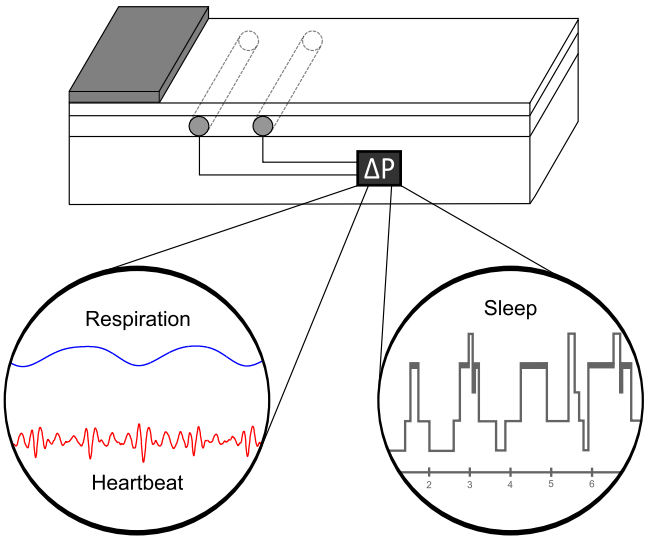
KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Biomechanics based analysis of sleep



Tim Willemen

Supervisor:
Prof. dr. ir. S. Van Huffel
Prof. dr. ir. J. Vander Sloten
Prof. dr. ir. B. Haex

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering

February 2015

Biomechanics based analysis of sleep

Tim WILLEMEN

Examination committee:
Prof. dr. ir. P. Van Houtte, chair
Prof. dr. ir. S. Van Huffel, supervisor
Prof. dr. ir. J. Vander Sloten, supervisor
Prof. dr. ir. B. Haex, supervisor
Prof. dr. ir. H. van Lenthe
Prof. dr. phys. B. Buyse
Prof. dr. ir. B. Vanrumste
Prof. dr. ir. R. M. Aarts
(TU Eindhoven)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering

February 2015

© 2015 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Tim Willemen, Celestijnenlaan 200C box 2419 - Kasteelpark Arenberg 10 box 2446, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

ISBN 978-94-6018-962-3

D/2015/7515/18

Dankwoord

Beste lezer,

Dank voor uw interesse om dit doctoraatswerk ter hand te nemen. Ik hoop dat de tekst u kan inwijden in de visies en conclusies die ik de afgelopen vier jaar heb opgebouwd. Weet echter dat het geen sinecure was om dit werk te voltooien. Ook ik heb zoals velen van mijn voorgangers met de gekende motivatieachtbaan binnen een doctoraat moeten afrekenen.

Je begint vol optimisme aan je nieuwe uitdaging, je verwachtingen zijn groot. Jouw doctoraatswerk gaat een verschil maken! Helaas realiseer je na verloop van tijd hoe groot en uitgebreid jouw onderzoeksdomein wel is, en hoe moeilijk het zal worden om hier een significante bijdrage aan toe te voegen. Je voelt je gevangen in een doolhof van onafgewerkte onderzoeksprojecten die volgens jou onsamenhangend, nutteloos en een tijdsverspilling waren. Je hebt op dit moment dringend steun nodig, of je raakt de weg voor eeuwig kwijt. Weeral een doctoraatsthesis die dreigt naar de prullenmand verwezen te worden. Maar omwille van de juiste mensen om je heen, weet je de draad weer op te pakken! Mensen die op de hoogte zijn van jouw onderzoeksdomein, mensen die op de hoogte zijn van wat een doctoraat juist inhoudt, mensen die je de achtbaan zien nemen en maar al te graag samen met jou de achtbaan willen trotseren. Dankzij hen begin je opnieuw samenhang te zien. Door de piramide van mensen die je ondersteunen, lukt het je hoofd boven de wanden van de doolhof uit te steken. Je ziet een uitgang! Het is niet die glorieuze uitgang die je bij jezelf in het begin van je reis had voorgesteld, maar het is een uitgang. Je weet alle eindjes aan elkaar te knopen, de laatste barre maanden van de reis te volmaken, en de uitgang te vinden. Van zo dichtbij bekeken, heeft deze uitgang ook wel iets.

Met de uitgang in zicht zou ik graag die mensen willen bedanken die hebben bijgedragen tot de realisatie van dit doctoraat. In de eerste plaats denk ik hierbij aan mijn promotoren, Prof. Sabine Van Huffel, Prof. Jos Vander Sloten

en Prof. Bart Haex, voor de kansen die zij mij hebben gegeven. Bart en Jos, bedankt voor het vertrouwen om mij op te nemen binnen het laatste jaar van het EASI-project. Dankzij de aangename werksfeer binnen de afdeling biomechanica duurde het niet lang vooraleer ik mij er volledig thuis voelde. Jos, mijn oprechte dank om mij het vertrouwen en de vrijheid te geven om mijn doctoraatswerk vervolgens binnen de afdeling verder te zetten. Zeker niet vanzelfsprekend, vermits de inhoud ervan niet rechtstreeks aansloot bij de afdeling haar directe onderzoekslijnen. Daarom sprong Sabine vervolgens ter hulp als derde promotor om mij verder te introduceren in de wereld van de biomedische signaalverwerking. Ik besefte dat ik nog veel te leren had, en dat heb ik binnen haar onderzoeksgroep en onder haar begeleiding zeker kunnen doen. Ook binnen de Biomed-groep voelde ik mij vrij snel thuis. Sabine, bedankt hiervoor, en voor het vertrouwen om mij in het project NXT_SLEEP op te nemen om mijn doctoraat tot een goed einde te kunnen brengen. Bart, Jos, Sabine, ik weet dat ik jullie niet altijd even uitgebreid en even gelijkwaardig betrokken heb bij mijn werk. Daarom reden te meer dank om mij ondanks alles toch te blijven ondersteunen. Jullie kritische en constructieve opmerkingen waren onontbeerlijk voor de kwaliteit van dit werk, en elke vorm van input gaf mij opnieuw de moed om er tegenaan te gaan!

Hartelijk dank ook aan Prof. Harry van Lenthe, Prof. Bertien Buyse, Prof. Bart Vanrumste, Prof. Ronald Aarts en voorzitter Prof. Paul Van Houtte om deel uit te maken van de examencommissie. Jullie opmerkingen en inzichten, klinisch, algoritmisch of structureel, waren onontbeerlijk voor de kwaliteit van dit werk. Zij gaven mij eveneens een nieuwe stimulans om de laatste stappen van deze reis te voltooien. Ik hoop dat de finale wijzigingen aangebracht na de preliminaire verdediging jullie kunnen bekoren. Prof. Ronald Aarts, extra dank om tot twee maal toe naar Leuven te komen voor de beoordeling van mijn werk.

Dank aan alle academische en industriële partners binnen het EASI-sleep en NXT_SLEEP project voor de aangename samenwerking, zinvolle discussies en leuke momenten. Het zijn jullie die deze projecten, en dus ook indirect dit doctoraat, mogelijk hebben gemaakt. Uiteraard ook dank aan IWT en iMinds voor de financiële steun van deze projecten.

Vervolgens wil ik uitgebreid mijn collega's van de afdeling biomechanica en de Biomed-groep bedanken. Ik had de eer om deel uit te maken van twee onderzoeksgroepen, wat betekende: dubbel zoveel verjaardagen, dubbel zoveel kerstfeestjes, dubbel zoveel teambuildings! Ik was zeker niet de persoon die voor het meeste animo zorgde binnen elk van beide groepen, maar ik hoop dat mijn aanwezigheid toch af en toe een glimlach op jullie mond bracht, of een kans gaf om jullie hart te luchten. Omgekeerd was dit alleszins het geval! Met velen van jullie heb ik tal van leuke en interessante conversaties kunnen voeren, op het werk, op congres, buiten de uren. People from BMe and Biomed, you

will be missed! In het bijzonder dank aan enkele collega's met wie ik lange tijd een bureau heb mogen delen en nauw heb mogen mee samenwerken. Bart, Dorien en Vincent, bedankt voor alle fijne momenten op veruit de meest hip geschilderde bureau van heel de KU Leuven! Carolina and Alex, thanks for all the amazing laughs we have shared, and will continue to do! Bedankt aan Stefan en Ivo voor de technische input bij het ontwikkelen van de meet setup, en aan Rita, Ida en John voor het regelen van alle administratieve taken. Ook jullie allen hebben dit doctoraat mee mogelijk gemaakt.

Verder wil ik mijn vrienden uitermate bedanken voor hun steun. Mijn oude schoolvrienden, Dries, Jeroen, Laurent, Maarten, Niels en Steven, waarmee ik nog steeds een superhechte groep vorm. Ik ken jullie al zo lang dat ik jullie eigenlijk bijna als familie kan beschouwen. Mijn oude atletiekvrienden, Maxime en Koen, waarmee ik ook al ettelijke jaren terugga. Mijn studievrienden, Bart, Chris, David, Johan, Maarten, Quentin en Thomas, waarmee ik de laatste jaren veel heb doorstaan. Mijn klimvrienden, atletiek- en fietsvrienden, boardgame-vrienden, studie- en doctoraatsvrienden, bedankt om er al deze jaren voor mij te zijn en voor de nodige ontspanning te zorgen. Weet dat jullie ook altijd op mij kunnen rekenen!

Ten slotte wil ik natuurlijk van harte mijn familie bedanken. Op de eerste plaats mijn ouders. Ik begin meer en meer te beseffen welke opofferingen jullie allemaal voor me hebben moeten maken. Jullie hebben de keuzes die ik doorheen mijn leven heb gemaakt steeds ten volle gesteund. De afgelopen vier jaar waren niet altijd even gemakkelijk om verschillende redenen. Ik ben dan ook blij dat ik bij jullie thuis op elk moment tot rust kon komen, dat jullie altijd voor me klaar stonden. Op de tweede plaats mijn twee broers, Bart en Pieter, en mijn zus Annemie. Jullie hebben vele jaren in Leuven met mij doorgebracht, zowel tijdens mijn studies als mijn doctoraat, wat het nog eenvoudiger maakte om mij in Leuven thuis te voelen. Ten slotte de hele familie Willemen, Thiré, Verbruggen, Castelyns en Ceulemans. Ik mag mij gelukkig prijzen om deel uit te maken van zo'n grote, gezellige bende! De omringing met zoveel liefde heeft zeker bijgedragen tot het verwezenlijken van dit doctoraat.

Tim
Februari 2015

Abstract

The fact that a third of a human life is spent in a bed indicates the essential character of sleep. While some people might opt voluntarily for sleep deprivation, others don't get to choose. Their healthy pattern of sleep is disrupted due to sleep disorders such as sleep apnea, insomnia and restless legs syndrome. Most clinical diagnoses revolve around complaints of excessive daytime sleepiness. People usually wait quite long however before contacting professional help, and might only do so when complaints have gone from minor to serious. It can be argued that people with minor complaints will have negligible compliance to rather obtrusive therapies, and should not be treated with pharmaceuticals. However, cognitive and behavioral therapy has proven its effectiveness for clinically diagnosed patients in different domains, and might thus also enhance the quality of life for people with minor complaints. Contrary to the rather invasive therapies, cognitive and behavioral therapies can be quite low-cost (and thus cost-effective). Current methods for objective diagnosis of sleep disorders are however too costly, impractical and intrusive, or lack sufficient information and/or accuracy, to be used for long-term screening or follow-up after diagnosis. This PhD work hypothesizes that automated cardiac, respiratory and movement-based analysis could be able to bridge this gap, especially when all signals are monitored off-body in a mechanical way.

The first part of this work investigates the ability to use cardiac, respiratory and movement activity for sleep monitoring in healthy subjects. Using a dataset of 85 nights, several classification models were built to distinguish between Wake, REM, light sleep (N1-N2) and deep sleep (N3). The models were trained with and validated against gold standard polysomnography annotations, derived by sleep experts. Relevant characteristics of the cardiac, respiratory and movement activity are outlined, and limitations and prerequisites of the model discussed. The large amount of variability in cardiac and respiratory functioning among different subjects led to difficult-to-avoid misclassifications. Still, agreement values around 80% and kappa values around 0.60 confirm the potential of the method.

The second part of this work investigates the ability to use cardiac and respiratory activity for sleep monitoring in subjects with sleep apnea. Using a dataset of 70 nights, a classification model was built to distinguish apneic from stable breathing episodes. This model was again trained with and validated against gold standard polysomnography. Agreement values around 90% and kappa values around 0.80 show great potential of the method. Using a second dataset of 25 nights, containing a high proportion of hypopneic events, a similar classification model led however to agreement values around 75% and kappa values around 0.30, stressing the possible need of pulse oxymetry for reliable detection of hypopneic events. The presence of apneic events also proofs to have a significant impact on sleep stage classification model performance, with agreement values dropping to around 70% and kappa values to around 0.40. While the proposed probabilistic representations of the classification model outputs (referred to as the hypnocorrogram and apneacorrogram) are a valuable addition by visualizing uncertainty of correspondence, further improvements might be necessary to make cardiac and respiratory based sleep monitoring in subjects with sleep apnea reliable.

The third part of this work investigates the ability to monitor cardiac, respiratory and movement activity in an off-body mechanical way. A pressure-based ballistocardiographic setup was implemented inside a bed, measuring fluctuations in pressure difference between two air volumes underneath the chest area of the subject. While respiratory and movement activity can be easily detected, cardiac signal quality varies between postures as well as persons. An adaptation of the Pan-Tompkins algorithm is proposed for accurate detection of cardiac inter beat intervals. The setup and algorithm were evaluated on a dataset of fifteen subjects lying in the four standard sleeping postures. In between episodes of movement activity, an average correspondence of 97% with respect to inter beat intervals detected by electrocardiogram proofs the potential of the method.

In conclusion, this PhD thesis supports the opportunity of a biomechanics-based approach in sleep analysis, although its accuracy might be insufficient towards detection of minor sleeping problems. Future work should evaluate whether overall accuracy of sleep stage analysis and apneic breathing detection is sufficient for night-based diagnosis of potential (early) sleep disorders, based on information from multiple consecutive nights. In addition, more specific analysis should be performed towards other common sleep disorders such as insomnia and periodic limb movement disorder, which was not included in this thesis.

Beknopte samenvatting

Het feit dat een derde van een mensenleven zich afspeelt in bed wijst op het essentiële karakter van slaap. Hoewel sommigen vrijwillig kiezen voor slaaptekort, krijgen anderen niet eens de keuze. Hun gezond slaappatroon is verstoord door slaapstoornissen zoals slaapapneu, insomnie en rusteloze benen syndroom. De meeste klinische diagnoses worden gesteld rond klachten van overmatige vermoeidheid gedurende de dag. Mensen wachten echter meestal vrij lang vooraleer beroep te doen op professionele hulp, en doen dit mogelijk pas wanneer hun klachten zijn verergerd van miniem tot ernstig. Het valt natuurlijk te beargumenteren dat mensen met minieme klachten een verwaarloosbare belangstelling en gebruik zullen vertonen ten opzichte van vrij ingrijpende behandelingen, en dat deze mensen ook niet noodzakelijk moeten behandeld worden met medicatie. Cognitieve gedragstherapie heeft echter in verschillende domeinen zijn effectiviteit bewezen bij klinisch gediagnosticeerde patiënten, en zou mogelijk eveneens de levenskwaliteit van mensen met minieme klachten kunnen verbeteren. In tegenstelling tot de vrij ingrijpende behandelingen, kan cognitieve gedragstherapie vrij goedkoop (en dus kosteneffectief) aangeboden worden. De huidige methodes voor objectieve diagnose van slaapstoornissen zijn echter te duur, omslachtig en ingrijpend, of hebben een gebrek aan voldoende informatie en/of accuraatheid, voor identificatie of opvolging van slaapstoornissen over een langere periode. Dit doctoraatswerk stelt de hypothese dat een automatische analyse, gebruikmakend van hart-, ademhalings- en bewegingsactiviteit, de huidige bestaande leemte kan overbruggen, zeker wanneer al deze signalen op een mechanische wijze geregistreerd kunnen worden zonder sensoren aan te brengen op het lichaam zelf.

Het eerste deel van dit werk onderzoekt de mogelijkheid om hart-, ademhalings- en bewegingsactiviteit te gebruiken voor het opvolgen van slaap bij gezonde personen. Gebruikmakend van een database van 85 nachten werden verschillende classificatiemodellen opgebouwd ter onderscheiding van Waak, REM, lichte slaap (N1-N2) en diepe slaap (N3). De modellen werden getraind en gevalideerd gebruikmakend van gouden standaard polysomnografische

annotaties, aangebracht door slaapexperten. Relevante karakteristieken van de hart-, ademhalings- en bewegingsactiviteit worden opgelijst, en limitaties en voorwaarden besproken. De grote verscheidenheid in hart- en ademhalingswerking tussen verschillende personen leed tot moeilijk te vermijden misclassificaties. Doch bevestigen de overeenkomstwaarden rond de 80% en kappa-waarden rond de 0.60 het potentieel van de gebruikte methodiek.

Het tweede deel van dit werk onderzoekt de mogelijkheid om hart- en ademhalingsactiviteit te gebruiken voor het opvolgen van slaap bij personen met slaapapneu. Gebruikmakend van een database van 70 nachten werd een classificatiemodel opgebouwd ter onderscheiding van periodes van gezonde ademhaling en verstoorde ademhaling omwille van apneus. Dit model werd opnieuw getraind en gevalideerd gebruikmakend van gouden standaard polysomnografische annotaties. Overeenkomstwaarden rond de 90% en kappa-waarden rond de 0.80 geven blijk van groot potentieel van de methode. Gebruikmakend van een tweede database van 25 nachten, welke een groot aandeel hypopneus bevat, leidde een gelijkaardig classificatiemodel echter tot overeenkomstwaarden rond de 75% en kappa-waarden rond de 0.30. Dit benadrukt de mogelijke noodzaak voor gebruik van pulsoximetrie om tot een betrouwbare detectie van hypopneus te komen. De aanwezigheid van apneus bewees ook een significante invloed te hebben op de performantie van de slaapstadia classificatiemodellen, met overeenkomstwaarden dalend tot rond de 70% en kappa-waarden tot rond de 0.40. Hoewel de voorgestelde probabilistische voorstellingswijze van de uitkomst van de classificatiemodellen (genaamd hypnocorrogram en apneacorrogram) een waardevolle toevoeging is omwille van het visualiseren van onzekerheid in overeenkomst, lijken verdere verbeteringen toch vereist om, gebruikmakend van hart- en ademhalingsactiviteit, het opvolgen van slaap bij personen met slaapapneu betrouwbaar te maken.

Het derde deel van dit werk onderzoekt de mogelijkheid om hart-, ademhalings- en bewegingsactiviteit op een mechanische wijze te registreren, zonder sensoren aan te brengen op het lichaam zelf. Een druk-gebaseerde ballistocardiografische opstelling werd in de matras van een bed geïmplementeerd, wat toeliet om fluctuaties in drukverschillen tussen twee luchtvolumes (gepositioneerd ter hoogte van de borstkas) te registreren. Hoewel ademhalings- en bewegingsactiviteit eenvoudig kan gedetecteerd worden, varieert de kwaliteit van het hartactiviteits-signaal van houding tot houding, en van persoon tot persoon. Een aangepast Pan-Tompkins algoritme wordt voorgesteld voor het nauwkeurig detecteren van het hartslagritme. De opstelling en het algoritme werden geëvalueerd op een database van 15 personen, liggend in de vier standaard slaaphoudingen. Met uitzondering van periodes van bewegingsactiviteit, bewijst een gemiddelde overeenkomstwaarde van 97% ten opzichte van elektrocardiografie het potentieel van de vooropgestelde methode.

