. The MPU-9250 provides accelerometer, gyroscope, magnetometer, all 3-axis inertial sensors, and a digital motion processor in a single package. The inertial sensors were AD-converted into 16 bits. Sensitivity can be configured to optimise accuracy depending on the application. For the accelerometer the ranges $\pm$ 2 g, 4 g, 8 g, and 16 g can be selected. The gyroscope offers $\pm$ 250 dps, 500 dps, 1000 dps, and 2000 dps. The magnetometer range is fixed to $\pm 4912\,\mu$T.

For light intensity measurements, a TCS3472 from ams was used. The TCS3472 senses light intensities for the red, green, blue, and clear light spectra separately. The sensor has an integrated IR filter and was connected to the MCU through the I^2C interface.

To obtain heart rate we added a PulseSensor [44] which is based on the reflexive photoplethysmography (PPG) principle. We found the temple to be a good location for measuring pulse as it eliminated the need for an ear clip.

3.4 Data collection

3.4.1 Study methodology

To evaluate our approach, we conducted a study with nine participants (3 female, 6 male, between 20 and 27 years). Participants either did not require prescription glasses or wore contact lenses. Each participant was given a pair of smart glasses. For heart rate reference measurements, participants were fitted with a CamNtech Actiwave Cardio ECG device. Annotations were performed by an observer using the ACTLog application for Android [85]. Figure 6.3 shows the sensor setup for a sample of the activities considered.

Figure 3.3: WISEglass worn during a subset of the activities considered in our full-day study protocol. In the protocol a variety of daily activities were covered and subsequently grouped into nine activity clusters for the recognition analysis. During our study we collected 66.08 hours of data from nine participants.

We recorded up to two participants per day. Participants received three complimentary meals during the recording day and a 25 Euro Amazon voucher as compensation.

During recordings we configured the accelerometer to a range of ± 4 g and the gyroscope to a range of ± 500 dps. Accelerometer, gyroscope, and magnetometer were sampled at a a rate of 50 Hz. Light measurements were integrated with a time constant of 154 msec yielding a sampling rate of 6.5 Hz and were upsampled to 50 Hz using a latest value strategy. The pulse sensor voltage was sampled at a rate of 50 Hz. The ECG reference was sampled at 200 Hz. From the study, we acquired a total 66.08 hours of data.

3.4.2 Study protocol

Participants were first introduced to the study protocol and signed an informed consent form. Participants then performed a scripted study protocol as shown in Figure 3.4. During the exit interview, participants evaluated their experience using the system usability scale (SUS) [21]. The protocol was designed to cover typical daily activities. Table 3.1 shows how activities were combined to activity clusters for further evaluation.

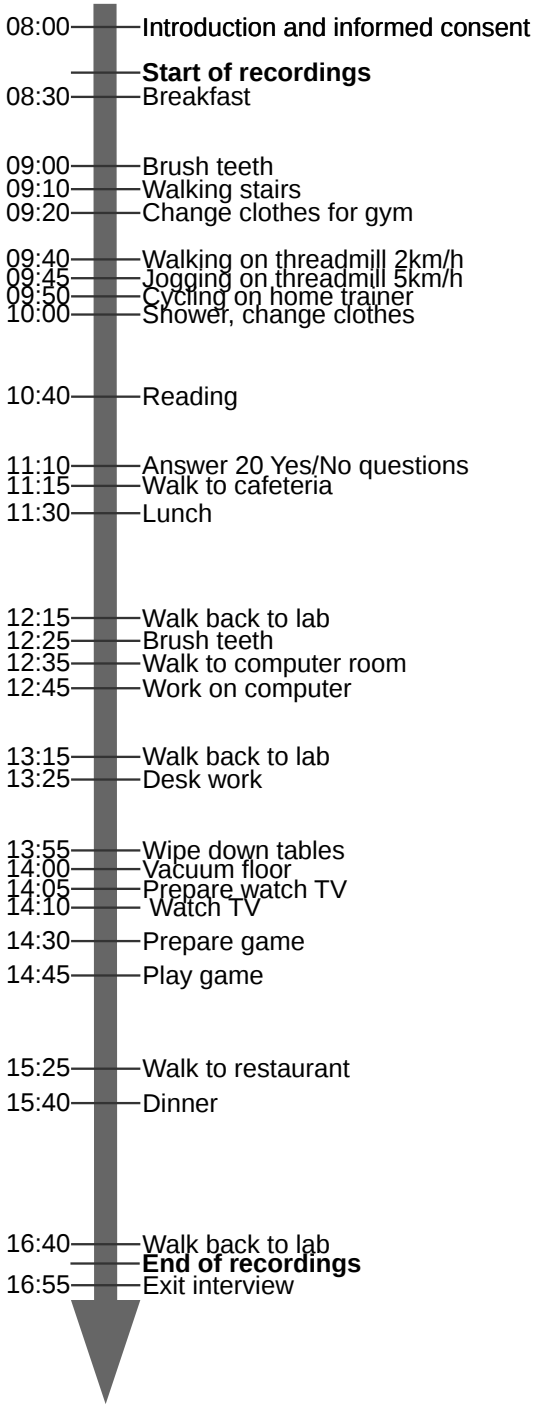


Figure 3.4: Protocol of the daily activity study. Recordings started before breakfast and ended before the exit interview. Sensors were temporarily removed while participants were in the gym locker room.

Table 3.1: Activity clusters and total duration.

No	Cluster	Activity & total duration [min:sec]
1	Eat	Breakfast [84:43]
		Lunch [156:49]
		Dinner [181:53]
2	Walk	Lab to Bathroom [17:37]
		Lab to gym [36:27]
		On treadmill 2km/h [47:06]
		Gym to lab [35:56]
		Lab to cafeteria [20:11]
		Queuing for lunch [6:28]
		Picking up lunch [13:20]
		Cafeteria to lab [25:28]
		Lab to restaurant [58:36]
		Restaurant to lab [55:26]
3	Brush	Teeth [39:17]
4	Stairs	Walking [19:49]
5	Jog	On treadmill 5km/h [45:27]
6	Cycle	On gym trainer [91:14]
7	Read	A book [276:31]
		Desk work [273:34]
8	Screen	Computer work [253:50]
		Watching movie [187:56]
9	Cleaning	Vacuuming [45:22]
		Wiping tables [45:02]

3.5 Evaluation methodology

3.5.1 Data preprocessing

Data streams from the WISEglass, ECG device, and smartphone were aligned according to their time stamps, merged, resampled to 50 Hz, and cropped to a common time axis. Subsequently, annotations were post-processed by visual inspection using the ACTLab MARKER toolbox for MATLAB.

3.5.2 Feature extraction

Features were extracted using pandas [63] and numpy [91] libraries of Python. For recognition of daily activities and the detection of screen-use, a sliding window over data samples with a window size of $n = 1500\,\text{sa}$ (30 s) and a step size of $s = 50\,\text{sa}$ (1 s) was used. Feature vectors were standardised and normalised before further processing. The sliding windows were assigned to the class that was most represented in the window. Motion features were computed over each of the three axis of the accelerometer (a_x, a_y, and a_z) and the gyroscope (g_x, g_y, and g_z). We computed the L^2 norm for the acceleration axis a_{norm} for each sample. Features listed in Table 3.2 were subsequently calculated for a_{norm}, a_x, a_y, a_z, g_x, g_y, and g_z per window.

In order to detect screen-use based on light sensor data, we derived features on the RGB and clear light intensity values named r, g, b, and c respectively. We calculated the median med() and number of median crossings mcr() for each color channel per window. To express

Table 3.2: Motion features calculated for a vector of data samples w with the window size n with elements $1 \dots N$, where w_i is the i-th sample in the window. w' and w'' are the first and second derivative of w determined between two neighbouring samples. zcr() denotes the number of zero crossings.

1. $\bar{w}$	2. $\sum_{i=1}^{n} w_i$, if $w_i > 0$	3. $\sum_{i=1}^{n} w_i^2$
4. $\sigma^2(w)$	5. $\sum_{i=1}^{n} w_i$, if $w_i < 0$	6. $\sum_{i=1}^{n} w_i$
7. $\max(w)$	8. $\sum_{i=1}^{n} w'_i$, if $w'_i > 0$	9. $\sum_{i=1}^{n} \lvert w_i \rvert$
10. $\min(w)$	11. $\sum_{i=1}^{n} w'_i$, if $w'_i < 0$	12. $\mathrm{zcr}(w)$
13. w_1	14. $\sum_{i=1}^{n} w''_i$, if $w''_i > 0$	15. $\sum_{i=1}^{n} \lvert w_i \rvert'$
16. w_N	17. $\sum_{i=1}^{n} w''_i$, if $w''_i < 0$	18. $\frac{\bar{w}}{\sigma(w)}$
19. $w_N - w_1$	20. $\max(w) - \min(w)$	21. $\sqrt{\sum_{i=1}^{n} w_i^2}$
22. $\sum_{i=1}^{n} w'_i$	23. $\sum_{i=1}^{n} \lvert w'_i \rvert$	24. $\sum_{i=1}^{n} \lvert w''_i \rvert$
25. $\sum_{i=1}^{n} w''_i$		

spectral composition of light, we computed the median of the ratio for each combination of two color channels c_1, c_2 defined as $\mathrm{crm}(c_1, c_2) = \mathrm{med}(\frac{c_1}{c_2})$.

Heart rate estimation was performed by counting the number of positive peaks in the raw voltage signal of the pulse sensor. For peak detection, we used an existing peak detector [34]. We experimented with different minimum peak distances, ranging from 0 to 25 sa in steps of 5 sa, but found 20 sa to be the best for our needs. As reference heart rate, the estimate provided by the ECG sensor was used. We computed the mean of the heart rate estimate from the ECG sensor per window. The calculation of the heart rate error was performed on windows with no overlap using window size $n = 1500$ samples.

3.5.3 Feature selection

We applied principle component analysis (PCA) to reduce the number of features. We configured the PCA algorithm to select the minimal number of features, such that at least 99.9 % of the variance in the training data was explained resulting in the selection of 78 out of 175 features for recognising daily activities. For screen-use detection, 9 out of 14 features were selected.

3.5.4 Cross-validation

To ensure generality and stability of our statistical model, we used Leave-One-Out (LOO) cross-validation on a per-participant basis. In every cross-validation fold the classifier was trained on all available feature instances from eight participants. To evaluate model performance the data of the remaining participant was used.

3.5.5 Classification with reject option

Classification of the motion-based activity and light-based screen detection was performed using a Gaussian Mixture Models (GMM) classifier as described in [14]. After training one GMM for each class, we computed the probabilities for each test sample. The classifier then selected the class for which the highest probability occurred. To also predict unlabelled data we implemented classifier reject to avoid making assumptions on the data in the null class. First we obtained the likelihoods for each test sample from the GMM model. If the maximum likelihood for a test sample was below the threshold, the classifier rejected it as a null class instance and otherwise selected the class of the GMM yielding the maximum likelihood. A sweep search was performed to find the optimal threshold. We used all maximum likelihoods, which occurred during training in each LOO cross-validation fold and picked the threshold yielding the best F1 score. Results per fold were averaged to show the performance for an optimal F1 score.

3.5.6 Heart rate estimation performance

To validate heart rate estimation from the pulse sensor, we computed the root-mean-square error (RMSe) between the ECG reference and the pulse sensor. Both signals were filtered using a bandpass filter before peak detection was applied. RMSe computation was limited to data samples, where the ECG heart rate was between 40 bpm and 150 bpm according to the reference estimate. We calculated RMSe for three data subsets to investigate the effect of motion artefacts: (1) All available data. (2) Data from stationary clusters eat, read, and screen. (3) Data from classes stairs, walking, and jogging on treadmill where the participant is assumed to be moving.

3.6 Results

Here we show the results for recognising user activities using inertial sensor data, detecting screen-use using environmental light, and estimating heart rate using the WISEglass pulse sensor.

3.6.1 Daily activity recognition

We evaluated the recognition of the activity clusters (Tab. 3.1) using the GMM classifier and LOO cross-validation. GMM performance was analysed by varying the number of Gaussian mixture components between $c = \{1, 3, 5, 7\}$ per class, as well as analysing the covariance matrix configuration (full or diagonal). All configurations using diagonal covariance matrix yielded similar accuracies, except for the configuration with one mixture only. The full covariance matrices generally perform marginally better (3 %) than diagonal, for the same number of mixture components. For our further analysis, we resided on using a diagonal covariance matrices due to the increased modelling and computation complexity that a full covariance matrix would introduce. Using diagonal covariance matrices, highest accuracy was 77 % using 3 GMM mixture components. The lowest performance was 70 % using 1 GMM mixture component. Based on the results, we used 3 GMM mixture components per class for all subsequent analyses.

Daily activity recognition varied in accuracy between 70 % and 84 % depending on the participant considered for testing. The confusion matrix shown in Figure 3.5 depicts the class assignment averaged over all participants. For the confusion analysis LOO cross-validation results were averaged across folds. The confusion results indicate per-class performances between 80 % and 90 %, except for "cycle". The performance of the cycle cluster was only 50 % due to confusions with "eat", "walk", "brush", and "clean" clusters. As "cycle" took place at slow pace on a gym trainer (see Fig. 6.3), only minor head motion occurred, thus resulting in motion patterns similar to other classes. The activity clusters "read" and "screen" show the largest confusions. Based on motion data it seems plausible that "read" and "screen" clusters were confused as they showed similar head motions.

At a first glance, the classifier reject yields a null class accuracy of only 20 % and several confusions with other activity clusters. We found that the performance is related to our recording protocol and the diversity of motion in the activity clusters. Our activity clusters were comprehensive on the activities contained in it, thus there was simply not much unlabelled data left with other motion patterns than what was represented in the activity clusters. In the unlabelled episodes, participants were, e.g., walking, sitting at a desk, or remained in another state that often matched a modelled activity cluster.

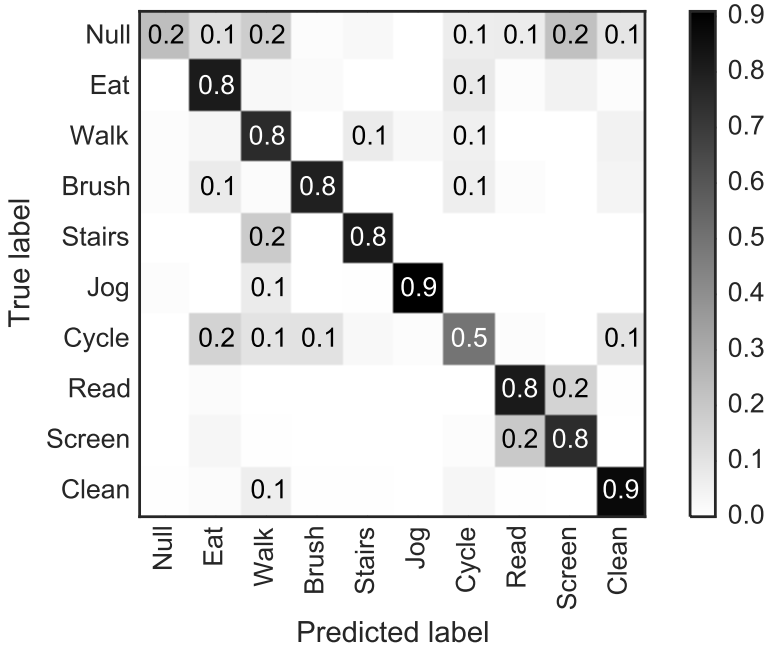


Figure 3.5: Daily activity recognition: confusion matrix across cross-validation folds. With the exception "cycle" cluster, all activity cluster perform at an accuracy of at least 80 %. Null class performance was low due to the nature of the study protocol, where many modelled activities occurred also as unlabelled data, thus resulting in confusions.

3.6.2 Screen-use detection

We detected screen-use from the light sensor data. Screen-use and its timing are important factors in circadian rhythm entrainment, as light emitted from screens and received through

eyes at night may hamper falling asleep later.

Figure 3.6 shows mean accuracies for the light-based screen detection per participant using LOO cross-validation and a GMM classifier with diagonal covariance matrices using 3 mixtures per class. An accuracy between 64 % and 93 % was obtained here, with a mean of 80 %. While most folds achieve performance around the mean score, evaluation for participant 4 showed as an outlier.

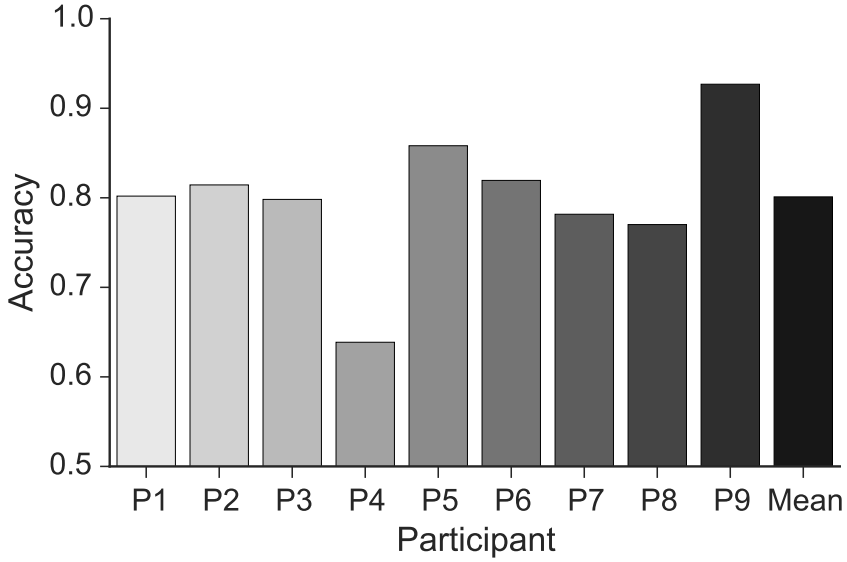


Figure 3.6: Screen-use detection: Normalised accuracy per participant (cross-validation fold). Mean accuracy was 80 %. Results for participant 4 seems to be an outlier.

3.6.3 Heart rate estimation

Heart activity was measured in WISEglass using a optical pulse sensor worn at the temple. Similar to the daily activity recognition, there are many context applications that benefit from physiological estimates. Here we focus on heart rate as one important indicator. Figure 3.7 shows the RMSe for all recordings as well as activity subsets "motion" and "stationary" per participant and averaged. For all activities combined, heart rate estimation RMSe varied between 10 bpm and 19 bpm with a mean of 13 bpm. Motion artefacts contributed to an increased RMSe. For motion activities the mean RMSe increases to 14 bpm while for stationary activities it is reduced by 40 % compared to motion activities, to 9 bpm. Motion artefacts were prominent in the heart data from both WISEglass pulse sensor and the ECG reference. During motion it could not be determined which of the systems (WISEglass or ECG reference) provides more accurate data. Thus during motion, the error reported here must be considered as a difference between the two systems only.

3.6.4 Usability evaluation

The SUS evaluation of all nine participants yielded a mean score of 66.4 of 100. Participants found the system easy to use with 4.1 of 5.0 points on average. The overall score may be

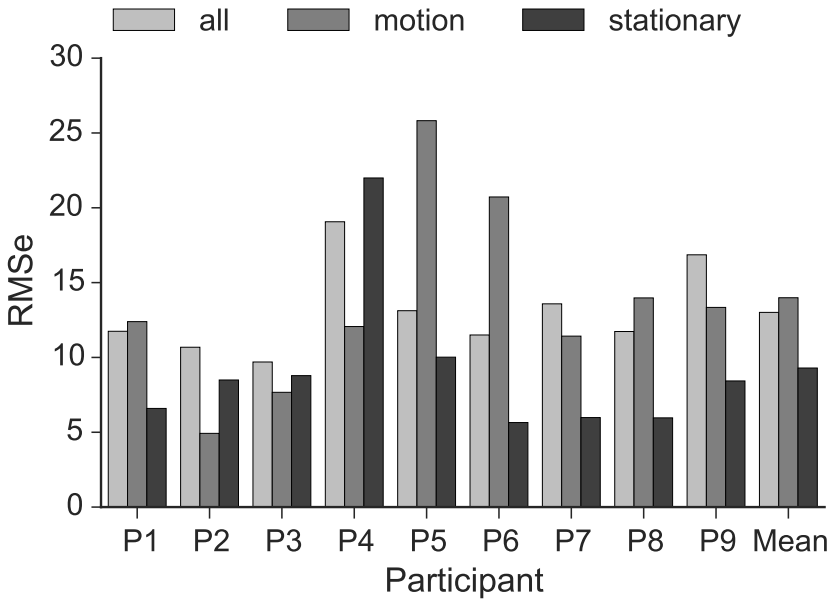


Figure 3.7: Heart rate estimation: RMSe analysis per participant, where the reference heart rate was between 40 and 150 bpm. RMSe was analysed for three activity clusters: all, motion ("stairs", "walk", and "jog"), and stationary ("eat", "read", and "screen").

negatively biased due to the experiment process (e.g. placing ECG electrodes) and hence consider the easy to use question the most relevant result of the SUS evaluation.

3.7 Discussion

The results of this work clearly confirm idea behind WISEglass: smart eyeglasses can provide multi-purpose sensing functions for different context recognition applications. Instead of clip-on designs, as Google Glass and others, we emphasise that eyeglasses could integrate sensing and processing functions into typical eyeglasses frames. The performance for activity recognition, screen-use detection, and heart rate estimation in this work show that the head-worn measurement positions and the used pattern processing methods are adequate. Compared to a phone, the smart eyeglasses are always at the same body position and in addition can measure HR continuously. Compared to a watch-like devices, smart eyeglasses can measure the light shining into the users eyes. Another advantage is that the activities a watch can detect must somehow include the arm (e.g. watching TV might be difficult to detect), so eyeglasses can cover a broader spectrum of activities. Nevertheless, for daily activity recognition improvements of the null class modelling seem necessary. It is clear that our recordings, although made over full days and including diverse activities do not reflect all conditions under which smart eyeglasses could be used. As a consequence, we had little data that was not already modelled by the activity clusters and thus unknown activities were under-represented in the dataset. This effect shows that activity clusters are robust enough to generalise onto unknown data. In a next step, WISEglass could be used in free living conditions. Then however, less accurate annotations must be expected.

The GMM classification approach demonstrated that user-independent models with good

average performance (approx. 80 %) can be obtained. Additional GMM parameter selection could be performed on a per class basis to further improve the model fit. Specifically, for the screen-use detection improvements could be realised by combining light and motion data. While we intended to span a wide range of possible applications of smart eyeglasses, improvements to the heart rate estimation method should be investigated, too. Here, a more controlled setting may be helpful to provide detailed RMSe for different movements. The ECG reference data showed motion artefacts due to insufficient electrode attachment. The present work presents a pilot study to confirm benefit of smart eyeglasses. In further investigations, additional participants and additional scenarios and applications for smart eyeglasses shall be investigated to further expand the use cases for smart eyeglasses.

We are aware that our current WISEglass prototype is yet not fully resembling the vision of regular eyeglasses. While we made important steps, e.g. by integrating the light sensor and wiring into the eyeglasses frame, further miniaturisation and integration work is necessary. Moreover, to improve unobtrusiveness, the baseboard and battery could be moved to the temple hinges. The pulse sensor could be made smaller and integrated in the temple such that the foam block disappears. Given the reliable data from the pulse sensor, we are confident that such improvements are feasible.

3.8 Conclusion

WISEglass is the first multi-modal sensor system retrofitted to off-the-shelf eyeglasses. The platform offers motion, light, and pulse sensors for use in a wide range of context-aware applications. All sensor modalities were successfully evaluated in three application scenarios, demonstrating that the smart eyeglasses serve as single-position context measurement device. For daily activity recognition, an average accuracy of 77 % for distinguishing 9 activity clusters using LOO cross-validation was achieved. The recognition rate is comparable to previous work [51, 52], while in our work more activities and user-independent models were considered. However, most importantly, WISEglass integrates all sensors into one commonly used accessory, thus does not require to wear head-mounted displays, cameras, etc. Screen-use detection achieved a mean accuracy of 80 %. Heart rate estimation was evaluated against a chest-worn ECG reference with an average RMSe of 13 bpm for all available data and 9 bpm RMSe during stationary activities confirming that the temple-worn sensor provides an adequate wearing position.

Acknowledgements

We thank Timo Spinde and Christoph Weidemeyer for collecting the dataset, Jakob Kasbauer for contributing to the light data analysis, and all study participants. This research was supported by the Dutch Technology Foundation STW under grant number 12184.

4

Smart eyeglasses detecting computer screen use

Chapter originally published as: F. Wahl, J. Kasbauer, and O. Amft. Computer Screen Use Detection Using Smart Eyeglasses. *Front. ICT*, 4, 2017.

4.1 Introduction

People spend a large portion of the day looking at computer, television, or tablet screens. For example, in 2015 the average US adult spent 9 hours and 52 minutes in front of different screens every day [36]. Many office workers spend more than 6 hours in front of a computer screen [71]. Extended screen use often causes eye strain, the most common repetitive strain injury. For example, in the United States 65 % of the population suffer from eye strain [87]. In addition, screen use can influence the circadian phase. The circadian clock is entrained by timing and intensity of light exposure. Morning light exposure advances circadian clock phase, light exposure in the evening delays it [74]. The circadian system is most sensitive in the blue range of the light spectrum [20]. LED backlit screens emit high energy in the blue light spectrum compared to other wavelengths and other indoor light sources. Thus screen use could shift circadian phase [22] leading to e.g. difficulty to fall asleep at night and consequently sleep deprivation. A fundamental requirement for guidance and intervention to prevent eye- or sleep-related health problems is therefore to detect screen use when it actually occurs.

Ambient sensors, e.g. proximity sensors [53], do not suffice for screen use detection as they cannot distinguish mere presence from looking at a screen. To detect screen use, light measurements have to be taken as close as possible to the eye. Only few head-worn wearables have been proposed for context recognition that could be used to robustly detect light received at the eyes. Eyeglasses appear to be practical, everyday accessories and could house a light sensor, without changing their main function as eyeglasses or substantially modifying the eyeglasses appearance. Considering the sensor's typical field of view of 60 degrees, the best measurement of light actually received at the eyes may be between the eyes, i.e. at the eyeglasses bridge.

Light reaching the eyes originates from a mixture of sources, which differ in the intensity and spectral distribution. Common light intensities range over several orders of magnitude, e.g. from 500 lux at a office desk to 100.000 lux outside on a sunny day. When exposed to large amounts (above 1.000 lux) of natural or artificial light, screens may contribute a negligible share. Consequently, screen use matters if there is only dim ambient light, e.g. during evening or night hours when ambient light intensity is low. Due to the relevance of blue light energy of LED-based screens, spectral irradiance patterns could help to discriminate screen use from other light sources. Thus, a spectral decomposition of the incident light measurement is requires at the sensor. The detection algorithm has to cope with noise added by head motion. In addition, a broad range of screens, content types, and ambient lighting conditions complicate the detection of screen use as screen light emissions vary.

In this paper, we introduce an approach to detect screen use with smart eyeglasses that provide a colour light sensor embedded into the eyeglasses bridge. Electronics to store, process, and transmit measured light data were integrated into the eyeglasses frame. Our approach involves three steps to investigate the challenges related to screen use recognition: A Test bench environment was used to investigate screen use detection in a fully controlled environment, ambient light intensity, and without head motion. In a Lab study screen use was investigated with participants wearing smart eyeglasses in different ambient light intensities. Finally, a study of typical activities of daily living (ADLs) evaluated detection of screen use in unconstrained daily life situations. Using features derived from the ratios between colour channels we detect when screen use occurs.

In particular, this paper provides the following contributions:

1. We analysed screens, content type, and typical viewing distance to derive device-specific light intensity at the user's eyes. We found that computer screens provide highest light intensities due to their size, displayed content, and typical viewing distance.

2. We analysed screen use detection performance under different ambient light sources and intensities in a Test bench environment, which indicated a perfect screen use detection.

3. We evaluated our screen use detection approach using smart eyeglasses worn by 14 participants in a Lab study. For light intensities below 200 lux a ROC AUC of above 0.9 was reached using participant-independent models.

4. We applied our approach to data recorded in a study investigating different ADLs. In seven participants we detected screen use with an average ROC AUC of 0.83 when assuming 30 % screen use.

4.2 Related work

According to Duffy and Wright [35] light is the dominant factor in entrainment of the circadian rhythm. Phase shifts of -2 to +3 hours per day are possible depending on the intensity and timing of light exposure. The authors report that a misaligned circadian clock can lead to impaired performance, alertness, and upset gastrointestinal functions. Kantermann and Roenneberg [56] found that light during night can damage DNA and thus lead to cancer putting those in shift work especially at risk. Brainard et al. [20] found spectral sensitivity of the circadian system to peak at 464 nm wavelength. Being exposed to blue light in the evening delays the production of melatonin and thus delays circadian phase. Screen use potentially causes circadian phase shift, e.g. when working late nights in front of a computer screen. As misaligned clocks negatively influence health, episodes of screen use could be detected automatically to support behavioural change.

The effect of screen use on the circadian rhythm has been analysed in several studies. Cajochen et al. [22] found a significant delay in dim light melatonin onset (DLMO) from evening exposure to LED-backlit computer screens compared to cold cathode fluorescent lamp (CCFL)-backlit computer screens. Prior to their regular bedtime, 13 participants performed a five hour screen use episode during which melatonin was sampled every 30 minutes. Wood et al. [106] conducted an experiment with 12 participants using a tablet 2 hours prior to bedtime. They found that melatonin was suppressed by 23% on average after two hours of tablet use in the evening compared to no tablet use. After one hour of tablet use the effect was smaller with only 7% melatonin suppression compared to no tablet use. Chang et al. [25] investigated the influence of using light-emitting eReaders at night instead of reading print books. In their crossover protocol, twelve participants read for 4 hours prior to bedtime during five consecutive evenings per device. They found that DLMO was delayed by more than 1.5 h when using the backlit eReader over the paper book. These findings confirm circadian phase shift effects through screen use in the evening. Thus, an automatic detection of screen use could inform users about their behaviour and support them in implementing an effective compensation.

Studies investigating light influences on human physiology often used wrist-worn devices to record light exposure. Wahl et al. [100] compared wrist and head-worn light sensors and found substantial differences in measured light intensities. One core problem with wrist-worn light measurements was a frequent occlusion of the sensor by long-sleeve clothing. Another issue

is that light sensor sensitivity depends on angular displacement, typically 50 % sensitivity at ±60° displacement, which requires sensors to be worn in an orientation similar to the eyes. Others used head-worn devices such as the Daysimeter [13] which offers accurate recordings but were found impractical for continuous use in everyday life due to their form factor and clip-on to eyeglasses. Regular eyeglasses are the most common vision aid, worn by many. For example, in Germany 63.5 % of the population above 16 years and 92 % above 60 years old wear eyeglasses [50]. In our previous work, we embedded a multi-modal, multi-purpose sensing system into regular eyeglasses termed WISEglass. WISEglass was used for a wide variety of sensing applications ranging from daily activity recognition [99], dietary monitoring [107] to motion based video game control [96]. The smart eyeglasses were validated in a study with 9 participants performing typical ADLs [97, 99]. Nine activity clusters were detected from accelerometer and gyroscope data with 77 % accuracy. Screen use detection using a colour light sensor showed a mean accuracy of 80 %. This paper provides an in-depth evaluation of screen use detection with colour light sensors in three subsequent studies. We investigate different screen types, ambient light sources, and content types.

Intervention measures can be applied to minimise influence of screen use on the circadian phase. Heath et al. [45] studied if tablet use one hour prior to bedtime influences the circadian phase. In their study with 16 participants, unfiltered light was compared to light filtered with the f.lux application. They could not find sound proof that one hour of tablet screen had a significant impact on the circadian phase. Van der Lely et al. [89] investigated the effect of blue light-blocking glasses on the circadian phase. In 13 young males which spent three hours prior to going to bed in front of a computer screen for one week each, they found a significant increase in melatonin when using blue light-blocking glasses. There are suitable intervention measures based on either software or hardware to minimise the effect of screen use on the circadian phase. The work presented in this paper could support users by reminding them to wear blue light-blocking glasses or (automatically) enabling a software-based intervention to screen light profile and light intensity.