Als besluit ondersteunt dit doctoraatswerk de opportuniteit van een biomechanisch gebaseerde benadering van slaapanalyse, hoewel de accuraatheid mogelijk onvoldoende is voor detectie van beginnende slaapproblemen. Toekomstig werk zal moeten nagaan in hoeverre de accuraatheid voor onderscheiding van slaapstadia alsook de detectie van verstoorde ademhaling omwille van apneus voldoende is om een diagnose van een mogelijke (beginnende) slaapstoornis te kunnen stellen, gebaseerd op informatie van meerdere opeenvolgende nachten. Voorts zal meer specifiek werk verricht moeten worden ter analyse van andere veelvoorkomende slaapstoornissen zoals insomnie en het periodieke been bewegingen syndroom, wat niet werd opgenomen in deze thesis.

List of abbreviations

AASM	American academy of sleep medicine
ACT	actigraphy
AHI	apnea hypopnea index
ANS	autonomous nervous system
BCG	ballistocardiography
BMI	body mass index
BR	breathing rate
BRV	breathing rate variability
BS	breathing signal
CBT	cognitive and behavioral therapy
CPAP	continuous positive airway pressure
EASI	enhancing activity through sleep improvement
ECG	electrocardiography
EC	energy content
ECR	energy content ratio
EDR	ECG derived respiration
EEG	electroencephalography
EMG	electromyography
EOG	electrooculography
EPS	Epworth sleepiness scale
FA	feature averaging
FE	feature extension
FFT	fast fourier transform
FN	false negative
FP	false positive
GUI	graphical user interface
HF	high frequency (0.15 Hz - 0.40 Hz)
HIJ	three successive extrema in the BCG
HR	heart rate
HRV	heart rate variability

IDR	inter decile range
IE	inspiration to expiration ratio
IEEE	institute of electrical and electronics engineers
IQR	inter quartile range
kPCA	kernal principal component analysis
KSS	Karolinska sleepiness scale
LDA	linear discriminant analysis
LF	low frequency (0.04 Hz - 0.15 Hz)
LSSVM	least squares support vector machines
MAD	mean (or median) absolute deviation
MI	mutual information
MOV	movement
mRMR	minimum redundancy maximum relevance
MSLT	multiple sleep latency test
MT	movement time
MWT	maintenance of wakefulness test
N1	stage 1 NREM sleep
N2	stage 2 NREM sleep
N3	stage 3 NREM sleep
NREM	non rapid eye movement
OA	oral appliance
OSAS	obstructive sleep apnea syndrome
PLMD	periodic limb movement disorder
PNS	parasympathetic nervous system
PSG	polysomnography
PSQI	Pittsburgh sleep quality index
PVT	psycho-motor vigilance test
QRS	three successive extrema in the ECG
R&K	Rechtschaffen and Kales
RA	result averaging
RBF	radial basis function
REM	rapid eye movement
RF	random forests
RIP	respiratory inductance plethysmography
RLS	restless legs syndrome
RMS	root mean square
RT	reaction time
SFFS	sequential forward floating search
SNR	signal-to-noise ratio
SNS	sympathetic nervous system
SSS	Stanford sleepiness scale
STD	standard deviation

SVM	support vector machines
TF	total frequency (0.004 Hz - 0.40 Hz)
TN	true negative
TP	true positive
UCD	university college Dublin
VLF	very low frequency (0.004 Hz - 0.04 Hz)
WE	window extension

Contents

Dankwoord i

Abstract v

Beknopte samenvatting vii

Contents xv

List of Figures xxi

List of Tables xxv

1 Hypothesis statement 1

1.1 Thesis outline and research questions 2

1.1.1 Research question 1 3

1.1.2 Research question 2 3

1.1.3 Research question 3 4

2 General introduction 5

2.1 Societal aspects of sleep 5

2.2 The physiology of sleep 6

2.3 Current standards of sleep assessment 8

2.3.1 Polysomnography 8
2.3.2 Actigraphy 8
2.3.3 MSLT & MWT 9
2.3.4 Psycho-motor vigilance test 9
2.3.5 Subjective questionnaires 10
2.4 Most common sleep disorders and their treatment options . . . 11
2.4.1 Sleep apnea 11
2.4.2 PLMD & RLS 12
2.4.3 Insomnia 13
2.5 Limitations to the current standards 15
2.6 Surmounting the limitations 16

3 Automatic sleep stage classification based on easy to register signals 23
3.1 Introduction 25
3.2 Methodology 26
3.2.1 Subjects 26
3.2.2 Measurements 26
3.2.3 Data analysis 27
3.3 Results 29
3.4 Discussion 30
3.5 Conclusion 30

4 Cardiorespiratory and movement features in sleep-stage classification 33
4.1 Introduction 35
4.2 Materials and methods 36
4.2.1 Data collection 36
4.2.2 Data pre-processing 37
4.2.3 Feature extraction 40

4.2.4 Feature selection 42
4.2.5 Feature classification 44
4.3 Results 45
4.3.1 Wake vs Sleep at onset: < 60 min 45
4.3.2 Wake vs Sleep after onset: > 60 min 46
4.3.3 REM vs NREM 47
4.3.4 N3 vs N1N2 47
4.4 Discussion 48
4.5 Conclusion 51

5 Classifying episodes of sleep apnea based on single-lead ECG 53
5.1 Introduction 55
5.2 Methodology 55
5.3 Results 58
5.4 Discussion 60

6 Probabilistic cardiac and respiratory classification of sleep and apnea 63
6.1 Introduction 65
6.2 Methodology 66
6.2.1 Database 66
6.2.2 Data preprocessing 67
6.2.3 Feature extraction 70
6.2.4 Feature classification 71
6.2.5 Feature selection 78
6.2.6 Visual representation 80
6.3 Results 84
6.3.1 Feature selection 84
6.3.2 Classifier performance 85

6.3.3 Visual representation 88
6.4 Discussion 89
6.5 Conclusions and future work 93

7 Characterization of an air pressure-based ballistocardiographic setup 95
7.1 Introduction 97
7.2 Methodology 97
7.2.1 Setup of the measurement system 97
7.2.2 Experiments 98
7.2.3 Signal characterization 99
7.3 Results 101
7.4 Discussion 106

8 Pan-Tompkins based heart beat detection in the ballistocardiogram 109
8.1 Introduction 111
8.2 Materials and methods 112
8.2.1 Set-up of the measurement system 112
8.2.2 Experiments 113
8.2.3 Movement detection algorithm 114
8.2.4 Basic Pan-Tompkins algorithm 115
8.2.5 Feature-based approach 117
8.2.6 Template-based approach 118
8.2.7 Evaluation of the HRV accuracy 119
8.2.8 Optimization of the bandwidth from the BCG 121
8.3 Results 122
8.4 Discussion 127
8.5 Conclusion 129

9 General discussion **131**

9.1 Summary of chapters 131

9.2 Discussion of chapters 134

9.2.1 Choice of signals 134

9.2.2 Characterization of apneic events 135

9.2.3 Choice of feature selection and classifier algorithms . . . 137

9.2.4 The hypnocorrogram in healthy sleep 141

9.3 Evaluation of research questions 141

9.3.1 Research question 1 142

9.3.2 Research question 2 142

9.3.3 Research question 3 143

9.4 General conclusion 144

A Encompassing interdisciplinary research projects **145**

A.1 EASI-SLEEP 145

A.2 NXT_SLEEP 146

Bibliography **147**

Curriculum vitae **163**

List of Figures

1.1 Thesis content overview 2

2.1 The two-process model of sleep 6
2.2 Exemplary hypnogram of healthy sleep 7
2.3 A patient subjected to a clinical polysomnography 8
2.4 A wrist-worn actigraph 9
2.5 The psycho-motor vigilance test 10
2.6 Visualization of obstructive apnea 12
2.7 Visualization of restless legs syndrome 13
2.8 Visualization of insomnia 14
2.9 Autonomic nervous system connections 17
2.10 Visualization of the principle of ballistocardiography 19

4.1 Visualization of the movement registration methodology 37
4.2 Examples of feature-class and inter-feature density functions . . 44

5.1 Neighboring segment combinations 57
5.2 Overlapping segments with 30-second shifts 58
5.3 Overview of training and test agreement values 59
5.4 Training and test agreement values for overlapping windows . . . 61

6.1 Example effect of detrending on density function 74
6.2 Example effect of Box Cox transform and normalization 74
6.3 Epoch class determination from classifier output 77
6.4 Triple layer validation scheme 78
6.5 From binary to multiclass posterior probabilities 81
6.6 An example hypnocorrogram 82
6.7 An example hypnocorrogram excluding WAKE 82
6.8 An example binary hypnocorrogram 83
6.9 An example apneacorrogram 83
6.10 Sorted feature scores for classification problems 84

7.1 Visualisation of the measurement setup 98
7.2 Eight second fragment of filtered pressure sensor signal 100
7.3 Visual representation of the variability in the heart beat signal 103
7.4 Heart beat templates for the three different base air pressures . 104
7.5 Heart beat templates for the three different mattress top layers 105

8.1 Visualisation of the measurement setup 113
8.2 Visualization of the result of the movement detection algorithm 114
8.3 Decomposition of raw pressure signal 115
8.4 Visualization of the Basic Pan-Tompkins algorithm 117
8.5 Visualization of the feature-based approach 118
8.6 Visualization of the template-based approach 120
8.7 HRV accuracy for all postures and methods 125
8.8 Template-based HRV accuracy for all postures and subjects . . 126

9.1 An example thorax breathing signal of a mixed apnea 136
9.2 An example thorax breathing signal of an obstructive apnea . . 136

9.3 An example heart rate signal of an obstructive apnea 137
9.4 An example of a more difficult to conceive hypopnea 138
9.5 Classifier comparison APNEIC vs HEALTHY 139
9.6 Feature selection comparison APNEIC vs HEALTHY 140
9.7 An example hypnocorrogram for chapter four 141

List of Tables

2.1 Changes in SNS and PNS activity 18

3.1 Overview of different training and test set combinations 26
3.2 Results for the different training and test set combinations . . . 29

4.1 Overview of accepted intervals in the final dataset 39
4.2 Overview of defined features 41
4.3 Results of the different classification problems 46
4.4 Overview of similar classification algorithms in literature 49

5.1 List of 10 greedy forward selected features 60

6.1 General properties on subjects and recordings 67
6.2 Overview of epochs within each category 68
6.3 Overview of time domain features 72
6.4 Overview of frequency domain features 73
6.5 Pseudo-code preprocessing step of training vectors 76
6.6 Overview of highest scoring features 85
6.7 Classifier performance 87
6.8 Confusion matrix of the combined sleep stage problem 88

6.9 Classifier performance of the combined sleep stage problem . . 88
6.10 Overview of similar classification algorithms in literature 90

7.1 Overview of the different experimental conditions 99
7.2 Signal parameters for all experimental conditions 102

8.1 HRV accuracy values for all methods and postures 123
8.2 HRV accuracy for standard and optimal bandwidth 124

9.1 Classifier comparison REM vs NREM 138

Chapter 1

Hypothesis statement

The fact that a third of a human life is spent in a bed indicates the essential character of sleep. While some people might opt voluntarily for sleep deprivation, others don't get to choose. Their healthy pattern of sleep is disrupted due to sleep disorders such as sleep apnea, insomnia and restless legs syndrome. Most clinical diagnoses revolve around complaints of excessive daytime sleepiness. People usually wait quite long however before contacting professional help, and might only do so when complaints have gone from minor to serious. It can be argued that people with minor complaints will have negligible compliance to rather obtrusive therapies, and should not be treated with pharmaceuticals. However, cognitive and behavioral therapy has proven its effectiveness for clinically diagnosed patients in different domains, and might thus also enhance the quality of life for people with minor complaints. Contrary to the rather invasive therapies, cognitive and behavioral therapies can be quite low-cost (and thus cost-effective). Current methods for objective diagnosis of sleep disorders are however too costly, impractical and intrusive, or lack sufficient information and/or accuracy, to be used for long-term screening or follow-up after diagnosis. This PhD thesis, entitled *Biomechanics based analysis of sleep*, hypothesizes that automated cardiac, respiratory and movement-based analysis could be able to bridge this gap, especially when all signals are monitored off-body in a mechanical way.

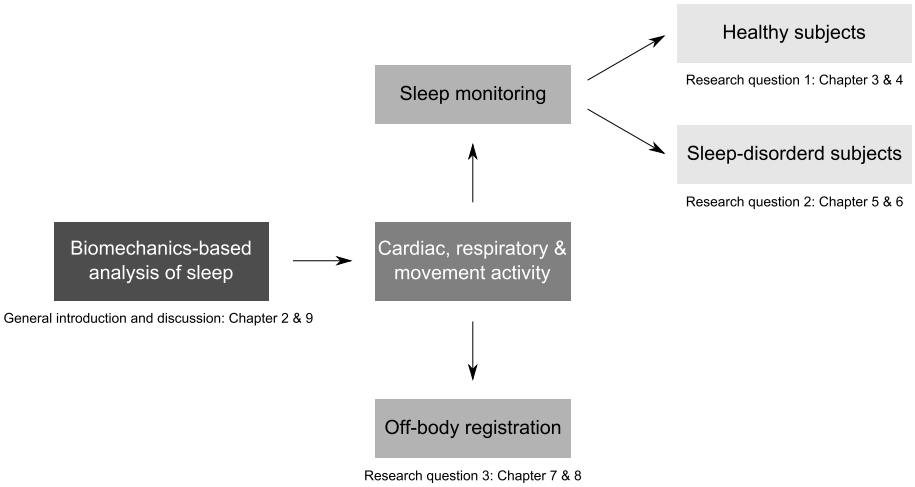


Figure 1.1: Overview of the content and the different chapters within this thesis.

1.1 Thesis outline and research questions

The general introduction of this thesis is found in chapter two. It starts by introducing the societal aspects of sleep, after which the physiology of sleep is illustrated, making use of the two-process and flip-flop models of sleep. Next an overview is given of the current standards of sleep assessment, the most common sleep disorders and some of their existing treatment options. Observed limitations to the current standards are highlighted, after which possible solutions and opportunities are revealed. The combination of the latter two allowed to formulate the hypothesis statement of this thesis.

The three main research questions within this thesis, as well as the contents of each of the corresponding chapters, are introduced in the remainder of this hypothesis statement chapter. While chapters three to eight are algorithmically oriented, their introduction, results and discussion section should be largely accessible to most readers. Although these chapters were written to be read as standalone publications (with limited page length), they have been adapted to improve general readability within this thesis. Furthermore, each chapter will be preceded by a short statement of relevance which will inform the reader of the chapter's content, and how his reading experience can be optimized.

The general discussion of this thesis is found in chapter nine. It starts by summarizing the content of the different chapters within this thesis. Next, this content is further discussed and supplemented by some new material. Finally,

the different research questions are specifically answered and the hypothesis statement concluded.

The two interdisciplinary research projects which formed the basis of this thesis are shortly summarized in appendix A.

1.1.1 Research question 1

Can sleep be monitored reliably in healthy subjects using cardiac, respiratory and movement activity?

Chapter three outlines the results of a proof-of-concept study distinguishing the combination of wake and active sleep (Wake-REM-N1) from passive sleep (N2-N3). It was performed on a small subset of a private database of polysomnographic nights from healthy subjects, as collected during a 4-year IWT-SBO project entitled *EASI-sleep*. Results have been presented with an oral presentation in the *18th World Congress on Ergonomics - 2012*, and published as a conference proceedings paper in a special issue of the journal *Work: A Journal of Prevention, Assessment & Rehabilitation* (Willemen et al. 2012).

Chapter four outlines the results of a larger follow-up study, distinguishing between Wake, REM, light sleep (N1-N2) and deep sleep (N3). It was performed on a larger subset of the same private database. Results have been published in the journal *IEEE Journal of Biomedical and Health Informatics* (Willemen et al. 2014a).

1.1.2 Research question 2

Can sleep be monitored reliably in sleep-disordered subjects using cardiac, respiratory and movement activity? More specifically in this thesis: can sleep and apneic events be monitored reliably in subjects with sleep apnea using cardiac and respiratory activity?

Chapter five outlines the results of a study distinguishing between apneic events and healthy breathing. It was performed on the public Apnea-ECG database, available at Physionet (Goldberger et al. 2000). The results of this study have been presented with an oral presentation in the *41st Annual Computing in Cardiology Conference - 2014*, and published online as a conference proceedings paper in the *IEEE Xplore Digital Library* (Willemen et al. 2014c).

Chapter six outlines the results of a study distinguishing between apneic events and healthy breathing, as well as between Wake, REM, light sleep (N1-N2)

and deep sleep (N3). It was performed on the public St. Vincent's University Hospital and University College Dublin Sleep Apnea Database, also available at Physionet. The results of this study have been submitted for publication to the journal *Physiological Measurement* (Willemen et al. 2014e).

1.1.3 Research question 3

Can cardiac, respiratory and movement activity be monitored reliably in an off-body mechanical way?

Chapter seven outlines the results of a study building and characterizing a ballistocardiographic setup, implemented inside of a bed. The results of this study have been presented with an oral presentation in the *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society - 2014*, and published online as a conference proceeding paper in the *IEEE Xplore Digital Library* (Willemen et al. 2014b).

Chapter eight outlines the results of a study presenting and benchmarking an algorithm for the accurate detection of inter beat intervals in the ballistocardiogram. The results of this study have been submitted for publication to the journal *IEEE Journal of Biomedical and Health Informatics* (Willemen et al. 2014d).

Chapter 2

General introduction

Sleep is for the weak! You can sleep when you're dead.
It's a mere symptom of caffeine deprivation.

2.1 Societal aspects of sleep

The fact that a third of people's lives are spent inside of a bed indicates the essential character of sleep. Nevertheless, the current evolution towards a 24-hour society is preventing people more and more from getting the necessary amount of sleep they require. Some call it the biggest social experiment in the history of man-kind: people sleep on average one to one and a half hours less per night compared to a century ago. Most people's life doesn't fit the evolved ideal of 8 hours of work, leisure time and sleep per day. People's jobs are holy. Agendas are packed with evening activities. Entertainment is never more than one click away, whether day or night. Time is running short, and it is usually sleep taking the hit.

While some might opt voluntarily for sleep deprivation, others don't get to choose. A 2011 published report from the Flemish government, describing the current condition of social welfare in Flanders and Belgium (Noppe et al. 2012), informs that 21% of Belgian respondents older than 15 year experience regular difficulties in sleeping. Sleep is however crucial for physical and mental health. Physically, it ensures musculoskeletal relaxation, lumbar disc rehydration (Haex 2004) and tissue recovery (Adam & Oswald 1984). Mentally, it ensures memory consolidation (Curcio et al. 2006) and preservation of neurocognitive capabilities

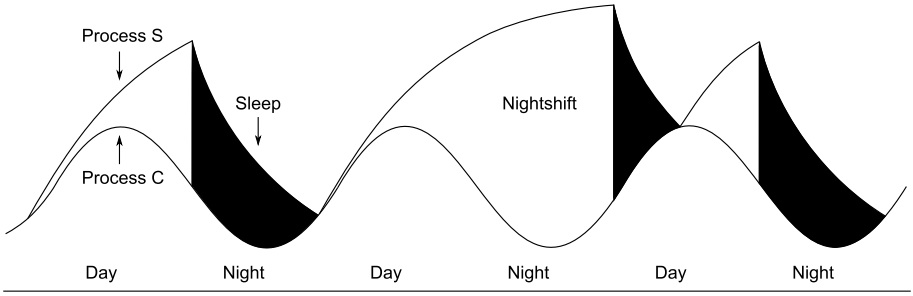


Figure 2.1: The two-process model of sleep, comprising of the homeostatic process S and the circadian process C.

(Meerlo et al. 2009). The 2008 Better Sleep Council Survey reported that sleep deprivation impacts people their ability to socially interact (44% of correspondents), their quality of work (31%) and proper judgment (31%) (Better Sleep Council 2008). The catastrophes of Chernobyl, Three Mile Island, and the Challenger space shuttle have all been linked to sleep deprived personnel (Mitler et al. 1988).

2.2 The physiology of sleep

The two-process model of sleep is based on the following two observations: "The longer people are awake, the more likely they are to fall asleep" and "People normally sleep in a specific time frame within the 24-hour day" (Borbély & Achermann 1992). The former is called the homeostatic process of sleep, (process S), whereas the latter is referred to as the circadian rhythm (process C). The two-process model is visualized in Fig. 2.1. After waking up in the morning, people start building up homeostatic sleep pressure. It's however only towards the night-phase of the circadian rhythm that the build-up of sleep pressure causes sleepiness, visualized by the height difference between the two curves. Sleep allows to get rid of the build-up sleep pressure, waking us up when both curves meet. The effect of a nightshift is also visualized, showing an additional build-up of sleep pressure, the ability to fall asleep in the morning, but the reduction in total sleep time. Jetlag has a similar effect due to the shift of process S with respect to process C. Sleepiness will be increased during the day, but reduced during the night.

The flip-flop model of sleep explains how the wake system and sleep system inhibit each other to make sure only one is active at the same time (Saper et al. 2001). The posterior hypothalamus contains the major wake-promoting

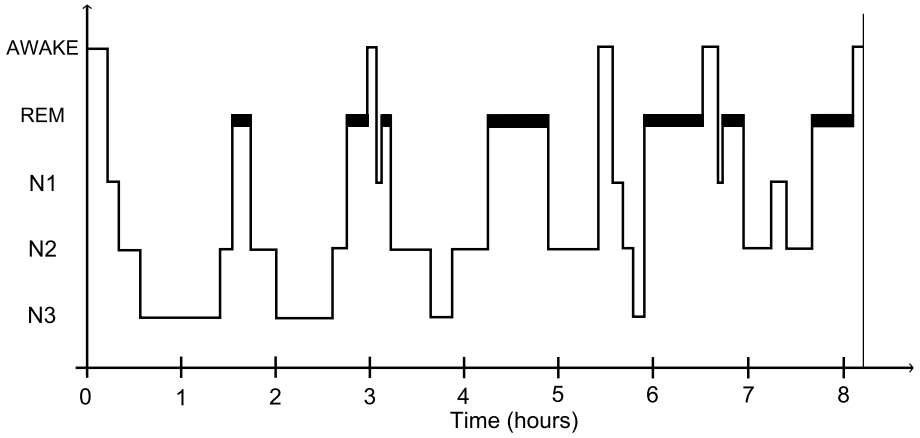


Figure 2.2: Exemplary hypnogram of a healthy distribution of sleep stages during the night.

area, while the preoptic region in the hypothalamus contains the major sleep-promoting area. Although many smaller important brain areas exist, these two suffice as a simplification. They are responsible for the release of wake- and sleep-promoting neurotransmitters, inhibiting each other. Those neurotransmitters who have the upper hand decide the state of the flip-flop.

Sleep comprises however not a single state. At certain times during the night, brain wave activity resembles the one perceived during wake, combined with an intermittent presence of rapid eye movements and almost complete states of muscle atonia due to motor neuron inhibition (Hobson et al. 2000). The presence of this so-called Rapid-Eye-Movement (REM) sleep follows an ultradian rhythm with an average cycle length of 100 minutes. Its proportion within each cycle increases during the night. The periods in between are referred to as non-REM (NREM) sleep, characterized by brain wave oscillations of different amplitude and frequency. Current standards divide this NREM sleep in three discrete stages N1, N2 and N3, with increasing depth of sleep (Berry et al. 2012). Stage N3 is therefore also referred to as deep sleep, and is more prominent during the early hours of sleep. Healthy sleep should contain enough sleep depth and continuity to accommodate on average 20% REM sleep and an equal amount of deep sleep (although exact amounts change significantly with age (Feinberg 1974)). An exemplary healthy distribution of sleep is shown in Fig. 2.2, using a representation referred to as a hypnogram.



Figure 2.3: A patient subjected to a clinical polysomnography. [Adapted from source: nhlbi.nih.gov/health/]

2.3 Current standards of sleep assessment

2.3.1 Polysomnography

The gold standard in assessment of sleep is polysomnography (PSG). The method comprises the overnight registration of brain wave activity using eight Electroencephalography (EEG) electrodes, eye-movement activity using two Electrooculography (EOG) electrodes, and muscle activity at the level of the chin using three Electromyography (EMG) electrodes. Sleep stages are annotated manually in 30-second intervals (referred to as epochs) according to international standards (Berry et al. 2012). In clinical use of polysomnography, extra sensors are added for diagnosis of specific sleep disorders related to heart activity (Electrocardiography (ECG)), breathing activity (nasal airflow, respiratory effort breathing belts at chest and abdomen, peripheral pulse oximetry) or movement activity (EMG at both arms and legs). When annotation of sleep stages is deemed unnecessary, cardio-respiratory polygraphy can be used, which monitors only heart and breathing activity (Chesson Jr et al. 2003).

2.3.2 Actigraphy

Actigraphs are usually wrist-worn devices which make use of accelerometers to detect movement activity. They are commonly used in combination with computerized algorithms to estimate total sleep time when PSG is deemed unnecessary (assessment of circadian rhythms), impractical (long-term follow



Figure 2.4: A wrist-worn actigraph. [Source: `actigraphy.respironics.com/devices/`]

up) or difficult (specific population groups such as patients with dementia or infants) (Morgenthaler et al. 2007). Although most studies indicate a reasonable accuracy in sleep time estimation for healthy sleepers, its reliability for sleep disordered patients is questionable (Sadeh 2011). Accuracy reduces with increased amounts of wakefulness, due to the erroneous practice of scoring episodes of immobility as sleep (Paquet et al. 2007).

2.3.3 MSLT & MWT

The multiple sleep latency test (MSLT) and maintenance of wakefulness test (MWT) are two tests designed to measure levels of sleepiness and alertness during the day (Littner et al. 2005). The tests consist of four to five trials of sitting or lying down in a dark room during which subjects are monitored by PSG. Trials are spread out over the day with two hour intervals. While during the MWT, subjects are asked to stay awake for as long as possible (up to 40 minutes), during the MSLT, they are asked to fall asleep (within 20 minutes). Average sleep onset times below 8 minutes are considered abnormal.

2.3.4 Psycho-motor vigilance test

The psycho-motor vigilance test (PVT) can be used as an alternative for assessing sleepiness and attention, when MSLT and MWT are impossible due to unavailability of EEG, EOG and EMG for sleep-wake assessment. During a 10-minute trial, it monitors a subject's reaction times (RT) to a recurring

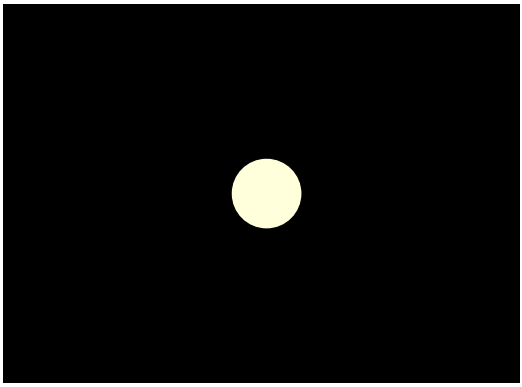


Figure 2.5: The psycho-motor vigilance test evaluates a subject's reaction times to a recurring visual stimulus, e.g. pushing a button after the appearance of a white dot on a black screen.

visual stimulus, e.g. pushing a button after the appearance of a white dot on a black screen. The use of shorter trials (< 10 min) has a negative effect on the sensitivity of the test with respect to sleepiness (Loh et al. 2004). Average response speed (1/RT) and the amount of lapses (RT > 500 ms) were proven to be the most sensitive measures (Basner & Dinges 2011).

2.3.5 Subjective questionnaires

Subjective questionnaires are normally used when objective measures are unavailable or are deemed impractical. Both the Karolinska Sleepiness Scale (KSS) (Åkerstedt & Gillberg 1990) as well as the Stanford Sleepiness Scale (SSS) (Hoddes 1972) ask for a single measure of subjective sleepiness at the current time, and can thus be used for measuring changes in sleepiness during the day. The Epworth Sleepiness Scale (ESS) (Johns et al. 1991) asks for the likelihood of dozing off in a number of day-to-day situations, and is reflecting a subject's state of sleepiness during recent times (i.e. the previous days or weeks). The Pittsburgh Sleep Quality Index (PSQI) (Buysse et al. 1989) is the most extensive questionnaire, and was devised to evaluate general sleep quality and disturbances over the last month. While the KSS, SSS and ESS are the subjective counterparts to the MSLT, MWT and PVT, the PSQI can be regarded as the subjective counterpart of a complete clinical PSG. It even takes into account input from a possible roommate or bed partner in order to be as complete as possible. Nevertheless, even when objective measures of sleep and sleepiness are available, subjective measures can have their value too, especially

when discrepancies exist between both (i.e. people over- or underestimating the quality of their sleep and their current state of sleepiness).

2.4 Most common sleep disorders and their treatment options

2.4.1 Sleep apnea

Sleep apnea is a sleep-related breathing disorder which causes frequent arousal from sleep, preventing patients to obtain sufficient sleep depth and continuity for healthy sleep. It also has a severe impact on the cardiovascular system, increasing the risk of cardiovascular disease (Leung & Douglas Bradley 2001). Clinical diagnosis of sleep apnea used to have an estimated prevalence of 4% in men and 2% in women (Young et al. 1993). Following amongst other things the escalation in the prevalence of obesity, a more recent update estimates prevalence values to have increased to 14% in men and 5% in women (Peppard et al. 2013).

An apneic event is defined as a clear decrease from baseline in breathing volume of at least 10 seconds (Berry et al. 2012). Apneic events can be obstructive (complete or partial obstruction of upper airway with surrounding soft tissue (Fig. 2.6)) or central (absent or reduced breathing effort). Central events are much less common than obstructive events, with an estimated prevalence in the general population of less than 1% (Bixler et al. 1998). Mixed events are a combination of both, usually starting as a central event, but continuing as an obstructive event once breathing effort is reinitiated. Partial obstructions and reduced breathing effort are referred to as hypopneas.

The Apnea-Hypopnea-Index (AHI) indicates the average amount of apneic or hypopneic events per hour. Clinical diagnosis is based on an AHI larger than five, in combination with excessive day time sleepiness that cannot better be explained by other factors (Quan et al. 1999). Without the requirement of excessive daytime sleepiness, the estimated prevalence of sleep apnea would increase from 14% to 34% in men, and from 5% to 17% in women (Peppard et al. 2013). Although snoring can act as an indication for sleep apnea, only 18% of snorers were found to meet the clinical diagnosis of sleep apnea (Boudewyns et al. 1997). This percentage increased to 35% when snoring was present in combination with complaints of excessive daytime sleepiness.

Standard treatment options for sleep apnea include continuous positive airway pressure (CPAP), oral appliances (OA) and behavioral treatment. As the term



Figure 2.6: Obstructive apnea, characterized by the collapse of soft tissue in the upper airway, blocking airflow. [Adapted from source: Wikimedia Commons]

states, CPAP provides a continuous positive airway pressure to prevent the upper airway from collapsing. This pressure can be applied through the nose, mouth or the combination of both. Custom made OA's hold the lower jaw or the tongue in an advanced position. This enlarges the upper airway, in order to decrease its chance of collapse. Behavioral treatment consists of weight loss (towards a BMI < 25 kg/m^2), limiting the intake of alcohol and sedatives before bedtime, and avoidance of supine posture which can affect airway size in some patients. Objective or subjective follow-up of treatment efficacy is recommended (Epstein et al. 2009).

2.4.2 PLMD & RLS

Periodic limb movement disorder (PLMD) is characterized by the involuntary repetitive movement of limbs (usually in the lower extremities) during sleep (Aurora et al. 2012). These movements can be associated with arousals from sleep, leading to sleep disruption. PLMD occurs in approximately 80% to 90% of patients with restless legs syndrome (RLS). RLS is characterized by an urge to move limbs which is partially relieved by physical activity. Symptoms usually become worse during rest, such as during the evening or at night. PLMD and RLS have an estimated prevalence of respectively 3.9% and 5.5% (Ohayon & Roth 2002). Prevalence values are higher in women than in men, and increase significantly with age. Clinical diagnosis of PLMD is made through PSG or actigraphy. A PLM series is defined as a minimum of 4 consecutive leg movements, with an interval length between 5 and 90 seconds (Berry et al. 2012).



Figure 2.7: Restless legs syndrome, which symptoms become worse during rest (e.g. at night). Co-morbidity with periodic limb movement disorder is estimated at 80% to 90%. [Source: `psychneuro.wordpress.com`]

Standard treatment options for RLS are mainly restricted to pharmacotherapy. Some other therapies exist but with minor evidence compared to pharmacotherapy. Examples are the daily use of pneumatic leg compression devices, regular aerobic exercise (3 days/week) and cognitive behavioral therapy (mindfulness-based exercises, stress reduction strategies and medical education). Regarding to subjects with only PLMD (and no RLS), there is insufficient evidence for recommendation of pharmacotherapy (Aurora et al. 2012).

2.4.3 Insomnia

Insomnia is characterized by the struggle to either initiate or maintain sleep, by the tendency of waking up too early, or by sleep which is felt as insufficiently restorative (Morgenthaler et al. 2006). Depending on the frequency of the reported subjective complaints, a distinction is made between situational and chronic insomnia. The former has an estimated prevalence rate around 30%, while the latter affects about 10% of adult population (Roth 2005). Insomnia in the elderly population is high, mainly towards sleep maintenance (McCall 2004). Individuals with situational insomnia show a consistent response to various types of stress (new sleep environment, circadian shifts, caffeine), and are more likely to develop chronic insomnia later on in their lives (Bonnet & Arand 2003). Patients with insomnia report impaired daytime functioning, increased loss of time from work and higher health costs. They are also much more likely to develop a major depression during their life (31.1% of insomniacs, versus 2.7% of normals) (Roth 2005). Clinical diagnosis is usually made on the basis of a



Figure 2.8: Insomnia, usually characterized by the struggle to initiate or maintain sleep. [Source: `whywesuffer.com`]

thorough analysis of sleep history, as reported by the patient and his or her social environment. PSG is only used when other underlying sleep disorders might be suspected as the primary cause (i.e. secondary insomnia). Actigraphy is sometimes used for an objective characterization of the reported problems, although the use of subjective sleep diaries (both prior as after diagnosis) is most common.

The majority of individuals who have trouble sleeping try to tackle the problem on their own, without looking for appropriate medical attention. Most of them however don't have the time or knowledge to tackle it in a safe and constructive way, and thus seek the use of non-prescribed pharmaceuticals all too fast. Standard clinical treatment of insomnia includes various forms of cognitive and behavioral therapy (CBT), or the use of hypnotics. Clinical reassessment is suggested every few weeks until symptoms appear stable or resolved. Short or long term pharmacotherapy is always recommended to be supplemented with some form of CBT (Schutte-Rodin et al. 2008). Cognitive therapy tries to change the patient's beliefs and attitudes towards sleep and insomnia, whereas behavioral therapy usually consists out of various forms of stimulus control (re-associate the bedroom with sleep, and re-establish a consistent sleep-wake schedule), relaxation training (both physically as well as mentally) and sleep restriction (reducing time in bed to actual time spent asleep). General education on sleep hygiene towards health (diet, exercise, substance abuse) and environmental factors (temperature, noise, light) is always supplemented, but has insufficient evidence to be effective on its own (Morgenthaler et al. 2006). Regular bi-yearly follow-up of insomnia patients is necessary as relapse rate is high (Schutte-Rodin et al. 2008).

2.5 Limitations to the current standards

Most clinical diagnoses revolve around the presence of complaints of excessive daytime sleepiness. People however usually wait quite long before contacting professional help, and might only do so in case of serious complaints (Leger & Poursain 2005). The prospect of an overnight stay in the hospital, and the usually quite long waiting lists before being able to see a sleep expert, are not helping either. Current standards are thus designed for improving sleep for people with serious complaints. Screening for sleep disorders in the natural home environment might be able to locate sleep disordered patients who would otherwise never seek medical attention, or would only do so when complaints have gone from minor to serious.

It can be argued that people with minor complaints will have negligible compliance to rather obtrusive therapies such as CPAP and OA's in case of sleep apnea. People with minor symptoms should also normally not be treated with pharmaceuticals in case of insomnia or PLMD. Cognitive and behavioral therapies have however proven their effectiveness for clinically diagnosed patients, and might thus also enhance the quality of life for people with minor complaints. Contrary to the rather invasive therapies, cognitive and behavioral treatments can be quite low-cost (and thus cost-effective), especially when offered online over the internet, such as Somnio (Somnio 2010) and Sleepio (Espie et al. 2012).

Due to the intermittent presence or daily variation in severity of some sleep disorders, screening over multiple nights would be advised (especially when symptoms would only be minor, and thus more difficult to detect) (Roth 2005, Levendowski et al. 2009). The use of PSG or respiratory polygraphy for multiple night observation in the home environment would however be too costly, impractical and intrusive. Cheap unobtrusive measurement tools such as actigraphy and subjective questionnaires could be used instead. They might however lack sufficient information and/or accuracy towards certain sleep disorders: actigraphs have a questionable reliability towards sleep-wake estimation in sleep disordered patients (Sadeh 2011), while none of the tools can give a direct estimation of apneic events, amount of REM sleep and amount of deep sleep.

The above limitations are not restricted to screening. Firstly, regular clinical reassessment or follow-up of treatment efficacy in diagnosed patients is recommended for all sleep disorders. Current standards limit this however towards actigraphy and mainly subjective questionnaires, which might lack sufficient information and/or accuracy. Secondly, research towards the positive or negative effects on sleep depth and continuity of environmental or behavioral manipulations requires long-term observations, since their effects could only

become visible after a couple of weeks of habituation. This makes large double-blind clinical trials in this field of research too expensive, impractical and intrusive, if limited to current standards of PSG. Thirdly, in recent years, people have shown an increased amount of interest in the continuous self-management of their health, including the quality of their sleep. The amount of information they can currently gain on a long-term basis, is however limited.

2.6 Surmounting the limitations

As mentioned in the previous section, current standards are designed for improving sleep towards people with serious complaints. They might thus be inadequate for multiple night screening towards people with minor complaints, for follow up of treatment efficacy in diagnosed patients, for long-term objective research towards the positive or negative effects on sleep of environmental or behavioral factors, or for continuous self-management of people's quality of sleep. The present limitations could be surmounted however by a cheap unobtrusive measurement tool able to give a reliable estimation of sleep-wake, combined with a direct estimation of apneic events, periodic limb movements, amount of REM sleep and amount of deep sleep.

Research has shown a significant influence of sleep and its disorders on the autonomic nervous system (ANS), which is in control of unconscious visceral functions such as activity of heart and lungs (Fig. 2.9). The ANS is characterized by the presence of two interacting systems: the sympathetic nervous system (SNS; in control of swift 'fight or flight' mechanisms) and the parasympathetic nervous system (PNS; in control of more gradual 'rest and digest' mechanisms). Variations in activity of the SNS, as directly measured using microneurography, have been linked to changes in sleep and its disorders. SNS activity decreased from 100% during wakefulness to 41% during deep sleep, and increased to 215% during REM sleep. The latter increase was however only present during tonic REM sleep (characterized by loss of muscle tone), and not within phasic REM (characterized by the presence of muscle twitches) (Somers et al. 1993). Elevated amounts of SNS activity were found in patients with sleep apnea, especially near the occurence of apneic events, but also during wake (Somers et al. 1995).