4.3 Light monitoring in eyeglasses

The WISEglass project aims to embed a multi-modal sensor system in regular eyeglasses that could be worn as an everyday accessory, just like eyeglasses are worn today. Our first prototype was based on regular eyeglasses and is shown in Figure 4.1a.

Subsequently a 3D printed version was built as shown in Figure 4.1b. Both prototypes were used for the analyses presented in this paper. We added compartments inside the temple ends of the eyeglasses to house the baseboard and battery. Compared to our first prototype the compartments improve balancing the component weight. The compartments disappear behind a wearer's ears when attached, thus WISEglass appears as regular eyeglasses to bystanders. We included a small compartment on the bridge of the eyeglasses to embed a colour light sensor. The bridge location can be used to unobtrusively measure light. 3D printed eyeglasses could be fitted to the wearers head to achieve similar wearing comfort as in existing eyeglasses models.

Integrated sensors are controlled and sampled by the on-board micro-controller. Data can be stored in flash memory for later download or streamed directly via Bluetooth Smart. The battery lasts for 32 hours when sampling light sensor data at 50 Hz. Sampling rates can be configured depending on the application.

(a) First prototype based on regular spectacles.

(b) 3D printed prototype.

Figure 4.1: Eyeglasses prototypes used in this work. The eyeglasses prototypes integrate a colour light sensor with red, green, blue, and clear channels into the eyeglasses bridge. In the prototype shown in Figure 4.1b electronics for recording, transmitting, and storing sensor data were integrated into the eyeglasses frame.

4.4 Evaluation methodology

In this section we describe goals, sensing approach, recording protocol, and evaluation methodology for each of the three analysis studies.

Starting from the illuminance at the screen during PC work and TV use, we calculated the light intensity at the user's eyes for different devices as described in Section 4.4.1. A Test bench was used to derive a baseline of the irradiance provided by screens at regular viewing distance and under controlled environmental lighting conditions. The Test bench analysis is further described in Section 4.4.2. In a Lab study (described in Section 4.4.3), data was collected using the same desk setup and viewing distances for every participant while ambient light intensity varied due to weather and time of the day. The ADL study (Section 4.4.4) investigates a natural variability of screen use, where activities were suggested, but the execution left to participants.

In this work we obtain data from a TCS34725 [3] light sensor made by ams. Light was measured in four different spectra: red, green, blue, and clear. After downloading the raw data, virtual light channels were computed from multiple raw colour channels by deriving channel ratios. As an example, the ratio of the blue to clear light channel expresses how much blue light is measured in relation to the total amount of light.

4.4.1 Light intensity calculations

The light intensity at the user's eyes is necessary to quantify the impact of screen use on the circadian phase. We calculated the light intensity at the user's eyes for different device types during PC and TV use at typical viewing distances. The area A was computed from the aspect ratio $w : h$ and the diagonal length l of the screen as

$$A = \frac{w \cdot h \cdot l^2}{w^2 + h^2}. \tag{4.1}$$

The luminance L for each screen was researched online and used to compute luminous intensity I as

$$I = L \cdot A. \tag{4.2}$$

Different screen activities yield different light intensity levels at the user's eyes. During high illuminance screen activities, e.g. editing documents at a PC screen, a majority of pixels is typically bright. In contrast, low illuminance screen activities, e.g. watching a movie, feature dark backgrounds. To reflect screen activity dependent light intensity differences, we used two activity prototypes to describe typical screen activities: PC work for high and watching TV for low illuminance screen activities. To incorporate the effect of content into the textbook irradiance computations, we added the content factor μ_{PC} and μ_{TV} in Equations 4.3 and 4.4, respectively. Our Test bench measurements found TV content to produce one third of the light intensity at the user's eyes compared to PC content. Content factors were therefore set to $\mu_{PC} = 1$ and $\mu_{TV} = \frac{1}{3}$. Finally, we used a typical viewing distance d for each device to compute light intensity at the user's eyes during PC use E_{PC} and TV use E_{TV} as

$$E_{PC} = \frac{I}{d^2} \cdot \mu_{PC}, \tag{4.3}$$

$$E_{TV} = \frac{I}{d^2} \cdot \mu_{TV}. \tag{4.4}$$

For the light intensity analysis, the distance d between screen and eyes is essential. However, viewing distances differ per device type. We used the following viewing distance settings in this analysis. Smartphone users maintain an average viewing distance of 32.2 cm when web browsing [10]. Tablets and eBook readers are typically used at a 50 cm distance [23, 83]. For PC use the German Social Accident Insurance (DGUV) recommends a viewing distance of 50 and 65 cm [32]. In this investigation, we assumed 50 cm for notebooks, and 60 cm for desktop PC use. Investigations on TV use found the average viewing distance to be above 200 cm [62, 67].

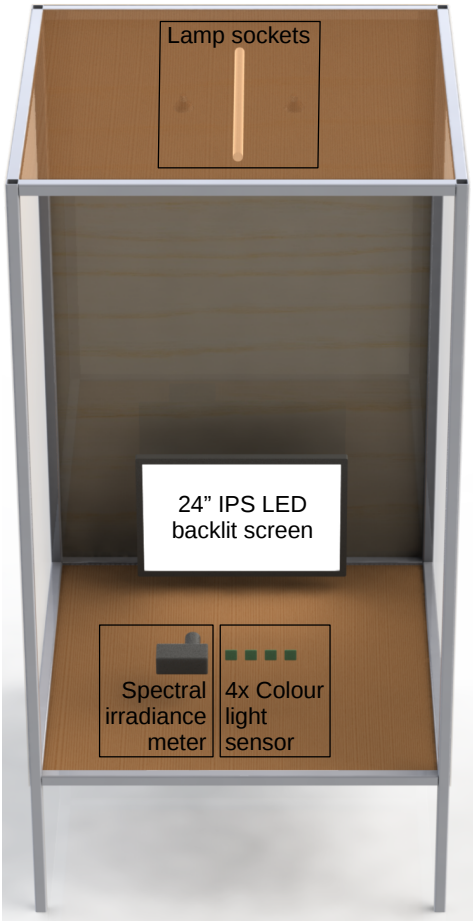
4.4.2 Test bench

We built a Test bench to analyse light irradiance from screens under controlled ambient lighting conditions. The Test bench included four different types of dimmable ambient light sources. Movie or PC use content was displayed on a computer screen to simulate different content types. Light irradiance was measured at a distance of 50 cm, where the recommended distance for screen use at the workplace is between 50 and 65 cm [32].

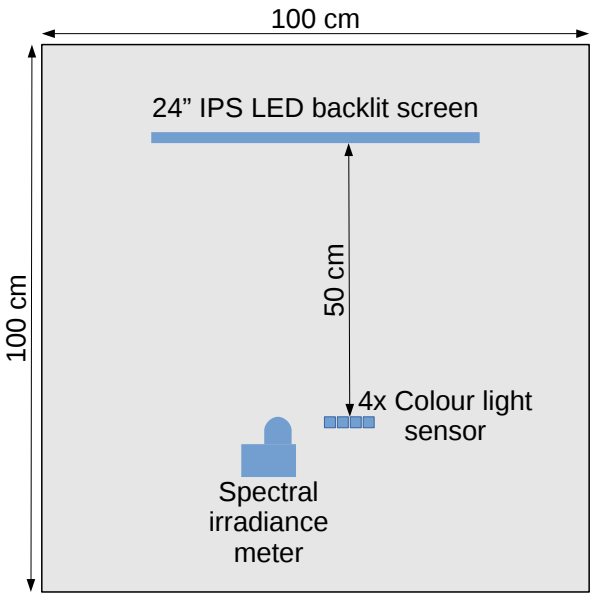
Sensing approach

The Test bench was constructed out of wooden boards and aluminium profiles forming a cuboid of 200 cm height with a footprint of 100 cm x 100 cm. The floor panel was mounted at a height of 40 cm from the ground leaving a distance from floor to ceiling of 160 cm. The height resembles the typical distance between table surface and ceiling lamps in office spaces. The screen was mounted at the rear end of the floor panel. Irradiance was measured at 50 cm distance, resembling the typical viewing distance of computer screens. A model of the Test bench is shown in Figure 4.2a.

Initially, we analysed the spectral irradiance of a typical LED backlit screen. Here we used the SpectraRad Xpress BSR112E-VIS miniature spectral irradiance meter by BWTEK. The recorded spectrum ranges from 380 - 750 nm wavelength with a resolution of 3 nm. Connected to a PC running the BWSpec software, samples were recorded at 26.3 Hz. Spectral measurements were averaged per second resulting in a sampling rate of 1 Hz. Dark calibration was conducted prior to recordings.



(a) 3D Model of the Test bench.



(b) Schematic of device arrangement in the horizontal plane at the floor panel.

Figure 4.2: Test bench analysis. Different ambient light sources were mounted on the ceiling of the construction. On the floor panel screen use was simulated on a 24 inch IPS LED backlit screen. Four colour light sensors were placed at 50 cm distance from the screen. Light irradiance reference measurements were recorded using a spectral irradiance meter.

For Test bench recordings a HP EliteDisplay E241i 24 inch IPS panel LED backlit screen was used. The screen configuration was reset to factory default prior to recordings. Figure 4.3 depicts the spectrum of the screen with all pixels set to white. The three spectral peaks reflect the composition of the three primary colours red, green and blue. The intense blue peak is characteristic for LED backlit screens [22].

Subsequently, we investigated the screen irradiance measured by a typical commercial colour light sensor under different lighting conditions. Data from four TCS34725 [3] colour light sensors, the same as used in WISEglass, were recorded. We used four sensors to accommodate for inter-sensor variation. Each light sensor was connected to an Arduino via I^2C bus and triggered an interrupt when a new measurement was ready. Sensor integration time was set to 50 ms resulting in a sampling rate of 20 Hz. Arduinos were connected to a computer running the CRNT software to record data [11]. Sensor gain was set to 60 × resulting in a 27.6 % saturation of the clear light channel at 1000 lux indoors, a typical indoor light intensity on a

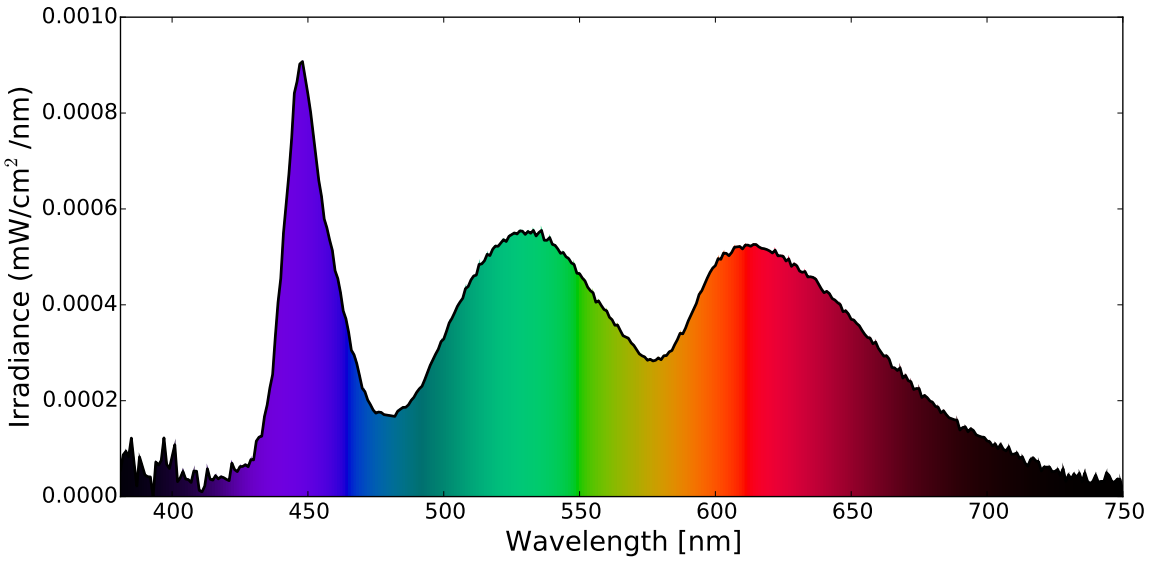


Figure 4.3: Test bench analysis. Spectrum of a HP EliteDisplay E241i 24 inch LED backlit IPS panel displaying a full white image. The spectrum was computed by averaging 26 samples from the spectral irradiance meter. It can be seen that the blue spectral component is more prominent than the green and red components.

sunny day. Figure 4.2b shows the installation on the floor panel of the test bench.

Three lamp sockets were mounted at the ceiling of the test bench for LED lamps, halogen lamps, and fluorescent lamps respectively. We evaluated screen use detection using four different artificial light sources described in Table 4.1. The spectra of the lamps at full brightness are shown in Figure 4.4.

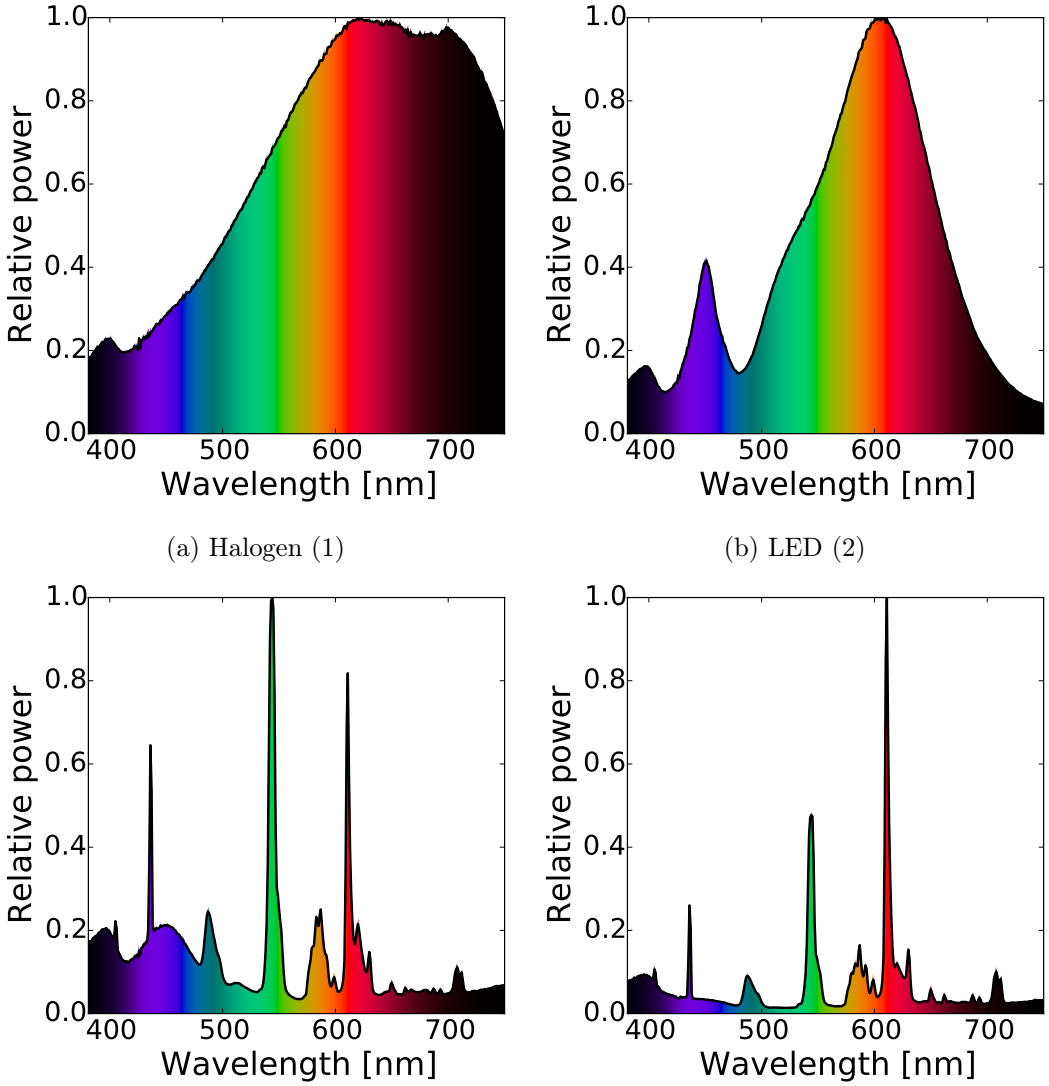
Table 4.1: Test bench analysis. Artificial light sources and their core properties according to manufacturer specifications.

No	Lamp type	Colour temperature	Energy consumption
1	Halogen	3000 K	35 W
2	LED	2700 K	5.9 W
3	Cold CCFL	8000 K	18 W
4	Warm CCFL	2700 K	18 W

Recording protocol

The Test bench recording protocol contained three independent variables: Screen state S, lamp type T, and lamp intensity D.

Screen state was set to one of the three following states: (1) Off for not screen use $S = \mathit{off}$, (2) Displaying TV content $S = \mathit{tv}$, or (3) Displaying PC use content $S = \mathit{pc}$. To switch between $S = \mathit{off}$ and $S = \mathit{tv}$, pc a *xrandr* script was used. The James Bond Spectre movie



(a) Halogen (1)

(b) LED (2)

(c) Cold CCFL (3)

(d) Warm CCFL (4)

Figure 4.4: Test bench analysis. Spectra of the four different ambient light sources used. Spectral distribution differs per lamp type. CCFL lamps show few, relatively high energy peaks.

trailer was used to simulate TV content[1]. To simulate PC content a video tutorial on iPython notebooks was played[2]. Videos were played in an endless loop.

The lamp type was either one of the four different lamps listed in Table 4.1 or darkness: T = *{Halogen, LED, Cold CCFL, Warm CCFL, dark}*. Lamp intensity was set to D = *{100 % ... 10 %}* in steps of 10 %. We recorded 10 minutes of data for a total of $S \times T \times D = 141$ states (T = *dark* has no steps for D).

[1] `https://www.youtube.com/watch?v=z4UDNzXD3qA`
[2] `https://www.youtube.com/watch?v=HaS4NXxL5Qc`

Evaluation procedure

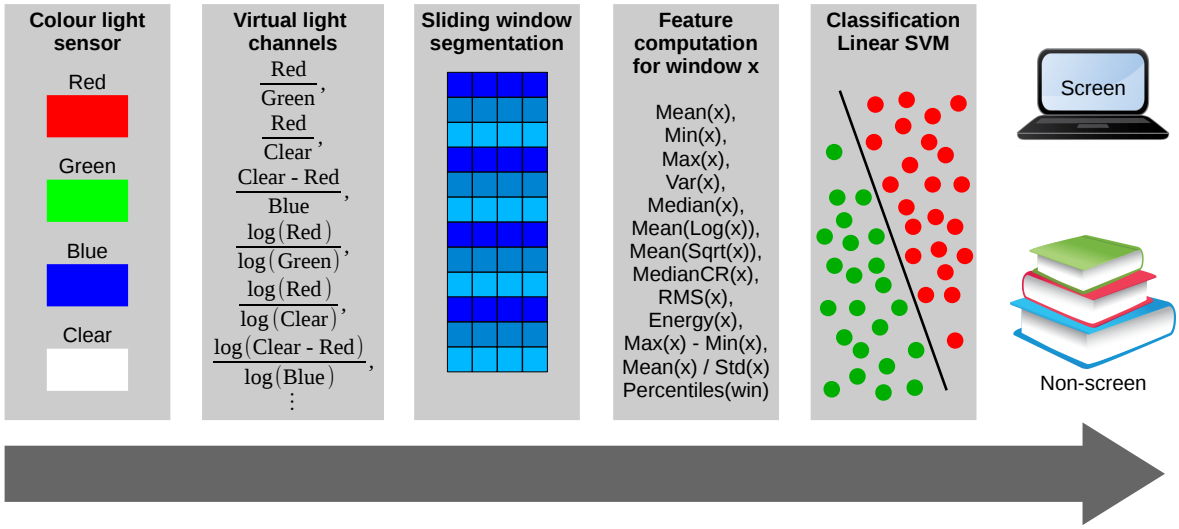


Figure 4.5: Data processing pipeline for light sensor data from raw colour channel data to classifier output. Virtual light channels were computed from colour light sensor data. After sliding window segmentation, features were extracted. In addition to features listed, the 5 %, 10 %, 25 %, 75 %, 90 %, and 95 % percentiles were calculated. Subsequently a linear support vector machine was used to classify screen use episodes.

Starting from light measurements, we computed ratios between multiple colour light sensor channels to become independent from absolute values. Each ratio represents one virtual light channel. Virtual light channels put each colour channel in relation to others. For example, the virtual channel $\frac{blue}{clear}$ indicates the ratio of blue light vs. overall light intensity. We computed the virtual light channels $\frac{\{R,\ G,\ B\}}{C}$, $\frac{R}{G}$, $\frac{B}{R}$, $\frac{B}{G}$, $\frac{C-\{R,\ G,\ B\}}{\{R,\ G,\ B,\ C\}}$, $\frac{log(C-\{R,\ G,\ B\})}{log(\{R,\ G,\ B,\ C\})}$, $\frac{log(\{R,\ G,\ B\})}{log(C)}$, $\frac{log(B)}{log(R)}$, $\frac{log(B)}{log(G)}$, $\frac{log(R)}{log(G)}$, $\frac{\sqrt{\{R,\ G,\ B\}}}{\sqrt{C}}$, $\frac{\sqrt{R}}{\sqrt{G}}$, $\frac{\sqrt{B}}{\sqrt{R}}$, and $\frac{\sqrt{B}}{\sqrt{G}}$ where R, G, B, C represent the red, green, blue and clear channel of the sensor, respectively.

Virtual light channel data were segmented using a sliding window. Different window sizes between 3 and 120 s were investigated while maintaining a fixed overlap of 50 % of the window size. For each virtual light channel 18 time domain features were calculated as listed in Figure 4.5 resulting in a total of 756 features per window. Features were standardised by removing mean and dividing by their standard deviation.

A linear support vector machine (LSVM) was used to classify between screen use and no screen use. PC and TV states described in the protocol were considered as screen use class, thus making detection of screen use independent of the displayed content. After training the LSVM on data from all but one sensors, data from the remaining sensor was used for prediction. Figure 4.5 depicts the data processing pipeline from raw sensor data to classifier output.

4.4.3 Lab study

We evaluated screen use detection using WISEglass in the Lab study. For all participants we used the same viewing distances and devices, but did not control ambient light intensity. Fourteen participants (2 female, 12 male, between 20 and 39 years old) were asked to read a print magazine, watch a documentary, and use a PC for 20 minutes per activity.

Sensing approach

Colour light sensor data were sampled and stored in flash memory with a sampling rate of 6.5 Hz and a gain factor of 1. Data were downloaded after each participant completed the protocol.

Recording protocol

After arrival, the protocol was explained to participants in detail. Prior to recordings participants signed a consent form agreeing to the recording protocol and future use of the data.

Participants would perform the following activities for 20 minutes each: (1) Reading a print article. (2) Watching a documentary about coffee on a 27 inch Samsung SyncMaster P2770HD TV at a 140 cm distance. (3) Browsing the internet using a HP EliteDisplay E241i 24 in screen at a 70 cm distance. The typical ergonomic distance is between 50 and 70 cm, where 50 cm was used in the Test bench analysis. Figure 4.6 depicts the activities as performed by each participant.

Both screens were reset to factory defaults prior to recordings. Participants wore WISEglass during data recordings. Ambient light intensity was measured with a standard lux meter (AMPROBE LM-120) at the beginning of each recording. A total of 15.2 hours of data were recorded during the Lab study.

Figure 4.6: Lab study. Activities reading a print magazine, watching TV, and PC use were performed by participants for 20 minutes per activity.

Evaluation procedure

We applied the same evaluation procedure as described in Section 4.4.2 with the following changes. To reduce the number of features mRMR[69] feature selection was applied. Different numbers of features were evaluated as described in Section 6.5. We used the same LSVM classifier as in the Test bench but applied Leave-One-Participant-Out (LOPO) cross-validation jointly for feature selection and classification.

4.4.4 ADL study

To evaluate if screen use can be distinguished from other typical ADLs we used data from a previous study. In total, nine participants (3 female, 6 male, between 20 and 27 years old) were involved in the study. The ADL dataset was recorded to evaluate activity recognition using the WISEglass inertial measurement unit [99].

Sensing approach

Participants wore WISEglass for the duration of the recordings. Data from the light sensor was sampled and stored in flash memory at 6.5 Hz.

Recording protocol

The study protocol was designed to include many typical ADLs across a regular day. Participants executed a scripted protocol while being followed by an observer labelling the data during recording time using the ACTLog application for Android [85]. In total 66.08 hours of data were recorded. After arrival, the study protocol was explained in detail. Prior to recordings participants signed a written consent form agreeing to the protocol and future use of the data. Participants were compensated for their efforts with a 25 Euro Amazon voucher and three meals during the recordings. Data from two participants were excluded due to hardware issues during recordings leaving seven participants for the ADL analyses. Table 4.2 provides a listing of the performed ADLs, their total recording duration, and the activity cluster they belong to.

To evaluate the detection of screen use, we used the computer work activity as the screen use class. We omitted data of the unlabelled class and from the watching a movie activity, where participants looked at a projector screen. Projectors are different from regular screen use and was not the goal of this work.

Evaluation procedure

For the ADL study, we modified the evaluation procedure described in Section 4.4.2 and Section 4.4.3 as follows.

All results for the ADL dataset were computed using the best 200 features selected by mRMR feature selection. Feature selection was performed on training data only for each validation fold.

The percentage of screen use in the ADL dataset was approximately 11 % of the entire dataset. We applied downsampling to adapt the ratio of screen use per participant from 10 % - 90 % in steps of 10 %. Downsampling was performed by randomly removing windows of 5 s of the desired class. To ensure stability of our results we repeated the selection process 10 times per screen use ratio and averaged the results.

4.5 Results

In this section we first analyse the impact of screen use depending on content and screen type. Subsequently we present screen use evaluation results for Test bench recordings, Lab study, and ADL study.

Table 4.2: ADL study. Activities and their recorded total duration for nine participants.

No	Cluster	Activity	Duration [min:sec]
1	Eat	Breakfast	84:43
		Lunch	156:49
		Dinner	181:53
2	Walk	Lab to Bathroom	17:37
		Lab to gym	36:27
		On treadmill 2km/h	47:06
		Gym to lab	35:56
		Lab to cafeteria	20:11
		Queuing for lunch	6:28
		Picking up lunch	13:20
		Cafeteria to lab	25:28
		Lab to restaurant	58:36
		Restaurant to lab	55:26
3	Brush	Teeth	39:17
4	Stairs	Walking	19:49
5	Jog	On treadmill 5km/h	45:27
6	Cycle	On gym trainer	91:14
7	Read	A book	276:31
		Desk work	273:34
8	Screen	Computer work	253:50
		Watching movie	187:56
9	Cleaning	Vacuuming	45:22
		Wiping tables	45:02

4.5.1 Impact of screen content and devices

For typical PC content, backgrounds are often white, resulting in a majority of pixels emitting light. For TV content, backgrounds are rather dark, resulting in a majority of pixels emitting little to no light. We measured irradiance using the reference spectrometer as described for the Test bench investigation (Section 4.4.2) during PC use and watching TV activities and found average irradiance during PC use to be 3 times higher than for watching TV with $0.03\,mW/cm^2$ for PC use and $0.01\,mW/cm^2$ for watching TV. Thus, when using the same screen for watching TV at 200 cm distance, light exposure is 48 times lower compared to PC use at 50 cm distance.

Table 4.3 shows light intensity at the user's eye for typical devices and content types. According to our results PC use with a desktop setup has the highest light intensity at the user's eyes and was thus investigated further. Light intensity is a primary consideration in this investigation, as intensities above 50 lux have substantial impact on the circadian phase [106]. Overall, watching TV is less critical than PC use due to viewing distance and content type. Screen size was also important. While bright, smartphone screens are too small to critically influence circadian phase considering that light intensities are typically below 50 lux.

Table 4.3: Light intensity at the user's eyes during PC use and watching TV for typical devices. We used screen size, luminance information, and typical viewing distances to calculate light intensity at the user's eyes as described in Section 4.4.1.

Device	Aspect ratio $w{:}h$	Diagonal [cm] l	Area [m²] A	Luminance [cd/m²] L	Luminous intensity [cd] I	Distance [m] d	Light intensity PC [lux] E_{PC}	Light intensity TV [lux] E_{TV}
Galaxy S7	16:9	12.95	0.0072	350	2.51	0.32	24.51	8.17
iPhone 6s	16:9	11.94	0.0061	550	3.35	0.32	32.71	10.90
iPad air 2	4:3	24.64	0.0291	420	12.24	0.5	48.95	16.32
Dell XPS 13	16:9	33.78	0.0488	276	13.46	0.5	53.84	17.95
MacBook Pro 15	16:10	39.12	0.0688	315	21.66	0.5	86.65	28.88
HP monitor	16:10	60.96	0.1670	250	41.75	0.6	115.98	38.66
Samsung TV	16:9	101.60	0.4411	120	52.93	2.0	13.23	4.41

4.5.2 Test bench recordings

Using the extracted feature and LSVM classification, screen use could be detected independent of lamp type and ambient light intensity. Averaging over all lamp types, a receiver operating characteristic area under curve (ROC AUC) of ≈1 was achieved, indicating that screen use could be detected independent of the ambient light type. Table 4.4 summarises the recognition results for the Test bench recordings. For screen use approx. 199 hours (66.6 %) and for non screen use approx. 99.5 hours (33.3 %) of 5 s windows were used.

Table 4.4: Test bench. Results for recordings using different ambient light sources. Results show that screen use can be detected independent of of ambient light type.

Metric	Halogen	LED	Warm CCFL	Cold CCFL	Dark
ROC AUC	0.99	1.0	1.0	1.0	1.0

4.5.3 Lab study

We investigated the impact of window size on screen use detection performance for the Lab study data. For all window sizes ROC AUC was between 0.85 and 0.90. Windows of 5 s produced the largest ROC AUC at the shortest recognition delay. For an online implementation of screen use detection recognition delay is crucial. Thus, we chose for 5 s windows for further analysis.

To analyse the impact of feature reduction on detection performance we applied mRMR feature selection. Figure 4.7 shows the number of features used vs. ROC AUC performance. Using ≈70 or more features increased performance. For 200 features results were stable with an average ROC AUC above 0.9.

Figure 4.8 shows ROC AUC over ambient light intensity. For 10 out of 14 participants, a ROC AUC above 0.9 was achieved. In all 7 samples with an ambient light intensity below 200 lux the ROC AUC was above 0.9. In 3 out of 14 participants ambient light intensity was above 500 lux, which is the recommended light intensity for working environments [32], and ROC AUC dropped below 0.8.

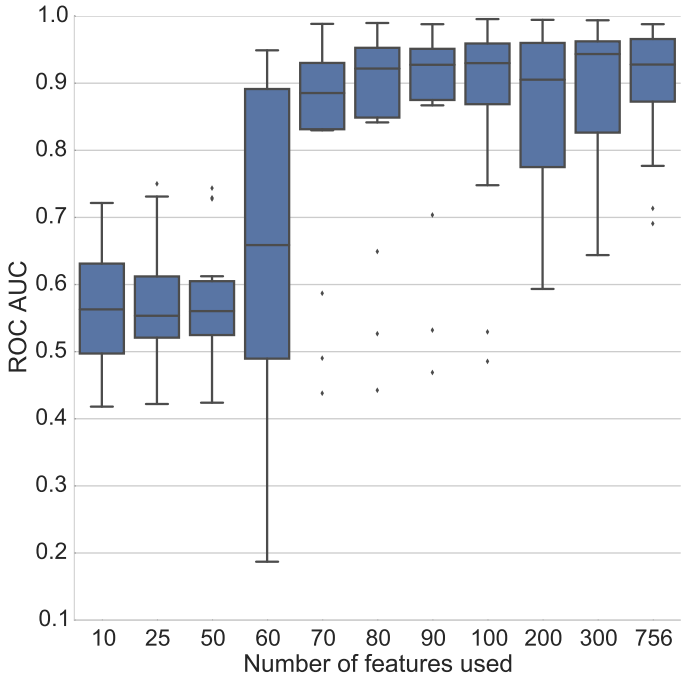


Figure 4.7: Lab study. ROC AUC vs. feature number is plotted where grey diamonds represent outliers. Performance drops when using less than ≈70 features.

4.5.4 ADL study

We evaluated screen use detection within a large variety of ADLs. Figure 4.9a depicts positive predictive value (PPV), true positive rate (TPR), and false positive rate (FPR) for different screen use percentages. It can be seen that PPV increases with screen use percentage from 0.45 to 0.98 while TPR and FPR did not show relevant changes.

Figure 4.9b shows ROC AUC per participant for 30 % screen use. For all participants a ROC AUC was between 0.69 and 0.91 with an average of 0.83. ROC AUC performance was stable per participant.

4.6 Discussion

Today computers are on almost every desk and the amount of work being digitalised is still growing. Further growth of screen use during the day and into the evening may lead to circadian misalignment and an increase of people suffering from eyestrain. We thus expect that screen use above 30 % is reasonable and potentially growing in the future. Our screen use detection results during ADL showed relevant performance with an average ROC AUC of 0.83 at 30 % screen use.

Previous study protocols administered continuous screen use for one to five hours [22, 25, 45, 89] when investigating circadian phase shifts due to screen use. Wood et al. [106] found a measurable melatonin suppression after 2 hours of tablet use. Our proposed detection only required 5 s of data, thus detecting screen use well before a measurable suppression of melatonin. Depending on real time detection requirements, majority voting could be applied over

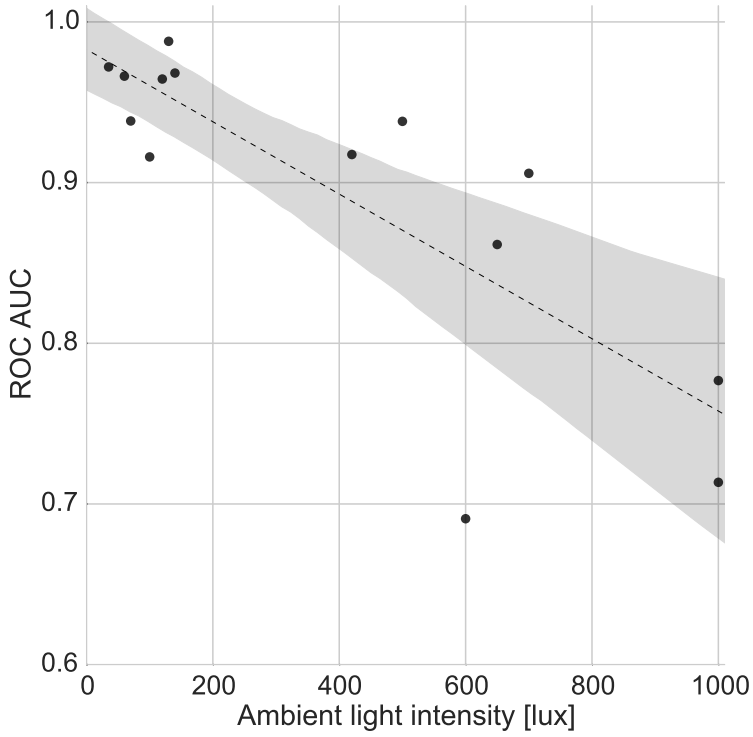


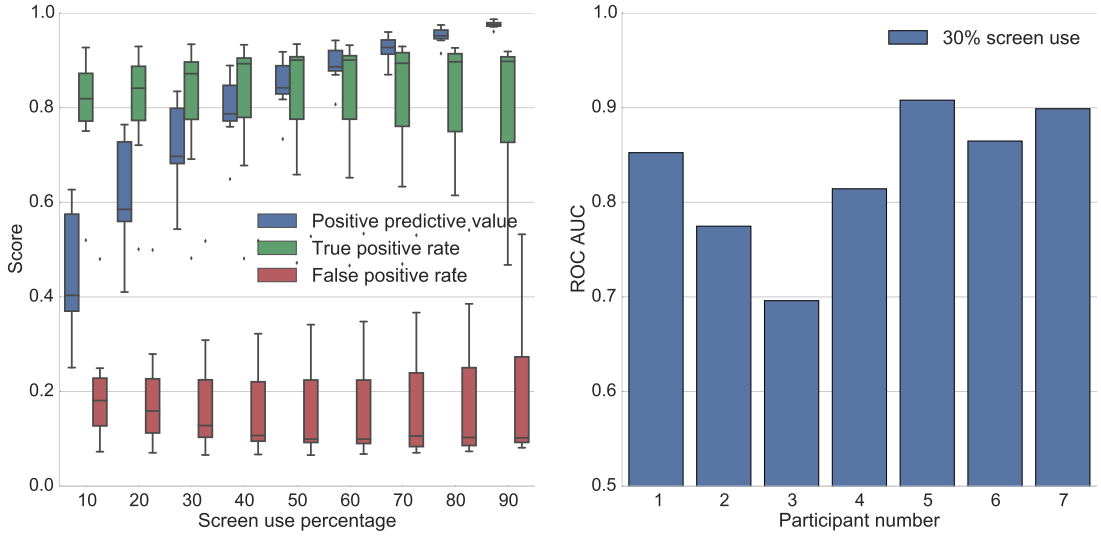
Figure 4.8: Lab study. Recognition performance over ambient light intensity. Each participant is represented by a dot. The dashed trend line shows the relationship between ambient light intensity and ROC AUC. The 95 % confidence interval around the trend line is shaded grey. It can be seen that screen use detection performance decreased when ambient light intensity increased.

multiple windows to smoothen the classifier result.

Circadian phase shift can be induced by light sources other than screens. For example, looking at a high energy blue light source could also shift the circadian phase but may not be detected by our screen use detector. Smart eyeglasses measure light intensity for each spectral component and thus can measure different sources of phase shift. In contrast to other light sources it is possible minimise the risk of phase shift due to screen use by adapting the screen colour profile. Therefore the detection of screen use is of interest.

4.6.1 Performance metrics

For screen use detection a low false alarm rate is crucial as false alarms tend to frustrate users. While ROC AUC is a good performance measure for balanced datasets, we additionally derived precision of the screen use detection (PPV), correctly detected screen use (TPR), and falsely reported screen use (FPR). TPR reports the ratio of correctly detected screen use. FPR describes the ratio of falsely reported screen use. ROC AUC is computed from TPR and FPR only and is thus insensitive to class skew [38]. In unbalanced data, where screen use instances are rare (e.g. for 10% screen use), FPR can be low and TPR can be high, even



(a) PPV, TPR, and FPR vs. screen use percentage.

(b) Per participant ROC AUC scores for 30 % screen use.

Figure 4.9: ADL study. Left: PPV, TPR, and FPR for 10 % - 90 % screen use. Grey diamonds represent outliers. Right: Per participant ROC AUC score at 30 % screen use.

though PPV is low.

4.6.2 Ambient light

Our lab study showed that screen use detection works best for ambient light intensities below 200 lux. Interpreting light emitted by the screen as signal and ambient light as noise, an increase in ambient light adds noise, thus decreases signal to noise ratio. Increased ambient light however eliminates the need for screen use detection. For example, circadian phase shift induced by screen use typically occurs at low ambient light intensity only. Circadian phase response curves express impact of light exposure on circadian phase depending on its timing.

Light intensities span multiple orders of magnitude ranging from 0 - 1.000 lux indoors to more than 100.000 lux outdoors on a sunny day. Our screen detection approach is meant for indoor screen use. When outdoors, the sunlight already influences the circadian phase. ADL recordings included outdoor episodes of walking and our system did not falsely report screen use due to the high brightness values.

Absolute light intensities differ depending on the environment, lamp type, viewing distance, content, time of day, and weather. Virtual light channels based on ratios of the different light colour components made the feature space independent from absolute light intensities. Further calibration of the colour light channels due to their different energy-related response was not needed, since the subsequent pattern analysis weights the features with respect to the classification task.

4.6.3 Screen use activities

In the Test bench study, we combined PC and TV content into one screen use class to ensure screen use could be detected independent of displayed content. For both content types

distance between screen and sensor was 50 cm. The combination of video content and close viewing distance could happen in deployment, e.g. when watching video content online using a notebook. However, our calculations (Table 4.3) showed that TV use, even at a short viewing distance, has no substantial impact on the circadian phase.

Section 4.5.1 reported on light intensity at the user's eyes for different device and content types. Previous research investigated the influence of screen use on circadian phase [22] and found that screen use can induce circadian phase shifts. Wood et al. [106] suggested 50 lux as the critical threshold for the circadian system. Our calculations showed that TVs, while typically being the largest screens used, produce less than 50 lux at the user's eyes due to the large viewing distance. Typical TV content caused one third of the irradiance compared to PC use. Heath et al. [45] found that one hour of tablet use prior to bedtime did not significantly impact circadian phase. While used at a close distance, smartphones and tablets are not large enough to reach the 50 lux threshold. PC use was the only combination of distance, content type, and screen size that could produce a light intensity of over 50 lux at the users eye. In contrast, the work of Chang et al. [25] found significant melatonin suppression after 5 days of using a backlit eReader for 4 hours prior to bedtime at 30 lux only. However, their control was reading a paper in ambient light conditions of less than 1 lux, the amount of ambient light in deep twilight. Such low light intensities might make reading difficult. We chose to investigate PC use with a desktop setup because it has the strongest light intensity at the user's eyes.

4.6.4 Assumptions

We used PC work as the prototype activity for high irradiance screen use activities as we assumed it was the activity with yielding the highest light intensities at the users eyes. Typically during PC work backgrounds are white, e.g. when editing a document or browsing the web, resulting in a majority of bright pixels.

Test bench results showed that screen use was detected well when sensors were mounted in a fixed position independent of the amount or type of artificial light, or the content being displayed on the screen. In reality, people move their head during screen use and sometimes look away from the screen for brief moments. Such movements are hard to annotate during an observational study and thus introduced noise to our screen detection. Head motion was clearly visible in the raw light sensor signal of the ADL study. Additional sensor information, e.g. from a motion sensor could be used to detect head motion and thus interruptions of screen use.

4.6.5 Practical applications

Possible intervention measures have been investigated [45, 89]. The latest update of Apple iOS introduced a software feature to adapt screen colour profiles to minimise unwanted circadian phase shifts. In this work, we showed that detecting screen use is feasible with the light sensor of smart eyeglasses. The screen use detection could be used to control the screen colour profile.

Ambient monitoring methods could detect presence in front of a desktop screen, e.g. camera-based face detection or ultra-sound based proximity detection [53]. However, screen-based light intensity must be assigned to an individual user to implement alerts, where wearable systems are advantageous. In addition with wearable systems privacy concerns could be easily addressed. An ambient monitoring solution requires one setup per screen and presence information alone is not sufficient to identify relevant screen illuminance as users may already

use a software intervention tool, e.g. f.lux to adapt the screen's spectral composition. In addition, a presence detector cannot distinguish between being next to and actually looking at a screen. With the light sensor embedded into the smart eyeglasses bridge, the wearable system captures a wearer's field of view.

We intentionally chose eyeglasses as a sensing platform because regular eyeglasses are worn by many. Smart eyeglasses, like regular eyeglasses, are worn throughout the day. In contrast to head-worn light measurement devices such as the Daysimeter [13], smart eyeglasses may be used for multiple applications [5]. In contrast to other smart eyewear, where the focus is on interaction and displaying information, e.g. Google Glass, our work is focused on sensing. While the camera of Google Glass could be used as a colour light sensor substitute, the larger power consumption may limit continuous use.

Detecting screen use is challenging as the screen-emitted light is distributed over the visible spectrum, with frequency components close to other light sources. Screen use was found to suppress evening rise in endogenous melatonin significantly and thus misaligning circadian rhythms [22]. Screen use can also cause repetitive strain injuries (RSI), such as eye strain [87]. With our screen use detection it is possible to support users by suggesting intervention measures when needed, including adapting a screen's colour profile to prevent circadian phase shifts and taking regular breaks to reduce RSI risks.

Screen use detection can be used beyond detecting impact on the circadian phase. Regular breaks are important to prevent eye strain, the most common repetitive strain injury [87]. Screen use detection can suggest breaks during computer work and remind users to implement the 20-20-20 rule: Every 20 minutes of screen work take a 20 second break and look at something at least 20 feet away.

4.7 Conclusion

We introduced an approach for screen use detection based on a colour light sensor embedded in smart eyeglasses and evaluated it on three studies. Our evaluation showed perfect results for the Test bench analysis. Lab and ADL study results introduced noise due to head motion and ambient light variation.

Lab study results revealed that screen use detection performance is related to ambient light intensity. Screen use was detected with over 0.9 ROC AUC at an ambient light intensity below 200 lux. A dataset of typical ADLs was used to further evaluate screen use detection. Screen use was detected with an average ROC AUC of 0.83 for 30 % screen use. Detection performance was evaluated on person independent models for Lab and ADL datasets.

Our work could be applied to other wearables as few hardware components are required. A colour light sensor, battery, and wireless interface could be embedded into smart jewellery, e.g. a brooch or necklace. However it is essential that the light sensor's field of view is aligned with the wearer's eyes. Our screen use detection algorithm is independent of the sensor position.

The proposed system detected screen use quickly due to the short window size of 5 s. Screen use information could be used to prevent eye strain by reminding users to take regular breaks. Undesired impact of screen use, e.g. circadian phase shift, could be minimised by either notifying the user or (automatically) activating a software intervention measure, e.g. f.lux. We thus conclude that smart eyeglasses are a feasible platform for screen use detection.

5

Sleep timing and chronotype estimation from smartphone context

Chapter originally published as: F. Wahl and O. Amft. Data and Expert Models for Sleep Timing and Chronotype Estimation from Smartphone Context Data and Simulations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3):139:1–139:28, Sept. 2018.

5.1 Introduction

Almost a third of our life is spent sleeping, making sleep one of the most important human activities. Sleep is needed for physical and mental recovery from time spent awake. For adults a duration of sleep of 7 to 9 hours was recommended by the National Sleep Foundation [48]. Getting sufficient sleep quantity and quality is challenging as 29% of adults in the United States sleep less than 7 hours. In a survey of the Centers for Disease Control and Prevention, 28% of 403,981 people reported to not have gotten enough rest or sleep on at least 14 out of 30 days [24].

While some prefer sleeping in, others naturally wake up early. Sleep timing preference is described by chronotype on a scale ranging from early types (larks) to late types (owls). Chronotype is derived from the midsleep points on free and work days and can be assessed using the Munich ChronoType questionnaire [78, 77]. Due to social obligations, many live against their internal (circadian) clock and thus suffer from social jet lag and circadian misalignment. Long-term circadian misalignment poses health risks, e.g. obesity and cancer [75, 56].

With ownership rates of 72% in the US and median of 68% in developed countries [70], smartphones are the most predominant multi-modal sensing platform that continuously accompanies people. With their sensors, smartphones could analyse the environment and continuously estimate the current situation, i.e. context. Thus, instead of wearing activity trackers, smartphones may be used to detect sleep times and replace the current practice of sleep diaries and questionnaires. As our study showed, sleep diaries are challenging to maintain, thus evaluating alternative sleep estimation methods is an important research goal.

Sleep timing estimation using smartphone sensors may be accomplished by data-driven models, i.e. data pattern learning and classification. In the related work section, we review several approaches using smartphones and data models. However, when deployed in real life, the performance of data models often deteriorates due to unforeseen conditions, e.g. smartphone power outages. Clearly, there are many other reasons for lacking or misleading sensor data. For a cyclic process or regular routine, such as the sleep/wake rhythm, it seems adequate to use time of the day as a complementary model to sensor data. However, human routines are affected by shifts, e.g. due to social norms. In this challenge, expert knowledge may support the data models.

Our approach in this paper is to capture expert knowledge in algorithms that can be used as filters and fusion methods applied to the results of the data models. In our approach, sleep opportunity is determined from various smartphone sensors. Using data recordings to evaluate sleep detection is challenging because capturing specific sleep behaviour cannot be elicitated. To evaluate how specific sleep behaviour, e.g. sleep regularity, influences performance, we developed a simulation methodology. Our simulation engine can generate arbitrary sleep schedules and fill in evaluation study data. For the smartphone-based estimation, we require that the phone is kept in the same room with the user. For personalisation, users can provide feedback on the estimated sleep times. In particular, this article provides the following contributions:

1. To estimate sleep timing, we developed data models and a personalisation strategy based on user feedback. To further improve estimation performance, we designed an expert model comprising an abstraction of the human circadian and homeostatic processes. We demonstrate how the expert model can be used to filter and fuse data model results.

2. To investigate specific sleep patterns on a larger data set, we developed a simulation

engine to generate arbitrary sleep patterns based on our recorded data set. Using the simulation methodology we generated a total of 98280 days of data with different sleep regularity and social jet lag patterns.

3. We compare the data and expert models on recorded and simulated data sets. Our results show that the fusion of data and expert models (*Time+Sensor+TPM*) outperforms data models (*Time+Sensor*). Absolute errors for estimated sleep onset were 40±48 min and 76±108 min, respectively and 42±57 min and 101±153 min for wake up.

4. We present a personalisation approach based on user feedback. After estimating sleep opportunity, users have the option to confirm or adjust detected sleep timing. We used feedback to perform retraining on our estimation algorithm. Our results showed an average improvement of 59% for sleep onset and 57% for wake up using user feedback from just two days.

5.2 Related Work

This section provides an overview of different approaches to sleep timing estimation in free living, based on wearable and ambient devices, and smartphones. Table 6.1 provides a summary of key figures from our analysis. In addition, insight from selected literature combining data models and expert models is discussed.

5.2.1 Wearable Sleep Detection

Fleury et al. [41] combined data from an actimeter and smart home sensors to detect seven different groups of activities of daily living including sleeping. Sleeping was detected with 93.9% accuracy using a support vector machine with a Gaussian kernel and leave-one-participant-out (LOPO) cross validation in data recorded from 13 participants. Borazio and van Laerhoven [17] used a wrist worn accelerometer to detect sleep onset and wake up times in a data set of 141 consecutive days from a single participant. Their threshold model detected sleep onset and wake up times with 83% and 85% accuracy, respectively. Borazio and van Laerhoven [18] used a wrist worn device to detect sleep episodes, sleep posture changes, and myoclonic twitches during sleep. A hidden Markov model detected sleep episodes from time, light intensity, and accelerometer data of eight participants. In across subject evaluation sleep episodes were detected with 92% precision and 94% recall on average. Infrared cameras were used to determine ground truth information. In contrast, our work goes beyond reporting instance-based sleep detection performance. In this work, we estimated sleep timing and investigated benefits of enriching data models with expert knowledge. Rofouei et al. [79] integrated a pulse oximetry sensor, microphone, and accelerometer into a neck-cuff system to monitor sleep. In a study with three participants they performed sleep apnea detection. In contrast, we use data collected from smartphones only. Our application detects sleep opportunity without requiring any special user behaviour, such as placing the smartphone on the mattress during sleep. A smartphone-only approach could avoid compliance issues that may arise with additional devices.

Table 5.1: Summary of related approaches and results. Cross validation approaches are noted in the CV column as leave-one-participant-out (LOPO), leave-one-day-out (LODO), N-fold over the entire data set, and 3-fold within-subject (WS). Support vector machines (SVM), Threshold models, Hidden Markov Models (HMM), convolutional (CNN) and recurrent (RNN) neural networks, linear models, rule-based models, Bayesian networks, and random forests (RF) were used to model sleep. Sleep and wake were estimated as sleep/wake classification, onset/wake timing, sleep stages, duration of sleep (DS), midsleep timing (MS), and wake up timing (WU). Results metrics were Accuracy (Acc), precision and recall (P/R), median error (ME), average error (AE), average midsleep error (MSe), and mean median absolute deviation (mMAD).

Work	Hardware	N	Days	CV	Model	Estimate	Metric	Result
Wearable sleep detection								
Fleury [41]	Wrist-worn + bedroom	13	0.46	LOPO	SVM Gaussian	Sleep/Wake	Acc	94%
Borazio [17]	Wrist-worn	1	141	LODO	Threshold	Onset/Wake	Acc	83/85%
Borazio [18]	Wrist-worn	8	184	LOPO	HMM	Sleep/Wake	P/R	92/94%
Rofouei [79]	Neck-worn	3	3	-	-	-	-	-
Environmental sleep detection								
Kay [57]	Bedroom	4	68	-	-	-	-	-
Rahman [73]	Bedroom	8	16	LOPO	Threshold	Sleep/Wake	P/R	89/89%
Zhao [108]	Bedroom	25	100	4-fold	CNN + RNN	Sleep stage	Acc	80%
Smartphone sleep detection								
Nandakumar [66]	Smartphone	37	298	-	Threshold	DS	ME	27 min
Chen [27]	Smartphone	8	56	5-fold	Linear	DS	AE	±42 min
Abdullah [2]	Smartphone	9	625	-	Rule-based	MS	MSe	±24 min
Min [64]	Smartphone	27	30	LOPO	Bayesian Net	SO/WU/DS	AE	44/42/64 min
Saeb [81]	Smartphone	207	10649	3-fold WS	RF + HMM	SO/WU/DS	mMAD	38/36/55 min

5.2.2 Ambient Sleep Detection

Kay et al. [57] developed a system to measure sleep environment and detect sleep episodes. The system comprised an infrared camera, two motion detectors, two light sensors, a microphone, a temperature sensor, and a touchscreen interface for controls. System feasibility was evaluated in a study with four participants and found that privacy was a key issue when deploying sensors in the user's bedrooms. Our proposed smartphone app was designed with user privacy in mind. We did not use cameras and sample microphone amplitude only once in 2 min. In addition, users can pause recordings through a button in the main menu.

Rahman et al. [73] used a bedside radar transceiver to detect sleep, sleep apnea, and sleep stages. In an evaluation study 8 participants were recorded for two nights each. A random forest classifier detected sleep with 89% precision and 89% recall in LOPO cross validation. Zhao et al. [108] used convolutional and recurrent neural networks to detect sleep based on reflections of radio frequency signals emitted by hardware installed in the bedroom. They distinguished awake, deep, light, and REM sleep with an average accuracy of 80% in 4-fold cross validation. Sleep stages and apnea can be monitored using sensors in bedrooms. In contrast, our smartphone-based approach uses existing hardware to estimate sleep timing.

5.2.3 Smartphone-based Sleep Detection

Nandakumar et al. [66] developed an application to emit audible sonar waves and sense their reflections to detect sleep, apnea, and breathing rate. They collected 298 days of data from 37 participants and detected sleep duration with a median error of 27 min. Our smartphone application detects sleep opportunity rather than focusing on distinguishing sleep/wake while in bed. Furthermore, our application does not emit audible sounds.

Chen et al. [27] developed a model to detect sleep using smartphone sensors only. Sleep detection was performed using information about whether the phone was in the dark, locked, turned off, charging, stationary, or in silent mode. In a study with eight participants they estimated duration of sleep with an average error of 42 min in 5-fold cross validation. In contrast, we used participant-independent classification models. We excluded light sensor information, as smartphone light sensors are often occluded, e.g. due to flip cases.

Choe et al. [28] conducted a literature review about sleep aiding devices and applications in combination with a survey of 230 participants investigating their sleep habits. The survey found 27% of participants to have irregular sleep habits. When asked if interested in using technology to support them with sleeping, 62% of participants answered *Yes* or *Maybe*. To account for dynamicity of sleep behaviour, our approach uses personalisation to retrain classification models on a daily basis through user feedback.

Abdullah et al. [2] inferred sleep episodes from smartphone usage patterns. Using screen on-off patterns, their rule-based algorithm searches for the longest period of non-usage starting between 10PM and 7AM. A personalised offset is computed using the first two weeks of ground truth to minimise the error. In an evaluation study with nine participants between 19 and 25 years who reported to use their phone within the first 10 min after waking up, the midsleep point was estimated with an average error of 23.8 min. In contrast to requiring two weeks of reference information upfront, our classification models dynamically adapt to user feedback over time. Our approach can detect sleep opportunity at any time of day instead of allowing sleep episodes to start within a predefined time window. Furthermore, our work used simulations to show how the proposed method can detect even irregular sleep, e.g. social jet

lag.

Min et al. [64] used smartphone state and sensor data to detect sleep episodes and quality. Alarm app use, phone charging state, screen status, accelerometer, sound amplitude, and light intensity data were collected for one month from 27 participants. Using participant-specific models and a Bayesian network classifier, sleep onset, wake up, and duration of sleep were estimated with an average relative error of ±35, ±31, and ±49 min, respectively. Saeb et al. [81] recorded smartphone sensor data for sleep monitoring in a study with 207 participants for 6 weeks amounting to 10649 nights for analysis. A random forest was used to predict sleep/wake state. Sleep timing was estimated using a hidden Markov model (HMM) with 10 min resolution. Sleep onset and wake up times were estimated with an mean median absolute deviation (mMAD) of 38 min and 36 min, respectively. In contrast, we extended our participant recordings with simulations of 98280 nights to evaluate performance depending on sleep parameters, including sleep regularity and social jet lag. In addition, we propose a model personalisation method, which adapts to users by learning from their feedback.

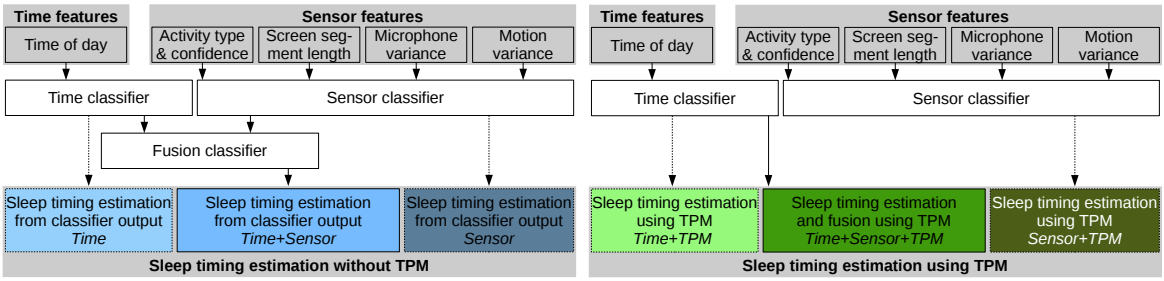
5.2.4 Use of Expert Knowledge

In recent years, data mining and machine learning algorithms have advanced considerably. Effort was mainly put into improving algorithms, rather than on incorporating domain expert knowledge. Fayyad et al. [39] found the inclusion of domain knowledge important for extracting useful insight from large volumes of data. In this work, we compared sleep timing estimation results of a data model and an expert model, which incorporates knowledge about the human homeostasis and circadian processes. Solomatine and Ostfeld [84] compared the advantages and disadvantages of data models for river basin management. They describe the inclusion of human expert knowledge as an important challenge of the modelling process. Aside from improving model accuracy, they found that feature selection could benefit from expert knowledge.

Bohanec and Delibašic [15] compared an expert-only model to a combined expert and data model for predicting skiing injury risks. They found a performance improvement of 10-15% when combining a classification model with expert knowledge. Azkune et al. [8] used expert knowledge as a starting point for incremental data-driven learning. Assuming expert models are incomplete, they incrementally enhanced their prediction performance using data-driven learning. Flouvat et al. [42] used expert models to reduce computational effort in pattern mining and improve mining results. They used expert models to select the most interesting soil erosion patterns. In our work, expert knowledge was used to filter classification results and as a fusion method to combine different data classification models. We compared the benefit of adding expert knowledge to enrich a data model and found substantial performance improvements over inferring sleep timing from data only.

5.3 Sleep Estimation Methods

In this work, we compared a data-driven models for sleep detection with a combined data and expert model, which incorporated expert knowledge about sleep timing in humans. Figure 5.1a shows the pure data model. Time of day was used to compute sleeping probability based on time of day only. Sensor features were used by the sensor classifier to predict sleeping probability from context information. For comparison we computed sleep timing estimations based on time (*Time*) or sensor (*Sensor*) data only. Time and sensor classifier outputs were



(a) Data models. (b) Data-expert combination models.

Figure 5.1: Overview of sleep timing estimation methods. In data-driven models sleep timing was extracted from classifier output directly. For *Time+Sensor* an additional classifier was derived to perform fusion of time and sensor classifier outputs. Expert models incorporated expert knowledge into the sleep timing estimation process using the two process model (TPM). In *Time+Sensor+TPM* time and sensor classifier outputs were fused with the TPM.

combined by the fusion classifier to obtain the fusion data model (*Time+Sensor*). Figure 5.1b depicts the expert model. Outputs of time and sensor classifiers were used by the expert model (*Time+Sensor+TPM*) to estimate sleep timing. *Time+Sensor+TPM* extended the pure data model using the two process model (TPM) developed by Daan et al. [31] to include expert knowledge about sleep/wake timing in humans. We investigated fusion approaches of the expert model with time and sensor classifier outputs separately (*Time+TPM* and *Sensor+TPM*) and in combination (*Time+Sensor+TPM*).

5.3.1 Sleep Opportunity Detection

Three classifiers were used to detect sleep opportunity. The time classifier used time of day as the only feature and the sensor classifier used smartphone sensor data. Time and sensor classifier outputs were fused by the fusion classifier for *Time+Sensor* and TPM fusion for *Time+Sensor+TPM*.

Preprocessing

Data were split into days, where each day starts at 6 pm in the evening. Around 6 pm was the time of the day where the least number of participants slept according to a smartphone-based survey of 8070 persons around the globe [103]. Days on which the standard deviation of the accelerometer x-axis samples was zero were removed from the data set. Data were resampled to 5 min intervals before feature extraction to match the sleep diary resolution.

Time Feature Extraction

For time-based sleep opportunity classification only time of day was used. Time of day t was transformed into two continuous features t_1 and t_2 as described in Eq. (6.2).

$$t_1 = \begin{cases} 1, & \text{when } t > 12.0 \\ 0, & \text{when } t \leq 12.0 \end{cases} \quad \text{and} \quad t_2 = \begin{cases} 24 - t, & \text{when } t > 12.0 \\ t, & \text{when } t \leq 12.0 \end{cases} \tag{5.1}$$

Sensor Feature Extraction and Selection

Initially 5 min averages of the following sensors were derived and considered as potential classification features: System status, WiFi connection, air plane mode, silent mode, do not disturb mode, proximity sensor, light sensor, step detector information, GPS information, charging status, activity and activity confidence from the Android ActivityRecognitionAPI, screen state, microphone amplitude, and accelerometer. GPS information was discarded due to privacy concerns. We observed that charging often took place during sleep episodes. However, charging usually started before sleep onset or ended after wake up, thus did not support classification and was omitted.

Activities were ordered descending by motion intensity, e.g. the still activity was represented by the highest number. Time between screen activation events was used to compute inactivity segment length. Microphone amplitude was used to incorporate ambient noise levels. Variance of accelerometer x-axis samples was computed.

Feature selection was performed by computing ANOVA F-scores between each feature and labels. Feature rankings led to removal of features representing system status, WiFi connection, air plane mode, silent mode, do not disturb mode, proximity sensor, light sensor, and step detector information due to low ANOVA F-scores. All removed features had ANOVA F-scores below 550. Selected features had ANOVA F-scores between 1207 and 11249. The surprisingly low F-score (101) of the light sensor can be explained by the frequent sensor occlusion. In many participants, smartphone light sensors were found to rarely deliver usable data due phone storage, e.g. flip cases. Consequently, the light sensor information was omitted.

Detected activity and confidence from the Android ActivityRecognitionAPI were among the most important features. After feature selection, sensor-based sleep opportunity classification used five features extracted from smartphone sensors. Figure 5.2 depicts selected sensor features for four days of recorded data.

Classification

The extreme gradient boosting classifier from the XGBoost library [26] was used for classification. XGBoost uses ensembles of CART classifiers and applies boosting through additive training to increase importance of misclassified samples. We chose a learning rate of 0.3 to prevent over-fitting. Each tree was limited to a height of six. Minimum loss was set to zero. Samples of the sleeping class were weighted with a factor of two over samples of the wake class to compensate for the sleep/wake ratio of approximately 1:2. Classifiers reported probability outputs per class. As per-class probabilities for binary problems always add up to one, we used only the sleeping probability, i.e. the probability of belonging to the sleeping class.