Since the ANS controls the activity of both heart and lungs, SNS and PNS variations are also reflected by cardiac and respiratory fluctuations. This indirect analysis of SNS and PNS variations can be performed using standard time and frequency domain measures as defined in the field of Heart Rate Variability (HRV) (Malik et al. 1996). Cardiac PNS activity showed a 24-hour variation (in synchrony with the circadian rhythm) independent from the presence of sleep,

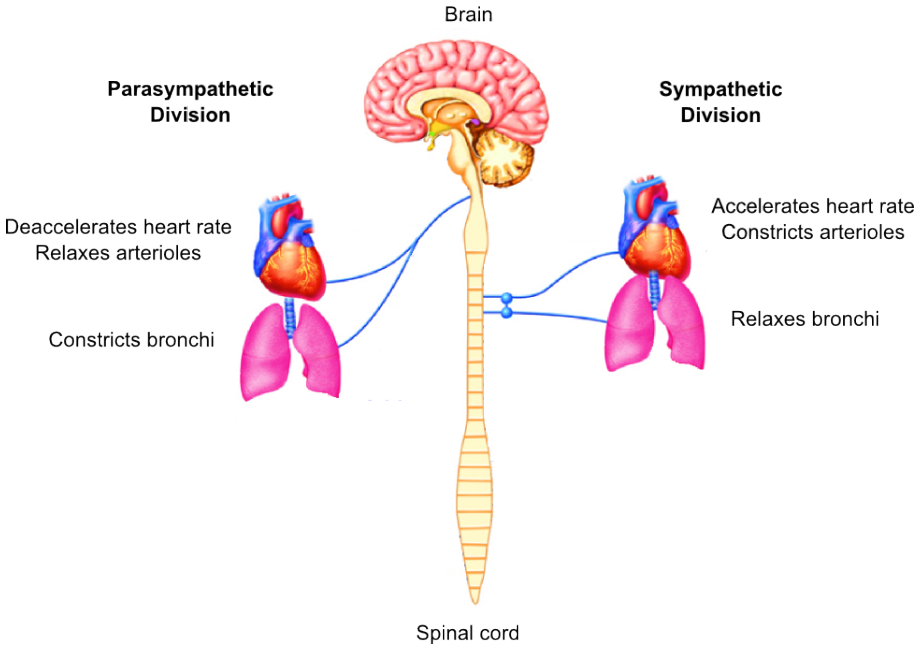


Figure 2.9: Connections of the autonomic nervous system towards our heart and lungs. [Adapted from source: `savingstudentsmoney.org`]

whereas cardiac SNS activity was only influenced during sleep itself. Cardiac PNS activity increased progressively from light sleep to deep sleep, while values in REM sleep were lower than those in NREM sleep. Cardiac SNS activity decreased only slightly from wakefulness to NREM sleep. Differences in REM sleep were inconclusive with respect to SNS activity, related to the presence of both tonic and phasic REM sleep (Burgess et al. 1997). Patients with insomnia showed an increased cardiac SNS and decreased PNS activity with respect to controls across all stages of sleep (Bonnet & Arand 1998). Table 2.1 gives an overview of the changes in SNS and PNS activity across sleep and its disorders.

The current gap in clinical standards towards assessment of sleep and its pathologies could thus be bridged by supplementing the movement-based analysis of actigraphy with a cardiac and respiratory-based analysis of sleep and its disorders; especially when all of these signals can be monitored off-body without attaching sensors directly to the skin. During the past decade, numerous methods and techniques have been developed to record cardiac and respiratory activity, as well as general body movement, from within or around the bed. Some methods directly measure electrical activity of the heart, while others

Table 2.1: Overview of the changes in SNS and PNS activity across sleep and its disorders.

	Microneurography	**Heart rate variability**	
	SNS activity	*SNS activity*	*PNS activity*
WAKE	Baseline	Baseline	Baseline
N1,N2	↓	↓	↑
N3	↓↓	↓	↑↑
REM	↑↑ (tonic phase)	↑/↓	↓
Healthy	Baseline	Baseline	Baseline
Apnea	↑		
Insomnia		↑	↓

focus on the mechanical activity of heart, lungs and general body movement. The main idea is however to perform all measurements off-body, i.e. without attaching sensors directly to the skin. This will make sure that people are not disturbed before and during their sleep, both physically and mentally. Physical disturbance can occur due to skin irritation on places where sensors were attached, or due to the presence of wires connecting the sensors to a registration or transmitting device. Mental disturbance on the other hand can be due to people being aware of the monitoring, or due to the nuisance of attaching the sensors before being able to go to sleep.

Electrical activity of the heart can be registered by placing textile electrodes into the bed sheets (Peltokangas et al. 2012). Similarly to how a standard ECG is obtained, it registers the voltage difference between the different electrodes that are in contact with the subject's body. Wearing pyjamas will of course limit the possible contact area between the body and the electrodes embedded in the bedding textile, making the registration difficult. Although these textile electrodes can also be implemented in pyjamas themselves, it violates the idea of having all sensors off-body. It also brings up issues related to battery status and wireless transmission.

Mechanical activity of the lungs causes obvious thorax and abdominal movement, whereas the pulsation of blood with every heart beat gives a subtle but detectable oscillation of the body. The registration of the latter is studied in the field of ballistocardiography (BCG) (Starr et al. 1939). The oscillation is different from the movement felt when placing a hand on someone's chest, the so-called seismocardiogram. Where the seismocardiogram is a direct result of the contraction of the heart, the ballistocardiogram reflects the movement of the blood. Its principle can be compared to the recoil when firing a gun, where

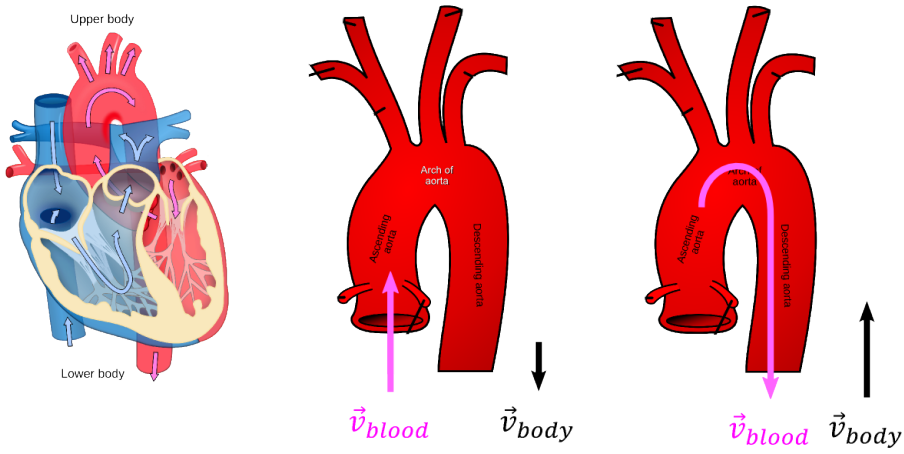


Figure 2.10: Simplified visualization of the recoils the body experiences due to the up- and downward movement of the blood through the aorta after every contraction of the heart. [Adapted from source: Wikimedia Commons]

the gun will gain velocity opposite to the direction of the bullet. In the case of the ballistocardiogram, the blood would be the bullet, and the human body would be the gun. At every heartbeat, blood is ejected from the heart through the aorta, first moving a bit upwards, after which most of the blood moves down towards the lower body. Due to these movements, the human body will experience a consecutive down- and upward recoil with every heartbeat (Fig. 2.10).

The recoils can be explained using the mechanical principle of conservation of momentum. A rigid body simplification of the problem is illustrated by equations 2.1 through 2.4,

$$\vec{p}_{t_1} = \vec{p}_{t_2} \tag{2.1}$$

$$t_1 \leftrightarrow \vec{p}_{t_1} = \vec{0} \tag{2.2}$$

$$t_2 \leftrightarrow \vec{p}_{t_2} = m_{blood}\vec{v}_{blood} + m_{body}\vec{v}_{body} \tag{2.3}$$

$$t_2 \leftrightarrow m_{blood}\vec{v}_{blood} = -m_{body}\vec{v}_{body} \tag{2.4}$$

with $\vec{p}_{t_1}$ the momentum at time step t_1 (before contraction of the heart), $\vec{p}_{t_2}$ the momentum at time step t_2 (after contraction of the heart), m the mass and $\vec{v}$ the velocity of the respective centers of mass. Since $\vec{p}_{t_2}$ should also be equal to

zero, the direction of the velocity vectors of blood and body should be opposite. The resulting longitudinal oscillations of the body can for example be registered on a weight scale (Inan et al. 2009), or from within or around a bed (Jansen et al. 1991). The mechanical activity of thorax and abdominal movement, as well as general body movement, can be measured in a similar way.

Automated cardiac, respiratory and movement-based analysis using off-body mechanical registration could thus very well enable multiple night cost-effective screening of people's sleep, in order to diagnose and follow-up both people with serious as well as minor complaints. Early research also points out significant opportunities to improve sleep depth and continuity not only in subjects with complaints, but also in healthy sleepers. Over the past 10 years, ambient intelligence has infiltrated our lives in the form of various home applications, enabled by the ever decreasing size and cost of computing technology. While in transport and industry its presence has become second nature, some areas have remained fairly untouched. Our bedroom is one of these areas, despite the fact that we spent almost a third of our lives accommodating it. A biomechanics-based approach in sleep analysis could be central to the intelligent bedroom of the future, which should assist us in the recovery and preparation of our past and future daily activities. Some environmental factors are already known to have a beneficial or detrimental impact on sleep, although their body of evidence is in most cases still lacking due to the limitations of the current standards. A biomechanics-based approach in sleep analysis might offer a solution for this type of long-term research, of which some examples are given in the next two paragraphs.

The world health organization recommends a maximum level of noise inside the bedroom at night of 45 dB, with a mean recommended level of 30 dB. Higher levels of noise can cause awakenings but also unconscious autonomic effects during sleep, such as heart rate changes and vasoconstrictions (Muzet 2007). White noise showed however a beneficial effect on sleep induction of neonates (Spencer et al. 1990), and on sleep in subjects exposed to ICU noise (Stanchina et al. 2005). High illumination levels late at night (especially blue light) are known to affect melatonin secretion levels, and to express phase delays in the circadian rhythm. Morning and day-time light therapy might however aid people frequently exposed to circadian shifts originating from jet lags or shift work (Chesson et al. 1999).

Extreme temperatures, outside of the thermo-neutral zone, require the body to be active, negatively effecting sleep depth and continuity. In case of low temperatures, metabolism levels are increased and vasoconstriction induced; in case of high temperatures, sweat is secreted and vasodilation induced (Gilbert et al. 2004). Heating up distal areas of the skin (hands and feet) led however to decreased sleep onset times in healthy young subjects (Kräuchi et al. 1999)

and to improved sleep depth and continuity in a healthy elderly population (Raymann et al. 2008). Poor bedding systems showed a detrimental impact on sleep quality (Verhaert et al. 2011b). Active control of bedding support properties, ensuring continuous optimal back support, already showed subjective improvements in sleep and day-time quality in young healthy subjects, combined with a trend towards increased amounts of deep sleep, even after only a single night of observation (Van Deun et al. 2012).

Further on, a biomechanics-based approach to sleep analysis could enable the possibility of continuous self-management of sleep quality. The latter could be compared to the use of a weight scale, which allows for long-term follow-up of diet and exercise level. Self-weighing frequency has proven to be associated with weight gain prevention and reduction among working adults (Van Wormer et al. 2012). Self-management of sleep quality could lead to similar gains in sleep improvement. It is however important to prevent people from making wrong interpretations, by informing them about the limitations. A person's weight is divided over body fat, muscle mass, water and bones, thus limiting the information a weight scale might give. Similarly, a cardiac, respiratory and movement-based approach is only an indirect measurement of sleep. Its diagnostic evidence should be handled accordingly.

Chapter 3

Automatic sleep stage classification based on easy to register signals as a validation tool for ergonomic steering in smart bedding systems

Tim Willemen[1], Dorien Van Deun[1], Vincent Verhaert[1], Sandra Pirrera[2], Vasileios Exadaktylos[3], Johan Verbraecken[4], Bart Haex[1], Jos Vander Sloten[1]

[1] Division of Biomechanics, KU Leuven, Celestijnenlaan 300C, 3001 Heverlee, Belgium

[2] Biological Psychology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

[3] M3-BIORES, KU Leuven, Kasteelpark Arenberg 30, 3001 Heverlee, Belgium

[4] Multidisciplinary Sleep Disorders Center, Universitair Ziekenhuis Antwerpen, Wilrijkstraat 10, 2650 Edegem, Belgium

As adapted from *Work: A Journal of Prevention, Assessment and Rehabilitation*, vol. 41, pp. 1985-1989, 2012

Statement of relevance

This chapter is the first of two chapters addressing research question 1. It outlines the results of a proof-of-concept study distinguishing the combination of wake and active sleep (Wake-REM-N1) from passive sleep (N2-N3). It was performed on a small subset of a private database of polysomnographic nights from healthy subjects, as collected during a 4-year IWT-SBO project entitled *EASI-sleep*. Results have been presented with an oral presentation in the *18th World Congress on Ergonomics - 2012*, and published as a conference proceedings paper in a special issue of the journal *Work: A Journal of Prevention, Assessment & Rehabilitation* (Willemen et al. 2012). Readers who went through the general introduction of this thesis (chapter two) are able to skip the introduction of this chapter. Readers with limited time can skip this chapter entirely and narrow down their reading to chapter four, which holds most relevant information related to research question 1.

Abstract

Ergonomic sleep studies would benefit from long-term monitoring in the home environment to cope with daily variations and habituation effects. Polysomnography allows to asses sleep accurately, but is costly, time-consuming and possibly disturbing for the sleeper. Actigraphy on the other hand is cheap and user friendly, but for many studies lacks accuracy and detailed information. This proof-of-concept study investigates the use of three easy to register signals (i.e. heart rate, breathing rate and movement) for distinguishing the combination of wake and active sleep (Wake-N1-REM) from passive sleep (N2-N3), using a Least-Squares Support Vector Machines automatic classifier. The algorithms are trained and validated using 16 polysomnographic nights from 10 healthy subjects. Two different training and test set combinations were analyzed, as to evaluate the benefit of person-specific training. Without person-specific training, an agreement of 78.56 % was obtained. Results improved to 91.48 % when the algorithm was trained in a person-specific way. The authors conclude that the proposed approach could serve its use in long-term sleep monitoring studies in the home environment, especially when prior person-specific polysomnographic training is available.

Keywords

LS-SVM, actigraphy, polysomnography

3.1 Introduction

Although sleep research has been around for almost a hundred years (Piéron 1913), there is still a huge need for studies looking at the influence of environmental effects on sleep, both from a physiological as a psychological point of view. In general, these studies would benefit from long-term monitoring. Polysomnography (PSG) (Rechtschaffen & Kales 1968, Berry et al. 2012) allows to asses sleep accurately, but is costly and time-consuming to apply in long-term studies. Due to the extensive amount of sensors attached to the sleeper's body, it could also disturb the normal sleep pattern, both before and during the night (Metersky & Castriotta 1996). Although habituation takes place after consecutive nights, letting the sleeper return to its normal sleep patterns, in most cases only the first night is disregarded from experimental analysis (due to cost and time). Habituation might however take longer than this so-called 'first-night effect' (Le Bon et al. 2001). Actigraphy (Morgenthaler et al. 2007) on the other hand is both cheap and user-friendly for long-term studies, but for many studies lacks not only detailed information due to its ability to only distinguish between sleep and wake states, but also accuracy due its low specificity in detecting wake states (Paquet et al. 2007). No other methods reached the point beyond prototype or have been sufficiently validated to fill the gap in between (van de Water et al. 2011).

This proof-of-concept study investigates the use of three easy to register signals (i.e. heart rate, breathing rate and movement) for distinguishing the combination of wake and active sleep (Wake-N1-REM) from passive sleep (N2-N3). Heart and breathing rate are coupled to the autonomic nervous system, of which variations in sympathetic and parasympathetic activation are known to correlate well with variations in sleep stages (Somers et al. 1993, Bonnet & Arand 1997, Burgess et al. 1999). The link of movement with sleep and wake states is extensively researched in the domain of actigraphy (Morgenthaler et al. 2007, Sadeh 2011).

Not only can all of these three signals be registered using a strap similar to the ones in classical heart rate monitors used for sports, they can even potentially be measured off-body in the bedding system itself, e.g. using upper body pressure mats (Kortelainen et al. 2010, Bader & Engdal 2000), force sensors in the bed legs (Brink et al. 2006) or optical fibers in the bed cover (Li et al. 2008). Firstly, this would avoid disturbing the person during or preceding his or her sleep, both physically and mentally. Secondly, it would allow for autonomous working, so that it can easily be used in long-term studies, even at home.

Table 3.1: Overview of different training and test set combinations, evaluating the effect of zero person-specific training (combination A) and complete person-specific training (combination B).

	A	**B**
# training nights	8	3
# test nights	4	1
# subjects in training nights	5	1
# subjects in test nights	4	1

3.2 Methodology

3.2.1 Subjects

During a four year project, about 100 subjects participated in different sleep studies, leading to a database of 605 different nights. All were volunteers free from medical problems that would interfere with normal sleep. All studies were approved by the Ethics Committee of the Vrije Universiteit Brussel. A subset of 16 nights was randomly selected for this proof-of-concept study, from a total of 10 subjects between the age of 19 and 30 years old (22.95 $\pm$ 2.74 y). From this subset, two different training and test set combinations were constructed to analyze the effect of person-specific training information. Table 3.1 lists an overview of the amount of different nights and different subjects used in both combinations. In combination A, the training set contained nights from different subjects compared to those from the test set. This was done to evaluate zero person-specific training. In combination B, all nights in the training set were from the same subject as the night in the test set, resulting in complete person-specific training.

3.2.2 Measurements

For all nights in the projects database, complete polysomnographic recordings were performed using the Medatec Dream System (Medatec nv., Brussels, Belgium), after which sleep stages were classified in 30-second epochs according to AASM rules (Berry et al. 2012). Additionally, heart rate was registered through ECG and a Pan-Tompkins R-peak detector (Pan & Tompkins 1985), breathing rate using a thorax respiratory belt and differential peak detector, and movement using the Custom8 DynaSleep system (Custom8, Leuven, Belgium). The latter measures the mattress indentation with a one second resolution,

allowing to relate changes in indentation to movement, as described in (Verhaert et al. 2009).

3.2.3 Data analysis

Heart rate (HR), breathing rate (BR) and movement (MOV) were scaled to unit mean for every night to ensure a better inter-person working of the algorithm. HR and BR were manually synchronized with the MOV signal by making use of the large amplitude spikes in the raw breathing signal when subjects were moving. Since sleep stages in a hypnogram only have a resolution of an epoch of length 30 seconds, this manual synchronization was sufficiently precise.

In order to distinguish the combination of wake and active sleep (Wake-N1-REM) from passive sleep (N2-N3), a binary Least-Squares Support Vector Machine (LS-SVM) classifier was used (Suykens et al. 2002). SVM theory is a supervised non-linear learning method that analyzes data and recognizes patterns (Cristianini & Shawe-Taylor 2000). Per data point in time or space, called a vector, a set of predefined variables is collected, called features. Based on these so-called feature vectors, the trained model predicts which of the two possible classes the input is a member of. It does this with the help of a set of training examples, each belonging to one of the two classes.

A multitude of different time and frequency domain features were defined based on the HR and BR signal: mean, variance, percentiles, approximate entropies, frequency intervals and ratios between frequency intervals. Frequency intervals were not limited to the standard Very Low Frequency (VLF), Low Frequency (LF) and High Frequency (HF) intervals, as defined in (Malik et al. 1996). Instead an exhaustive combination of all possible intervals between 0.01 and 0.40 Hz was defined (with a step size of 0.01 Hz). Features were also averaged over different symmetrically defined intervals of different epoch sizes, varying between 3 epochs (1.5 minutes) and 61 epochs (30.5 minutes). For the MOV signal, features related to the time to previous and to future movements were defined, using different movement thresholds to distinguish between different movement sizes. Due to the large amount of frequency interval combinations and averaging interval sizes, a total feature set of 122844 features was obtained.