Personalisation

Sleep timing and environment were dynamic, even within participants. To continuously adjust to changes in sleep behaviour we applied model personalisation. Starting from a generalised, participant-independent classification model, sleep timing of past days was used to personalise the classification model by retraining the model after each day. For example, predictions of the first day were based on a generalised model only, i.e. a model trained on data from other participants only. Predictions of the second day used a model trained on data from the first day of that participant and the data of all other participants.

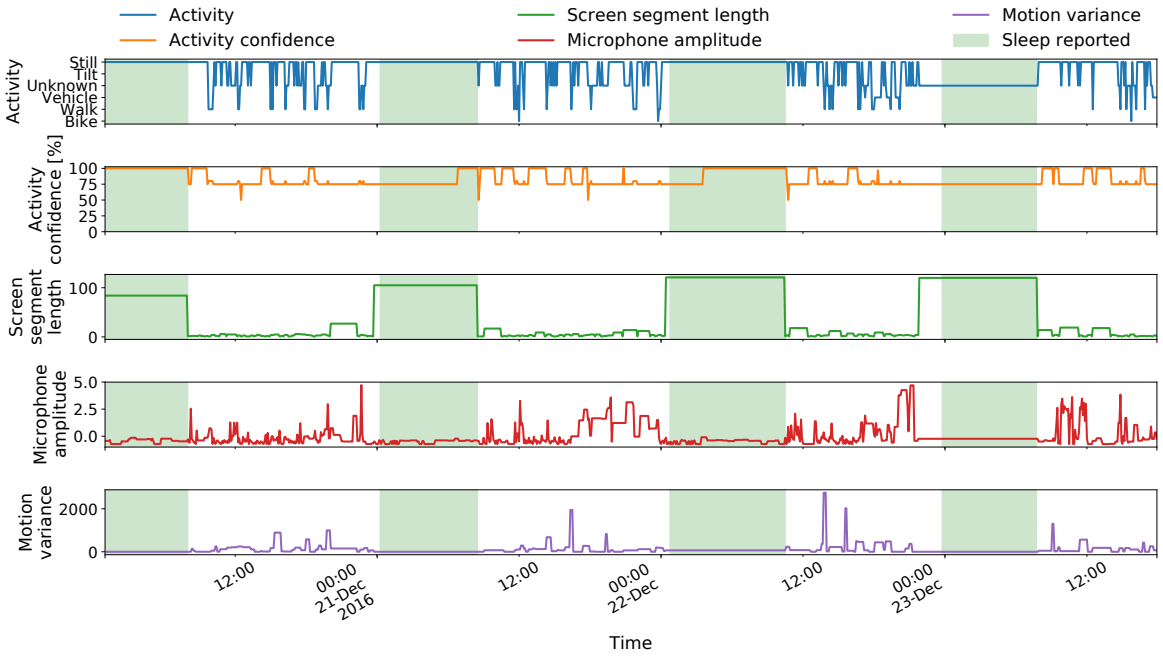


Figure 5.2: Example of selected sensor features for four days of recorded data. Activity type and confidence were obtained from the ActivityRecognitionAPI. Screen segment length was computed as the time between screen activation events. Microphone amplitude was used to incorporate ambient noise levels. Variance of accelerometer x-axis samples was used as a measure of motion intensity.

In this work, we used the sleep diary information of previous days as ground truth. In deployment, responses to daily notifications of the Chronify application may be used. Users could be asked to confirm or correct the detected sleep timing information. Through personalisation user-specific differences were incorporated into the classification model over time, e.g. ambient noise levels.

5.3.2 Data Model Sleep Timing Estimation

In data models sleep onset and wake up were estimated from the time and sensor classifier outputs by converting classifier probability p into a crisp classification result y using a threshold as described in Eq. (5.2).

$$y = \begin{cases} 1, & \text{when } p > 0.5 \\ 0, & \text{when } p \leq 0.5 \end{cases} \tag{5.2}$$

Sleep onset SO and wake up WU times were estimated from the first derivative $\dot{y} = \frac{dy}{dt}$. SO was estimated as the time of the first occurrence where $\dot{y} = 1$ within each day. WU was estimated as the time of the last occurrence where $\dot{y} = -1$ within each day.

Figure 5.3 depicts an example of the detection results for *Time* and *Sensor* models. Personalisation effects are clearly visible in the slopes of p_{time}. On the first day p_{time} was based on the generalised model and thus appears blurry and bell shaped. The slopes of p_{time} become steeper with each additional day of personalised training data.

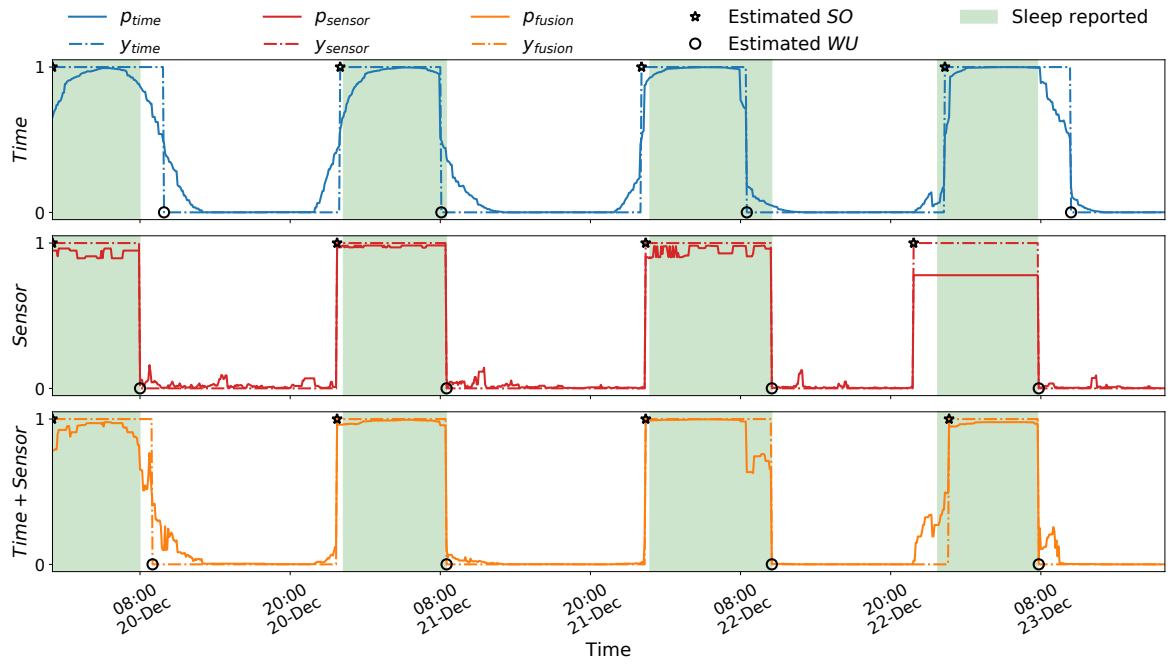


Figure 5.3: Example of *Time+Sensor* sleep timing estimation for four days of data. Each sleep probability $p_\star$ for time, sensor, and fusion classifiers were first binarised into a classification result y using Eq. (5.2). Subsequently, sleep onset time SO and wake up time WU were estimated.

For the first three days in Figure 5.3, *Sensor* achieved good results. On the fourth day *Sensor* detected sleep onset too early (due to no screen activation events as shown in Figure 5.2), while *Time* estimated sleep onset well. As shown in this example, fusion of *Time* and *Sensor* can improve results.

Data Model Fusion Classifier

An additional XGBoost classifier was used to perform classifier fusion for the *Time+Sensor* model. The fusion classifier used time and sensor sleep probabilities p_{time} and p_{sensor} as feature inputs and was trained and personalised as described in Sections 5.3.1 and 5.5.2. Resulting sleep probabilities were binarised as described in Eq. (5.2).

5.3.3 TPM Sleep Timing Estimation

We enriched classification results with expert knowledge about sleep/wake behaviour to reduce sleep timing errors. The TPM by Daan et al. [31] captures expert knowledge about sleep and recovery in humans. TPM consists of a circadian process C and a homeostatic process S to estimate sleep onset and wake up times. S and C are dimensionless and express homeostatic and circadian sleep pressure, respectively. As depicted in Figure 5.4, C oscillates and is shifted to the upper circadian bound by adding H_m and the lower circadian bound by adding L_m. Sleep/wake transitions occurred when circadian and homeostatic processes meet as described in Eq. (5.3). Model formulas and parameters were adapted from Borbély and Achman [19].

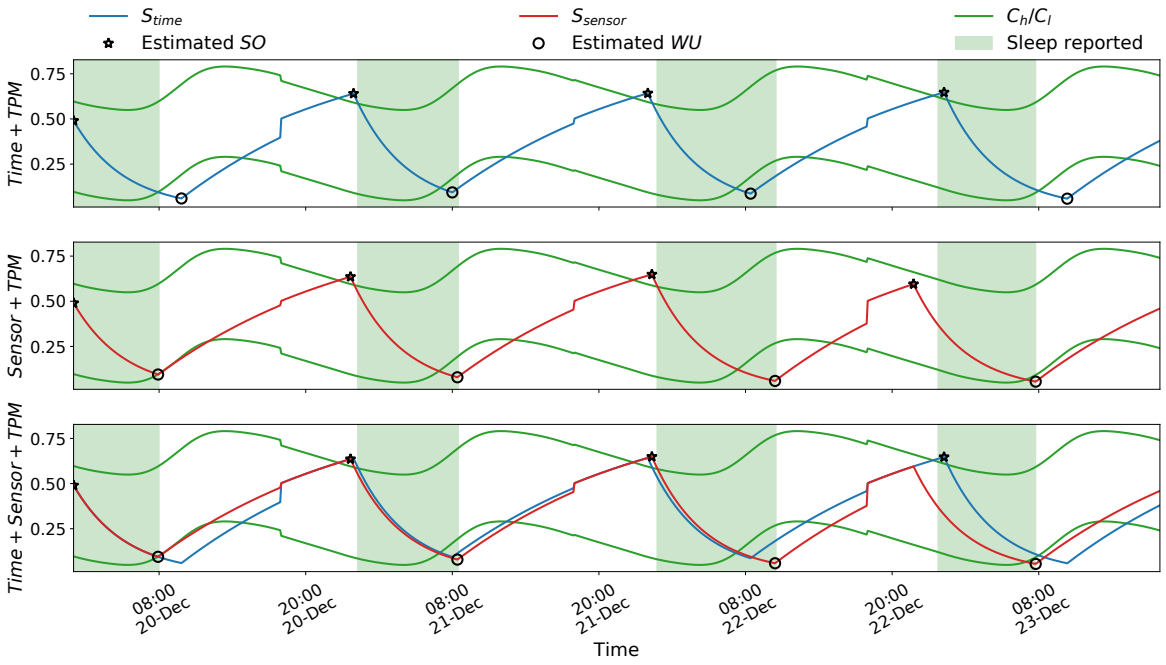


Figure 5.4: Example of TPM sleep timing estimation for four days of data. Sleep onset was detected at peaks of homeostatic sleep pressure above the upper circadian bound H_m and wake up for valleys below the lower circadian bound L_m. Homeostatic processes S_{time} and S_{sensor} were computed from the corresponding classification results. S and C are dimensionless and express homeostatic and circadian sleep pressure, respectively.

Our Python TPM implementation is available online[1].

$$\text{sleep initiation: } S > H_m + C; \qquad \text{sleep termination: } S < L_m + C \tag{5.3}$$

Circadian Process Model

Internal time of the circadian clock was modelled by C. The circadian clock is entrained through external cues, e.g. environmental light. According to Daan et al. [31] C was modelled as an addition of multiple fitted sinusoid functions as shown in Eq. (5.4), where $\omega = \frac{2\pi}{24h}$.

$$\begin{aligned} C = A \cdot [&0.97 \sin(\omega(t - t_0)) + \\ &0.22 \sin(2\omega(t - t_0)) + \\ &0.07 \sin(3\omega(t - t_0)) + \\ &0.03 \sin(4\omega(t - t_0)) + \\ &0.001 \sin(5\omega(t - t_0))] \end{aligned} \tag{5.4}$$

[1]Source code is available at `https://gitlab.com/fwahl/twoprocessmodel`. The Python package can be installed from PyPI by running `pip install twoprocessmodel`.

We parametrised the circadian process separately for each day. Sleep onset and wake up times of the previous day time classifier results y_{time} were used to compute t_0 as $t_0 = SO_{time} + (WU_{time} - SO_{time}) - 20$. Circadian amplitude A was set to $A = 0.1$ as it produced best results in a parameter sweep from 0.05 to 0.2 using steps of 0.01. Offset values were set to $H_m = 0.67$ and $L_m = 0.17$ following the recommendation of Borbély and Achman [19].

Homeostatic Process Model

Sleep pressure was modelled by the homeostatic process. The homeostat works similar to a thermostat: Sleep pressure pressure rises exponentially when awake and decays exponentially during sleep. Sleep pressure S was modelled as shown in Eq. (5.5) where Δt denotes rate at which the model was sampled in hours (here $\Delta t = \frac{5}{60}$ hours). Initial sleep pressure S_{init} was set to $S_{init} = 0.5$.

$$S_t = \begin{cases} S_{t-1} \cdot e^{-\frac{\Delta t}{4.2h}}, & \text{when sleeping at t} \\ 1 - (1 - S_{t-1}) \cdot e^{-\frac{\Delta t}{18.2h}}, & \text{when awake at t} \end{cases} \tag{5.5}$$

Homeostats were modelled using classification results y as sleep/wake information for sleep pressure S: For *Time+TPM* we used y_{time} to compute S_{time} and for *Sensor+TPM* we used y_{sensor} to compute S_{sensor}. In *Time+Sensor+TPM*, S_{time} and S_{sensor} were used.

TPM Estimation

Figure 5.4 shows the circadian and homeostatic processes for the data shown in Figure 5.2. Expert model sleep timing estimation was identical for *Time+TPM* and *Sensor+TPM* approaches.

TPM estimation of sleep onset time SO was performed using homeostatic process output S, circadian process output C, and upper circadian bound H_m. First, C_h was computed as shown in Eq. (5.6). Subsequently peak detection was performed on S using the BMC toolbox [34] to obtain a list of peaks P_a. Next, peak candidates were determined. Each peak p in P_a was added to the list of peak candidates P_c if S was at or above C_h, as described in Eq. (5.7), and p occurred before or at the minimum of C_h, as described in Eq. (5.8). If $P_c = \emptyset$ after the first search, the peak candidate search was repeated with relaxed requirements. In order to be added to P_c in the second search, only Eq. (5.8) needed to hold for p. SO was estimated as $SO = t(p_0)$, the time of the first peak p_0 in P_c.

$$C_h = C + H_m \tag{5.6}$$

$$S(p) \geq C_h(p) \tag{5.7}$$

$$t(p) \leq t(\min(C_h)) \tag{5.8}$$

TPM estimation of wake up time WU was performed in analogous fashion. First, C_l was computed as described in Eq. (5.9). Instead of detecting peaks, valleys must be detected, thus peak detection was performed on $-S$ to obtain a list of valleys V_a. Subsequently, valley candidates were determined. Each valley v in V_a was added to the list of valley candidates V_c if S was at or below C_l, as described in Eq. (5.10), and v occurred before or at the maximum of C_l, as described in Eq. (5.11). If $V_c = \emptyset$ after the first search, the valley candidate search was repeated with relaxed requirements. In order to be added to V_c in the second search,

only Eq. (5.11) needed to hold for v. WU was estimated as $WU = t(v_0)$, the time at the first valley v_0 in V_c.

$$C_l = C + L_m \tag{5.9}$$
$$S(v) \leq C_l(v) \tag{5.10}$$
$$t(v) \leq t(\max(C_l)) \tag{5.11}$$

All TPM methods used up to two peak/valley candidate searches. While in the first round, two requirements had to be met to be added to the candidate list, only one is required in the second, optional round. Relaxed requirements of the second search, which was only performed if the first round failed, helped to find peaks, which lied between C_l and C_h. In the original TPM, sleep/wake state transitions took place when sleep pressure S reached the upper circadian bound C_h. Since we used classifier outputs to compute S, we could not influence direction changes of S as Daan et al. [31] did in their simulations. Therefore, not all transition changes occurred above/at C_h or below/at C_l. Relaxed peak candidate requirements in a second peak search were needed to recover those peaks. An example is shown in Figure 5.4 for sleep onset on the last day of data of *Sensor+TPM*.

TPM Fusion

Time+Sensor+TPM handled information fusion and refinement of classifier results internally. Time and sensor classifier outputs were fused while priority was given to sensor classifier output because it is based on actual measurements. In contrast, time classifier results were based on sleep timing of other persons or previous nights of the same individual and thus invariant to behaviour.

Time+Sensor+TPM estimated sleep onset SO and wake up WU times from homeostatic process outputs S_{sensor} and S_{time}, circadian process output C, upper and lower circadian bounds H_m and L_m. First, C_h and C_l were computed as shown in Eq. (5.6) and (5.9). Subsequently, peak detection was performed on S_{sensor} to obtain a list of all sensor peaks P_s. Each peak p in P_s was added to the list of peak candidates P_c if Eq. (5.7) and Eq. (5.8) were true. If $P_c = \emptyset$ afterwards, peak detection on S_{time} was performed to obtain a list of all time peaks P_t. Each peak p in P_t was added to the list of peak candidates P_c if Eq. (5.7) and Eq. (5.8) were true. If $P_c = \emptyset$ after the initial search through P_s and P_t, the peak candidate search was repeated with relaxed requirements as described above: First, P_s and P_t were concatenated to a list of all peaks P_a and sorted in descending order of time. Subsequently, each peak p in P_a was added to P_c if Eq. (5.8) held true. SO was estimated as $SO = t(p_0)$, the time at the first peak p_0 in P_c. WU estimation was performed in analogous fashion.

5.4 Sleep Simulation Methods

A simulation engine was developed to evaluate our proposed sleep estimation methods across a wide range of sleep patterns. Sleep schedules generated by the simulator were defined by a set of simulation parameters. Recording specific sleep behaviour, e.g. highly regular sleep, in sufficient quantities in a participant study is difficult as many factors determine when sleep occurs, e.g. outdoor noise. The simulation engine performed two steps as shown in Figure 5.5.

First, arbitrary sleep schedules were generated. Subsequently, recorded data was resampled to fill generated sleep schedules with feature data from the evaluation study.

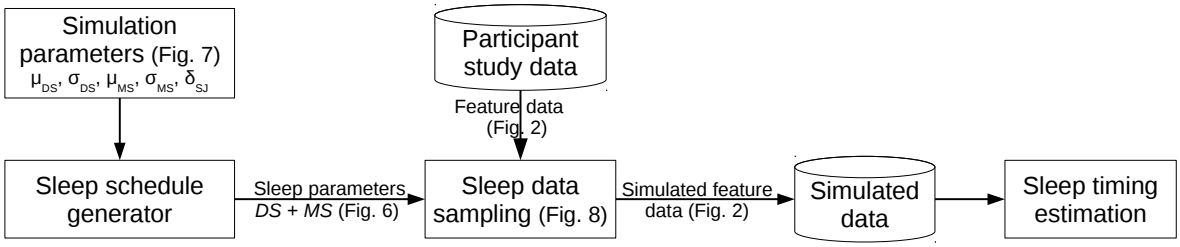


Figure 5.5: Sleep simulation engine. Simulation of sleep data was performed in two steps: First, sleep schedules were generated based on simulation parameters as described in Section 5.4.1. Subsequently, generated sleep schedules were filled with feature data by resampling participant study data as described in Section 5.4.2. Sleep parameters are shown in Figure 5.6. Influence of simulation parameters on sleep timing is illustrated in Figure 5.7.

5.4.1 Sleep Schedule Generation

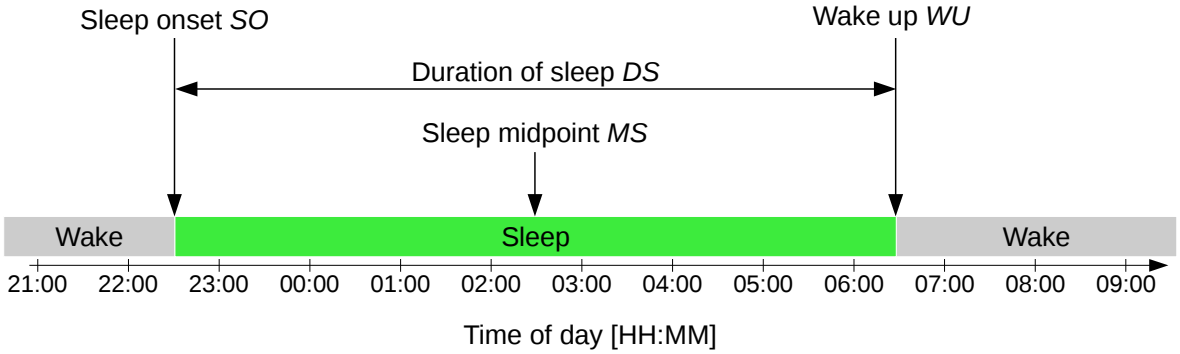


Figure 5.6: Sleep/wake model, comprising sleep onset time SO, wake up time WU, duration of sleep DS and sleep midpoint MS. between SO and WU marked the midpoint of a sleep episode. We used DS and MS to describe and simulate sleep episodes.

We modelled sleep through sleep onset time SO and wake up time WU resulting in the duration of sleep DS. The sleep midpoint MS between SO and WU marked the midpoint of a sleep episode. A sleep episode can be completely described by either SO and WU or DS and MS. Our simulation engine used DS and MS to describe sleep episodes. Figure 5.6 shows sleep parameters of a sleep episode. Sleep parameters duration of sleep DS and sleep midpoint MS were sampled from normal distributions as described in Eq. (5.12).

$$DS \sim \mathcal{N}(\mu_{DS}, \sigma^2_{DS}) \qquad \text{and} \qquad MS \sim \mathcal{N}(\mu_{MS}, \sigma^2_{MS}) \tag{5.12}$$

Simulated sleep patterns were characterised by the simulation parameters. μ_{DS} controlled average duration of sleep and σ_{DS} defined duration of sleep regularity. Chronotype was

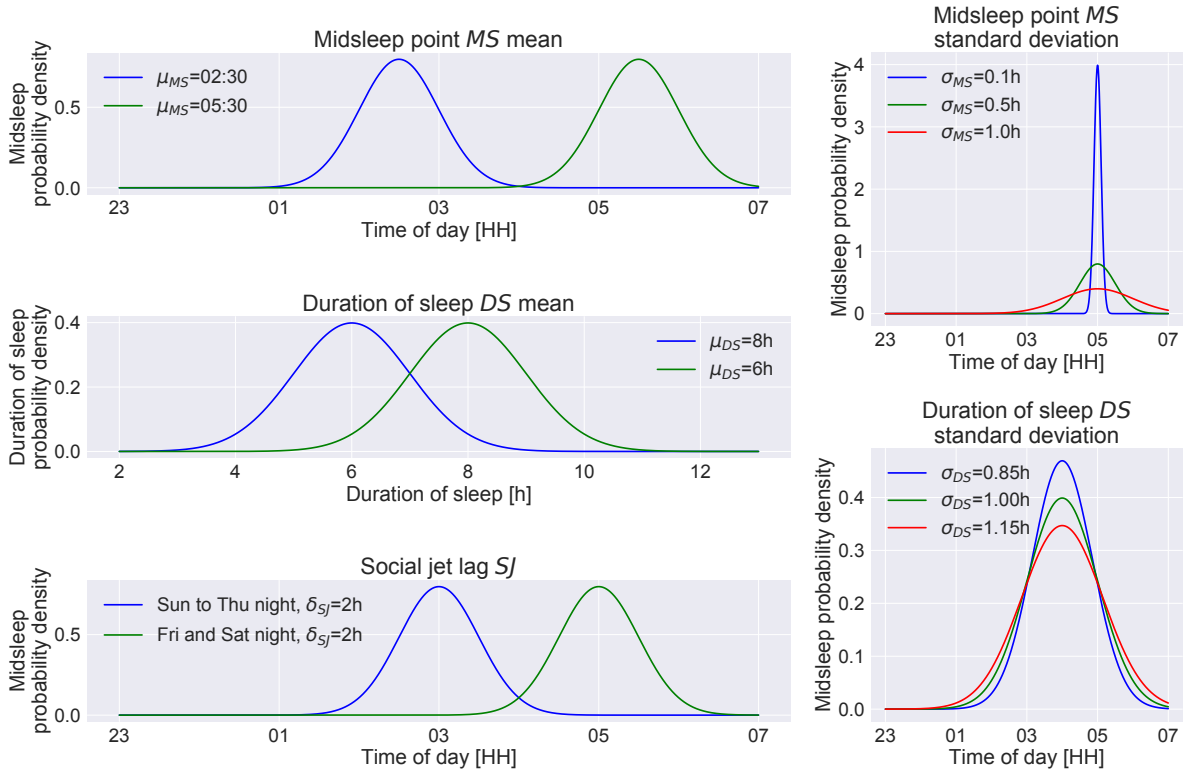


Figure 5.7: Impact of simulation parameters on sleep patterns: μ_{MS} defines chronotype, μ_{DS} regulates the mean duration of sleep, δ_{SJ} adds social jet lag by shifting sleep midpoints on weekends, and σ_{MS}, σ_{DS} control the regularity of sleep timing.

controlled by μ_{MS}. Sleep timing regularity was expressed by σ_{MS}. Social jet lag SJ was controlled with δ_{SJ}. SJ occurs when social and internal times are not aligned, e.g. on weekends. SJ was simulated by adding δ_{SJ} to MS on Saturdays and Sundays. Figure 5.7 illustrates the influence of each simulation parameter.

5.4.2 Sleep Data Sampling

Daily schedules were simulated by resampling and randomly shuffling sensor data segments from the evaluation study data set. The resampling procedure was the following: First, recorded data was split into data segments. A segment was defined as a consecutive number of samples during which the first derivative of all feature values was zero. Subsequently, simulated schedules were split into days, where each day started at 6 pm. Finally, each simulated day was filled data by randomly sampling data segments of evaluation study data without replacement and separately for wake and sleep periods according to the ground truth labelling. Figure 5.8 shows an example result of sleep data segment resampling and shuffling procedure.

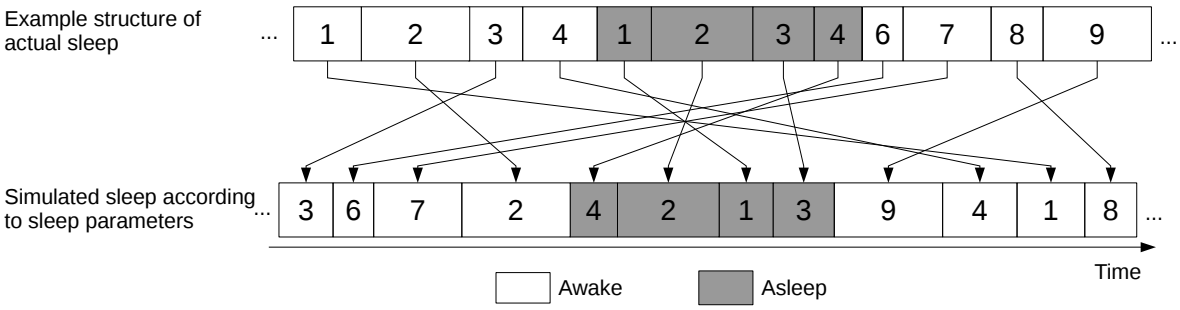


Figure 5.8: Illustration of the sleep data segment resampling and shuffling. Simulated days were filled with randomly sampled data segments from recorded study data without replacement for wake and sleep periods separately.

5.5 Evaluation Methodology

In this section, we introduce Chronify, an Android application to collect smartphone context data for sleep opportunity detection. Subsequently, we describe the data collection of our evaluation study. Finally, the sleep simulation generated sleep/wake schedules.

5.5.1 Chronify Smartphone Application

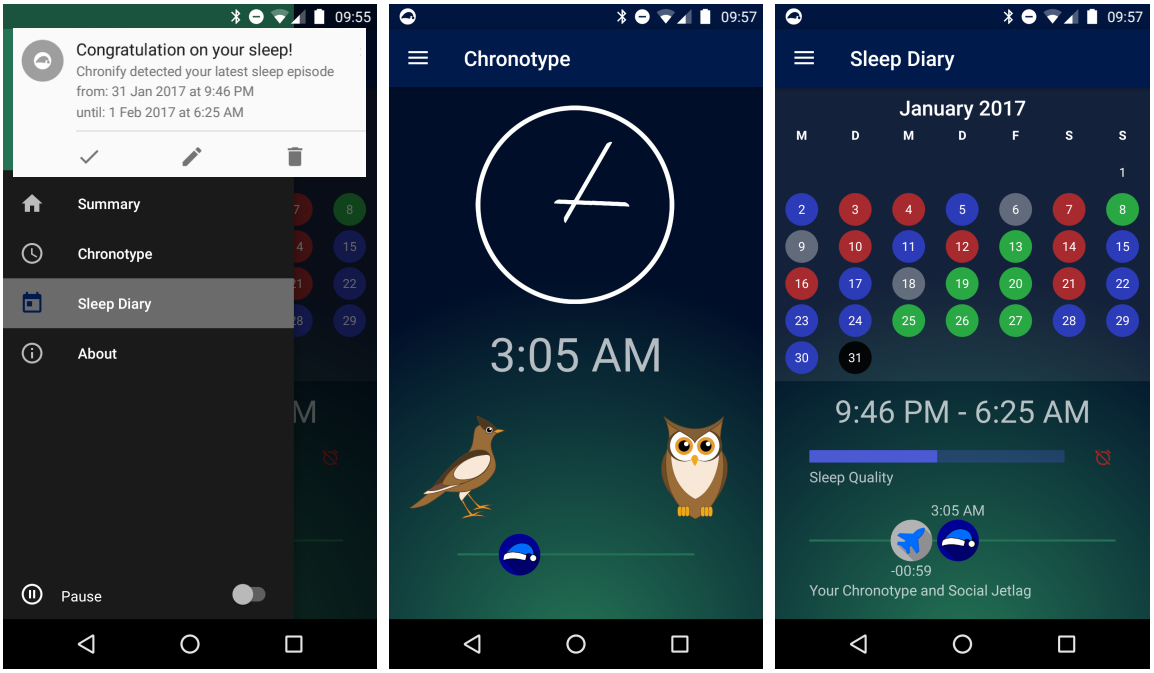
User Interface

Figure 5.9 shows the user interface. Users can choose different data visualisations from a drawer style menu shown in Figure 5.9a. A pause option is available to preserve privacy. Figure 5.9b shows the chronotype information screen. Once enough data is available, chronotype is displayed on a scale from early (lark) to late (owl). Figure 5.9c depicts the sleep diary view, where users can explore the temporal patterns of their subjective sleep quality. Days in the calendar can be selected to show detailed sleep information, such as sleep times, social jet lag, and alarm clock use. Users receive a notification after sleep timing was extracted and are then asked to confirm or adjust estimated sleep timing and add annotations such as subjective sleep quality.

Chronify continuously recorded context information to a SQLite database and transferred the encrypted database to a remote server. The sampling interval was set to 2 min. At the end of each interval per axis accelerometer signal range (max - min) was reported. An audio sample was taken from the integrated microphone at the end of each sampling interval to measure ambient noise level while preserving privacy. If the screen was activated during the sampling interval it was logged as active. Activity information and confidence of the current activity were obtained from the Android ActivityRecognitionAPI. Status information of system status, WiFi connection, charging state, air plane mode, silent mode, and do not disturb mode were recorded. Also data from proximity sensor, light sensor, step detector, and GPS were recorded.

5.5.2 Evaluation Study

We recruited 30 participants (13 female, between 20 and 60 years old) online using social media from Southern Bavaria. After filling in an online recruitment form, participants received detailed written instructions. Once written consent was given, participants completed an



(a) Main menu and daily notification.

(b) Chronotype view.

(c) Sleep diary and daily report.

Figure 5.9: Chronify user interface. The Android application has a drawer menu and displays a notification upon sleep timing extraction. Once enough sleep information is available, chronotype is displayed. Users can annotate subjective sleep quality information, which is displayed in a calendar overview. Detailed sleep reports show sleep quality, alarm clock use, and social jet lag.

entry form and were supplied with the smartphone application. Participants were asked to use their smartphones as usual. A pen and a paper sleep diary was used as reference. After recordings were completed, participants handed in the sleep diary and completed an exit form.

We experienced substantial drop-out throughout the study. Out of 30 enrolled participants, only 16 handed in a complete sleep diary at the end of the recording period. Two out of those 16 participants did not complete the online exit form. From 14 remaining participants, one was excluded from the analysis as s/he was sick in bed for the majority of recording days. The remaining 13 participants provided a total of 196 recording days which were used in the analysis. The average number of consecutive recording days per participant was 15 days, including week days and weekends. Christmas holidays of 2016/2017 were included in the recording period.

Reference Sleep Diary

Participants completed a sleep diary during the recording period. The sleep diary recorded bed time, sleep onset time, wake up time, and getting out of bed time. Timing and use of alarm clocks and work/free day information was also recorded. Participants were asked to provide a subjective sleep quality indication from 1 (excellent) to 6 (bad) and could note any additional information in a free text field.

Cross Validation

Sleep timing estimation was performed starting from a generalised, leave-one-participant-out (LOPO) cross validation model. For each participant P_{test}, estimation models were trained initially, on all data from all days of all other participants P_{train}. Subsequently, sleep timing of the first day $P_{test}(D_1)$ was estimated. Afterwards, estimation models were retrained using data from all days of all participants P_{train} and data of the first day of $P_{test}(D_1)$ to predict the second day $P_{test}(D_2)$. Sleep timing estimation of the third day $P_{test}(D_3)$ was performed after retraining the estimation models on P_{train}, $P_{test}(D_1)$, and $P_{test}(D_2)$. Sleep diary information was used to personalise the model on a daily basis as described in Section 5.3.1.

5.5.3 Simulated Data Set

We generated a simulated data set to explore our method on a wide variety of sleep behaviours. Each sleep schedule is described by sleep midpoint MS and duration of sleep DS, and describes the sleep timing of one day. The mean duration of sleep μ_{DS} and duration of sleep standard deviation σ_{DS} were chosen based on the findings of a smartphone-based sleep survey by Walch et al. [103]. Sleep midpoint mean μ_{MS}, sleep midpoint standard deviation σ_{MS}, and social jet lag δ_{SJ} values were chosen based on the findings of Wittmann et al. [104]. Table 5.2 shows simulation parameter value ranges. Recorded data set parameter values were included for comparison.

Table 5.2: Simulation parameters and value ranges. For each combination of simulation parameters 28 days of sleep schedules were generated resulting in 7560 simulated sleep schedules. Simulated values for μ_{DS} and σ_{DS} were chosen based on the findings of Walch, et. al [103]. Recorded data set sleep parameters ($\mu \pm \sigma^2$) were included for comparison.

Simulation Parameter Name		**Controls**	**Simulated values**	**Recorded values**	**Unit**	**Results**
μ_{DS}	Duration of sleep mean	Average duration of sleep	8	7.85±0.49	hours	
σ_{DS}	Duration of sleep standard deviation	Sleep regularity	0.85, 1, 1.15	1.37±0.41	hours	Sec. 5.6.4
μ_{MS}	Sleep midpoint mean	Chronotype	02:30, 04:00, 05:30	05:08±00:41	HH:MM	Sec. 5.6.4
σ_{MS}	Sleep midpoint standard deviation	Sleep regularity	0.1, 0.25, 0.5, 1, 1.5	1.12±0.3	hours	Sec. 5.6.4
δ_{SJ}	Sleep shift (added to μ_{MS} on Sat,Sun)	Social jet lag	0, 1, 2, 3, 4, 5	0.76±0.6	hours	Sec. 5.6.5

Table 5.3: Sleep estimation error metrics overview.

Metric	Description	Metric	Description
E_{SO}	relative sleep onset error [min]	$\|E_{SO}\|$	absolute sleep onset error [min]
E_{WU}	relative wake up error [min]	$\|E_{WU}\|$	absolute wake up error [min]
E_{MS}	relative midsleep point error [min]	$\|E_{MS}\|$	absolute midsleep point error [min]

Simulated Data Set Generation

For each combination of simulation parameters 28 sleep schedules were simulated by drawing sleep midpoint MS and duration of sleep DS from normal distributions. Subsequently, sleep data segment resampling from the recorded data set was performed for each of the 7560 simulated sleep schedules. Data from 13 participants were treated as separate data sets, i.e each of the 13 participant data sets was used for resampling data to each simulated sleep schedule separately, resulting in a total of 98280 simulated nights. LOPO cross validation was performed identical as for the recorded data set described in Section 5.5.2.

Simulated Data Set Validation

To ensure validity of our simulation approach, we performed a cross evaluation of data sets. Instead of using the same data set for training and testing, we used the simulated data set including the sleep parameter ranges of Table 5.2 with $\delta_{SJ} = 0$ for training and tested on the recorded data set. Results were similar to training and testing on the same data set. Full validation results are available in Table 5.5.

Statistical validation was performed using the Kolmogorov-Smirnov test (KS-test). KS-test scores measure the maximum distance between the cumulative distribution functions of two distributions. We computed KS-test scores for each feature of the sleep and wake classes of each simulation and the recorded data set and found their average score to be 0.04. In comparison, scores were 0.36 on average, testing opposing classes against each others, e.g. comparing the recorded wake vs. the simulated sleep data. Full KS-test results are available in Table 5.6.

5.6 Results

In this section, we analyse sleep estimation performance on simulated and recorded data. First, per participant sleep estimation performance is presented, before we investigate the impact of personalisation on sleep estimation performance. Our personalisation methodology was applied for all results presented. Subsequently, impact of typical sleep parameters on sleep estimation performance is presented. Lastly, we analyse the impact of social jet lag on sleep estimation performance. Results are presented using error metrics summarised in Table 5.3.

5.6.1 Simulations vs. Recordings

Sleep estimation performance of simulated data without social jet lag ($\delta_{SJ} = 0$, $N = 16380$) was compared to recorded data ($N = 196$). Figure 5.10 depicts absolute sleep onset $|E_{SO}|$

and wake up $|E_{WU}|$ estimation errors for both data sets and methods. For both data sets, *Time+Sensor+TPM* performed significantly better than *Time+Sensor*.

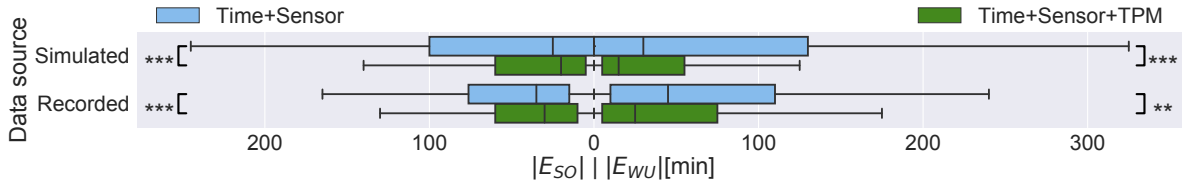


Figure 5.10: Absolute error for sleep onset and wake up estimations $|E_{SO}|$ and $|E_{WU}|$ estimation results using *Time+Sensor* and *Time+Sensor+TPM* for simulated ($\delta_{SJ} = 0$, $N = 16380$) and recorded (N=196) data sets. $|E_{SO}|$ and $|E_{WU}|$ are plotted on the horizontal axis. *Time+Sensor+TPM* showed significant improvements over *Time+Sensor* for simulated and recorded data sets. ** indicates $p < 0.01$, *** indicates $p < 0.001$.

For simulated data, *Time+Sensor* had average $|E_{SO}| = 76 \pm 108$ min and $|E_{WU}| = 101 \pm 153$ min. *Time+Sensor+TPM* reduced error by 47% to $|E_{SO}| = 40 \pm 48$ min and by 58% to $|E_{WU}| = 42 \pm 57$ min. Relative estimation errors for *Time+Sensor+TPM* were $E_{SO} = -2 \pm 62$ min and $E_{WU} = -5 \pm 70$ min, which were a 97% and 95% improvement over *Time+Sensor* results ($E_{SO} = 61 \pm 117$ min and $E_{WU} = -93 \pm 159$ min).

In the recorded data set, *Time+Sensor+TPM* yielded average errors of $|E_{SO}| = 45 \pm 49$ min and $|E_{WU}| = 54 \pm 72$ min ($E_{SO} = 3 \pm 67$ min and $E_{WU} = -20 \pm 88$ min). In comparison, data-driven *Time+Sensor* yielded $|E_{SO}| = 73 \pm 104$ min and $|E_{WU}| = 84 \pm 106$ min ($E_{SO} = 26 \pm 124$ min and $E_{WU} = -35 \pm 131$ min).

Time+Sensor+TPM outperformed *Time+Sensor* in both data sets for onset and wake up estimation. Relative errors were lower on average compared to absolute errors for all results as over- and underestimations cancelled out for relative errors. In our opinion, only absolute errors should be compared as users may be interested in their daily sleep patterns. Thus, underestimating one night and overestimating the next, would be as undesirable as underestimating two nights in a row. For the remainder of the paper, we present absolute errors only. All following results were obtained from the simulated data set.

5.6.2 Per Participant Analysis

We evaluated sleep estimation performance per participant for *Time+Sensor* and *Time+Sensor+TPM* simulations without social jet lag ($\delta_{SJ} = 0$). Figure 5.11 depicts $|E_{SO}|$ and $|E_{WU}|$ per participant for *Time+Sensor* and *Time+Sensor+TPM* simulations. Across participants *Time+Sensor+TPM* significantly outperformed *Time+Sensor*.

Time+Sensor+TPM outperformed *Time+Sensor* because integrated expert model knowledge of circadian and homeostatic sleep pressures improved sleep timing estimates. Remaining performance differences between participants were due to data and label quality. Participants were asked to use their phone as they normally would. While some participants kept their phone in close range at all times, others placed it on a shelf when arriving home. Sleep diary accuracy also varied between participants and contributed to performance differences.

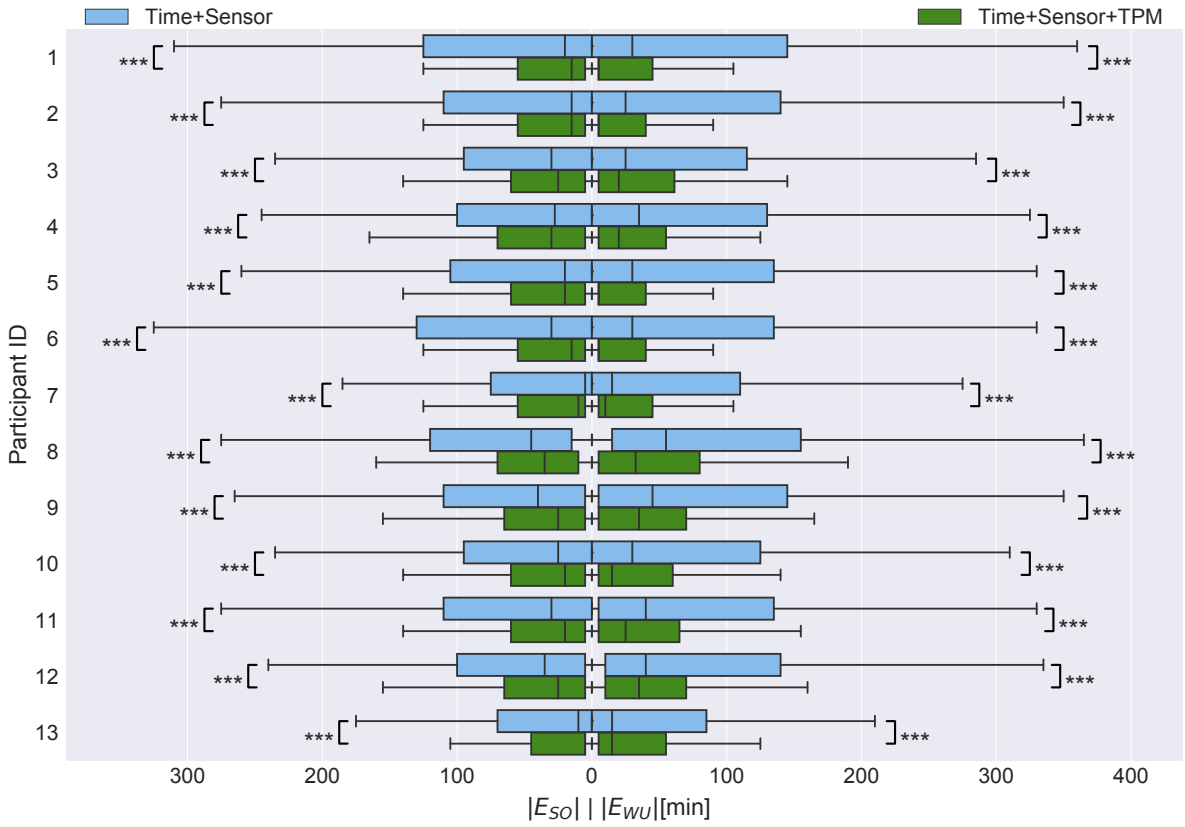


Figure 5.11: Per participant results for simulated ($\delta_{SJ} = 0$, $N = 16380$) data. Absolute sleep onset and wake up estimation errors $|E_{SO}|$ and $|E_{WU}|$ are plotted on the horizontal axis for each participant. *Time+Sensor+TPM* had significantly less $|E_{SO}|$ and $|E_{WU}|$ over *Time+Sensor* for all participants. *** indicates $p < 0.001$.

5.6.3 Personalisation Performance Improvements

We investigated the effect of personalisation on sleep timing estimation performance for simulated data without social jet lag ($\delta_{SJ} = 0$, $N = 16380$) using *Time+Sensor+TPM*. Personalisation was applied by retraining the classifiers using labels of previous days. Figure 5.12 shows $|E_{SO}|$ and $|E_{WU}|$ estimation errors over the number of days used for model personalisation. Without personalisation, estimation errors were $|E_{SO}| = 93 \pm 69$ min and $|E_{WU}| = 98 \pm 92$ min. With a single day of data used for model personalisation, absolute errors were reduced to $|E_{SO}| = 66 \pm 74$ min and $|E_{WU}| 62 \pm 81$ min. Using two days for personalisation reduced errors further to $|E_{SO}| = 38 \pm 41$ min and $|E_{WU}| = 42 \pm 56$ min resulting in a total error reduction of 59% for sleep onset and 57% for wake up timing compared to no personalisation. Adding further days for personalisation did not improve sleep onset performance, but wake up estimation was further improved. Using 8 to 9 personalisation days showed an improvement for wake up estimation, suggesting a weekend days' effect.

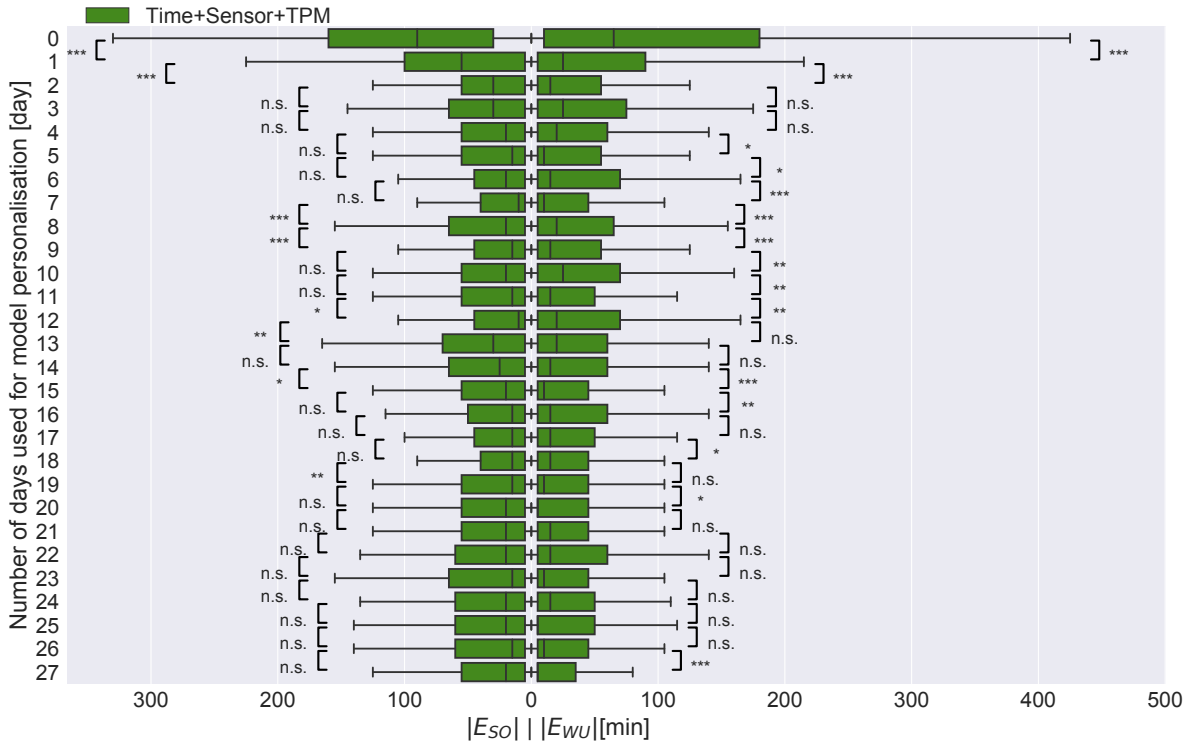


Figure 5.12: Absolute sleep onset and wake up estimation errors $|E_{SO}|$ and $|E_{WU}|$ over number of days used for model personalisation for simulated ($\delta_{SJ} = 0$, $N = 16380$) data. Sleep onset and wake up timing estimation performance was significantly improved through applying personalisation. Using one day of data for personalisation over no personalisation yielded the largest performance improvement. *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$, n.s. indicates $p \geq 0.05$.

5.6.4 Impact of Sleep Parameters

Mean Sleep Midpoint

Sleep midpoint is related to chronotype, thus we investigated if sleep midpoint affected sleep timing estimation performance in simulated data without social jet lag ($\delta_{SJ} = 0$). Sleep estimation performance should not depend on the mean sleep midpoint μ_{MS} as it would yield better results for certain chronotypes. Figure 5.13 depicts absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ over mean sleep midpoint μ_{MS} values. *Time+Sensor+TPM* results were independent of μ_{MS} while *Time+Sensor* results improved slightly for $\mu_{MS} \geq$ 04:00. *Time+Sensor+TPM* performed significantly better than *Time+Sensor* for all μ_{MS}.

Sleep Regularity

We investigated how sleep regularity impacted sleep timing estimation performance by varying the standard deviation of sleep midpoints σ_{MS} from 6 to 90 min. Results in Figure 5.14 showed increased $|E_{SO}|$ and $|E_{WU}|$ with increased σ_{MS} for *Time+Sensor* and *Time+Sensor+TPM* simulations without social jet lag ($\delta_{SJ} = 0$). With increased σ_{MS}, $|E_{SO}|$ and $|E_{WU}|$ increased

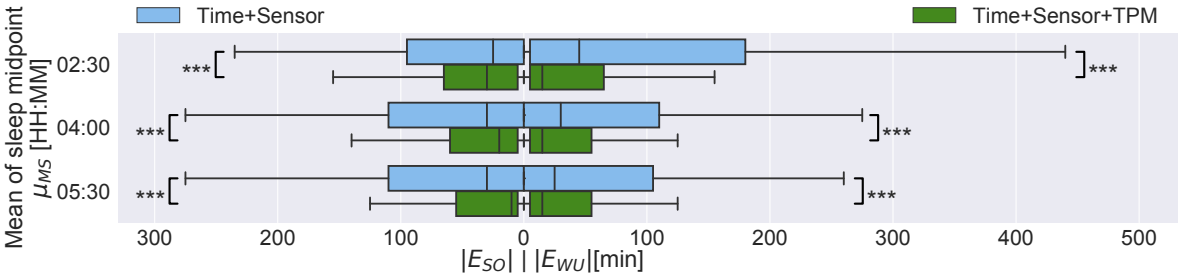


Figure 5.13: Absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ over mean sleep midpoint μ_{MS} for simulations ($\delta_{SJ} = 0$, $N = 16380$). In *Time+Sensor+TPM*, $|E_{SO}|$ and $|E_{WU}|$ were independent of mean sleep midpoint μ_{MS}. *Time+Sensor+TPM* showed significant improvements over *Time+Sensor* for all μ_{MS}. *** indicates $p < 0.001$.

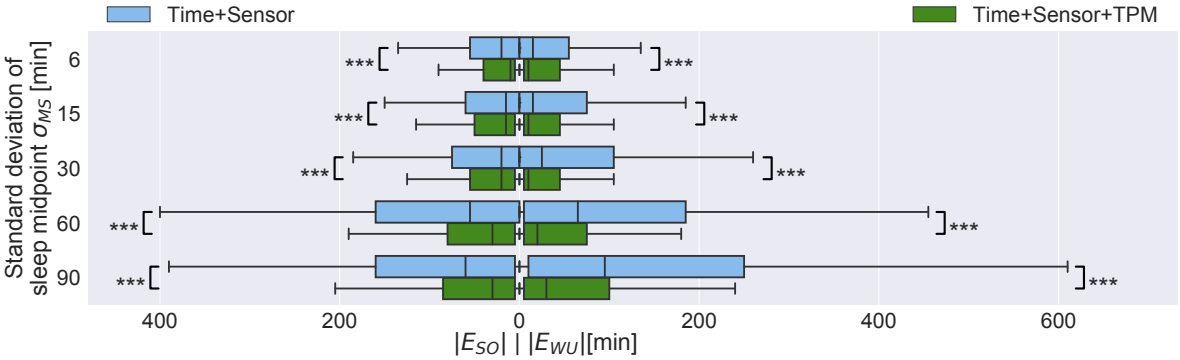


Figure 5.14: Absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ over sleep midpoint standard deviation σ_{MS} for simulations ($\delta_{SJ} = 0$, $N = 16380$). $|E_{SO}|$ and $|E_{WU}|$ increased with σ_{MS}. *Time+Sensor+TPM* showed significant improvements over *Time+Sensor* for all σ_{MS}. *** indicates $p < 0.001$.

more rapidly for *Time+Sensor* than for *Time+Sensor+TPM*. For all σ_{MS} *Time+Sensor+TPM* performed best with significant performance differences.

The time classifier component benefits from sleep timing regularity, thus $|E_{SO}|$ and $|E_{WU}|$ were expected to increase with σ_{MS}. *Time+Sensor+TPM* combined time and sensor classifier outputs using internal expert knowledge of homeostatic and circadian sleep pressures while prioritising the sensor classifier. Thus, *Time+Sensor+TPM* coped better with sleep irregularity than *Time+Sensor*.

We analysed how *Time+Sensor+TPM* and its components, *Time+TPM* and *Sensor+TPM*, were influenced by the sleep midpoint standard deviation σ_{MS} to understand how sleep regularity impacted sleep timing estimation performance. Figure 5.15 shows absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ for *Time+TPM*, *Sensor+TPM*, and *Time+Sensor+TPM* simulations without social jet lag ($\delta_{SJ} = 0$). $|E_{SO}|$ and $|E_{WU}|$ were significantly different between *Time+TPM* and *Time+Sensor+TPM* for all σ_{MS}. Absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ depended on the mean sleep midpoint σ_{MS} for *Time+TPM* and *Time+Sensor+TPM*. For *Sensor+TPM* $|E_{SO}|$ and $|E_{WU}|$ were independent of σ_{MS} because *Sensor+TPM* relies on sensor data only, thus is independent from time of day. In

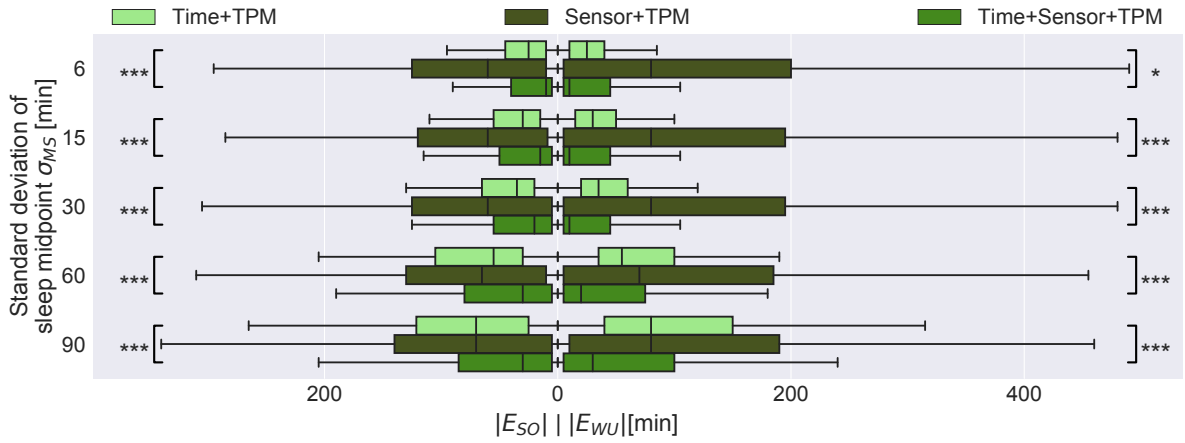


Figure 5.15: Absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ over sleep midpoint standard deviation σ_{MS} for simulations ($\delta_{SJ} = 0$, $N = 16380$). $|E_{SO}|$ and $|E_{WU}|$ depended on σ_{MS} for *Time+TPM* and *Time+Sensor+TPM*, while *Sensor+TPM* performance was independent of σ_{MS}. * indicates $p < 0.05$, *** indicates $p < 0.001$.

contrast, *Time+TPM* depended entirely on time of day, thus its performance was coupled to σ_{MS}. Results demonstrated the benefits of combining time and sensor approaches in *Time+Sensor+TPM* when $\sigma_{MS} \geq 30$ min.

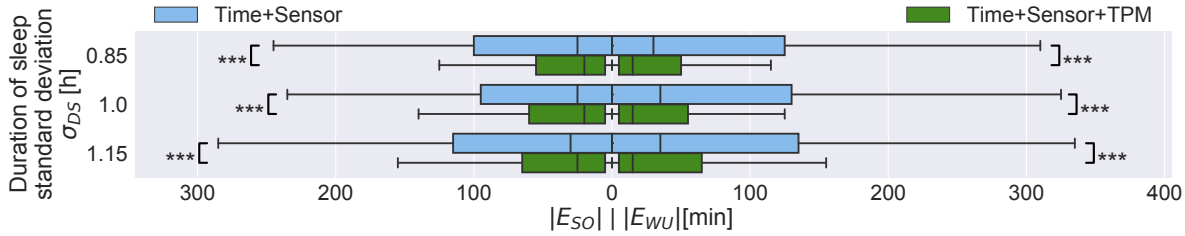


Figure 5.16: Absolute sleep estimation errors $|E_{SO}|$ and errwu over duration of sleep standard deviation σ_{DS} for simulations ($\delta_{SJ} = 0$, $N = 16380$). Results showed no dependency between sleep midpoint estimation performance and σ_{DS}. Performance differences between *Time+Sensor+TPM* and *Time+Sensor* were significant for all σ_{DS}. *** indicates $p < 0.001$.

In our simulations the duration of sleep DS was sampled from a normal distribution characterised by the mean duration of sleep μ_{DS} and the duration of sleep standard deviation σ_{DS}. According to the sleep duration variance reported in in [103] we chose three different values for σ_{DS} ranging from 0.85 h to 1.15 h for our simulations. Figure 5.16 shows $|E_{SO}|$ and $|E_{WU}|$ for different σ_{DS}. *Time+Sensor+TPM* outperformed *Time+Sensor* significantly for all σ_{DS}. In simulations without social jet lag ($\delta_{SJ} = 0$), changes in σ_{DS} did not show performance impact. The selected range of σ_{DS} was small compared to the range of simulated sleep midpoint standard deviation σ_{MS}. We expect that the impact of σ_{DS} on $|E_{SO}|$ and $|E_{WU}|$ would be similar to the one of σ_{MS}, if σ_{DS} would be varied in the same wide range as σ_{MS}.

5.6.5 Impact of Social Jet Lag

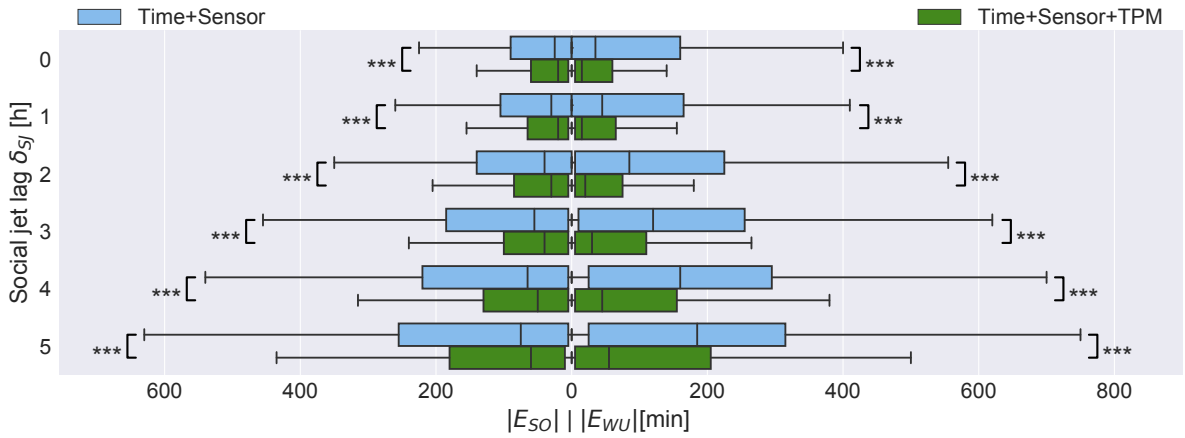


Figure 5.17: Absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ over social jet lag intensity δ_{SJ} for simulations ($N = 98280$). Increased δ_{SJ} led to an increase of $|E_{SO}|$ and $|E_{WU}|$ for *Time+Sensor* and *Time+Sensor+TPM* while $|E_{SO}|$ and $|E_{WU}|$ increased more rapidly for *Time+Sensor*. *Time+Sensor+TPM* significantly outperformed *Time+Sensor* for all social jet lag intensities. *** indicates $p < 0.001$.

We investigated the impact of social jet lag intensity δ_{SJ} on absolute sleep timing estimation errors $|E_{SO}|$ and $|E_{WU}|$ for *Time+Sensor* and *Time+Sensor+TPM*. Figure 5.17 depicts $|E_{SO}|$ and $|E_{WU}|$ for different δ_{SJ} ranging from 0 to 5 h. $|E_{SO}|$ and $|E_{WU}|$ increased with δ_{SJ} for *Time+Sensor* and *Time+Sensor+TPM* with a more rapid increase for *Time+Sensor*. *Time+Sensor+TPM* significantly outperformed *Time+Sensor* for all social jet lag intensities.

Social jet lag is a special case of sleep irregularity due to a conflict of internal and social clocks. We expected the impact of social jet lag to be similar to increased sleep irregularity because increased δ_{SJ} implies an increase of sleep midpoint standard deviation σ_{MS}. *Time+Sensor+TPM* outperformed *Time+Sensor* since TPM balanced time and sensor classification results using expert knowledge.

5.7 Discussion

5.7.1 Smartphone Sleep Detection

Smartphones are used by many today for tasks far beyond phone calls and text messages, combining our music players, remote controls, and print media into a single device. According to the Cisco world report 54% of young people aged 18 to 30 years from 15 countries look at their smartphone before their loved one as a first thing in the morning [29]. As ubiquitous devices, smartphones could be used to track sleep timing. Existing smartphone sleep apps require users to place the phone on the mattress or under the pillow, or rely on manual sleep timing input. While the latter is no improvement over a pen and paper sleep diary, the former poses a fire hazard [92]. In contrast, Chronify uses smartphone context information and does not require users to change their behaviour.