In order to suppress noise on the future classification results (due to the use of insignificant features), and to learn which features are more relevant than others, a subset of semi-uncorrelated features was extracted. Using the training set data, the mean and variance of every feature was calculated, separately for the two classes. This allows to estimate the significance by which the means of those two classes differ from each other using an independent two-sample t-test. Second, using Pearson's correlation coefficient, inter-feature correlation

was determined between all features. Third, features were added one by one to the subset in order of t-test significance, as long as their significance p-value was lower than 0.05, and their inter-feature Pearson's correlation coefficient between all the previously selected features lower than 0.70.

A set of 1000 epochs from each of the two classes was used to train the LS-SVM algorithm's parameters using grid search and 10-fold cross validation. Epochs were selected based on having maximum entropy in the complete set of 2000 feature vectors. This ensured a valid representation of the total feature space without the computational load of using all epochs. The specifics of the method can be found in the documentation of the LS-SVMlab software library (De Brabanter et al. 2010). The same entropy-based epoch selection was done for the test sets, although in the case of test set combination B, the amount of epochs for each class was reduced to the amount of the least present class.

Agreement, sensitivity and specificity parameters were calculated using an epoch-by-epoch comparison between the AASM-scored epochs and the resulting test set classifications by the trained LS-SVM algorithm. Agreement was calculated as the percentage total correspondence,

$$\text{agreement} = \frac{TP + TN}{TP + FP + TN + FN}, \tag{3.1}$$

sensitivity as the percentage correctly identified Wake-N1-REM epochs,

$$\text{sensitivity} = \frac{TP}{TP + FN}, \tag{3.2}$$

and specificity as the percentage correctly identified N2-N3 epochs,

$$\text{specificity} = \frac{TN}{TN + FP}, \tag{3.3}$$

where a True Positive (TP) was defined as a correctly identified Wake-N1-REM epoch, a True Negative (TN) as a correctly identified N2-N3 epoch, a False Positive (FP) as a N2-N3 epoch incorrectly identified as a Wake-N1-REM epoch, and a False Negative (FN) as a Wake-N1-REM epoch incorrectly identified as a N2-N3 epoch. Agreements on the complete training and test nights (which do not necessarily have a balanced distribution between both classes) were also calculated.

Table 3.2: Results for the different training and test set combinations.

	A	**B**
Agreement training set	94.20%	94.11%
Agreement test set	75.96%	94.25%
Sensitivity test set	68.38%	90.80%
Specificity test set	83.55%	97.70%
Agreement training nights	94.47 ± 2.31%	93.15 ± 1.53%
Agreement test nights	78.56 ± 10.57%	91.48 ± 0.00%

3.3 Results

Table 3.2 lists the results from both training and test set combinations A and B. First, the 10-fold cross-validated agreement on the 2000 feature vectors of the training set. Second, the agreement, sensitivity and specificity on the 2000 feature vectors of the test set. Third, the agreements on the complete nights of training and test set.

The results of combination A in Table 3.2 show a very high accuracy on the training set (94.20%) as on the complete training nights (94.47%). This is not surprising since epochs from the same nights are used for training the classifier, thus for every epoch in a training night, there will always be an epoch close by that was used in the training set itself. Feature values from that epoch won't be too much different from the ones nearby, thus giving rise to the good results. However, this is still a good indicator that the feature variables defined are able to separate the multidimensional feature space between Wake-N1-Rem and N2-N3 well, thus confirming that HR, BR and MOV are not just randomly changing over sleep stages but have a certain pattern towards different sleep stages. A reasonable accuracy was also found for the test set (75.96%) as on the complete test nights (78.56%), mainly due to a good specificity (83.55%), but lacking sensitivity (68.38%). The high standard deviation on the accuracy of the full test nights (10.57%) indicates that some nights or some subjects are better or worse for automatic classifications than others.

The results of combination B in Table 3.2 indicate that training person-specific leads to better classification. The fact that the agreement on the test set (94.25%) becomes equal to the training set value (94.11%) indicates that not much difference in feature values must have been present between training and test nights for the same sleep stages.

3.4 Discussion

This proof-of-concept study presented some initial results in distinguishing the combination of wake and active sleep (Wake-REM-N1) from passive sleep (N2-N3) using easy to register signals, i.e. heart rate, breathing rate and movement. Compared to other classification algorithms found in literature, converting their listed results to reflect the same binary Wake-N1-REM vs N2-N3 classification, the results described here are on the same level. A HR-based classification algorithm led to an accuracy of 73.27% (Canisius et al. 2010), whereas an algorithm using peripheral arterial tone and actigraphy reached an accuracy of 80.56% (Herscovici et al. 2006).

While results are not on the same level as gold standard PSG, it is important to consider the inter- and intra-scorer variation present within the latter. Schwaibold et al. (Schwaibold et al. 2003) reports an accuracy between two expert scorers of only 85.93% on the same binary classification. This lowers the threshold of comparison between gold standard PSG and the technique presented in this study. Automatic algorithms also benefit from reproducible results and are consistent in mistakes they make, avoiding inter- and intra-scorer variation.

In the current algorithm both information from past, present and future epochs is used, since feature variables are averaged over symmetric interval lengths around the current epoch. Limiting the algorithm to only use past and present epoch information will enable it to control and direct different devices depending on the sleep state of the person. Future work will investigate how much this limitation would deteriorate current classification results, and with which time resolution transitions between different sleep stages can be detected. Future work will also consist of increasing the data set for better validation and extending the sleep classification towards a full hypnogram.

3.5 Conclusion

This proof-of-concept study presented some initial results in distinguishing the combination of wake and active sleep (Wake-REM-N1) from passive sleep (N2-N3) using easy to register signals, i.e. heart rate, breathing rate and movement. Results obtained look promising towards use in long-term studies, especially when person-specific training on the basis of polysomnography is possible. Future work will increase the data set for better validation, and extend the sleep classification towards a full hypnogram.

Acknowledgements

This research was carried out thanks to the financial support of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT).

Chapter 4

An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification

Tim Willemen[1,2,3], Dorien Van Deun[1], Vincent Verhaert[1], Marie Vandekerckhove[4], Vasileios Exadaktylos[5], Johan Verbraecken[6], Sabine Van Huffel[2,3], Bart Haex[1,7], Jos Vander Sloten[1]

[1] Biomechanics Section, Mechanical Engineering Department, KU Leuven, 3001 Heverlee, Belgium

[2] STADIUS, Electrical Engineering Department, KU Leuven, 3001 Heverlee, Belgium

[3] Future Health Department, iMinds, 3001 Heverlee, Belgium

[4] Biological Psychology, Vrije Universiteit Brussel, 1050 Brussel, Belgium

[5] Division of Measure, Model and Manage Bioresponses, Bio Engineering Department, KU Leuven, 3001 Heverlee, Belgium

[6] Multidisciplinary Sleep Disorders Center, Antwerp University Hospital, University of Antwerp, 2650 Edegem, Belgium

[7] Imec, 3001 Heverlee, Belgium

As adapted from *IEEE Journal of Biomedical and Health Informatics*, vol. 18, pp. 661-669, 2014

Statement of relevance

This chapter is the second of two chapters addressing research question 1. It outlines the results of a follow-up study to the previous chapter. While the previous chapter only distinguished the combination of wake and active sleep (Wake-REM-N1) from passive sleep (N2-N3), this chapter distinguishes between Wake, REM, light sleep (N1-N2) and deep sleep (N3). The study was also performed on a larger subset of the same private database of polysomnographic nights from healthy subjects, as used in the previous chapter. Results have been published in the journal *IEEE Journal of Biomedical and Health Informatics* (Willemen et al. 2014a). Readers having read the general introduction of this thesis (chapter two) are able to skip the first two paragraphs of the introduction of this chapter.

Abstract

Polysomnography is considered the gold standard to assess sleep accurately, but it can be expensive, time-consuming and uncomfortable, specifically in long-term sleep studies. Actigraphy, on the other hand, is both cheap and user-friendly, but depending on the application lacks detail and accuracy. Our aim was to evaluate cardio-respiratory and movement signals in discriminating between Wake, REM, light (N1N2), and deep (N3) sleep. The dataset comprised 85 nights of polysomnography from a healthy population. Starting from a total of 750 characteristic variables (features), problem-specific subsets of 40 features were forwardly selected using the combination of a wrapper method (Cohen's Kappa statistic on RBF-kernel Support Vector Machine (SVM) classifier) and filter method (minimum Redundancy Maximum Relevance criterion on Mutual Information). Final classification was performed using an RBF-kernel SVM classifier. Non subject-specific Wake versus Sleep classification resulted in a Cohen's kappa value of 0.695, while REM versus NREM resulted in 0.558 and N3 versus N1N2 in 0.553. The broad pool of initial features gave insight in which features discriminated best between the different classes. The classification results demonstrate the possibility of making long-term sleep monitoring more widely available.

Keywords

Biomedical signal processing, data analysis, medical information systems, sleep research, supervised learning

4.1 Introduction

Sleep can roughly be divided into two main states, labeled Rapid-Eye-Movement (REM) and NREM (N1-N2-N3) sleep, which alternate in cycles of about 90 minutes (Rechtschaffen & Kales 1968). NREM sleep, especially deep sleep (N3), is more prominent during the first hours of sleep, while REM sleep, linked to dreaming, is more prominent during the last hours. Both are essential towards the recovery of our physical condition (Haex 2004, Adam & Oswald 1984) and mental state (Meerlo et al. 2009). Demographics show that up to 24% of the population is faced with regular sleep problems (The National Sleep Foundation 2012), due to e.g. insomnia (the inability to initiate and maintain sleep), obstructive sleep apnea syndrome (OSAS; upper airway collapse during sleep), or a mere lack of sleep hygiene.

Polysomnography (PSG) (Berry et al. 2012) is considered the gold standard in sleep research and allows to assess most aspects of sleep accurately. Unfortunately, it requires at least one night in a specialized sleep lab. Expert technicians apply an extensive amount of sensors to the patient (possibly affecting sleep) and evaluate the collected data manually in 30 second intervals. Due to the complexity of sleep, dependent on physiological as well as psychological factors, great inter-night variability can be present, requiring multiple recording nights to prevent high incidence of false positives and negatives (Carlile & Carlile 2008). These factors make the method costly and time-consuming, limiting its application on a large scale, especially when long-term monitoring is considered. The addition of automatic sleep classifiers on PSG signals can lessen the burden of manual evaluation for a small drop in accuracy (Anderer et al. 2010). Actigraphy (ACT) (Morgenthaler et al. 2007) on the other hand, which makes use of movement information to differentiate between Wake and Sleep, is both cheap and user-friendly. Unfortunately, for many cases, it lacks detail (such as the REM-NREM distinction) and accuracy (Paquet et al. 2007, Bulckaert et al. 2010). No other methods reached the point beyond prototype or have been sufficiently validated to fill the gap between PSG and ACT (van de Water et al. 2011).

Over the years however, extensive research has been performed on changes in heart rate (HR) and breathing rate (BR) across sleep stages and other related events. In 1923, MacWilliams et al. (MacWilliam 1923) were an early pioneer in noting the influence of sleeping and dreaming on blood pressure and heart activity. In 1956, Brooks et al. (Brooks et al. 1956) published the first definitive measurements of HR variation during sleep in six individuals, each observed for 50 nights. In 1964, Snyder et al. (Snyder et al. 1964) studied 12 subjects across 30 nights, reporting significant differences in blood pressure, HR and BR with varying sleep depth. Already in 1973 and 1976, attempts were made to

distinguish active REM-N1 sleep from quiet N2-N3 sleep using HR data (Welch & Richardson 1973, Lisenby et al. 1976). HR and BR are currently known to be linked to the autonomic nervous system (ANS), from which sympathetic and parasympathetic variations correlate well with changes in sleep architecture (Somers et al. 1993, Bonnet & Arand 1997, Burgess et al. 1999). HR and BR can be measured reliably using electrocardiography (ECG) and respiratory inductance plethysmography (RIP), limiting the amount of sensors attached.

This paper evaluates the use of heart rate, breathing rate and movement information in discriminating between between Wake, REM, light (N1N2) and deep (N3) sleep. It presents methods and results on data preprocessing, the extraction and selection of relevant features and the final classification step.

4.2 Materials and methods

4.2.1 Data collection

The dataset comprises a total of 85 nights from a population of 36 healthy volunteers (age 22.1 $\pm$ 3.2 years), with one to four nights per subject. Participants were recruited through advertisement. Inclusion criteria were a good general health condition, regular sleep-wake schedule and an average of 6 to 9 hours of sleep per night. Exclusion criteria were sleeping disturbances, intake of medication or drugs influencing sleep, smoking habits or intake of more than three beverages containing caffeine or alcohol. The study was approved by the Ethics Committee at the 'Vrije Universiteit Brussel'.

For all nights, complete polysomnographic recordings were performed (Dream system, Medatec NV, Brussels, Belgium), after which experts classified sleep stages (Wake, REM, N1, N2, N3) in 30 second epochs according to AASM rules (American Academy of Sleep Medicine) (Berry et al. 2012). Additionally, HR was registered through electrocardiography (ECG), BR using respiratory inductance plethysmography (RIP) and movement (MOV) by the DynaSleep system (Custom8, Leuven, Belgium). The latter continuously measures the perpendicular indentation of the mattress surface in a 2D-grid of 170 points. This allows for a sensitive and accurate registration of body movements, twitches and sleep postures, as described by Verhaert et al. (Verhaert et al. 2011a).

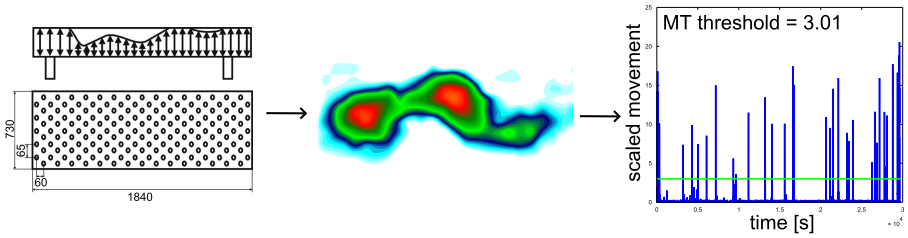


Figure 4.1: Visualisation of the 2D measuring grid of the DynaSleep system (Custom8, Leuven, Belgium) on the left. In the center, an example of the measured indentation pattern from a person lying in the right lateral position. On the right, the extracted movement signal for one night, with the MT threshold indicated as a horizontal line.

4.2.2 Data pre-processing

HR was extracted from 200 Hz ECG measurements using the Pan-Tompkins algorithm (Pan & Tompkins 1985). A search-back post-processing algorithm was applied to identify and correct false positive and false negative R-peak detections by looking for abnormally large or short interbeat intervals. BR was extracted by first applying a cubic spline interpolation to the raw 200 Hz RIP data, smoothing the breathing signal. This allowed for an easy extraction of the valleys and peaks by differentiation, identifying inspiration, expiration and total breathing length intervals. A similar search-back post-processing algorithm was applied as for the R-peaks, in order to correct false positive and false negative peak and valley detections. The 1 Hz movement signal was evaluated by 2D integration of the derivative of the continuously measured indentation over the mattress surface. Normalization was performed by scaling this movement signal with the inverse of the subject's Body Mass Index (BMI). Significant body movement events occur when the signal raises above the threshold value 3.01, as heuristically determined in (Verhaert et al. 2011a). Fig. 4.1 visualizes the measurement system and shows an example of the movement signal for one night, including the MT (movement time) threshold.

The interval size of the feature vectors to be classified was chosen to be 60 seconds, which is a trade-off between feature quality and time resolution. Especially for the breathing rate, smaller intervals can lead to unreliable ventilatory features due to the slow rhythm of the breathing signal. Also, since the lower boundary of the Low Frequency (LF) interval in Heart Rate Variability (HRV) analysis is put at 0.04 Hz, and the HRV Task Force (Malik et al. 1996) recommends an interval length of at least 10 times the wavelength of the lowest frequency

bound (which would require an interval length of 250 seconds), the chosen 60 second interval is already quite short for HRV analysis. An artificial increase in time resolution can always be accomplished afterwards using a 60 second moving window, with a step size of e.g. 15 seconds. To transform the 30 second interval PSG-scored hypnogram values to 60 second interval values, the following decisions were made (in order of priority):

- Intervals containing a significant body movement event were stored as MT
- Intervals containing at least one PSG-scored Wake epoch were stored as Wake
- Intervals containing at least one PSG-scored REM epoch were stored as REM
- Intervals containing at least one PSG-scored N3 epoch were stored as deep sleep (N3)
- The remaining intervals were stored as light sleep (N1N2)

For example, a 60 second interval containing 30 seconds of Wake and 30 seconds of light sleep (N1N2) was stored as Wake; a 60 second interval containing 30 seconds of MT and 30 seconds of REM was stored as MT.

Parts of the data had to be excluded due to bad signal registration (e.g. a loose ECG electrode or bad tension of an RIP breathing belt) in order to ensure proper training and validation data. Given the size of the dataset, two automatic signal quality evaluation methods were implemented for automatic exclusion analysis. For the ECG, the robust Mahalanobis distance between the low frequency energy content of every 60 seconds of ECG data was calculated. Energy content was estimated by integration of the squared wavelet transform coefficients of 10 levels of Daubechies-6 wavelets (db6) (Daubechies et al. 1992), representing the logarithmically spaced frequency spectrum between 0.3 and 10 Hz. The robust Mahalanobis distance is calculated as

$$D_M(x) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu_r})^T \, \boldsymbol{S_r}^{-1} \, (\boldsymbol{x} - \boldsymbol{\mu_r})}\,, \tag{4.1}$$

where $\boldsymbol{x}$ is the multivariate vector containing the 10 energy variables for that interval, $\boldsymbol{\mu_r}$ contains the robust average values of these 10 energy variables over all intervals of that night and $\boldsymbol{S_r}$ is the robust covariance matrix of the data. The robust Mahalanobis distance allows for an estimation of the similarity of the data in that interval to the complete set of intervals. A suitable rejection threshold for ECG intervals having a too large Mahalanobis distance was heuristically determined at a distance value of 30. Calculations were performed

Table 4.1: Overview of the amount and percentage of accepted intervals in the final dataset.

Sleep state	**Training set**		**Validation set**	
	amount	percentage	amount	percentage
MT	2107	88.00	987	87.91
Wake	1297	86.35	808	88.99
REM	3470	83.19	1595	85.48
N1N2	10396	90.76	5113	90.98
N3	6166	86.74	3045	84.23
Total	23550	88.00	11574	87.91

using Libra, a Matlab Library for Robust Analysis (Verboven & Hubert 2005), more specifically using the mcdcov function.

For the RIP, the signal to noise ratio (SNR) of every interval was estimated by taking the ratio of the signal and noise variances,

$$SNR = 10 \log_{10} \left(\frac{\text{var}(\boldsymbol{rip})}{\text{var}\left(\boldsymbol{rip} - \boldsymbol{srip}\right)} \right), \tag{4.2}$$

were $\boldsymbol{rip}$ is the RIP signal and $\boldsymbol{srip}$ is the cubic spline interpolated RIP signal. A suitable rejection threshold for RIP intervals having a too low SNR was heuristically determined at a dB value of 5. At lower SNR values, irregularities start to arise in the breathing waveforms due to insufficient tension in the breathing belt.

In the following sections, 57 nights were used for feature selection and training of the classification model parameters, while 28 nights were used as validation set. No nights from the same subject were present in both training and test set to ensure that there was no person-specific training possible. Overall, 12.00% of the training set 60 second intervals and 12.09% of the test set 60 second intervals were not included in the analysis based on the above described signal quality evaluation methods. Table 4.1 gives an overview of the amount and percentage of accepted intervals for the different sleep states. For the MT intervals, never rejected based on the signal quality evaluation methods, an equal percentage (respectively 12.00% and 12.09% for training and test set) was excluded from analysis to maintain a balanced data set. Since from every interval type a relatively equal amount of data was rejected, a realistic distribution of sleep data was ensured.