Systems presented in previous work often aimed at measuring more detailed sleep information, e.g. apnea, sleep stage, sleep quality. Measuring such parameters often requires specific smartphone placement or additional hardware, which is mechanically coupled to the body during sleep. With Chronify, our aim was to provide a ubiquitous, population scale, long-term service for sleep/wake information.

Our approach does not use any additional sensors and relies on smartphone context information only. Smartphones are often mechanically decoupled from the body while in bed, making it difficult to distinguish sleeping from lying in bed awake. Thus, we detect sleep opportunity rather than sleep. We chose to rely on smartphone sensors only due to their wide availability compared to wearables. Data from additional sensors, e.g. smartwatches could be included, if available, to improve detection results in the future.

5.7.2 Comparison with Related Work

Wearable sleep detection as performed by Fleury et al. [41], Borazio and van Laerhoven [17, 18], and Rofouei et al. [79] requires users to wear additional hardware. To obtain sleep timing estimates, users must have access to the hardware, ensure it is charged and worn. Studies with fitness trackers, e.g. FitBit devices, have found users to often abandon devices after a while [30]. In contrast, smartphone-based sleep detection relies on smartphones only and does not require additional hardware.

Ambient sleep detection as performed by Kay et al. [57], Rahman et al. [73], and Zhao et al. [108] depends on additional hardware in the user's bedroom. Beside the potential privacy concern, the solution must be installed for each bedroom and transported, e.g. when travelling. Smartphone-based sleep detection automatically travels with the user as smartphones are by definition mobile devices.

Nandakumar et al. [66] used a sonar-like approach with smartphones to detect sleep. Essentially, the approach involved probing the user presence to determine sleep, apnea, and breathing rate. Aside from being in the audible range, such measurements depend on specific smartphone placement, e.g. distance between user and smartphone. In contrast, our approach did not depend on specific placement.

Relative sleep timing estimation errors were frequently reported instead of absolute errors. While relative errors, e.g. over two weeks of data, can be close to zero, daily errors can remain large. Min et al. [64] used participant-specific models to predict sleep onset and wake up with an average relative error of ± 35 min and ± 31 min, respectively. In the present study, our Time+Sensor+TPM models were initially participant-independent and personalised during operation based on up to 20 days of user feedback, yielding average relative errors of $E_{SO} = -2 \pm 62$ min and $E_{WU} = -5 \pm 70$ min for the recorded data set.

Although previous studies did rarely report sleep variance, our results showed large variance across recorded and simulated data. Personalisation results in Section 5.6.3 indicated a reduction of absolute estimation error variance from 41 min to 37 min for sleep onset and 56 min to 30 min for wake up with increased model personalisation.

Table 5.4: Comparison of $|E_{SO}|$, $|E_{WU}|$, and mMAD for sleep onset, wake up, and sleep duration results. We implemented the algorithm described by Abdullah [2] and evaluated it on our recorded and simulated data sets. Our *Time+Sensor+TPM* algorithm performed best overall. Saeb [81] reported mMAD only. (*) Abdullah [2] reported a relative average sleep duration error of <45 min across participants and a relative average midsleep error of 23.8 min. In comparison, *Time+Sensor+TPM* had an absolute midsleep error of $|E_{MS}| = 34 \pm 42$ min ($E_{MS} = -4 \pm 54$ min) for the simulated data set and $|E_{MS}| = 39 \pm 43$ min ($E_{MS} = -8 \pm 58$ min) for the recorded data set.

Algorithm	**Data set**	**N**	**Days**	$\lvert E_{SO} \rvert$ **[min]**	$\lvert E_{WU} \rvert$ **[min]**	**mMAD SO/WU/DS [min]**
Saeb [81]	Saeb [81]	208	8736	not reported	not reported	38/36/55
Abdullah [2]	Abdullah [2]	9	625	not reported (*)	not reported (*)	not reported (*)
Abdullah [2]	This paper recorded	13	172	60±68	61±99	32/37/58
Abdullah [2]	This paper simulated	13	15960	98±136	73±88	44/36/110
Time+Sensor+TPM	This paper recorded	13	196	45±49	54±72	31/34/45
Time+Sensor+TPM	This paper simulated	13	16380	40±48	42±57	22/15/41

We compared our *Time+Sensor+TPM* algorithm with Saeb et al. [81] and Abdullah et al. [2] and summarised our findings in Table 5.4. Saeb et al. [81] estimated sleep onset, wake up, and sleep duration with a mean median absolute deviation (mMAD) of 38 min, 36 min, and 55 min, respectively. Unfortunately, we could not test their algorithm on our data sets due to missing features and lower sampling rates of our recorded dataset. On our simulated data set, *Time+Sensor+TPM* estimated sleep onset, wake up, and sleep duration with mMADs of 22 min, 15 min, and 41 min, respectively. Due to lower sampling rates our approach required ∼3 MB of storage for two weeks of data. In comparison to Saeb et al. [81], our approach did not require a constant wireless network connection, is more power efficient, and was designed with user privacy in mind. For example, we took a single audio sample once every 2 min, while Saeb et al. [81] recorded audio for 30 sec every 5 min. They also recorded GPS location information, call, and text message activity of their participants.

Saeb et al. [81] used random forrest classifiers to detect sleep/wake and a hidden Markov model (HMM) to filter classifier outputs. In comparison, we adapted the TPM of Daan et al. [31], which introduces expert knowledge on sleep and wake timing in humans. In particular, the homeostatic signal served as guidance to select physiologically meaningful transition points between sleep and wake. The benefit of the TPM becomes particularly evident with varying sleep schedules as observed in our recorded dataset. Saeb et al. [81] considered that the constant wireless connection required in their investigation resulted in a participant selection bias, which in turn may have lead to regular sleep patterns of their participants.

Abdullah et al. [2] reported an average sleep duration error of <45 min for all participants. We implemented their algorithm, tested it on our data, and computed mMADs for sleep onset, wake up, and sleep duration. Their algorithm used screen on/off events as the only feature. Some days in the simulated and recorded data sets had no screen on/off events. On those days, their algorithm could not estimate sleep timing and days were therefore omitted from the evaluation, resulting in a lower days count. Results of their algorithm on our recorded data set were only slightly worse than *Time+Sensor+TPM* results. On the simulated data set however, which included a wide variety of sleep behaviours, performance of their algorithm dropped to $|E_{SO}| = 98 \pm 136$ min and $|E_{WU}| = 73 \pm 88$ min. One reason is in the logic of their algorithm, which detects sleep onset only between 10 pm and 7 am. In comparison, *Time+Sensor+TPM* results improved for simulations as the algorithm is robust to a wide variety of sleep behaviours. The comparison illustrates the benefit of our simulation approach to explore sleep parameter ranges and thus helped us to identify the advantage of our data and expert model combination under naturally varying sleep schedules.

Daan et al. [31] developed the TPM to model the human recovery process and used it in simulations. We adapted the TPM to perform sleep estimation. Running the TPM with the daily parametrisation of t_0 only (see Section 5.3.3), would resemble the *Time* classifier's behaviour, i.e. neglecting daily variation of sleep patterns. In Fig. 9, the *Time+TPM* model's performance is displayed, which closely resembles the *Time* classifier: As the plot shows, increasing variation, here expressed by σ_{MS}, raises error of *Time+TPM* profoundly. In our recorded data set, we determined $\sigma_{MS} = 1.12 \pm 0.3$ hours, illustrating that variation of the sleep midpoint is a usual phenomenon.

5.7.3 Simulation Engine

During our data collection we experienced a participant drop out of over 50%. The majority of dropouts occurred due to missing pen and paper sleep diaries. At this rate, more than two

participants would have to be included in the study to obtain one complete participant data set. Saeb et al. [81] experienced less drop out, yet for 20% of their participants more than half of the recorded data were missing. The difference in drop out might be explained by participant compensation. In our evaluation study, participants did not receive any compensation, while in their study, compensation was based on the number of sleep diary entries returned.

Given the substantial participant effort in maintaining a sleep diary across 15 days, we believe that simulations represent an important, complementary analysis strategy to synthesise sleep patterns. We resampled recorded study data to simulate a total of 98280 days. Sleep schedules were generated for each combination of simulation parameter values listed in Table 5.2 and subsequently filled with data from each participant.

Results in Section 5.6.3 showed convergence for the absolute sleep estimation error after few days confirming that the 28 simulated days per parameter configuration suffice. While our study data contained considerable variation in sleep behaviour already, the simulation engine allowed us to explore algorithm behaviour across several sleep-related variables, including sleep regularity and average duration of sleep.

Resampling of data segments may have broken correlations, e.g. creating sharper than natural boundaries between sleep and wake phases. At the same time, boundaries between sleep and wake could be hardly reported via sleep diaries, which suggests that our data segment resampling helped to create more accurate labels around sleep and wake transitions. While both aforementioned aspects may partly explain estimation error reduction for simulated data, another reason are the sleep parameter ranges covered by our simulation. As the literature-derived ranges did not cover values observed in our recorded data for variation parameters δ_{DS} and δ_{MS}.

5.7.4 Evaluation Study Recordings

Obtaining accurate sleep timing reference in free living is challenging. We used pen and paper sleep diaries with 5 min resolution as reference information due to their practicality. Reliability of self-reported sleep diaries has been investigated, e.g. by Rogers et al. [80]. They compared polysomnography recordings against sleep diaries with 15 min resolution and found a mean discrepancy of 39 min for sleep transitions in 12 healthy subjects. Sleep timing estimation using *Time+Sensor+TPM* achieved an average absolute sleep timing estimation error $|E_{MS}| = 39 \pm 43$ min for the recorded data set, thus reaching the accuracy of sleep diaries.

Managing good sleep hygiene has become difficult. Alarm clocks are common for many, even on free days due to social obligations. Chronotype estimation relies on data from days with and without alarm clock use [77] to correct for sleep debt on days with alarm clock use. However, no alarm clock use does not guarantee self-determined sleep either as users may wake up due to bed partners or a noisy environment. We monitored alarm clock use in the sleep diary. In deployment, users would be asked to fill in their usual alarm clock behaviour when starting the Chronify app for the first time. Alarm clock use could be modified in daily notification feedback.

We chose the Christmas holiday period of 2016/2017 as our recording period. For many, the time around Christmas is the only time of the year, where no alarm is used. However, the recording period also included New Years Eve, which was an outlier to regular sleep timing of most participants. Some participants completely neglected their smartphone during the Christmas holidays to spend time with their family, which could explain some part of the data loss and participant exclusion.

In the study, we intended to avoid the smartphone's relevance to estimate sleep timing. Therefore, participants were not instructed on nor queried about smartphone placement. With the study design, we intended to avoid creating awareness on the smartphone location or influence its location during nighttime. However, we assume all participants kept the phone in the same room.

5.7.5 Chronify App Design

The Chronify app was designed to be resource efficient. Critical computing tasks were outsourced to the cloud and samples were collected only every 2 min. Although Chronify ran continuously during the recording period it did not show relevant energy consumption in the energy consumption monitor of Android.

To minimise mobile data consumption, Chronify uploaded new data to the cloud only when the phone was connected to a power supply and a wireless network. Clearly, cloud computing requires the smartphone to upload data at least daily, ideally in the morning hours to provide feedback on estimated sleep times soon after wake up.

To receive feedback early after wake up, the notification could be displayed as soon as wake up was detected. While we consider that the algorithms proposed in this investigation are generally suitable for an implementation on the smartphone, an anonymous cloud-based processing offers opportunities to elaborate population-based models that are the basis for personalisation in our approach.

Due to an increase in sensing modalities and the omnipresence of smartphones, privacy concerns have been on the rise. We incorporated privacy concerns into the Chronify app design and data analysis by recording the minimal number of audio samples required by Android OS from the device once every 2 min. No location-based information was stored for sleep opportunity detection. By our intention, the features selected for sleep/wake estimation do provide minimal cues on an individual's identity. In addition, users have the option to pause recordings either for a predefined duration or indefinitely.

5.7.6 Model Personalisation

Previous work often used participant dependent models when estimating sleep timing. During our study the majority of drop out was due to missing sleep diaries. Thus, we believe using participant independent models are key. Abduallah et al. [2] used the first two weeks of data to compute a participant-specific correction factor. In contrast, we used participant independent models, which were personalised with user feedback.

Chronify learns from user feedback because sleep timing is both personal and dynamic. By retraining, the sensor classifier adapts feature weights to different sleep environments and sleep behaviour. Sleep timing is dynamic, even within users (e.g. changing shift schedules), which the time classifier adapts to. *Time+Sensor+TPM* started based on a generalised model in our analyses.

Each time users respond to notifications of estimated sleep timing as shown in Figure 5.9a, time and sensor models are retrained to incorporate new data. Further investigations are needed to confirm if continuous retraining as applied here or training according to specific context variables, e.g. location, day of the week, etc. is most suitable.

We assumed users would react to sleep timing notifications every day as we used entries from the pen and paper sleep diaries to simulate reactions to the notifications. In deployment,

users might not react to sleep timing notifications on a daily basis. If no reaction is given, no retraining takes place. Results in Section 5.6.3 showed an average sleep timing estimation error reduction of 59% for sleep onset and 57% for wake up using *Time+Sensor+TPM* with just two days of data used for personalisation. We consider that obtaining accurate user feedback for a few times, i.e. two times, is feasible.

5.8 Conclusion

In this work we investigated sleep timing estimation from smartphone context data by integrating expert knowledge into data-based estimation models. Our *Time+Sensor+TPM* model, which combines data-based classification and expert knowledge, showed best sleep onset, midpoint, and wake up estimation results compared to data-driven models and the related approaches, with an absolute sleep timing estimation error of $|E_{SO}| = 40 \pm 48$ min for sleep onset and $|E_{WU}| = 42 \pm 57$ min for wake up (relative errors $= -2 \pm 62$ min and $= -5 \pm 70$ min) over 16380 days of simulated data ($\delta_{SJ} = 0$). Our expert model improved absolute estimation results by 47% for sleep onset and 58% for wake up over the data models. Our combined data and expert model was robust against irregular sleep patterns and social jet lag. To cope with sleep timing changes and individual sleep hygiene, we incorporate user feedback to personalise estimation algorithms. Personalisation improved estimation results by 59% for sleep onset and 57% for wake up with just two days of data used for personalisation. Finally, simulations allowed us to explore the sleep timing estimation performance for varying sleep behaviour, encoded in key sleep parameters and wide parameter ranges. The simulation approach was confirmed by cross dataset tests that yielded similar performances and statistical testing of each feature.

Our approach could be used to track sleep timing and chronotype on an everyday basis. Long-term recordings could reveal social jet lag and thus support users in reducing their circadian misalignment and living a healthier lifestyle. We plan to publicly release the Chronify smartphone application soon.

We conclude that using expert knowledge can improve data-driven modelling results and smartphones are a suitable platform for unobtrusive sleep timing and chronotype monitoring. The expert model integration as fusion technique to refine the data model performance showed the best performance.

5.9 Appendix

Table 5.5: LOPO cross validation results for *Time+Sensor+TPM* using different data sets for training and testing.

Training data set	**Testing data set**	$\lvert E_{SO}\rvert$ **min**	E_{SO} **min**	$\lvert E_{WU}\rvert$ **min**	E_{WU} **min**	$\lvert E_{MS}\rvert$ **min**	E_{MS} **min**
Recorded	Recorded	45±49	3±67	54±72	-20±88	39±43	-8±58
Simulated	Simulated	40±48	-2±62	42±57	-5±70	34±42	-4±54
Simulated	Recorded	50±55	6±74	58±76	-19±93	42±48	-6±64

Table 5.6: Kolmogorov-Smirnov test scores (mean and standard deviation). We tested distribution similarity of the recorded and simulated data sets for wake and sleep classes separately for each feature. Results show that feature data is very similar between recorded and simulated data within each class and dissimilar between opposite classes.

Feature	**Recorded wake vs. simulated wake**	**Recorded sleep vs. simulated sleep**	**Recorded wake vs. simulated sleep**	**Recorded sleep vs. simulated wake**
Activity	0.01±0.01	0.03±0.03	0.15±0.08	0.15±0.08
Activity confidence	0.01±0.01	0.03±0.04	0.17±0.12	0.16±0.14
Screen segment length	0.04±0.04	0.09±0.07	0.78±0.17	0.74±0.2
Microphone amplitude	0.02±0.01	0.05±0.04	0.28±0.2	0.29±0.2
Motion variance	0.04±0.03	0.08±0.05	0.42±0.1	0.42±0.12

6

Circadian timing estimation from smartphone context data

6.1 Introduction

As time is a scarce resource, clocks are used to synchronise our activities with others. Wall clock time synchronises social activities, e.g. a regular time for dinner with the family or events with friends, and organise our work activities, e.g. meetings. In addition, our body also has an internal clock, called the circadian clock. With many social and professional obligations, it becomes difficult to keep internal and external clocks synchronised. Teenagers for example have a later circadian clock phase on average, yet are expected to study in schools starting early in the morning [90].

Light exposure to the eyes synchronises the circadian clock to the environment. Timing and intensity of light exposure determine phase shift intensity [76]. Before artificial light was available, humans could perform outdoor activities, e.g. hunting natural enemies or finding food, only during daytime. Thus, nighttime was used to find shelter and recover.

We have a natural preference towards an earlier or later circadian phase, called chronotype [78]. Chronotypes range from early (larks) to late (owl) types and are naturally distributed amongst the population. Chronotype drifts throughout the lifespan of a person [77]. During our teenage years we drift towards a later chronotype and back to an earlier chronotype as we age. Studies have indicated serious health risks for living against the chronotype for extended periods of time, including diabetes, heart diseases, and cancer [35, 56].

With artificial light we can work anywhere at anytime. However, exposure to artificial light can lead to a desynchonisation of wall and circadian clocks, called circadian misalignment. Our circadian system is especially influenced by light in the blue part of the visual spectrum [20]. Current display technology often uses light emitting diode (LED) technology as backlight. LEDs are know to emit blue-rich light, making screen use at night one of the main sources for circadian misalignment [22].

Measuring light exposure in daily life is cumbersome. Measurements at wrist level are common but known to be inaccurate due to their on-body location [100]. In addition, wrist-worn light sensors are frequently occluded by long sleeved clothing, resulting in underestimations of light exposure. Another option is to measure light exposure at eye-level. One solution are the LightWatcher devices from the EUCLOCK research project. When mounted to an eyeglasses frame they offer accurate measurements but were often found intrusive and stigmatising. WISEglass is another research solution, where a light sensor is integrated into the bridge of a 3D-printed eyeglasses frame [99]. However, such measurement devices are not available to the masses yet.

In turn, smartphones have become our ubiquitous companions in daily life with ownership rates of 72% in the US and a median of 68% in developed countries [70]. Todays smartphones are cloud connected, multi-modal sensor platforms which are already carried by their users. Smartphone light sensors cannot be used for continuous light measurements due to frequent occlusion, e.g. due to flip cases or when the phone is stored in a pocket. Instead, information from other integrated sensors could be used to infer light exposure continuously.

Our approach in this paper is to infer light exposure from smartphone sensors and compare light estimations to measurements taken at eye and wrist level. Subsequently, we estimate circadian phase shift from light estimations using the Kronauer model. In particular, this article provides the following contributions:

- We present a method for continuous light exposure estimation from smartphone sensor information and cloud-sourced weather reports and compare them against light mea-

surements taken at eye and wrist level.

- A method to estimate daily circadian phase shift from light exposure using the Kronauer model.
- We compare circadian phase shift estimation results from wrist measurements with estimations from our smartphone-based method.

6.2 Related work

Circadian rhythm research began in the 20th century and was awarded with the 2017 Nobel Prize for physiology or medicine. Duffy and Wright [35] found light to be the dominant synchonizer in human circadian systems and observed that the maximum daily clock shift was between -2 h and +3 h. They state multiple negative health effects due to improper circadian entrainment, e.g. upset of gastrointernal function, impaired performance, and impaired alertness. Kantermann and Roenneberg [56] suggest that light exposure at night time can damage DNA and thus lead to cancer. Shift workers are especially at risk, as they constantly live against the natural day/night rhythm.

Prior efforts were made to estimate human circadian phase in ambulatory settings. Table 6.1 provides an overview. Kolodyazhniy [59] et al. used an ambulatory multi-channel monitoring system and least-squares regression to estimate human circadian phase. They recorded core body temperature, skin temperature at 11 locations, ECG, respiration, accelerometer, and light data from 16 participants for one week. Afterwards, participants underwent a 32 h constant routine protocol to measure circadian phase from their melatonin profile. Their regression model could predict human circadian phase with an error of 12±41 min and a mean absolute deviation (MAD) of 36 min. Following up on their original paper, Kolodyazhniy et al. [60] extended the data set to 25 participants and used artificial neural networks. Their best model was able to predict circadian phase with an error of -3±23 min and a MAD of 19 min. Gil et al. [43] estimated human circadian phase using autoregressive models from light measurements at the wrist and ECG recorded using standard 3-lead ECG placement.

Table 6.1: Summary of related approaches and results. Reference techniques are noted in the reference column as constant routine melatonin assessment (CRMA), dim light melatonin onset (DLMO) assessment, and head light measurement (HLM), respectively. Least squares (LSQ), artificial neural networks (ANN), autoregressive moving average with exogenous inputs (ARMAX), extreme gradient boosting regression (XGB reg.), and the Kronauer model (KM) have been used to estimate circadian phase.

Name	**N**	**days**	**In-/On-/Off-Body Sensors**	**Reference**	**Model**	**Error**
Kolodyazhniy 2011 [59]	16	7	1/14/1	CRMA	LSQ	12±41 min
Kolodyazhniy 2012 [60]	25	7	1/14/1	CRMA	ANN	-3±23 min
Gil 2013 (Light only) [43]	38	3	0/1/0	DLMO	ARMAX	37±56 min
Gil 2013 (Light + ECG) [43]	38	1	0/2/0	DLMO	ARMAX	2±39 min
Bonmati-Carrion 2014 [16]	13	10	0/3/0	DLMO	Correlation	46±7 min
Our phase shift est.	12	6	0/0/1	HLM	XGB reg. + KM	35±25 min

Dim light melatonin onset (DLMO) assessments were used as reference techniques. Using light measurements only, they were able to predict circadian phase with an error of 37±56 min for 38 participants. In combination with ECG measurements, prediction error was reduced to 2±39 min. Bonmati-Carrion et al. [16] assessed correlations between DLMO and multiple ambulatory variables. They monitored motor activity, wrist temperature, body position, and light for 10 days in 13 participants. Recorded signals were transformed to different features, e.g. wrist temperature onset, activity offset, etc. and compared to DLMO reference measurements. They predicted DLMO from single features with mean absolute errors between 46 and 58 min. They found wrist temperature increase onset to be the closely correlated to DLMO and simple to obtain. In contrast to prior work, our approach does not require any body worn hardware, but instead works with a smartphone only.

Screen use at night leads to circadian misalignment. Wahl et al. [101] performed screen use detection using smart eyeglasses with a colour light sensor embedded into the bridge. Their system detected screen use in environments with ambient light intensities below 200 lux with a ROC AUC of above 0.9. In this work, we estimate light exposure throughout the day to estimate circadian phase shift. Woelders et al. [105] estimated clock phase and period of 20 participants. They measured light exposure at the wrist for 9 consecutive days and used the revised Kronauer model to estimate clock phase and period. In this work we estimate daily circadian phase shift from smartphone estimated light exposure and compare our results against light measurements at the wrist.

Kronauer et al. [61] described a mathematical model to estimate circadian phase shift. They used two van der Pol type oscillators to model core body temperature and rest-activity. Their initial implementation has no input for light but estimates core body temperature as circadian phase marker based on rest-activity rhythm. In 1990, Kronauer updated the model to take effects of light on circadian amplitude and phase into account. Jewett et al. [55] revised Kronauers model by replacing the van der Pol oscillator with a higher order limit cycle oscillator. Furthermore, the model was adapted to be in accordance with Aschoff's rule by adding a direct effect of light on the circadian period. In this work we use their model and parameter settings.

Our previous work on light estimation [100] serves as the starting point for smartphone-based light estimation in this work. In this previous paper, data recording and light estimation using CART decision tree classification were described.

6.3 Methodology

In this section we present an approach to estimate circadian phase shift from smartphone sensors and cloud-sourced weather reports. Figure 6.1 shows the methodology used to estimate circadian phase shift from raw sensor and weather data. First, sensor and weather data were recorded with a smartphone app. Subsequently, we extracted features and estimated light intensity with our light estimation algorithms. Finally, circadian phase shift was estimated from light intensity estimations.

6.3.1 Feature extraction

We extracted two kinds of features from recorded data: Sensor features and cloud-sourced weather information. Table 6.2 provides an overview of extracted features.

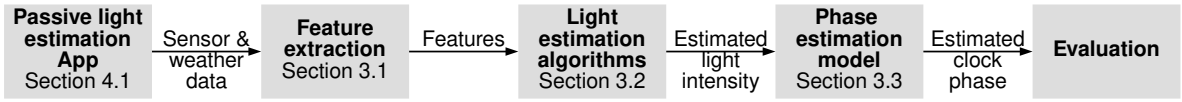


Figure 6.1: Overview of circadian phase shift estimation methods. First, sensor and weather data were recorded with a smartphone app. After feature extraction, light intensity was estimated. Finally, circadian phase shift was estimated from light intensity estimations.

Table 6.2: Smartphone sensor and weather features used to estimate light intensity. All features were resampled to a sampling rate of 0.2 Hz.

	Sensor features			**Weather features**	
	Description	**Unit**		**Description**	**Unit**
M	Motion intensity	$\frac{\mathrm{m}}{\mathrm{s}^2}$	C	Cloud coverage	%
S	Satellite count	1	W	Weather code	–
A	Location accuracy	m	T	Time of day	hh:mm
L	Light intensity	lux			
P	Proximity	1			

Sensor features

We computed the acceleration magnitude M_{acc} of three axes accelerometer axes (x, y, and z) as $M_{acc} = \sqrt{x^2 + y^2 + z^2}$ for each accelerometer sample. Accelerometer samples $M_{acc,1} \dots M_{acc,n}$ were summed up to compute motion intensity M as $M = \sum_{i=1}^{n} M_{acc,i}$ for each 30 s window.

Location accuracy A in meters was provided by the Android Location API. GPS reception is typically better outdoors and therefore A should decrease when the smartphone is in an outdoor environment. Satellite count S measured how many GPS satellites were used for the location estimate. We expected S to behave similar to A as the number of satellites in reception range was expected to increase outdoors.

Smartphone light sensors reported ambient light intensity L in lux. Light sensors often reported zero lux readings due to occlusion, e.g. when using flip cases. To compensate for occlusion, we also recorded proximity sensor information P as explained below. Proximity sensors are typically located next to the smartphone screen, together with light sensors, and were used to detect light sensor occlusion. P measured whether the sensor was occluded or not as shown in Eq. (6.1) and was also used as a feature.

$$P = \begin{cases} 0 & \text{if proximity sensor was not occluded} \\ 1 & \text{if proximity sensor was occluded} \end{cases} \tag{6.1}$$

Cloud-sourced weather features

GPS location was used to obtain weather reports at smartphone location from Open Weather Map[1], a cloud weather service. Weather code W described current weather conditions,

[1] `https://openweathermap.org`

e.g. rain, clear sky, clouds, etc. A full list of weather codes is available at `https://openweathermap.org/weather-conditions`. Cloud coverage C indicated the percentage of cloud coverage in the sky. Time of day T was transformed into two continuous features T_1 and T_2 as described in Eq. (6.2).

$$T_1 = \begin{cases} 1, & \text{when } T > 12.0 \\ 0, & \text{when } T \leq 12.0 \end{cases} \qquad \text{and} \qquad T_2 = \begin{cases} 24 - T, & \text{when } T > 12.0 \\ T, & \text{when } T \leq 12.0 \end{cases} \tag{6.2}$$

6.3.2 Light estimation algorithms

Light estimation algorithms estimated light intensity from extracted feature data. We compared three different approaches: CART Decision tree classification, extreme gradient boosted decision tree (XGB) classification, and XGB regression. All three approaches used the feature set described in Section 6.3.1. We added a 5 lux and a 500 lux dummy estimator as baseline comparison. Dummy estimators always returned the same light intensity, independent of feature values. Dummy estimator light intensities were chosen to resemble staying in a constant dark environment (5 lux) and a constant bright indoor environment (500 lux).

All algorithms were trained to estimate light intensity I_{pred} as the dependent variable from independent variables described in Section 6.3.1 weighted with individual weights ω_i as shown in Equation (6.3).

$$I_{pred} = \omega_1 * M + \omega_2 * S + \omega_3 * A + \omega_4 * L + \omega_5 * P + \omega_6 * C + \omega_7 * W + \omega_8 * T \tag{6.3}$$

Light intensity classes

For classification algorithms, light intensities were sorted into five light intensity classes as shown in Table 6.3. Classes were selected according to the Illuminating Engineering Society lighting handbook [33]. We performed light intensity estimation by deriving the measurement average for each of the five illuminance categories from the reference measurements and assigning them to the estimated light class, respectively. In a practical application, illuminance levels could be derived from smartphone light sensor measurements.

Table 6.3: Light intensity classes used for classification algorithms. The classes were derived from IES lighting handbook [33].

Class	min [lux]	max [lux]	Description
1	0	119	Dark room
2	120	249	Dim room
3	250	999	Bright room
4	1000	4999	Cloudy
5	5000	∞	Sunny

Decision tree classification

Decision tree classification was used as comparison to our prior work [100]. We used the `DecisionTreeClassifier` implementation of scikit-learn [68], which implements an optimised

version of the CART algorithm. We configured the classifier to use a minimum number of three samples to split an internal node.

XGB classification

The XGB classifier implementation from the xgboost [26] package was used with default settings to classify light intensity. Using an ensemble of smaller decision trees instead of a single tree was expected to improve results.

XGB regression

For regression, the XGB regressor the xgboost [26] package was used with default settings. We included a regressor to investigate whether direct light intensity estimations offer an advantage over light class estimations, where each class is represented by a single light intensity value.

6.3.3 Circadian phase shift estimation

Circadian phase shift was estimated from light intensity estimations using the revised version of the Kronauer model by Jewett et al. [55]. Kronauer's model outputs clock oscillation, from which daily circadian phase shift was calculated.

Kronauer model

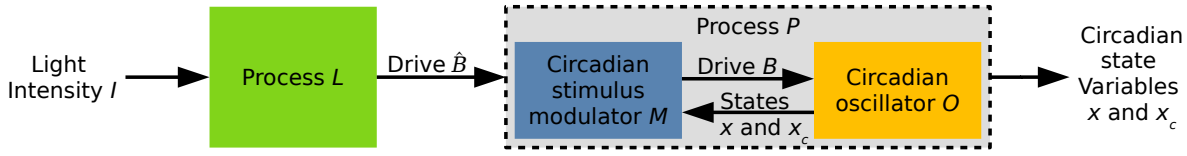


Figure 6.2: Kronauer model overview. Process L transforms light intensity at eye level I into drive $\hat{B}$. Process P consists of circadian stimulus modulator M and circadian oscillator O. M modulates drive $\hat{B}$ into drive B depending on phase and amplitudes of circadian oscillator O. O is a self-sustaining oscillator, which outputs two state variables x and x_c. x closely follows the circadian body temperature rhythm.

The Kronauer model consists of two processes, process L and process P. Figure 6.2 shows how processes L and P are used to model the influence of light at eye level on the circadian oscillator O. Circadian phase and amplitude are described by circadian state variables x and x_c. Process L converts light intensity at the eye to drive $\hat{B}$. First, the forward drive α is computed by multiplying the forward rate constant α_0 with the normalised light intensity. Light intensity is normalised to 9500 lux, as it is the intensity at which the maximum response takes place. Next, the change in flux rate $\dot{n}$ is computed. Drive $\hat{B}$ is proportional to the flux rate n.

Process L acts as a dynamic light stimulus processor and derives drive $\hat{B}$ from light intensity I. L is described in Equations (6.4 – 6.6), where α_0 =0.16 min^{-1}, β =0.013 min^{-1},

I_0 =9500 lux, p =0.6, and G=19.875, as described in Jewett et al. [55].

$$\alpha = \alpha_0 \left(\frac{I}{I_0}\right)^p \tag{6.4}$$

$$\dot{n} = 60 \cdot \left[\alpha(1-n) - \beta n\right] \tag{6.5}$$

$$\hat{B} = G\alpha(1-n) \tag{6.6}$$

Process P consists of circadian stimulus modulator M and circadian oscillator O, which models circadian phase and amplitude. M modulates $\hat{B}$, depending on current phase and amplitude of the circadian pacemaker, into drive B. P models the effect of drive B on oscillation speed and amplitude of the circadian pacemaker. O outputs two independent state variables x and x_c.

Depending on current circadian phase and amplitude, M modulates drive $\hat{B}$ into drive B according to Equation (6.7).

$$B = (1 - 0.4x)(1 - 0.4x_c)\hat{B} \tag{6.7}$$

Drive B influences state variables x and x_c as shown in Equations (6.8) and (6.9), where $\mu = 0.3$, $q = \frac{1}{3}$, $\tau_x = 24.2$, $k = 0.55$. Our implementation of the Kronauer model was developed using the description and parameters from Jewett et al. [55] and is available online[2].

$$\dot{x} = \left(\frac{\pi}{12}\right)\left[x_c + \mu\left(\frac{1}{3}x + \frac{4}{3}x^3 - \frac{256}{105}x^7\right) + B\right] \tag{6.8}$$

$$\dot{x_c} = \left(\frac{\pi}{12}\right)\left\{qBx_c - \left[\left(\frac{24}{0.99729\tau_x}\right)^2 + kB\right]x\right\} \tag{6.9}$$

Phase shift estimation

Daily circadian phase shift was estimated from measured and estimated light intensities, I_{true} and I_{pred}. We used the Kronauer model to compute state variables x and x_c. Daily minima of x were used as a phase marker and were extracted by applying peak detection on $-x$ using the peak detection implementation from the BMC toolbox [34]. Minimum distance between two peaks was set to 48 min to avoid false positives. Timing of daily minima t_{true} and t_{pred}, extracted from x_{true} and x_{pred}, were used as daily phase markers. Subsequently, daily phase shift error was defined as $e_i = t_{true,i} - t_{pred,i}$, where i denoted the day.

6.4 Evaluation

6.4.1 Passive light estimation app

Our passive light estimation app (PLEA) recorded smartphone sensor and weather data from Android smartphones. PLEA collected a sensor sample once every 30 s. PLEA implemented a service, which continuously ran in the background to ensure reliable recordings. Once enabled, PLEA was started automatically after a smartphone reboot. PLEA recorded current time, average acceleration magnitude, light and proximity sensor information, magnetometer data,

[2]Source code is available at `https://gitlab.com/fwahl/kronauermodel`.

screen state, WiFi SSID, and GPS information. GPS sensor information consisted of longitude, latitude, satellite count, signal-to-noise ratio (SNR). Longitude and latitude were used to retrieve local weather information, which consisted of weather code, cloudiness percentage, local sunrise and sunset timing. Furthermore, users annotated whether they were awake or asleep, and whether they were in an indoor or outdoor environment.

6.4.2 Study population

We recruited 12 participants (5 female, between 18 and 28 years old) from the University of Passau student population. Participants carried a Samsung Galaxy S III smartphone, wore a CamNtech MotionWatch 8, and a LightWatcher light sensor device mounted to an eyeglasses frame. Figure 6.3 depicts the sensor setup. Recordings were performed in three folds of four participants. Each participant recorded data for six consecutive days. We expected different light exposure patterns on weekends and weekdays due to different daily routines on work and free days. Participants were not asked to follow a script but to maintain their regular routines. During recordings, participants were unsupervised. We asked participants to charge the smartphone at least once per day. Participants were compensated with a 50 Euro Amazon voucher.

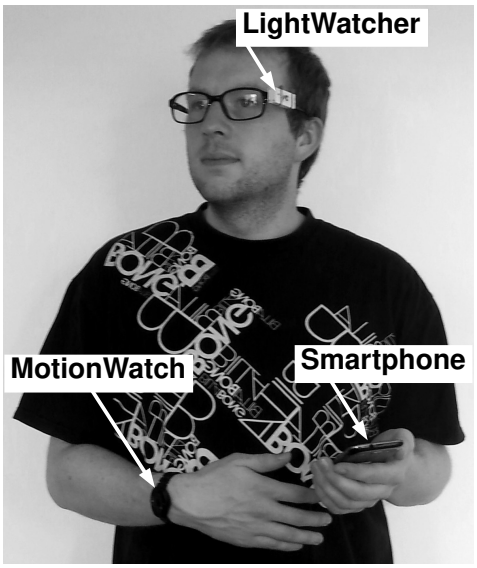


Figure 6.3: Study participant wearing Motionwatch, LightWatcher mounted to an eyeglasses frame, and carrying a smartphone.

6.4.3 Recording setup

Our passive light estimation smartphone app was used to record data as described in Section 6.4.1. Participants were reminded by a notification to send recorded data via email once per day. Light and activity were monitored at wrist level as baseline using a CamNtech MotionWatch 8, a common device for long-term circadian assessments in free living. Reference measurements were recorded using LightWatcher devices from the EUCLOCK research project. LightWatcher devices were mounted to an eyeglasses frame, as depicted in

Figure 6.3. Participants were outfitted with a pair of dioptre-free eyeglasses, if they did not wear eyeglasses usually. Participants were instructed to wear the eyeglasses throughout the day and place them next to their bed during the night. We converted LightWatcher irradiance measurements of each RGB channel to luminous intensity using the standardised luminosity function normalised to a 555 nm peak value. All devices were configured to record one sample every 30 s. In total, 1700 h of data were obtained.

6.4.4 Cross validation

All light estimation algorithms were evaluated using leave-one-participant-out cross validation. For each participant p, estimator training was performed on data of all other participants. Subsequently, estimator testing was performed using data from participant p.

6.4.5 Phase shift estimation error metrics

We computed two types of errors as evaluation metrics: Daily phase shift error and summed phase shift error. Daily phase shift was computed using daily minima of the Kronauer model state variable x as described in Section 6.3.3. Daily phase shift error describes the error made by an estimator compared to the head level reference measurements within one day of recordings and was computed as $e_i = t_{true,i} - t_{pred,i}$, where i denoted the day.

Summed phase shift error was computed to evaluate the estimation error over the entire recording period. Since daily errors add up, it is important to evaluate performance over multiple days. Summed phase shift error e_s was computed as $e_s = \sum_i e_i$. If daily estimation errors have a zero mean, they can cancel each other out, while one sided daily phase shift errors, e.g. always predicting a stronger shift forward than reference measurements, will add up over time.

6.5 Results

In this section, we analyse light intensity estimation performance of the PLEA app. We compare our estimations against measurements at eye level from the LightWatcher sensor, and wrist measurements from a MotionWatch 8. Subsequently, we compare daily and summed circadian phase shift estimations from light intensity estimations from the PLEA app against head and wrist measurements. Results from our 5 and 500 lux dummy estimators were included as baseline. Figure 6.4 shows average light intensity distribution over 24 hours for the entire data set.

6.5.1 Light intensity estimation

Light exposure information was required to estimate circadian phase shift. We estimated light exposure from phone context and weather information, and compared estimations against head reference measurements. Table 6.4 shows root mean square error (RMSe) across participants for each estimation method described in Section 6.3.2.

Extreme gradient boosted decision tree (XGB) classification showed best results with an RMSe of 3746 lux followed by XGB regression with 3818 lux RMSe. However, classification approaches used per-class means to obtain an estimate from light intensity class output. Regression directly estimates light intensities and did not need additional information. Dummy

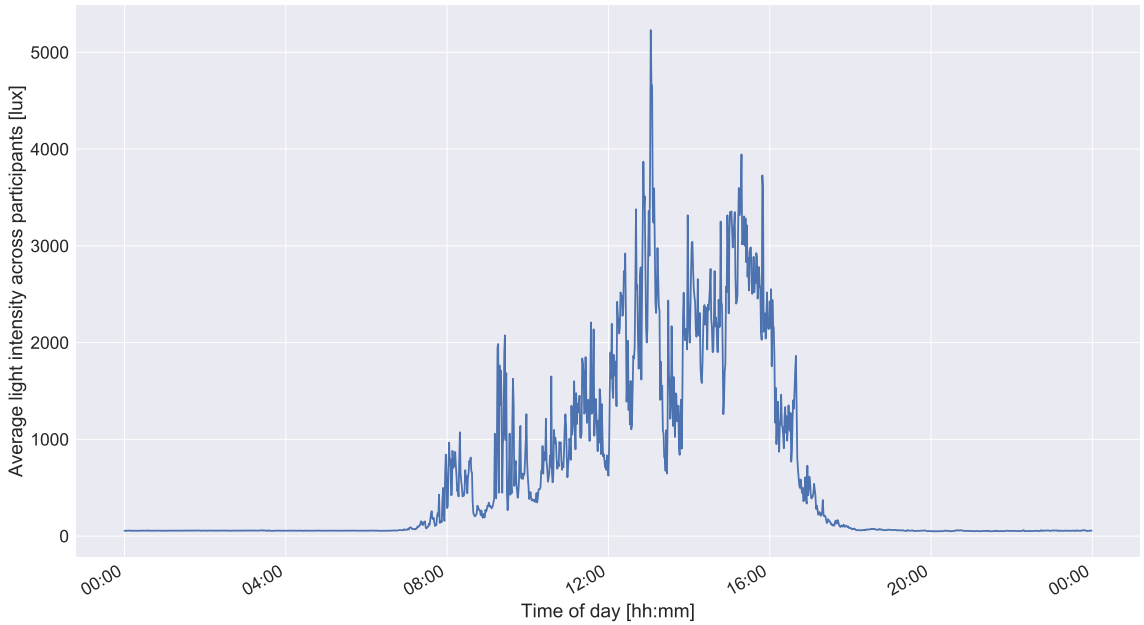


Figure 6.4: Average light intensity distribution over 24 hours for all participants. Light exposure peaked around 1300 hrs and in the afternoon around 1600 hrs.

Table 6.4: RMSe for all estimators. Extreme gradient boosted classification and regression approaches outperformed CART decision tree classification and wrist measurements.

Estimator	DT clf	XGB clf	XGB reg	Dummy 5	Dummy 500	Wrist
RMSe [lux]	4536	3746	3818	4046	4007	4238

estimators achieved RMSe results just above 4000 lux followed by wrist measurements and decision tree classification.

Wrist measurement RMSe was 12% higher compared to XGB classification. We found that wrist measurements were often lower than at head level. One reason might be sensor occlusion due to long sleeved clothing.

6.5.2 Daily circadian phase shift estimation

Figure 6.5 shows an example of light intensities measured at head level and estimated light intensities using the XGB regression approach in the top plot. XGB regression was able to follow the mean trend of measured light intensities. Light intensities and schedules in vary considerably in the three days displayed. The bottom plot shows the state variable x of the Kronauer model as described in Section 6.3.3. Short peaks in light intentisity were often missed by the regression model, which did not have large impact on the estimated phase shift.

After estimating light intensities, daily circadian phase shift was evaluated. Daily circadian phase shift errors were computed in minutes as described in Section 6.3.3. Figure 6.6 shows

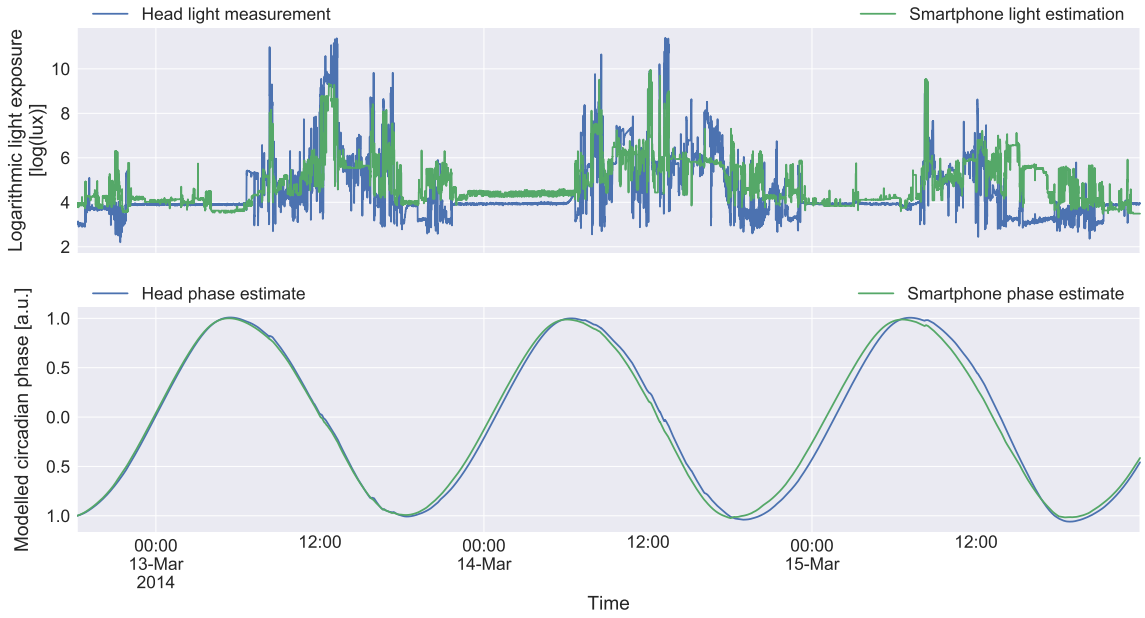


Figure 6.5: Light intensities measured at head level and estimated using our smartphone-only approach (top) over a three day period. Estimated phase according to the Kronauer model described in Section 6.3.3 (bottom) from light measurements and estimations. Smartphone light estimations reflect daily light exposure patterns of light reference measurements at head level well. Light estimation errors show only little effect on phase estimates.

daily phase shift error for all estimators described in Section 6.3.2. XGB regression performed best with 34.7±25.3 min phase shift estimation error. Decision tree and XGB classification had phase shift errors of 35.7±29.0 min and 35.7±28.2 min, respectively. While mean errors were almost equal, standard deviation was lowest for XGB regression. Wrist measurements showed phase estimation error of -52.4±40.0 min and a median error of -47.3 min. These large phase shift errors originate from frequent light intensity underestimations. Phase estimation errors of the 500 lux dummy estimator were more than twice on average compared to the 5 lux estimator, due to the impact of light on the circadian phase. Both did not perform well with phase shift errors of 63.5±42.9 min and 140.0±55.7 min for constant light estimations of 5 lux and 500 lux, respectively.

To investigate if daily phase shift error was similar for all participants, we evaluated daily phase shift estimation errors per participant for XGB regression and wrist measurements. We selected light estimation using XGB regression because it showed the lowest median error across participants. Wrist measurements were selected due to their popularity in circadian rhythm studies. Figure 6.7 shows absolute daily phase shift error per participant. Dashed lines indicate per estimator means. XGB regression had an absolute error of 35.5±24.2 min with a median of 30 min. For 10 out of 12 participants, light estimation with XGB regression median errors outperformed wrist measurements. Per participant absolute daily phase shift errors from light estimations using XGB regression were below the mean error of wrist measurements for

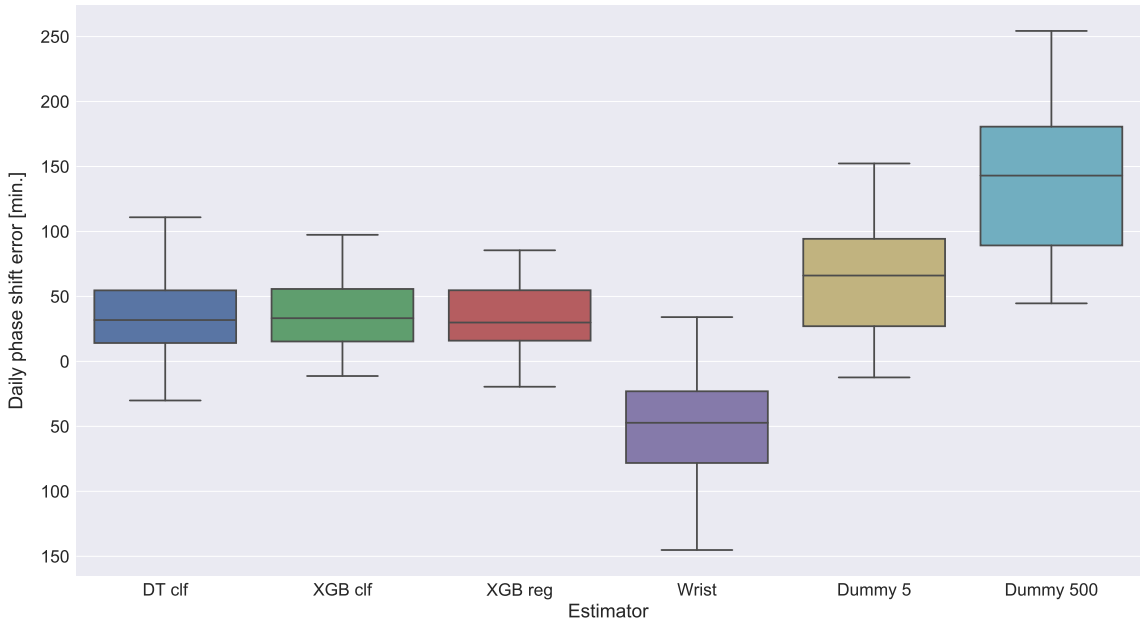


Figure 6.6: Daily phase shift estimation error in minutes for each estimator. All estimators using smartphone-based light estimation outperformed phase estimation from wrist measurements.

all participants. Absolute daily phase shift errors from wrist measurements had an error of 53.8±38.0 min, a 52% increase over XGB regression estimates. For participants 4-8 phase shift estimations from light measurements at the wrist were worse than for other participants. We assume those participants wore more long sleeved clothing than others resulting in a larger number of false measurements due to sensor occlusion.

6.5.3 Summed phase shift estimation

We evaluated summed phase shift estimation to investigate how phase shift estimation error accumulated over multiple days. To track circadian phase, daily circadian phase shifts need to be added up. Calculating summed phase shift estimation errors was important as small daily errors in the same shift direction could add up over time. Figure 6.8 shows summed phase shift error for each estimator. Phase shift estimation from light estimations based on decision tree classification performed best with an absolute error of 26.9±19.9 min. XGB classification and XGB regression performed similar with an absolute mean summed phase shift error of 38.0±23.5 min and 32.2±16.7 min, respectively. Phase estimations from wrist measurements had an absolute mean error of 57.2±24.3 min. Phase shift estimations from light estimations outperformed wrist measurement-based estimations because their daily error mean is less as shown in Figure 6.6. Both dummy estimators share the same problem with wrist measurements.

Finally, we investigated per participant phase shift error over the entire recording period. Figure 6.9 shows summed phase shift errors per participant for phase estimations from light

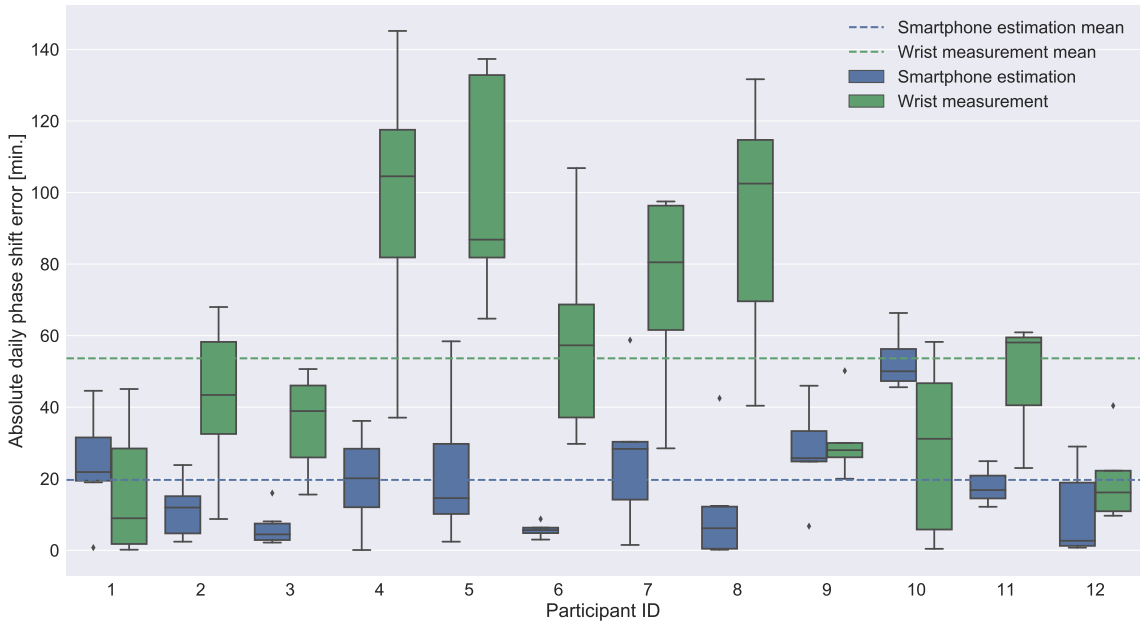


Figure 6.7: Absolute daily phase shift estimation error in minutes based on smartphone estimations and wrist measurements per participant. Phase estimations from smartphone light estimations had lower median errors compared those from wrist light measurement for 10 out of 12 participants.

estimations based on XGB regression and wrist measurements. XGB regression achieved an absolute mean error of 32.2 min over all participants, while wrist measurements had an absolute mean error of 57.2 min. Our smartphone-based light estimation outperformed wrist measurements for 11 out of 12 participants when analysing the total error over the recording duration. Low performance of phase shift estimations based on wrist measurements clearly show the advantage of estimations over measurements at a poor sensor location.

6.6 Discussion

6.6.1 Comparison with related work

Compared to related work summarised in Table 6.1, our evaluation study had the least number of participants. Our study protocol required participants to continuously wear the LightWatcher for head light reference measurements. Such a device can be stigmatising in public and made it difficult to recruit participants. In deployment, users would not have to wear any devices but just take their smartphone with them.

While related work relied on melatonin assessment for reference, we used head light measurements. Our focus was the impact of light intensity estimation errors on circadian phase estimations. Thus, we used the Kronauer model [55, 61] to estimate phase shift from light intensity for smartphone estimations, wrist and head measurements.

With a median ownership rate of 68% in developed countries, smartphones have become

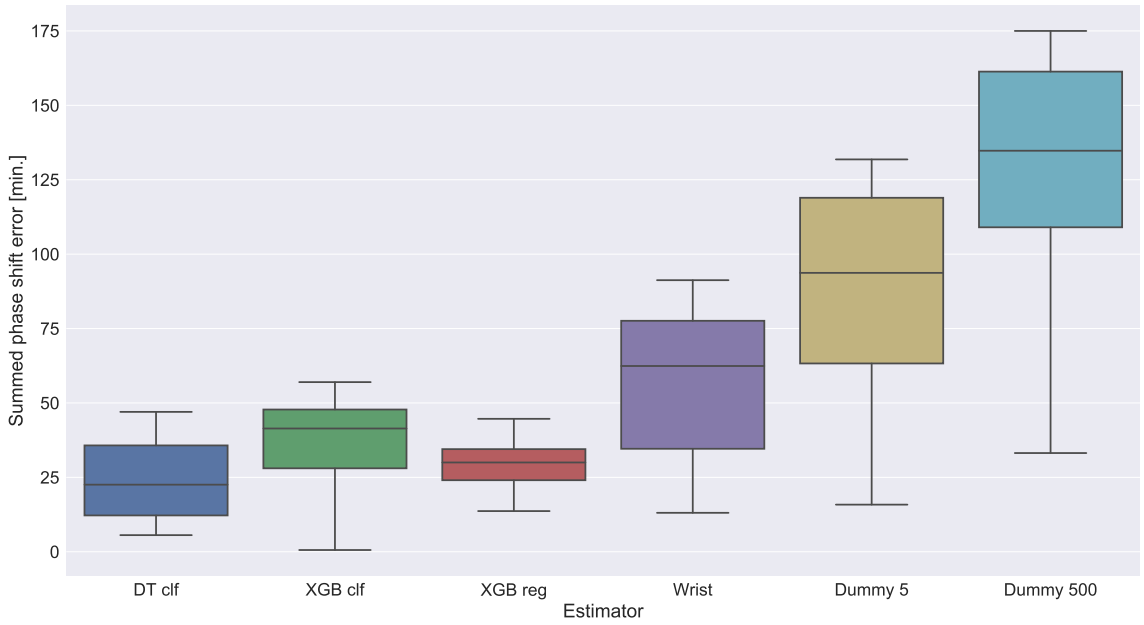


Figure 6.8: Absolute summed phase shift estimation error in minutes for each estimator. While CART decision tree classification achieved the lowest median error of all estimators, XGB regression estimates had smallest spread.

widely available [70]. In contrast, all approaches in Table 6.1 used proprietary research equipment limiting the availability to small study populations. In addition, our approach was the least intrusive in the comparison. Kolodyazhniy et al. [59, 60] achieved the best results but used 16 sensors, including a rectal probe. Gil et al. [43] used a combination of light measurements at the wrist and ECG to achieve similar results. For ECG measurements, three electrodes need to be mounted on the chest using sticky tape.

Compared to related work with less intrusive phase shift estimation methods, our approach has the lowest entry barrier as it could just be downloaded with no extra hardware required. However, as our approach estimates phase shift, it requires an initial phase estimate. Such an initial estimate could come from a questionnaire within the app or through another app, e.g. a sleep tracker. We investigated the estimation of sleep timing using smartphones [95]. To estimate circadian phase, sleep timing, alarm clock use, and work/free day information for at least one week are required. In combination, phase and phase shift estimation can reveal the influence of light exposure patterns on the circadian clock and support users in reaching a desired circadian phase.

6.6.2 Comparison with light measurements at wrist level

In circadian biology, actigraphy measurements at wrist level are commonly used in field studies. Manufacturers of actigraphs have added light sensors to their devices and thus, light measurements at wrist level have become common. Our results in Section 6.5 uncovered the shortcomings of light measurements at wrist level. Sensor occlusion due to long sleeved cloth-

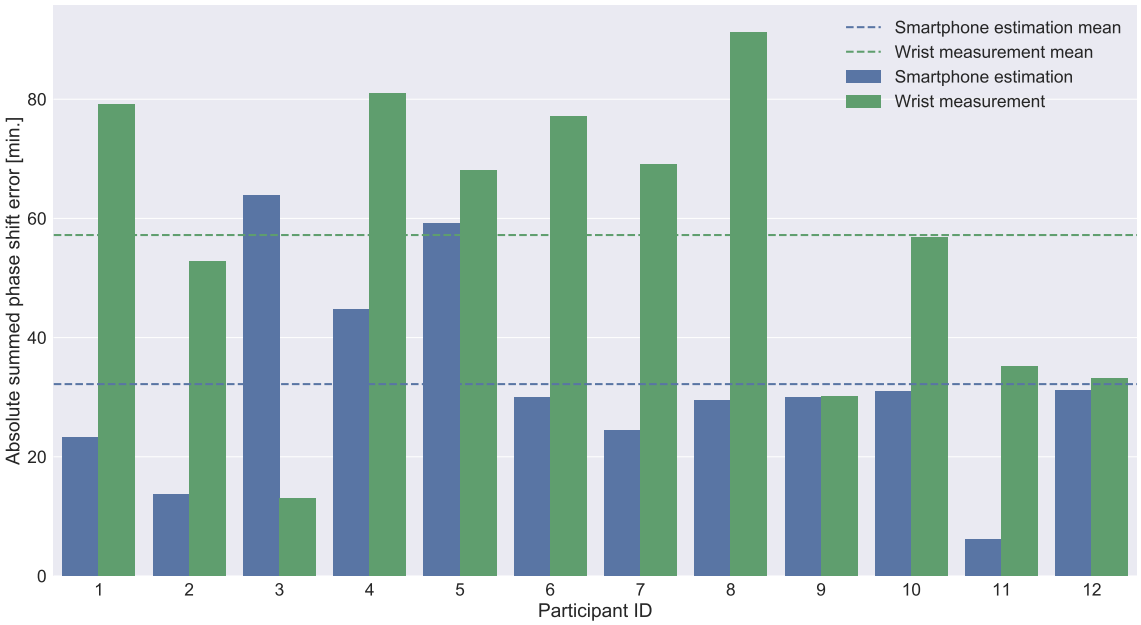


Figure 6.9: Absolute summed phase shift estimation error in minutes from XGB regression light estimations and wrist measurements. Our smartphone-only approach showed better phase estimation results than wrist measurements for all but one participant.

ing was responsible for large underestimation errors. In addition, wrist-worn sensors usually point in a different direction than the eyes, making measurements less accurate. In comparison, our smartphone-based estimations do not require additional hardware as smartphones are already widely available.

6.6.3 Limitations of smartphone light estimation

As shown in Section 6.5.2, circadian phase shift estimations from light estimations based on classification and regression-based light estimations performed similar. However, classification-based approaches use per-class mean light intensities derived from our light measurements at eye level to map classes to light intensities. For our study population, per-class means showed good results because they fitted the population well. If the same values were used in a entirely different population, e.g. from a different latitude, new per-class means would have to be recorded and used. Regression-based approaches do not need per-class means as they directly output estimated light intensities.

While smartphones are usually with their users, estimations were based on the location of the phone, not the user. If the phone would be in a different environment than its user for multiple hours, resulting estimations would not fit the user. With circadian shift estimations however, errors can add up over time as circadian shift intensity depends on the state of the circadian pacemaker. Smartphone motion could be analysed to ensure that the smartphone was with the user.

6.6.4 Benefits of smartphone light estimation

While smartphone-based light estimations cannot replace accurate measurements at eye level, they are an attractive alternative in large field studies or daily life. With light estimations from smartphones, participants could use their own devices in field studies, lowering per participant costs significantly. Compared to cheap but cumbersome pen and paper diaries or standardised questionnaires, smartphone sensor information provides more detailed insights.

For quantified self applications, circadian shift estimations could support users in improving their circadian alignment. In addition, gamification could be applied in order to motivate users to live a healthier lifestyle. By recommending light exposure at certain times, users could actively support circadian clock shift, e.g. when travelling to a different time zone.

6.6.5 Passive light estimation app

Our passive light estimation app (PLEA) was built as a prototype to facilitate sensor recordings. In a release version, the application would have to be optimised for energy consumption. Currently GPS sensor readings are taken every 30 s. Energy consumed per GPS sample depends on the time needed to aquire a position estimate. As most users are not constantly moving, power consumption could be reduced by dynamically adapting the frequency of GPS sensor readings. GPS sampling could be paused when accelerometer variance is low for extended periods of time or when the wireless network connection changes, e.g. from home WiFi to cellular network. Sensor updates would only be requested when a motion variance or network connection indicate a location change.

PLEA could ship with trained models and perform all computations on the smartphone in real-time. Privacy is essential for PLEA as it relies on GPS location data. With on-smartphone computations user privacy could be preserved, as no sensor data has to be sent to the cloud.

With on-smartphone estimations, real-time user feedback could be implemented. Similar to step counting applications, PLEA could provide users with circadian shift estimates and behaviour recommendations to shift their internal clock in either direction. For example, if users intend a backward shift, PLEA would recommend light exposure in the morning hours.

6.7 Conclusion

In this work we investigated circadian phase shift estimation based on continuous light exposure estimations from smartphone context data and cloud-sourced weather reports. Continuous light exposure was estimated using smartphone context and weather information. Estimated light exposure was compared against measurements at eye level. Subsequently, light exposure profiles were used as input to estimate circadian phase shift using the Kronauer model. Using XGB regression for light estimation, our approach could estimate daily phase shift with an error of 34.7±25.3 min resulting in a weekly error of 32.2±16.7 min compared to measurements at eye level. In addition, we compared our approach against measurements at the wrist, which are common in circadian biology field studies. For phase shift errors summed over the entire recording duration, our smartphone estimations outperformed wrist measurements for 11 out of 12 participants.