4.2.3 Feature extraction

After the data preprocessing step, the following signals were obtained, divided in 60 second intervals and labeled with their respective sleep states (as defined in section 4.2.2):

- Start time and length of each beat-to-beat RR-interval (HR)
- Start time and length of each breathing cycle (BR)
- Start time and length of the inspiratory phase of each breathing cycle, expiratory phase of each breathing cycle and the ratio between these lengths (BRin, BRout and BRinoutratio)
- The 1 Hz movement signal derived from the DynaSleep system (MOV)

For each of these intervals an extensive number of features was defined. The choice of evaluating a broad set of features instead of a limited one, will allow to find a more optimal set of features, with respect to the specific classification problem and population. An overview of the defined features is given in Table 4.2. For the HR signal, feature types 1 through 8 were extracted, giving a total of 81 HR features. The features within feature type 4 were calculated by first detrending the HR signal by substracting the mean HR value of the previous x seconds, with x equal to 150, 600, 1800 and 7200. The features within feature type 7 were calculated by integration of the squared wavelet transform coefficients of 40 levels of db6 wavelets representing the linearly spaced frequency spectrum between 0.01 Hz and 0.40 Hz. The wavelet transforms were executed on the 2 Hz cubic spline interpolated HR signal. The features within feature type 8 were calculated by taking the natural logarithm of '1 + the original feature value'.

For the BR, BRin, BRout and BRinoutratio signals, the same feature types were extracted, except for feature type 7 and its logarithmic transform within feature type 8, giving a total of 65 features each. For the MOV signal, feature types 9 through 13 were extracted, giving a total of 34 MOV features. Movement intensity within feature type 9 through 12 was calculated by integrating the MOV signal over each movement interval with all its values above a minimal threshold of 0.15 (to cancel out noise). The values of the different movement intensity thresholds were 0, 0.75, 2.5, 7.5 and 25. The window sizes used in feature type 13 were 60, 150, 300 and 600.

In total, this sums up to 375 features per 60 second interval. Finally, for each of the previously defined feature types, a night-specific normalized version of the feature type was defined. For every night in the dataset and for every feature, the robust mean and standard deviation were calculated (again using Libra (Verboven & Hubert 2005), more specifically the median and madc functions).

Table 4.2: Features defined on the signals HR (1-8), BR, BRin, BRout, BRinoutratio (1-6, 8) and MOV (9-13).

	Feature name	**Amount per signal**	**Short description**
1	mean	1	mean length of RR intervals or breath cycles
2	percentiles	7	2.5, 10, 25, 50 (=median), 75, 90 and 97.5 percentile
3	inter percentile ranges	3	range between percentile 2.5-97.5, 10-90, and 25-75
4	detrended of 1-2-3	44	by subtracting the mean value of the previous x seconds, ... with x $\in$ [150 600 1800 7200].
5	variance	1	variance of RR intervals/breath cycles
6	median absolute deviation	4	median(abs(xi-median)), with xi = data points of interval; ... ... variants by switching median with mean
7	frequency HRV	8	VLF (0.01 - 0.03), LF (0.04 - 0.15), HF (0.16 - 0.40), ... LF/HF (absolute and relative)
8	log of 5-6-7	13	natural logarithm of '1 + original feature value'
9	time to next movement	5	time in seconds until movement signal $\geq$ intensity threshold x, ... with x $\in$ [0 0.75 2.5 7.5 25]
10	time to previous movement	5	cfr. feature 9
11	time in between movements	5	cfr. feature 9
12	log of 9-10-11	15	natural logarithm of '1 + original feature value'
13	amount of movements	4	amount of movements within centered window $\in$ [60 150 300 600]

These parameters were then used to normalize the original feature values of every 60 second interval of that night, using the formula

$$feature_{norm} = \left(\frac{feature - \text{median}}{\text{madc}} \right). \tag{4.3}$$

This normalization should allow for a better interchangeability between features from different subjects and from different nights. The final amount of features thus sums up to 750.

4.2.4 Feature selection

In order to reduce computational cost, complexity and noise on the classification result, a subset of features was selected on the training set as being most descriptive for the targeted classification problem and population. For this purpose, a combination of a filter method and wrapper method was used. Filter methods are low in computational costs since they work independent from a classification algorithm, while wrapper methods require parameter training and effective classification. As classification algorithm, the libSVM C-implementation (Chang & Lin 2011) of an RBF-kernel (Radial Basis Function) Support Vector Machine (SVM) was used, which requires two parameters to train, namely a cost parameter and regularization parameter. To save on computation time, the wrapper method was thus only used for the evaluation of single feature classification performance by calculating a five-fold cross-validated value of Cohen's kappa (Landis & Koch 1977). The latter is defined as

$$\text{kappa} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \tag{4.4}$$

which corrects agreement,

$$\Pr(a) = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4.5}$$

for agreement occurring by chance,

$$\Pr(e) = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}, \tag{4.6}$$

where a True Positive (TP) was defined as a correctly identified class one instance, a True Negative (TN) as a correctly identified class two instance, a

False Positive (FP) as a class two instance incorrectly identified as a class one instance, and a False Negative (FN) as a class one instance incorrectly identified as a class two instance.

For evaluating synergies between features (constructing an optimized set of features that work well together), a filter method based on mutual information and a minimum Redundancy Maximum Relevance-criterion (mRMR) was used (Peng et al. 2005). The criterion is defined as,

$$\max_{x_j \in (\boldsymbol{X} - \boldsymbol{S}_{m-1})} \left[I\left(x_j; c\right) - \frac{1}{m-1} \sum_{x_j \in \boldsymbol{S}_{m-1}} I\left(x_j; x_i\right) \right], \tag{4.7}$$

a greedy forward selection algorithm with $\boldsymbol{X}$ representing the complete set of features x_j, $\boldsymbol{S}$ the subset of already selected features x_i (of size m), c representing the class identifier and I the mutual information, defined as

$$I\left(x; y\right) = \iint \boldsymbol{p}\left(x, y\right) \log \frac{\boldsymbol{p}\left(x, y\right)}{\boldsymbol{p}\left(x\right) \boldsymbol{p}\left(y\right)} \, \mathrm{d}x \, \mathrm{d}y \,, \tag{4.8}$$

with $\boldsymbol{p}\left(x\right)$, $\boldsymbol{p}\left(y\right)$ and $\boldsymbol{p}\left(x, y\right)$ probability density functions. These functions were approximated by an adaptive diffusion-based kernel density estimator (Botev et al. 2010). The left side of figure 4.2 shows an example of the probability density functions $\boldsymbol{p}\left(x_j, c_1\right)$ and $\boldsymbol{p}\left(x_j, c_2\right)$, used for the calculation of $I\left(x_j; c\right)$. The more overlap, the lower the mutual information will be. The right side of figure 4.2 shows $\boldsymbol{p}\left(x_j, x_i\right)$, used for calculating $I\left(x_j; x_i\right)$. The more the distribution follows the diagonal, the higher the mutual information between the two features.

The feature selection process was thus performed in two steps. In step one, the previously described wrapper method (RBF-kernel SVM) was used in order to obtain a fast reduction in feature amount, and this for each of the four signal types separately (HR, BR, BRin/BRout/BRinoutratio and MOV) in order to ensure the presence of features across all signals. As a result, 75% of features having the lowest values of Cohen's kappa were discarded from the feature set, as to only keep those features already reaching a decent classification accuracy on their own. In step two, a subset of 10 synergetic features per signal type was selected out of the remaining feature set using the previously described filter method (mRMR-criterion), leading to a final subset of 40 features. Further optimization of this set (e.g. by further reducing its size, while evaluating the impact on classification accuracy) was no part of this study.

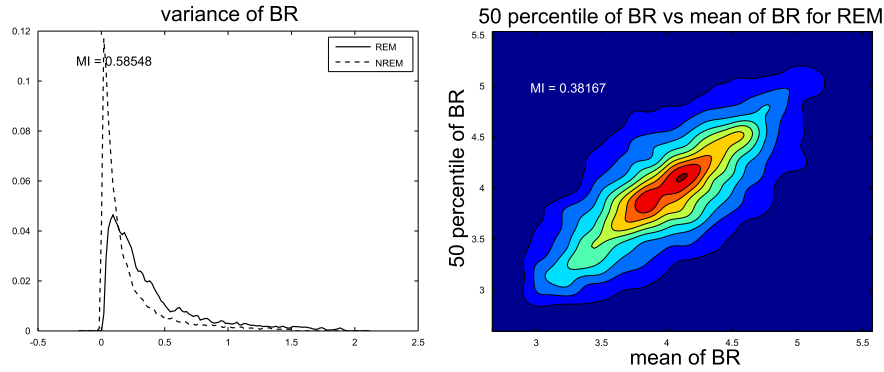


Figure 4.2: Examples of feature-class (on the left) and inter-feature (on the right) probability density functions used to calculate the mutual information (MI) values of the mRMR-criterion. More overlap in the feature-class distributions will result in smaller mutual information between the feature and the classes. The more the inter-feature distributions follow the diagonal, the higher their mutual information will be.

4.2.5 Feature classification

In order to distinguish Wake from Sleep, both at sleep onset (the first hour of every night) and during the night (the remaining hours of the night), and to distinguish NREM sleep from REM sleep and light sleep (N1N2) from deep sleep (N3), four separate binary classification problems were assessed. For this purpose, four separate feature sets were selected, one for each classification problem. The results of the different binary classification problems were also combined in a decision tree format to evaluate the complete Wake-Sleep, Wake-REM-NREM and Wake-REM-N1N2-N3 classification problems. In all classification problems, MT intervals were scored as Wake.

As already stated in the previous section, classification was performed using the libSVM C-implementation of an RBF-kernel SVM. Its two parameters, the cost (penalizing value for misclassifications during training of the classifier) and the RBF-kernel parameter gamma, were optimized using five-fold cross-validation and gridsearch on the training set. The optimal parameters were selected based on a maximal averaged value of Cohen's kappa over the five folds. Besides agreement and Cohen's kappa, also precision and recall were calculated, defined as

$$\text{precision} = \frac{TP}{TP + FP}, \tag{4.9}$$

$$\text{recall} = \frac{TP}{TP + FN}. \tag{4.10}$$

Recall quantifies the amount of class one instances (i.e. WAKE, REM, N3) identified within the complete set of instances, while precision quantifies how precise the classifier was in this identification. Hence, there is normally a trade-off between precision and recall.

The cost penalizing value parameter of the SVM can have a different value for each of the two classes, in order to cope with unbalanced datasets in different ways. Increasing the cost value for class one will cause an increase in precision (less false positives), but a drop in recall (more false negatives). Tuning the difference between the cost value of class one and two allows to maximize Cohen's kappa, although for some applications one could be more interested in maximizing precision or recall itself. During the five-fold cross-validated training phase of our classifiers, we first optimized this ratio between the cost parameter of the two classes, using the default recommended values for cost (=1) and gamma (=1/feature amount = 1/40). In the next step, using this optimal ratio, the five-fold cross-validated gridsearch was performed over cost and gamma parameters.

4.3 Results

Table 4.3 gives an overview of the averaged cross-validated training results and the test set results for all four classification problems. The next subsections will only hold information about which features were selected, and thus deemed distinctive, for all classification problems.

4.3.1 Wake vs Sleep at onset: $<$ 60 min

As features, mean and percentile features (25th, median, 75th) were the most important for the HR and BR signal, with their average values being higher in Wake compared to Sleep. Normalization was important here, proven by the fact that 7/10 selected HR and 9/10 selected BR features were normalized. Detrending was less important (6/10 HR and 3/10 BR) mainly because no data was present to detrend with from before subjects went to bed, thus detrending was done with only the data from bed time on. Since no BRinoutratio features

Table 4.3: Results of the different classification problems (in percentages).

	Wake-Sleep < 60 min[a]	**Wake-Sleep** > 60 min[b]	**REM-NREM**	**N3-N1N2**
Training set				
Agreement	90.21	91.14	84.21	78.59
Kappa	74.10	66.08	49.27	53.56
Precision	78.13	61.54	53.40	72.78
Recall	83.33	84.37	65.71	67.88
Validation set				
Agreement	89.55	91.96	86.24	79.48
Kappa	76.23	66.55	55.81	55.25
Precision	86.75	65.96	55.78	74.99
Recall	81.36	77.32	75.30	67.55

[a]Only using data from the first hour of every night.

[b]Only using data after first hour of every night.

were selected, the perceived trend in increasing BRout interval lengths during sleep onset compared to BRin interval lengths was not significantly present enough for all subjects. The selected BRin and BRout features followed the same characteristics as the selected BR features. As for movement features, the logarithmic transform was important (7/10 selected MOV features) to make its distribution of values Gaussian. Selected features were time between higher intensity movements (2.5, 7.5, 25; no twitches) and time to the last higher intensity movement.

4.3.2 Wake vs Sleep after onset: $>$ 60 min

As HR features, a raise in heart rhythm was found distinctive, characterized by mean and upper percentile (median, 75th, 90th, 97.5th) features being selected, coupled to detrending over 150 seconds (selected for 7/10 features; 9/10 features were detrended). For BR features, it was primarily the higher variance (also its variants Median Absolute Deviation (MAD) and Inter Quartile Range (IQR)) that was found distinctive, with again detrending found important (8/10; 4/10 was detrended over 150 seconds). Overall, normalization was less important here (5/10 HR and 1/10 BR), probably because longer periods of Wake (like with insomnia patients) were not present in the dataset, and periods of Wake were thus primarily coupled to a higher amount of movement activity,

leading to a more similar overall higher heart and breathing rhythm for all subjects. There were 7 BRinoutratio features selected, distributed over variance (and its variants) and extreme percentiles at both ends (thus also reflecting variance). This probably reflects the difference between controlling our breathing more consciously (sympathetic) or it being controlled completely unconsciously during sleep (parasympathetic). For movement features, again the logarithmic transform was important (10/10), and selected features were the time to the last movement or time to the next movement with at least an intensity of 2.5 (so again no twitches).

4.3.3 REM vs NREM

As HR and BR features, mainly the presence of short periods of higher heart and breathing rhythm were found distinctive for REM sleep, reflected in the selection of upper percentile features (75th, 90th, 97.5th) but no mean. For BR, also higher MAD and IQR in REM were characteristic. Normalization was again important (7/10 HR and 7/10 BR), together with detrending although this time over a longer timeframe of 7200 seconds (7/10 HR and 7/10 BR). There were also 3 BRinoutratio features selected (MAD and IQR), probably again reflecting the shifting balance between sympathetic (REM) and parasympathetic (NREM) activity. All other features were BRout features, and no BRin features, which seems to tell that the changes in breathing rhythm are mainly caused by changes in the expiration phase. Besides the necessary logarithmic transform (7/10), selected MOV features were all related to the time between movements of any intensity, which means also short twitches.

4.3.4 N3 vs N1N2

As HR and BR features, the presence of a lower heart and breathing rhythm within N3 sleep was found distinctive, reflected in the selection of lower percentile features (2.5th, 10th, 25th) but this time also mean and median for HR. For BR, also variance, MAD and IQR were selected as features, their values being significantly lower in N3 sleep. Normalization of features was again important (6/10 HR and 10/10 BR), as was detrending (9/10 HR detrended over 600 sec and 5/10 BR). The detrending length of 600 seconds is remarkable since most periods of N3 sleep lasted longer than 10 minutes, which would make a longer detrending period more logical. The next possible interval however (1800 seconds) could have been too large. Adding more possible interval lengths should be considered. No BRinoutratio features were selected; BRin and BRout features followed the same characteristics as the selected BR features. As for

movement features, the logarithmic transform (7/10) and the time to the last movement of any intensity was important. This means that longterm absence of movement is a good predictor for N3 sleep.

4.4 Discussion

An extensive study of feature extraction, selection and classification was presented for Wake-Sleep, REM-NREM and N3-N1N2 classification. Table 4.4 compares the obtained classification results to similar classification algorithms found in literature. The last row of the table lists the presented results of this study. It has by far the largest dataset, even with 10% of the data being excluded due to noise, mostly caused by the RIP-signal for the detection of breathing. The use of a nasal thermistor would have led to a better quality signal, but is less comfortable for the subject, with more interference of normal sleep. The large dataset allowed for a reliable feature selection process, algorithm training and classifier validation. The resemblance in Cohen's Kappa values on training and validation set (Table 4.3) proves the absence of overtraining, and thus the reliable interchangeability between different healthy subjects and nights. In the case of sleep-disordered patients however, separate training data sets will be required to classify for example breathing disorder related events as apneas or periodic leg movement disorder events. During these events, output from the currently presented classifier will not be meaningful and should be disregarded. For the case of insomnia, a separate training data set with a higher presence of wake after sleep onset would probably give a higher classification accuracy, since this information was barely present in the currently used data set.

The use of different cost parameter ratios within the classifier allowed to optimize the trade-off between false positives and false negatives, leading to the highest kappa values. In some applications, one can however prefer a higher sensitivity, or a higher precision, thus preferring a result with a lower kappa value. Some authors (Harper et al. 1987, Devot et al. 2010, Mendez et al. 2010a, Redmond & Heneghan 2006, Redmond et al. 2007, Canisius et al. 2008, Kortelainen et al. 2010, Migliorini et al. 2010) used time-dependent a priori probabilities to increase the accuracy of their classifiers, giving a higher chance to the classifier of scoring Wake, REM, N1N2 or N3 sleep in certain parts of the night. Although this leads in general to better results for normal healthy sleep, it can lead to distorted results when less than normal sleep characteristics occur. Others (Devot et al. 2010, Mendez et al. 2010a, Kortelainen et al. 2010, Migliorini et al. 2010) used smoothing algorithms such as median filters or averaging over e.g. 15 minute intervals, either on the classified hypnogram or on the feature values themselves. While this can improve overall accuracy, it gives rise to a

Table 4.4: Overview of similar classification algorithms found in literature. Agr = agreement. Nights: h = healthy; o = OSAS; i = insomniacs. Subj Spec = Subject specific.

Author	Signals	Classification	Agr	Kappa	Nights	Subj Spec	Epoch
(Harper et al. 1987)	HR, BR	Wake-REM-NREM	85%	-	25 h	No	60s
	HR		82%	-			
	BR		80%	-			
(Redmond & Heneghan 2006)	HR, BR	Wake-REM-NREM	79%	0.56	37 o	Yes[a]	30s
	HR, BR		67%	0.32		No	
	EEG		87%	0.75		Yes[a]	
	EEG		84%	0.68		No	
(Redmond et al. 2007)	HR, BR	Wake-Sleep	89%	0.60	31 h	No	30s
		Wake-REM-NREM	76%	0.46			
(Canisius et al. 2008)	HR	Wake-REM-NREM	76%	-	18 h	No	30s
(Devot et al. 2010)	HR, BR, MOV	Wake-Sleep	96%	0.70	9 h	No	30s
	HR, BR, MOV		85%	0.61	27 i		
	Actigraphy		94%	0.51	9 h		
	Actigraphy		78%	0.39	27 i		
(Kortelainen et al. 2010)	HR, MOV	Wake-REM-NREM	79%	0.44	18 h	No[b]	30s
(Mendez et al. 2010a)	HR	REM-NREM	79%	-	25 h	No	30s
(Migliorini et al. 2010)	HR, BR, MOV	Wake-REM-NREM	77%	0.55	17 h	No[b]	30s
This paper	HR, BR, MOV	Wake-Sleep	92%	0.69	85 h	No	60s
		REM-NREM	86%	0.56			
		N3-N1N2	79%	0.55			
		Wake-REM-NREM	81%	0.62			
		Wake-REM-N1N2-N3	69%	0.56			

[a]Test and training set information derived from the same night; possible overestimation of subject-specific classification potential

[b]Test and training set contain nights from the same individuals; possible overestimation of non-subject-specific classification potential

drop in time resolution and will likely distort the time location of sleep stage transitions. In some applications however, when only a rough overview of the sleep stage distribution of a person is required, these methods could significantly improve the overall correlation between an EEG-based and a cardio-respiratory based sleep assessment.