Compared to measurements at the eye, our smartphone-only approach was less invasive and more cost-efficient, as smartphones are already widely available. As result, studies using our proposed method could include more participants and cover longer study durations without

increasing their total cost. Furthermore, an end-user smartphone app could be built to enable users to track their circadian phase shift and support them in synchronising their internal and external clocks, resulting in a healthier lifestyle.

We conclude that smartphone context can be used to estimate daily light exposure profiles. In combination with models encapsulating expert knowledge, as the Kronauer model does for circadian phase shift, passive light exposure estimates can be used to obtain accurate estimates of circadian phase shift and thus enable users to live a healthier lifestyle.

Acknowledgements

This research was supported by the Dutch Technology Foundation STW under grant number 12184.

7

Conclusion

This thesis introduced devices, methods, and algorithms for monitoring circadian rhythms in free-living conditions. We provide three main contributions: (1) We integrated sensor technology into a customisable accessory for everyday use to improve social acceptance and reduce sensor noise, (2) performed context recognition on data from smart eyeglasses to infer daily routines and detect computer screen-use, (3) used existing domain expert models in combination with machine learning techniques to improve sleep timing estimation and estimate circadian phase shift from smartphone context information.

7.1 Personalised wearables

Social acceptance of wearables is key, as it influences wearer compliance. We deliberately chose to integrate sensors into an already socially accepted accessory, in our case eyeglasses. By building smart eyeglasses with a regular eyeglasses look, social acceptance can be maintained.

Wearing position of smart eyeglasses, on the head, is a prime sensor position: The head motion and orientation patterns during the day, can be used to infer activities of daily living. Subsequently, daily routines indicate phase and phase shift of the circadian clock. For light exposure recordings, light sensor position plays an important role. Sensor integration into eyeglasses allowed us to recorded light exposure information between the wearers eyes, yielding reliable light exposure information. The area behind the ears could be used to conveniently hide electronic and power components. In addition, health and fitness markers, such as chewing monitoring, heart rate, and others are also located on the head. Finally, the chance of neglecting eyeglasses are low, as eyeglasses wearers rely on their eyeglasses to enhance their vision.

In Chapter 2 we integrated sensor technology into a customisable eyeglasses frame model to create a novel wearable device. We provide a parameterised eyeglasses frame model, which can be fitted to the wearers head, minimising sensor noise by reducing mechanical play. By providing a parameterised model and using cost-efficient manufacturing techniques, our smart eyeglasses are personalised to the wearers head to increase acceptance. This is in contrast to existing wearables, which are usually one size fits all products. By using an additive manufacturing process, 3D printing, production cost is not affected. For smart eyeglasses we proposed a novel fitting method together with a parameterised 3D computer aided design (CAD) model. Our fitting method can be used to select optimal CAD model parameters with five head measurements. We printed fitted eyeglasses for four participants, which wear eyeglasses on a daily basis. Participants were asked to wear our smart eyeglasses for one day and rate wearing comfort and fit on a 5 point Likert scale. Overall, participants rated the eyeglasses fit at nose, ears, and frame width as comfortable and found them comfortable to wear overall (4.25 points out of 5). Compared to their own eyeglasses, participants reported similar wear comfort.

For circadian rhythm monitoring in free-living conditions, personalised smart eyeglasses offer a more suitable sensor location compared to wrist-worn actigraphs. Furthermore, our proposed smart eyeglasses record more detailed information compared to actigraphy. Knowledge of a persons context in addition to light exposure provides a holistic view beyond motion intensity.

7.2 Context recognition

We used smart eyeglasses to recognise activities of daily living (ADLs) in Chapter 3 and to detect computer screen use in Chapter 4. Our ADL detection was able to classify nine clusters

of typical ADLs with 77% average accuracy. Knowing when ADLs such as cleaning, screen work, reading, eating, etc. are performed gives a holistic insight into a subjects daily schedule. Thus, ADL recognition can be used to identify behaviour, which influences human circadian rhythms, e.g. the timing of food intake or outdoor activities. ADL recognition can raise the wearers awareness to schedule activities at specific time, e.g. to minimise social jet lag.

Screen use in the evening and night contributes to the suppression of melatonin production. Without melatonin, screen users will not feel tired in the evening and thus delay their sleep onset. With our smart eyeglasses, computer screen use was detected with over 0.9 ROC AUC in ambient light intensities below 200 lux. When screen use is detected, users could be informed and choose to turn off the screen or adjust the colour profile to minimise melatonin suppression.

7.3 Algorithms for circadian monitoring

In Chapters 5 and 6 we proposed novel algorithms for circadian monitoring in free-living conditions. In Chapter 5 we used smartphone context information to detect sleep opportunity. Sleep/wake schedules are key for computing chronotype and assessing social jet lag. Our proposed algorithm predicted sleep probability using a smartphone context classifier and a classifier based on time of day. For filtering and fusing sleep probabilities into sleep onset and wake up timing, we used the Two Process Model (TPM) by Daan et al. [31]. The integration of a domain expert model improved sleep onset time estimation results by 47% and wake up time estimation by 58% with absolute errors of 40±48 min for sleep onset and 42±57 min for wake up timing (relative errors of -2±62 min and -5±70 min, respectively). Our proposed method showed similar error as pen and paper sleep diaries but requires less effort, as the application just runs in the background. In addition, users can personalise the recognition model by providing feedback on their sleep timing.

In Chapter 6 we used smartphone context data in combination with cloud-sourced weather reports to estimate circadian shift. While smart eyeglasses and actigraphs can be used to measure light exposure, they can not match smartphone availability. We first estimated light intensity using classification and regression techniques. Subsequently, we used the Kronauer model by Jewett et al. [55] to estimate circadian shift from light exposure estimations. Our smartphone estimation approach had a circadian phase shift error of 32±17 min over a 6 day recording duration. Smartphone estimations outperformed wrist-worn actimetry measurements for 11 out of 12 participants. Phase shift estimation could help users understand how an activity will influence their circadian phase, e.g. users could see how a long walk in the afternoon would shift their circadian phase. Thus, a circadian phase shift estimation app could support users in living a healthier lifestyle by synchronising their internal and external clocks.

7.4 Outlook

The wearables, applications, and methods proposed in this thesis could be combined to help users and researchers to understand the internal clock and its Zeitgebers (German for "time givers"). While measurements from smart eyeglasses for the masses are not reality yet, smartphones are already widely available today. In the future, our calendar schedule might be

optimised to our circadian rhythm. For example, when we schedule a workout in the gym, our calendar app might suggest the time of day, during which we are at our physiological peak.

7.4.1 Personalised wearables

Wearables can integrate electronics into already existing accessories, e.g. eyeglasses. Personalised fitting will increase wearer compliance in the long term and reduce noise in recorded sensor data due to improved fit. In the future, personalised smart eyeglasses could replace one-size fits all eyeglasses frames. New techniques for embedding hardware in a production scale into individualised wearables need to be established. Furthermore, wearables could be extended beyond eyeglasses, e.g. by printing circuits onto glueable patches. Personalised, miniaturised wearables could be used for continuous monitoring of physiological signals in free-living conditions.

Smart eyeglasses could advance interdisciplinary research, e.g. with medicine or sociology, by using smart eyeglasses for ambient assisted living research. The majority of elderly already wear eyeglasses to enhance their vision, which should improve compliance compared to other wearable sensor solutions. Smart eyeglasses could be used to identify wearer context throughout the day. With continuous remote monitoring, elderly users could live independent in their homes for a longer time, reducing healthcare cost. As most eyeglasses wearers place their eyeglasses in the same place overnight, a wireless charging mat or charging case could be developed to avoid running out of battery.

To make the proposed 3D-printed smart eyeglasses mass compatible, further research on the combination of electronics and eyeglasses frame is needed. Coping with different wire lengths due to personalisation in mass manufacturing is an important research question. It may be feasible to print conductive and non-conductive materials, eliminating the need for a wire harness entirely. Printed wires would also make connectors obsolete, making the integration of electronics easier.

Wearables are closer to their user than smartphones, thus can capture more detailed information about their user, e.g. physiological data, such as heart rate. Further research on the integration of new sensor modalities into wearables will make smart accessories even more useful, e.g. smart eyeglasses or rings to monitor blood sugar levels for diabetic patients. Context recognition from wearable sensor data provides holistic insight into daily routines. However, consumers could benefit from a novel class of circadian applications running on the smartphones they already own.

7.4.2 Circadian applications

Circadian apps building on the work presented in this thesis could support users in tuning their circadian phase to a desired state, e.g. by suggesting behavioural adjustments to reduce social jet lag. The combination of sleep opportunity detection introduced in Chapter 5 and light estimation from smartphone context introduced in Chapter 6 could enable circadian phase tracking in free living conditions for a large user base.

Continuous monitoring of the circadian phase using smartphones would advance chronobiology research in free-living. The MCTQ database contains approximately 200.000 entries and has advanced chronobiology research from questionnaires only. While questionnaires provide a single frame only, continuous monitoring captures a motion picture. For example, the MCTQ asks for a subject's usual bedtime on weekdays, a smartphone application could continuously

estimate it from smartphone context. A database of continuously monitored individuals of similar size, e.g. from a smartphone application, would offer many possibilities for further research, e.g. understanding the relationship between circadian phase and performing certain ADLs.

Based on collected data, personalised feedback could be offered to the user in order to reach a desired circadian phase. Machine learning could be used to learn which type of feedback users respond to best. Another application could aim at scheduling activities in a best effort to synchronise internal and external clocks. A circadian alarm clock could wake users at an optimal time according to their current circadian phase. Fitness apps could also benefit from circadian awareness by recommending the best time of day for sports. Furthermore, a circadian fitness app could encourage users to perform outdoor activities at times, where exposure to sunlight would support shift circadian phase in the desired direction.

7.4.3 Expert models

Research on the inclusion of expert knowledge could be extended into other areas, e.g. kinetic models could be used in camera-based activity recognition. While current trends in machine learning seem to favour the exclusion of expert knowledge (e.g. deep learning), we believe that the inclusion of expert models as shown in Chapters 5 and 6 can be extended to other application areas. In contrast to neural networks, which attempt to learn the structure of the model from training data alone, the combination of classic machine learning algorithms and expert models yields comprehensible results.

In Chapter 5, expert models were used as filter technique for classifier outputs. Similarly, expert models could be used to compute a confidence measure for predictions from a deep learning model. Since it is challenging to validate deep learning models with reasonable effort, a confidence estimate could help to assess model output. As most existing expert models are computationally inexpensive to run, the overhead of an expert model would be small compared to the deep learning model. In addition, expert models could be used to fuse outputs of multiple deep learning models.

Training a deep learning model from scratch is computationally expensive and requires a substantial amount of training data. In transfer learning, an already trained deep learning model is adapted to a new data set by performing less training iterations. Expert models could be used to measure the adaptation rate of the model. During adaption of a deep learning model, the output should converge to an expert model output. Furthermore, expert models can be used to transform the results into another domain, as suggested in Chapter 6. With a transformation, it would be possible to set the stopping criterion during training in the correct domain.

Bibliography

[1] EYEGLASSES Trends 2016: What To Wear? http://thefashiontag.com/2016/03/15/eyeglasses-trends-2016-what-to-wear/, 2016.

[2] S. Abdullah, M. Matthews, E. L. Murnane, G. Gay, and T. Choudhury. Towards circadian computing: Early to bed and early to rise makes some of us unhealthy and sleep deprived. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 673–684. ACM, 2014.

[3] a. AG. TCS34725. http://ams.com/eng/Products/Light-Sensors/Color-Sensor/TCS34725, 2013.

[4] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster. Analysis of Chewing Sounds for Dietary Monitoring. In *UbiComp 2005: Proceedings of the 7th International Conference on Ubiquitous Computing*, volume 3660 of *LNCS*, pages 56–72. Springer Berlin, Heidelberg, Sept. 2005.

[5] O. Amft, F. Wahl, S. Ishimaru, and K. Kunze. Making Regular Eyeglasses Smart. *Pervasive Computing, IEEE*, 14(3):32–43, 2015.

[6] L. Atallah, B. Lo, R. King, and G.-Z. Yang. Sensor placement for activity detection using wearable accelerometers. In *Body Sensor Networks, International Conference On*, pages 24–29. IEEE, 2010.

[7] O. Aziz, B. Lo, R. King, A. Darzi, and G.-Z. Yang. Pervasive body sensor network: An approach to monitoring the post-operative surgical patient. In *Wearable and Implantable Body Sensor Networks. International Workshop On*, pages 4–pp. IEEE, 2006.

[8] G. Azkune, A. Almeida, D. López-de-Ipiña, and L. Chen. Extending knowledge-driven activity models through data-driven learning techniques. *Expert Systems with Applications*, 42(6):3115–3128, 2015.

[9] R. T. Azuma and others. A survey of augmented reality. *Presence*, 6(4):355–385, 1997.

[10] Y. Bababekova, M. Rosenfield, J. E. Hue, and R. R. Huang. Font size and viewing distance of handheld smart phones. *Optom Vis Sci*, 88(7):795–797, July 2011.

[11] D. Bannach, P. Lukowicz, and O. Amft. Rapid Prototyping of Activity Recognition Applications. *IEEE Pervasive Computing*, 7(2):22–31, April–June 2008.

[12] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, pages 1–17. Springer, 2004.

[13] A. Bierman, T. R. Klein, and M. S. Rea. The Daysimeter: A device for measuring optical radiation as a stimulus for the human circadian system. *Measurement Science and Technology*, 16(11):2292–2299, Nov. 2005.

[14] C. M. Bishop. *Pattern Recognition and Machine Learning*. springer, 2006.

[15] M. Bohanec and B. Delibašić. Data-mining and expert models for predicting injury risk in ski resorts. In *International Conference on Decision Support System Technology*, pages 46–60. Springer, 2015.

[16] M. A. Bonmati-Carrion, B. Middleton, V. Revell, D. J. Skene, M. A. Rol, and J. A. Madrid. Circadian phase assessment by ambulatory monitoring in humans: Correlation with dim light melatonin onset. *Chronobiol. Int.*, 31(1):37–51, Feb. 2014.

[17] M. Borazio and K. V. Laerhoven. Predicting Sleeping Behaviors in Long-Term Studies with Wrist-Worn Sensor Data. In *AMI 2011: Proceedings of the International Joint Conference on Ambient Intelligence*, volume 7040 of *Lecture Notes in Computer Science*, pages 151–156. Springer, 2011.

[18] M. Borazio and K. Van Laerhoven. Combining wearable and environmental sensing into an unobtrusive tool for long-term sleep studies. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 71–80. ACM, 2012.

[19] A. A. Borbély and P. Achermann. Sleep Homeostasis and Models of Sleep Regulation. *J Biol Rhythms*, 14(6):559–570, Dec. 1999.

[20] G. C. Brainard, J. P. Hanifin, J. M. Greeson, B. Byrne, G. Glickman, E. Gerner, and M. D. Rollag. Action Spectrum for Melatonin Regulation in Humans: Evidence for a Novel Circadian Photoreceptor. *J. Neurosci.*, 21(16):6405–6412, Aug. 2001.

[21] J. Brooke. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[22] C. Cajochen, S. Frey, D. Anders, J. Späti, M. Bues, A. Pross, R. Mager, A. Wirz-Justice, and O. Stefani. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. *Journal of Applied Physiology*, 110(5):1432–1438, May 2011.

[23] A. Campbell. How to hold your iPad. https://alastairc.ac/2013/02/how-to-hold-your-ipad/, 2013.

[24] Centers for Disease Control and Prevention (CDC). Perceived insufficient rest or sleep among adults - United States, 2008. *MMWR Morb. Mortal. Wkly. Rep.*, 58(42):1175–1179, Oct. 2009.

[25] A.-M. Chang, D. Aeschbach, J. F. Duffy, and C. A. Czeisler. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *PNAS*, 112(4):1232–1237, Jan. 2015.

[26] T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[27] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell. Unobtrusive sleep monitoring using smartphones. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference On*, pages 145–152. IEEE, 2013.

[28] E. K. Choe, S. Consolvo, N. F. Watson, and J. A. Kientz. Opportunities for computing technologies to support healthy sleep behaviors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3053–3062. ACM, 2011.

[29] Cisco. Cisco Connected World Technology Report. http://www.cisco.com/c/en/us/solutions/enterprise/connected-world-technology-report/index.html, 2014.

[30] J. Clawson, J. A. Pater, A. D. Miller, E. D. Mynatt, and L. Mamykina. No longer wearing: Investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 647–658. ACM, 2015.

[31] S. Daan, D. G. Beersma, and A. A. Borbély. Timing of human sleep: Recovery process gated by a circadian pacemaker. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 246(2):R161–R183, Feb. 1984.

[32] Deutsche Gesetzliche and Unfallversicherung e.V. (DGUV). DGUV Information 215-410 - Bildschirm- und Büroarbeitsplätze Leitfaden für die Gestaltung (DGUV Information 215-410), Sept. 2015.

[33] D. L. DiLaura, K. W. Houser, R. G. Mistrick, and G. R. Steffy. *The Lighting Handbook: Reference and Application*. Illuminating Engineering Society of North America New York, 2011.

[34] M. Duarte. Notes on Scientific Computing for Biomechanics and Motor Control. *GitHub repository*, 2015.

[35] J. F. Duffy and K. P. Wright. Entrainment of the human circadian system by light. *J. Biol. Rhythms*, 20(4):326–338, Aug. 2005.

[36] eMarketer. US Adults Spend 5.5 Hours with Video Content Each Day - eMarketer. http://www.emarketer.com/Article/US-Adults-Spend-55-Hours-with-Video-Content-Each-Day/1012362, Oct. 2016.

[37] FAA USDOT. Human Factors Design Standard (HFDS). *Rep. DOT/FAA/CT-03/05 HF-STD-001*, 2003.

[38] T. Fawcett. ROC Graphs: Notes and Practical Considerations for Researchers. Technical report, 2004.

[39] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.

[40] M. Figueiro, R. Hamner, A. Bierman, and M. Rea. Comparisons of three practical field devices used to measure personal light exposures and activity levels. *Lighting Research and Technology*, 45(4):421–434, 2013.

[41] A. Fleury, M. Vacher, and N. Noury. SVM-Based Multimodal Classification of Activities of Daily Living in Health Smart Homes: Sensors, Algorithms, and First Experimental Results. *IEEE Trans Inf Technol Biomed*, 14(2):274–283, 2010.

[42] F. Flouvat, J. Sanhes, C. Pasquier, N. Selmaoui-Folcher, and J.-F. Boulicaut. Improving pattern discovery relevancy by deriving constraints from expert models. In *Proceedings of the Twenty-First European Conference on Artificial Intelligence*, pages 327–332. IOS Press, 2014.

[43] E. A. Gil, X. L. Aubert, E. I. S. Møst, and D. G. M. Beersma. Human circadian phase estimation from signals collected in ambulatory conditions using an autoregressive model. *J. Biol. Rhythms*, 28(2):152–163, Apr. 2013.

[44] Y. Gitman and J. Murphy. Pulse Sensor-A wearable device to give your projects a live heartbeat. *Make-Technology on Your Time*, 29:52, 2012.

[45] M. Heath, C. Sutherland, K. Bartel, M. Gradisar, P. Williamson, N. Lovato, and G. Micic. Does one hour of bright or short-wavelength filtered tablet screenlight have a meaningful effect on adolescents' pre-bedtime alertness, sleep, and daytime functioning? *Chronobiology International*, 31(4):496–505, May 2014.

[46] J. Hernandez, Y. Li, J. M. Rehg, and R. W. Picard. BioGlass: Physiological parameter estimation using a head-mounted wearable device. In *2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth)*, pages 55–58, Nov. 2014.

[47] J. Hernandez and R. W. Picard. SenseGlass: Using Google Glass to sense daily emotions. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology*, pages 77–78. ACM, 2014.

[48] M. Hirshkowitz, K. Whiton, S. M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, P. J. A. Hillard, E. S. Katz, L. Kheirandish-Gozal, D. N. Neubauer, A. E. O'Donnell, M. Ohayon, J. Peever, R. Rawding, R. C. Sachdeva, B. Setters, M. V. Vitiello, and J. C. Ware. National Sleep Foundation's updated sleep duration recommendations: Final report. *Sleep Health: Journal of the National Sleep Foundation*, 1(4):233–243, Dec. 2015.

[49] J. A. Horne and O. Östberg. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology*, 1976.

[50] Institut für Demoskopie Allensbach. Brillenstudie 2014. Technical report, 2014.

[51] S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling. In the blink of an eye: Combining head motion and eye blink frequency for activity recognition with google glass. In *Proceedings of the 5th Augmented Human International Conference*, page 15. ACM, 2014.

[52] S. Ishimaru, K. Kunze, Y. Uema, K. Kise, M. Inami, and K. Tanaka. Smarter Eyewear: Using Commercial EOG Glasses for Activity Recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 239–242, New York, NY, USA, 2014. ACM.

[53] P. Jaramillo-Garcia, L. I. Lopera Gonzalez, and O. Amft. Using implicit user feedback to balance energy consumption and user comfort of proximity-controlled computer screens. *Journal of Ambient Intelligence and Humanized Computing*, 6(2):207–221, Feb. 2014.

[54] D. Jarchi, B. Lo, E. Ieong, D. Nathwani, and G.-Z. Yang. Validation of the e-AR sensor for gait event detection using the Parotec foot insole with application to post-operative recovery monitoring. In *Proceedings of 11th International Conference on Wearable and Implantable Body Sensor Networks*, pages 127–131. IEEE, 2014.

[55] M. E. Jewett, D. B. Forger, and R. E. Kronauer. Revised limit cycle oscillator model of human circadian pacemaker. *Journal of Biological Rhythms*, 14(6):493–500, 1999.

[56] T. Kantermann and T. Roenneberg. Is light-at-night a health risk factor or a health risk predictor? *Chronobiol Int*, 26(6):1069–1074, Aug. 2009.

[57] M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Consolvo, and J. A. Kientz. Lullaby: A capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 226–234. ACM, 2012.

[58] M. Koelle, M. Kranz, and A. Möller. Don'T Look at Me That Way!: Understanding User Attitudes Towards Data Glasses Usage. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pages 362–372, New York, NY, USA, 2015. ACM.

[59] V. Kolodyazhniy, J. Späti, S. Frey, T. Götz, A. Wirz-Justice, K. Kräuchi, C. Cajochen, and F. H. Wilhelm. Estimation of human circadian phase via a multi-channel ambulatory monitoring system and a multiple regression model. *J. Biol. Rhythms*, 26(1):55–67, Feb. 2011.

[60] V. Kolodyazhniy, J. Späti, S. Frey, T. Götz, A. Wirz-Justice, K. Kräuchi, C. Cajochen, and F. H. Wilhelm. An improved method for estimating human circadian phase derived from multichannel ambulatory monitoring and artificial neural networks. *Chronobiol. Int.*, 29(8):1078–1097, Oct. 2012.

[61] R. E. Kronauer, C. A. Czeisler, S. F. Pilato, M. C. Moore-Ede, and E. D. Weitzman. Mathematical model of the human circadian system with two interacting oscillators. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 242(1):R3–R17, 1982.

[62] D.-S. Lee. Preferred viewing distance of liquid crystal high-definition television. *Applied ergonomics*, 43(1):151–156, 2012.

[63] W. McKinney. Pandas: A Foundational Python Library for Data Analysis and Statistics. *Python for High Performance and Scientific Computing*, pages 1–9, 2011.

[64] J.-K. Min, A. Doryab, J. Wiese, S. Amini, J. Zimmerman, and J. I. Hong. Toss'n'turn: Smartphone as sleep and sleep quality detector. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 477–486. ACM, 2014.

[65] J. A. Mohawk, C. B. Green, and J. S. Takahashi. Central and peripheral circadian clocks in mammals. *Annual review of neuroscience*, 35:445–462, 2012.

[66] R. Nandakumar, S. Gollakota, and N. Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 45–57. ACM, 2015.

[67] J. G. Nathan, D. R. Anderson, D. E. Field, and P. Collins. Television viewing at home: Distances and visual angles of children and adults. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 27(4):467–476, 1985.

[68] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[69] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.

[70] J. Poushter. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center*, 22, 2016.

[71] PresseBox. Deutsche sitzen lange vor dem Computer - BITKOM - Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. - Pressemitteilung. http://www.pressebox.de/pressemitteilung/bitkom-bundesverband-informationswirtschaft-telekommunikation-und-neue-medien-ev/Deutsche-sitzen-lange-vor-dem-Computer/boxid/200275, 2008.

[72] S. A. Rahman, C. Merck, Y. Huang, and S. Kleinberg. Unintrusive Eating Recognition Using Google Glass. In *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '15, pages 108–111. ICST, 2015.

[73] T. Rahman, A. T. Adams, R. V. Ravichandran, M. Zhang, S. N. Patel, J. A. Kientz, and T. Choudhury. Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 39–50. ACM, 2015.

[74] V. L. Revell and C. I. Eastman. How to trick mother nature into letting you fly around or stay up all night. *J Biol Rhythms*, 20(4):353–365, Aug. 2005.

[75] T. Roenneberg, K. V. Allebrandt, M. Merrow, and C. Vetter. Social Jetlag and Obesity. *Current Biology*, 22(10):939–943, May 2012.

[76] T. Roenneberg, T. Kantermann, M. Juda, C. Vetter, and K. Allebrandt. Light and the Human Circadian Clock. In A. Kramer and M. Merrow, editors, *Circadian Clocks*, volume 217 of *Handbook of Experimental Pharmacology*, pages 311–331. Springer Berlin Heidelberg, 2013.

[77] T. Roenneberg, T. Kuehnle, M. Juda, T. Kantermann, K. Allebrandt, M. Gordijn, and M. Merrow. Epidemiology of the human circadian clock. *Sleep Medicine Reviews*, 11(6):429–438, Dec. 2007.

[78] T. Roenneberg, A. Wirz-Justice, and M. Merrow. Life between clocks: Daily temporal patterns of human chronotypes. *Journal of biological rhythms*, 18(1):80–90, 2003.

[79] M. Rofouei, M. Sinclair, R. Bittner, T. Blank, N. Saw, G. DeJean, and J. Heffron. A non-invasive wearable neck-cuff system for real-time sleep monitoring. In *Body Sensor Networks (BSN), 2011 International Conference On*, pages 156–161. IEEE, 2011.

[80] A. E. Rogers, C. C. Caruso, and M. S. Aldrich. Reliability of sleep diaries for assessment of sleep/wake patterns. *Nursing research*, 42(6):368–371, 1993.

[81] S. Saeb, T. R. Cybulski, S. M. Schueller, K. P. Kording, and D. C. Mohr. Scalable Passive Sleep Monitoring Using Mobile Phones: Opportunities and Obstacles. *J. Med. Internet Res.*, 19(4):e118, 04 18, 2017.

[82] J. Seiter. *Topic Models for Activity Discovery in Daily Life*. PhD thesis, ETH Zurich, 2015.

[83] K.-K. Shieh and D.-S. Lee. Preferred viewing distance and screen angle of electronic paper displays. *Applied Ergonomics*, 38(5):601–608, Sept. 2007.

[84] D. P. Solomatine and A. Ostfeld. Data-driven modelling: Some past experiences and new approaches. *Journal of hydroinformatics*, 10(1):3–22, 2008.

[85] G. Spina, F. Roberts, J. Weppner, P. Lukowicz, and O. Amft. CRNTC+: A smartphone-based sensor processing framework for prototyping personal healthcare applications. In *PH 2013: Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pages 252–255. IEEE, 2013.

[86] E. Tamaki, T. Miyaki, and J. Rekimoto. Brainy Hand: An Ear-worn Hand Gesture Interaction Device. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, pages 4255–4260, Boston, MA, USA, 2009. ACM.

[87] The Vision Council of America. Digital Eye Strain Report 2016. Technical report, 2016.

[88] The Vision Council of America. OTC Readers Report Q4 2015. Technical report, 2016.

[89] S. van der Lely, S. Frey, C. Garbazza, A. Wirz-Justice, O. G. Jenni, R. Steiner, S. Wolf, C. Cajochen, V. Bromundt, and C. Schmidt. Blue Blocker Glasses as a Countermeasure for Alerting Effects of Evening Light-Emitting Diode Screen Exposure in Male Teenagers. *Journal of Adolescent Health*, 56(1):113–119, Jan. 2015.

[90] V. van der Vinne, G. Zerbini, A. Siersema, A. Pieper, M. Merrow, R. A. Hut, T. Roenneberg, and T. Kantermann. Timing of examinations affects school performance differently in early and late chronotypes. *Journal of biological rhythms*, 30(1):53–60, 2015.

[91] S. van der Walt, S. C. Colbert, and G. Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.

[92] E. Vulliamy. These pictures show why you should never charge your phone under your pillow. http://www.independent.co.uk/news/world/americas/the-pictures-that-show-why-you-should-never-charge-your-phone-under-your-pillow-a6883796.html, 2016.

[93] F. Wahl and O. Amft. Using RFID tags as reference for phone location and orientation in daily life. In *Proceedings of the 4th Augmented Human International Conference*, pages 194–197. ACM, 2013.

[94] F. Wahl and O. Amft. Personalised phone placement recognition in daily life using RFID tagging. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference On*, pages 19–26. IEEE, 2014.

[95] F. Wahl and O. Amft. Data and Expert Models for Sleep Timing and Chronotype Estimation from Smartphone Context Data and Simulations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3):139:1–139:28, Sept. 2018.

[96] F. Wahl, M. Freund, and O. Amft. Using smart eyeglasses as a wearable game controller. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 377–380. ACM, 2015.

[97] F. Wahl, M. Freund, and O. Amft. WISEglass: Multi-purpose context-aware smart eyeglasses. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 159–160. ACM, 2015.

[98] F. Wahl, M. Freund, and O. Amft. WISEglass: Smart eyeglasses recognising context. In *Bodynets*, 2015.

[99] F. Wahl, M. Freund, and O. Amft. WISEglass: Smart eyeglasses recognising context. *EAI Endorsed Transactions on Pervasive Health and Technology*, 16(5), 2016.

[100] F. Wahl, T. Kantermann, and O. Amft. How much light do you get? Estimating daily light exposure using smartphones. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, pages 43–46. ACM, 2014.

[101] F. Wahl, J. Kasbauer, and O. Amft. Computer Screen Use Detection Using Smart Eyeglasses. *Front. ICT*, 4, 2017.

[102] F. Wahl, R. Zhang, M. Freund, and O. Amft. Personalizing 3D-printed smart eyeglasses to augment daily life. *IEEE Computer*, 50(2):26–35, 2017.

[103] O. J. Walch, A. Cochran, and D. B. Forger. A global quantification of "normal" sleep schedules using smartphone data. *Science Advances*, 2(5):e1501705, May 2016.

[104] M. Wittmann, J. Dinich, M. Merrow, and T. Roenneberg. Social Jetlag: Misalignment of Biological and Social Time. *Chronobiology International*, 23(1-2):497–509, Jan. 2006.

[105] T. Woelders, D. G. M. Beersma, M. C. M. Gordijn, R. A. Hut, and E. J. Wams. Daily light exposure patterns reveal phase and period of the human circadian clock. *Journal of Biological Rhythms*, Apr. 2017.

[106] B. Wood, M. S. Rea, B. Plitnick, and M. G. Figueiro. Light level and duration of exposure determine the impact of self-luminous tablets on melatonin suppression. *Applied Ergonomics*, 44(2):237–240, Mar. 2013.

[107] R. Zhang, S. Bernhart, and O. Amft. Diet eyeglasses: Recognising food chewing using EMG and smart eyeglasses. In *Proceedings of the International Conference on Wearable and Implantable Body Sensor Networks (BSN' 16)*, pages 7–12. IEEE, 2016.

[108] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi. Learning sleep stages from radio signals: A conditional adversarial architecture. In *International Conference on Machine Learning*, pages 4100–4109, 2017.

[109] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 193–202. IEEE, 2008.