Harper et al. (Harper et al. 1987) and Devot et al. (Devot et al. 2010) showed that a combination of signals (cfr. their results in Table 4.4) can improve classification accuracy significantly. The broad pool of initial features used on this combination of signals is one of the main factors that contributed to the good classification accuracies in this study. A thorough description of the selected features is discussed in the results section and can be used to optimize future feature starting sets. In the REM-NREM classification problem, the absence of features regarding the energy spectrum of Heart Rate Variability (HRV) is remarkable, since it was the main investigated feature in the study of Mendez et al. (Mendez et al. 2010a), and important in many others. While NREM sleep is associated with a decrease in both frequency and power of sympathetic bursts, no complete absence of this activity was observed. Somers et al. (Somers et al. 1993) and Bonnet et al. (Bonnet & Arand 1997) described that sympathetic bursts will occur e.g. at locations of K-complexes and arousals. Furthermore, during REM sleep, sympathetic activity is mainly concentrated during periods of rapid-eye movement and complete muscle atonia. In between, there is a marked decrease in activity, which makes them looking more like NREM periods.

The use of short intervals without smoothing or averaging could thus lead to false positive and negative classifications. While these sympathetic activity changes should also have an influence on heart and breathing rate rhythm and variance, their impact proves to be less. It could be hypothesized that the change in rhythm and variance is only a second order effect, thus inherently filtering out the more rapid changes in presence and absence of sympathetic activity. Further on, since the HRV Task Force (Malik et al. 1996) recommends interval lengths of at least 10 times the wavelength of the lowest frequency bound, the chosen 60 second intervals might just be too short and unreliable for frequency HRV analysis. Besides Harper et al. (Harper et al. 1987) and this study, all other authors made use of 30 second intervals instead of 60 second intervals, mainly because 30 second intervals are considered the gold standard in polysomnography, acting as a reference method. The predominant frequencies observed in our brain waves are however an order of two greater than the frequencies observed in our heart and breathing rhythm, making the choice for 30 second intervals in cardio-respiratory based sleep scoring less obvious. Future work must therefore investigate the influence of interval length on feature quality and classification accuracy.

Bonnet et al. (Bonnet & Arand 1997) also described that the shift in HRV (from low to high sympathetic activity) already starts a few minutes before PSG-scored REM onset, and also lasts a few minutes longer than the PSG-scored REM period. In PSG, the epochs after REM sleep are usually scored as light N1 sleep, and in over 50 percent of the cases the epochs before REM sleep too (Kishi et al. 2011). Interesting is the fact that, when only using EEG activity information, sleep stages N1 and REM are almost impossible to distinguish. This is one of the reasons why they used to be combined to describe a single sleep state. This REM-N1 correspondence could thus be the source of difficult to avoid REM-NREM misclassifications. A possible solution would be to combine sleep stages N1 and REM again in cardio-respiratory based sleep scoring.

Redmond et al. (Redmond & Heneghan 2006) showed that an automatic classifier based on EEG signals has a much smaller drop in accuracy when going from subject-specific to non-subject-specific, compared to a classifier based on HR and BR. This can be related to the greater variability in heart and lung strength and capacity among different subjects, compared to a greater similarity in brain functioning. While this reveals a possible limitation in HR and BR classification accuracy, the current results show that sufficiently accurately results are feasible whilst using the right features. It is important to mention that, due to inter-scorer variability present in expert polysomnography, the reference intervals are not always correct. Danker-Hopfe et al. (Danker-hopfe et al. 2009) reported an inter-scorer Cohen's Kappa of 0.7626 in a database of 56 healthy and 16 sleep disordered patients, when considering a 5-class Wake-REM-N1-N2-N3 classification according to AASM rules. Automatic classifiers, even on EEG signals, are still not capable of attaining or surpassing this accuracy. They are however consistent in the decisions and mistakes they make, which is not the case in humans.

4.5 Conclusion

An extensive study of feature extraction, selection and classification was presented, for Wake-Sleep, REM-NREM and N3-N1N2 classification. The broad pool of initial features gave insight in which features discriminated best between the separate classes. Attributes such as normalization and detrending proved vital. The use of different cost ratio parameters within the classifier allowed for a trade-off between sensitivity and precision. Future work will investigate the influence of interval length on feature quality and classification accuracy, and will evaluate the classification method on sleep-disordered patients if the necessary datasets can be acquired.

Acknowledgements

This work was supported in part by the Institute for the Promotion of Innovation through Science and Technology in Flanders IWT (SBO EASI-SLEEP), by iMinds (ICON NXT_SLEEP) and by the KU Leuven Research Council (GOA MANET and PFV/10/002 (OPTEC)).

Chapter 5

Assessment of Different Methodologies to Include Temporal Information in Classifying Episodes of Sleep Apnea Based on Single-Lead Electrocardiogram

Tim Willemen[1,2,3], Carolina Varon[2,3], Bart Haex[1,4], Jos Vander Sloten[1], Sabine Van Huffel [2,3]

[1] Biomechanics Section, Department of Mechanical Engineering, KU Leuven, 3001 Leuven, Belgium

[2] STADIUS, Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium

[3] Medical IT, iMinds, 3001 Leuven, Belgium

[4] Imec, 3001 Leuven, Belgium

As adapted from *Proceedings of the 41st Computers in Cardiology Conference*, Cambridge, pp. 1-4, Sep. 2014.

Statement of relevance

This chapter is the first of two chapters addressing research question 2. It outlines the results of a study distinguishing between apneic events and healthy breathing. It was performed on the public Apnea-ECG database, available at Physionet (Goldberger et al. 2000). The results of this study have been presented with an oral presentation in the *41st Annual Computing in Cardiology Conference - 2014*, and published online as a conference proceedings paper in the *IEEE Xplore Digital Library* (Willemen et al. 2014c). Readers who went through the general introduction of this thesis (chapter two) are able to skip the first two paragraphs of the introduction of this chapter. The details of the feature extraction and feature selection methodology of this chapter can be better understood by reading those respective subsections in the next chapter (chapter six). Readers with limited time can skip this chapter entirely and narrow down their reading to chapter six, which holds most relevant information related to research question 2.

Abstract

Automated analysis of sleep apnea based on single-lead electrocardiogram would make screening and diagnosis much more accessible. Over the years, several algorithms have been proposed in the literature. In most of them, one or several temporal averaging techniques are used to improve classifier performance. A comprehensive comparison between those techniques however has never been published. Four different temporal averaging techniques, as well as overlapping of segments, were independently assessed using a database of 70 night-time recordings, originally released for the Computers in Cardiology challenge in 2000. Classification was performed with an LDA classifier. Multiple problem-specific feature sets of 10 features were selected out of a complete set of 304 using a two-step approach. Averaging classifier input features over neighboring segments led to the highest agreement values on the test set, outperforming the best automatic entry during the original competition (90.4% vs 89.4%). When combining classifier output values, an odd amount of segments should be used. Calculating features on larger segments (> 1-min) led to the worst results, possibly explained by its higher susceptibility to noise. Overlapping of segments improved overall agreement by about 1%.

Keywords

Biomedical signal processing, supervised learning, sleep apnea, Physionet

5.1 Introduction

Sleep apnea is an under-diagnosed sleep-related breathing disorder which has an estimated prevalence of 14% in men and 5% in women (Peppard et al. 2013). Clinical diagnosis is based on a presence of five or more apneic events per hour in combination with excessive day time sleepiness that cannot better be explained by other factors. An apneic event is currently defined as a clear decrease from baseline in breathing volume of at least 10 seconds (Berry et al. 2012). Apneic events can be obstructive (complete or partial obstruction of upper airway with surrounding soft tissue) or central (reduced or absent breathing effort). Mixed events are a combination of both, usually starting as a central event, but continuing as an obstructive event once breathing effort is reinitiated. In the case of only a partial obstruction, or when breathing effort is still present but significantly reduced, an apneic event is also referred to as a hypopnea.

Current diagnosis is made using polysomnography or polygraphy, requiring the use of an extensive amount of sensors (Berry et al. 2012). Annotation of apneic events is performed manually, resulting in large inter- and intra-observer variability (Whitney et al. 1998). Automated analysis based on less obtrusive and expensive sensors would reduce costs, making diagnosis much more accessible and allowing for cost-effective population-wide screening.

This paper discusses automated analysis of sleep apnea based on single-lead electrocardiogram (ECG). More specifically, it focuses on the assessment of different methodologies to include temporal information, i.e. looking both backward and forward in time in order to improve classification performance. Many authors have used one or several of these techniques to improve their classifier's performance, but a comprehensive comparison between them has never been published.

5.2 Methodology

The database used in this study was released online at PhysioNet on the occasion of a scientific competition held during the Computers in Cardiology conference in 2000 (Penzel et al. 2000a). It contains 70 night-time recordings of single-lead ECG, simultaneously recorded with full polysomnography to provide expert annotations on the presence of apneic events according to clinical standards. These annotations were transformed by the competition organizers to reflect the presence of sleep apnea in 1-minute segments, leading to the final dataset comprising 34313 minutes of annotated data, split up in 35 training and 35 test nights.

First, RR interval signals were computed from the ECG signals using Pan-Tompkins (Pan & Tompkins 1985), while ECG-Derived Respiration (EDR) signals were calculated from the 200 ms median filtered ECG signal by integrating the area underneath the QRS complexes, changing under the influence of respiration. Next, for every 1-minute segment of RR and EDR data, both time and frequency domain features were calculated, similarly as in (Bsoul et al. 2011). Calculating features on detrended RR and EDR signals, as well as on the original traces, and adding optional logarithmic transformation and night-specific normalisation, led to a total feature set of 304 features per 1-minute segment.

Classification was performed using Linear Discriminant Analysis (LDA) (Ripley 1996), applying gridsearch in order to train its parameters (prior class probability and covariance matrix regularization), both having possible values between 0 and 1. In order to reduce computational cost, complexity and noise on the classification result, unique subsets of features were selected on the training set as being most descriptive for their respective targeted classification problems (i.e. the different temporal information methodologies). Feature selection was performed in two steps. In step one, features having a single-feature classification accuracy of Cohen's kappa (Landis & Koch 1977) smaller than 0.20 were disregarded from the feature set. In step two, a greedy forward selection algorithm extracted 10 features subsequently based on a maximal Cohen's kappa criterion after every step.

Cohen's kappa is defined as

$$\text{kappa} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \tag{5.1}$$

which corrects agreement,

$$\Pr(a) = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5.2}$$

for agreement occurring by chance,

$$\Pr(e) = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}, \tag{5.3}$$

where a True Positive (TP) was defined as a correctly identified apneic breathing segment, a True Negative (TN) as a correctly identified normal breathing segment, a False Positive (FP) as a normal breathing segment incorrectly

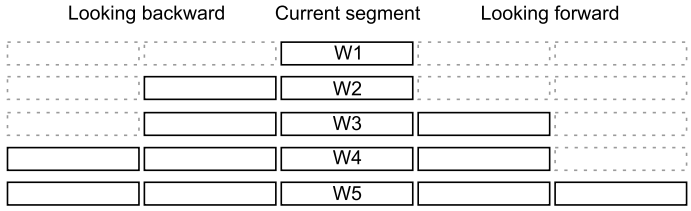


Figure 5.1: The five combinations of neighboring segments assessed for every temporal information methodology.

identified as an apneic one, and a False Negative (FN) as an apneic breathing segment incorrectly identified as a normal one.

To prevent overfitting on the training data, a double cross-validation scheme was introduced. The first cross-validation layer allowed for a robust selection of features, splitting up the training set 10 times at random into 2/3 training data and 1/3 validation data, averaging Cohen's kappa values over these 10 folds. A second similar cross-validation layer, within the training data from each of the 10 first cross validation splits, allowed for a robust gridsearch optimization of the LDA classifier parameters.

Four different temporal information methodologies (i.e. looking both backward and forward in time in order to improve classification performance) were investigated. First, feature values were calculated over wider segments (> 1-minute) around every 1-minute segment, shifting these wider segments with 1-minute steps (WE = window extension). Second, feature values were averaged over neighbouring segments (FA = feature averaging). Third, classifier output values (+1 and -1) were averaged over neighbouring segments using majority vote, prioritizing +1 (apnea) over -1 (normal) in case of tied results (RA = result averaging). Fourth, feature values of both current segment as well as neighbouring segments were used as input for the classifier (FE = feature extension). The amount of neighbouring segments used (looking backward and forward), varied between 0 and 4. Five different combinations were assessed for every temporal information methodology, as visualized in Figure 5.1.

Finally, apart from the other four temporal information methodologies, the effect of overlapping 1-minute segments was investigated by shifting 1-minute segments with 30-sec steps instead of 1-minute steps, as visualized in Figure 5.2. This to prevent misdetection of apneic events situated at the boundary between two segments. Classifier output values of these three overlapping segments were averaged by majority vote.

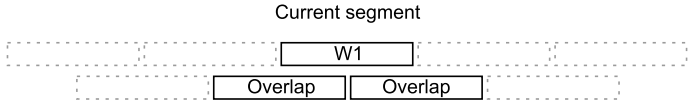


Figure 5.2: Overlapping segments with 30-second shifts.

5.3 Results

Box plots of the 10 fold classification agreement values on the training set (left) and agreement values on the test set (right) are displayed in Figure 5.3, for each of the four temporal information methodologies and for each of the five neighboring segment combinations. Looking at the training set results, an overall increase in agreement can be seen for all temporal information methodologies when the amount of incorporated neighboring segments is increased ($p < 0.01$ on both student t-test and Mann-Whitney u-test for all methodologies). For the window extension method, this increase is however saturated from W3 on, while for both the window extension method and the result averaging method, an even amount of segments (W2, W4) causes a drop in agreement compared to an odd amount of segments. Agreement values for the feature averaging and feature extension methods are the highest, with slightly higher agreement for the feature averaging method. Looking at the test set, results are similar as on the training set, except for the window extension method where a drop in agreement can be seen, especially for an even amount of segments. Agreement values are again the highest for the feature averaging method, reaching as high as 90.44% with W5.

Overall, single-feature classification accuracies (not shown) improved the most for result averaging, especially for an odd number of segments (with W5, 255 of 304 features improve), followed by feature averaging (163 of 304 features), feature extension (149 of 304 features) and window extension (137 of 304 features). For the latter three, mainly more robust features improved (percentiles, frequency content location values, logarithmic transform), while others degraded. Table 5.1 lists the 10 selected features for W1. The feature sets for W5 are similar in all temporal information methodologies, except for the absence of VLF energy features in the window extension method, which are replaced by more robust frequency location values of energy content. Classification agreement values on training and test set, with all methods using the same feature set W1 instead of their method-specific trained versions, show similar trends to those shown in Figure 5.3. Saturation of increasing agreement with amount of incorporated neighboring segments is now present however from W3 on for all temporal information methodologies, leading to slightly lower agreement values (about 1% for W3, about 2% for W5).

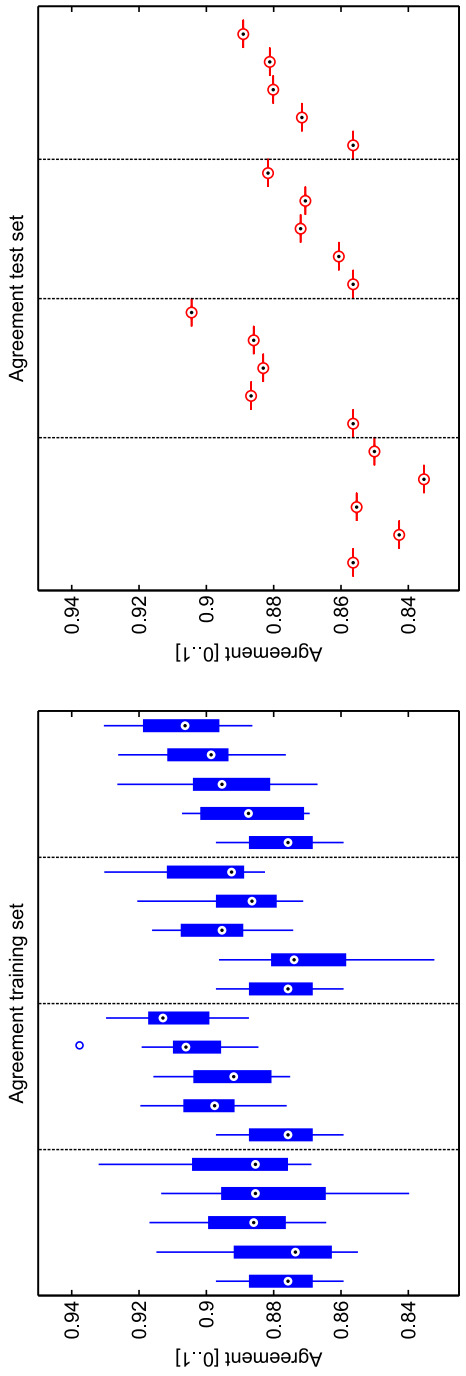


Figure 5.3: Box plots of the 10 fold classification agreement values on the training set (left) and agreement values on the test set (right) for each of the four temporal information methodologies (Window Extension, Feature Averaging, Result Averaging, Feature Extension). The five combinations of neighboring segments assessed (W1 through W5) are plotted from left to right.

Table 5.1: List of the 10 greedy forward selected features for standard W1 segments.

	W1(all methods)
1	log MAD EDR; mean(abs(x-mean))
2	EDR relative VLF energy (0.003-0.04 Hz)
3	RR serial correlation coefficient for k = 3
4	log MAD RR; mean(abs(x-mean))
5	norm 10 percentile of 600 sec detrended RR
6	log EDR relative VLF energy (0.003-0.04 Hz)
7	RR Fractal Alan factor for k = 10
8	norm log RR absolute VLF energy (0.003-0.04 Hz)
9	log RR absolute LF energy (0.04-0.15 Hz)
10	RR relative VLF energy (0.003-0.04 Hz)

log → ln(1+x); norm → (x-median)/mad

The result of the overlapping 1-minute segment methodology compared to standard W1 segments is shown in figure 5.4; left the boxplots of the 10 fold classification agreement values on the training set, right the agreement values on the test set. Overlapping of segments increases agreement in both training and test set by 1%, although the difference between both is not significant ($p = 0.07$ on a student t-test, $p = 0.12$ on a Mann-Whitney u-test).

5.4 Discussion

This paper discussed different methodologies to include temporal information in a classification procedure in order to improve overall performance of the classifier. Results on the training and test data show that the feature averaging method led to the highest agreement values, while the window extension methods led to the worst. Agreement values are similar as those reached by submitted entries during the scientific Computing in Cardiology competition, as reported in (Penzel et al. 2002). The test set result of the feature averaging method (with W5) even outperformed the highest results of all automatic classifiers submitted during the competition (90.44% vs 89.4%). While the amount of neighboring segments used was limited to four, it seems worthwhile to investigate even higher amounts, since in (Mendez et al. 2010b) a maximal accuracy was reached with a total window size of 7 segments using a similar feature averaging approach.

The worst performance of the window extension method can possibly be explained by a higher susceptibility to noise. While other methods average out

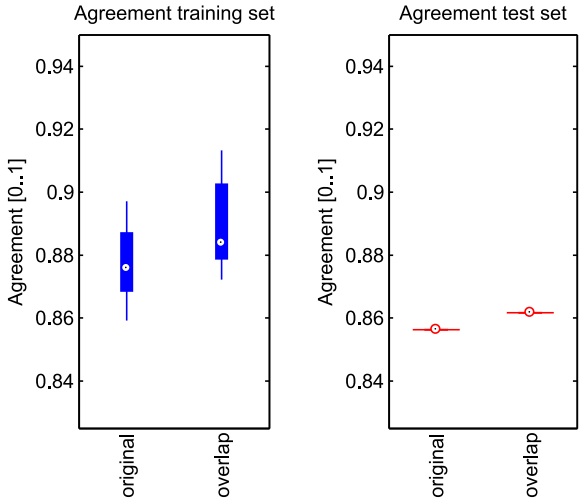


Figure 5.4: Box plots of the 10-fold classification agreement values on the training set (left) and agreement values on the test set (right) for standard W1 segments (original) and overlapping 1-minute segments (overlap).

noisy features or classifier output over possibly noise-free neighboring segments, in the window extension method, features of neighboring noise-free segments will also be affected. It explains why more robust versions of features are preferred when window width increased (up to W5). The window extension method also shows the largest difference between training and test set agreement, indicating higher overfitting and again higher susceptibility to noise present in the test set. The result averaging and window extension method showed drops in agreement when an even amount of segments was used. This can be explained by the possible ties in output class (+1 or -1) when averaging the classifier output over the neighboring segment outputs. Since priority is given to +1 (apnea) in these cases, it leads to an increased amount of false positives. Given priority to -1 (no apnea) would in the same way lead to an increased amount of false negatives. For this reason, the use of an odd amount of segments should be preferred.

While overlapping of segments did improve overall agreement slightly, it suffers from the same problem as described in the paragraph above. The final segment output needs to be calculated by averaging over three classifier output values (one from the segment itself, and two from the left and right overlapping segments). This can again lead to the introduction of false positives when neighboring segments both contain apnea but the middle one does not. However, since apneic events are mostly occurring in a repetitive pattern (Quan et al. 1999), the amount of false positives introduced this way is limited.

Future work will consist of improving the calculation of the RR interval and EDR signals by applying more robust methodologies (e.g. kernel Principal Component Analysis (kPCA) for EDR calculation (Widjaja et al. 2012)), using the LDA-based extracted feature sets in a more versatile LS-SVM classifier (De Brabanter et al. 2010), and increasing the amount of neighboring segments as mentioned earlier.

Acknowledgements

Research supported by Research Council KUL: GOA MaNet, PFV/10/002 (OPTEC), several PhD/postdoc & fellow grants; iMinds: SBO dotatie 2013, ICON: NXT_SLEEP, Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, 2012-2017).

Chapter 6

Probabilistic Cardiac and Respiratory Based Classification of Sleep and Apneic events in Subjects with Sleep Apnea

Tim Willemen[1,2,3], Carolina Varon[2,3], Alexander Caicedo Dorado[2,3], Bart Haex[1,4], Jos Vander Sloten[1], Sabine Van Huffel [2,3]

[1] Biomechanics Section, Department of Mechanical Engineering, KU Leuven, 3001 Leuven, Belgium

[2] STADIUS, Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium

[3] Medical IT, iMinds, 3001 Leuven, Belgium

[4] Maastricht University, 6200 Maastricht, The Netherlands

As adapted from a submission to *Physiological Measurement*, Jan. 2015.

Statement of relevance

This chapter is the second of two chapters addressing research question 2. It outlines the results of a study distinguishing between apneic events and healthy breathing, as well as between Wake, REM, light sleep (N1-N2) and deep sleep (N3). It was performed on the public St. Vincent's University Hospital / University College Dublin Sleep Apnea Database, available at Physionet (Goldberger et al. 2000). The results of this study have been submitted for publication to the journal *Physiological Measurement* (Willemen et al. 2014e). Readers who went through the general introduction of this thesis (chapter two) are able to skip the first three paragraphs of the introduction of this chapter.

Abstract

Sleep depth and continuity are vital for a healthy pattern of sleep. Disorders such as sleep apnea can prevent people from obtaining sufficient amounts of uninterrupted REM and deep sleep. Current clinical standards to assess sleep and its disorders lack either accuracy or user-friendliness. They are therefore difficult to use in cost-effective population-wide screening or long-term objective follow-up after diagnosis. In order to fill this gap, the use of cardiac and respiratory information was evaluated for detection of apneic breathing, and for discrimination between different sleep stages. An alternative probabilistic visual representation of the hypnogram was also presented, referred to as the hypnocorrogram. Analysis was performed on the UCD sleep apnea database, available on Physionet. The presence of apneic events proved to have a significant impact on the performance of a cardiac and respiratory based algorithm for sleep stage classification. WAKE versus SLEEP discrimination resulted in a kappa value of $\kappa = 0.439$, while REM versus NREM resulted in $\kappa = 0.298$ and light sleep (N1N2) versus deep sleep (N3) in $\kappa = 0.339$. While the probabilistic representation of the hypnocorrogram allows to put the classifier output in perspective, further improvements would be necessary to make the classifier reliable for use on patients with sleep apnea.

Keywords

Biomedical signal processing, supervised learning, medical information systems, sleep research, Physionet

6.1 Introduction

Sleep is not an on-off state process, but is distributed over many different stages of sleep, alternating between episodes of Rapid-Eye-Movement (REM) sleep and NREM sleep, occasionally interrupted by episodes of WAKE (Rechtschaffen & Kales 1968). Both REM and NREM sleep are essential towards the recovery of our physical condition (Adam & Oswald 1984) and mental state (Meerlo et al. 2009). NREM sleep is subdivided into three states N1, N2 and N3, depending on the depth of sleep, with N3 (deep sleep) being the most beneficial. A healthy sleep hypnogram should contain a sufficient amount of uninterrupted REM and N3 sleep, with N3 more prominent during the first hours of sleep, and REM more present near the end. For many people, this healthy pattern of sleep is however disrupted due to sleep disorders such as insomnia, sleep apnea, restless legs syndrome, or due to the effect of age on sleep depth and sleep continuity (Berry et al. 2012).

Sleep apnea is an under-diagnosed sleep-related breathing disorder which causes frequent arousal from sleep, preventing patients to obtain sufficient sleep depth and continuity for healthy sleep. It has an estimated prevalence of 14% in men and 5% in women (Peppard et al. 2013). Clinical diagnosis is based on a presence of five or more apneic events per hour (indicated with the AHI index), in combination with excessive day time sleepiness that cannot better be explained by other factors. An apneic event is currently defined as a clear decrease from baseline in breathing volume of at least 10 seconds. Apneic events can be obstructive (complete or partial obstruction of upper airway with surrounding soft tissue) or central (absent or reduced breathing effort). Mixed events are a combination of both, usually starting as a central event, but continuing as an obstructive event once breathing effort is reinitiated. In the case of only a partial obstruction, or when breathing effort is still present but significantly reduced, an apneic event is also referred to as a hypopnea (Berry et al. 2012).

Currently, there are two main standards to assess sleep and its pathologies. One is the current gold standard, polysomnography (PSG), but lacks ease of use towards both patient and physician due to the required attachment of numerous sensors. Annotations of sleep stages and/or apneic events are performed manually, resulting in inter- and intra-observer variability (Whitney et al. 1998). The second standard, actigraphy, is a very unobtrusive and user-friendly method, but lacks accuracy due to its simple analysis based on movement activity (Morgenthaler et al. 2007). Research has however shown a significant influence of sleep and its disorders on the autonomic nervous system, of which sympathetic and parasympathetic variations can be analyzed using cardiac and respiratory changes (Somers et al. 1993, Burgess et al. 1997). The current gap in

clinical standards towards assessment of sleep and its pathologies could thus be filled using cardiac, respiratory and movement based analysis of sleep. This is especially true when these signals can be monitored off-body (without attaching sensors directly to the skin) using the principle of ballistocardiography (Starr et al. 1939). Automated analysis based on less obtrusive and less expensive sensors would reduce costs and observer annotation variability, not only making diagnosis much more accessible and allowing for cost-effective population-wide screening, but also allowing for long-term objective follow-up of patients after a clinical diagnosis has been provided.

Our previous work has already shown promising results in using cardiac, respiratory and movement information to distinguish between different stages of sleep in healthy adults (Willemen et al. 2014a). Furthermore, on a dataset containing a wide range of adults with different levels of apnea severity, we reached an accuracy above 90% in distinguishing between apneic and healthy breathing episodes (Willemen et al. 2014c). Apneic events have however a significant impact on both the respiratory and cardiac system. Cardiac and respiratory based discrimination between different stages of sleep is therefore far from obvious in adults suffering from sleep apnea.

This study will evaluate the influence of apneic events on the ability to use cardiac and respiratory information for discriminating between WAKE, REM, light (N1N2) and deep (N3) NREM sleep. Discrimination between APNEIC and HEALTHY breathing will also be assessed. Moreover, an alternative visual representation of the classifier output is presented, exploiting the benefit of a machine learning approach compared to the gold standard discrete manual annotation of both sleep stages and apneic events.

6.2 Methodology

6.2.1 Database

The database used in this study is available online at Physionet, which offers free web access to large collections of recorded physiologic signals (Goldberger et al. 2000). The database is entitled St. Vincent's University Hospital/University College Dublin Sleep Apnea Database (UCD database) (St. Vincent's University Hospital/University College Dublin 2008), and contains 25 overnight PSG recordings of subjects suspected of sleep-disordered breathing (21 males and 4 females). An overview of the general properties on the subjects and recordings within the database can be found in Table 6.1. PSG recordings were obtained using the Jaeger-Toennies system (Erich Jaeger GmbH, Germany). Expert

Table 6.1: General properties on subjects and recordings within the UCD database.

	Mean	**Range**
Age	49.96 ± 9.55 years	28 - 68 years
BMI	31.60 ± 4.03 kg/m^2	25.1 - 42.5 kg/m^2
AHI	24.24 ± 20.29	2 - 91
Duration of measurement	6.93 ± 0.54 hours	5.9 - 7.7 hours
Sleep efficiency	77.2 ± 11.16 %	58 - 92 %

reference annotations were given on the presence of apneic events according to clinical standards (onset and duration of every apnea/hypopnea event), as well as on sleep stages for every 30-second epoch (window) according to Rechtschaffen and Kales (R&K) rules: WAKE, REM, NREM (Stage 1, Stage 2, Stage 3, Stage 4).

Since the more recent American Academy of Sleep Medicine (AASM) standard for NREM sleep stage annotations makes use of annotations N1 through N3, the original R&K Stage 1 through Stage 4 annotations were mapped to the new N1 through N3 annotations according to AASM guidelines (Stage 1 → N1, Stage 2 → N2, Stage 3+4 → N3). N1 and N2 annotations were combined to represent light sleep, whereas N3 represents deep sleep. The event based apneic annotations were transformed to 30-second epoch based annotations according to the following rule: if at least one second of the 30-second epoch has an overlap with an event based apneic annotation, the epoch is considered as APNEIC, whereas in the case of no overlap it is considered as HEALTHY. An overlap of only one second was chosen to make up for the fact that the event based annotations do not cover the sequential recovery breaths and autonomic influences, but only the periods of reduced breathing volume. An overview of the amount of epochs within each category can be found in Table 6.2. The total amount of epochs in the complete data set equals 20793.

6.2.2 Data preprocessing

Cardiac activity within the PSG recordings was monitored by Electrocardiogram (ECG), while respiratory activity was monitored by both oro-nasal airflow (thermistor) as well as ribcage and abdomen movements (breathing belts with uncalibrated strain gauges). The PSG recordings did not contain any signal directly related to movement information. Although the thermistor signal is usually of much better quality and significance compared to the breathing belt

Table 6.2: Overview of the amount of 30-second epochs within each category (absolute amount and relative percentage). Values given are the mean, standard deviation and range over the 25 recordings.

	Mean amount of epochs		**Range of the amount**	
	#	%	#	%
APNEIC	178.6 ± 134.1	21.9 ± 17.7	15 - 593	1.8 - 83.4
HEALTHY	653.1 ± 159.1	78.1 ± 17.7	118 - 856	16.6 - 98.2
TOTAL	831.7 ± 64.5	100 ± 0.0	711 - 925	100 - 100
SLEEP	642.7 ± 113.0	77.1 ± 11.0	460 - 833	58.4 - 91.8
WAKE	189.0 ± 85.4	22.9 ± 11.0	71 - 357	8.2 - 41.6
REM	120.6 ± 63.4	14.5 ± 7.6	19 - 246	2.3 - 29.4
N1N2	415.5 ± 87.1	50.0 ± 9.8	276 - 592	34.2 - 69.3
N3	106.5 ± 57.3	12.7 ± 6.7	0 - 248	0 - 30.7
NREM	522.0 ± 81.2	62.6 ± 7.1	392 - 664	47.5 - 73.4

signals, respiratory activity was extracted only from the ribcage breathing belt signal. This was done to simulate the idea of measuring respiratory activity in a less obtrusive or off-body way, where respiratory activity can only be monitored via chest or abdomen movements. Similarly for the cardiac activity, only the interbeat RR intervals were used, since those interbeat intervals can be detected in an off-body way using the ballistocardiogram.

The ECG was first stripped from baseline wander by subtracting a median filtered version of the original ECG: first a median filter over 200 ms, followed by a median filter over 600 ms (as proposed in (de Chazal et al. 2004)). Next, the RR intervals were extracted from the resulting ECG using the Pan-Tompkins algorithm (Pan & Tompkins 1985). To identify and correct false positive and false negative R-peak detections, a search-back post-processing algorithm was applied, also proposed in (de Chazal et al. 2004). To detect false positive detections, extracted RR interval values were compared with the median value of themselves and their surrounding four RR intervals (=median $([RR_{i-2}, RR_{i-1}, RR_i, RR_{i+1}, RR_{i+2}])$). If an RR interval becomes closer to this median by increasing it with one of its two neighbouring RR intervals ($RR_i + RR_{i-1}$ or $RR_i + RR_{i+1}$), it will be added to this neighbouring RR interval. For false negative detections, if an RR interval is larger than 1.8 times this median, it will be split in an amount of intervals equal to the amount of median RR intervals that would fit in the RR interval, rounded to the nearest integer (= round $(RR_i/\text{median}([...]))$). The resulting interbeat RR interval series will from now on be labelled as the RR time series.

The 8 Hz ribcage breathing belt signal was also stripped from baseline wander, by subtracting a 20 second median filtered version of the original breathing belt signal (an interval which corresponds to an average of five breathing intervals of four seconds each). The resulting signal will from now on be labelled as the BS signal. This BS signal was subsequently lowpass filtered with a second order forward and reverse Butterworth filter with upper frequency bound at 0.5 Hz. Remaining baseline wander was removed by applying a median filter over 8 seconds. Next, peaks and valleys were detected with the constraint that peaks had to be above the zero baseline, valleys had to be below the zero baseline, and peaks and valleys should follow up each other in turn. Exact starting locations of breathing in and breathing out intervals in the BS signal were extracted by looking for respectively the minimum and maximum within an interval of 2 seconds around the approximated locations of the detected valleys and peaks. From these starting locations of breathing in and breathing out intervals, two time series were extracted: Firstly the interbreath interval series (from now on labelled as BR time series), by extracting the distances between the sequential starting locations of the breathing in intervals; and secondly the inspiration to expiration ratio interval series (from now on labelled as IEratio time series), by extracting the distances between the sequential starting locations of the breathing in and breathing out intervals per breathing interval, and making the ratio between them.

In order to account for inter subject variability within the average heart rhythm, breathing rhythm and inspiration to expiration ratio, detrended versions of the RR, BR and IEratio were defined for the overnight data of every subject. Three different detrending lengths were chosen to account for different night-time variations. Firstly, 6000 seconds (100 minutes) corresponding on average to one NREM-REM sleep cycle, which ensures that changes over a single sleep cycle remain present in the respective time series. Secondly, 600 seconds (10 minutes) reflecting shorter changes between light and deep sleep, or light and REM sleep within each sleep cycle. And lastly, 120 seconds (2 minutes) reflecting short awakenings or arousals within each sleep cycle. Detrending was performed by subtracting a median filtered version of the respective window size from the different time series. To simplify this process, the original un-equidistant time series were interpolated to 4 Hz by cubic spline interpolation before detrending was applied. Afterwards, the detrended spline interpolated equidistant time series were re-interpolated to their original un-equidistant time points. For future frequency analysis of the respective time series by Fast Fourier Transform (requiring equidistant time points), the 120 seconds detrended spline interpolated equidistant time series were also stored.

The resulting time series (RR, BR, IEratio, their detrended and their interpolated versions) and the BS signal from the overnight data of every subject

were segmented in epochs of 60 seconds with 30 seconds overlap. An overlap of 30 seconds was chosen to enable the future classifier annotations on the 60-second epochs to be compared with the reference annotation on the 30-second epochs, which will be discussed in more detail later on. An epoch length of 60 seconds was chosen because it achieved the best results in distinguishing apneic from healthy breathing in a study comparing different epoch lengths, ranging from 15 seconds to 90 seconds (de Chazal et al. 2004). According to the authors, the epoch length of 60 seconds is also quite close to the natural time scale associated with an apneic event, its sequential recovery breaths and its corresponding autonomic influences on the cardiac and respiratory system. Moreover, in our previous work on distinguishing between different sleep stages in healthy adults, we also opted for an epoch length of 60 seconds instead of using the gold standard size of 30 seconds as used in the reference Electroencephalography (EEG) based sleep stage annotations (Willemen et al. 2014a). The predominant frequencies in our EEG brain waves are an order of two greater than the frequencies observed in our heart and breathing rhythm, making the choice for 30-second epochs in cardiac and respiratory based sleep stage scoring less obvious. The Heart Rate Variability (HRV) Task Force also recommends interval lengths of at least 10 times the wavelength of the lowest frequency bound (Malik et al. 1996). This already makes HRV measures characterizing oscillations slower than 0.166 Hz unreliable, while 30-second epochs would even make oscillations slower than 0.333 Hz unreliable (a frequency bound higher than an average healthy breathing rhythm).

6.2.3 Feature extraction

For every 60-second epoch, a multitude of time and frequency domain features were calculated. An overview of the feature type definitions can be found in Table 6.3 and Table 6.4. All of the listed time domain feature types (1-9) were calculated for the original RR time series (18 features). The first three feature types were also calculated on the three detrended RR time series (3 x 8 = 24 features). For the original BR and IEratio time series, feature type 9 was skipped due to the limited amount of points within each window (13 features each). The first 3 feature types were again also calculated on the three BR and IEratio detrended time series (3 x 8 = 24 features each). For the BS signal, feature types 5, 8 and 9 were skipped since they would mostly reflect technological characteristics instead of physiological characteristics (11 features). Frequency domain features (10-16) were calculated on the 120 seconds detrended 4 Hz interpolated RR and BR time series, and on the BS signal (14 features each). Before every Fast Fourier Transform (FFT) calculation, a Hanning window was applied to suppress spectral leakage (referred to as ‘hann’ in the definition of

feature 10). Frequency domain features were deemed irrelevant on the IEratio time series, and therefore skipped. Overall, this leads to a total amount of 184 features: 56 RR features, 51 BR features, 37 IEratio features and 25 BS signal features.

Next, to improve Gaussianity of features, Box Cox transformed features were defined for every feature type that can only be positive (thus for all feature types except 1, 2 and 9), defined as

$$\boldsymbol{d}_{BoxCox} = \begin{cases} \frac{\boldsymbol{d}^{\lambda}-1}{\boldsymbol{\lambda}}, & \lambda \neq 0 \\ \ln(\boldsymbol{d}), & \lambda = 0 \end{cases} \tag{6.1}$$

with a feature-specific optimized λ that maximizes the Log-Likelihood Function given the feature input values $\boldsymbol{d}$. To account for inter subject variability, night-specific normalized features were also defined for every feature (including the Box Cox transformed features),

$$\boldsymbol{d}_{norm} = \frac{\boldsymbol{d} - \text{median}(\boldsymbol{d})}{\text{mad}(\boldsymbol{d})}, \tag{6.2}$$

with median and median absolute deviation (mad) specific for every feature and overnight measurement. These two transformations increase the total amount of features significantly, leading to a total feature set of 510 features. An example of the effect of detrending on the probability density function is displayed in Fig. 6.1. Detrending over 6000 seconds enables to filter out the inter subject variability, as well as the daily circadian variation in average heart rhythm. An example of the effect of the Box Cox transform and normalization on the probability density function is displayed in Fig. 6.2. Both of them reduce the long tail in the distribution of the feature to some extent, where the former specifically targets the distribution characteristics and the latter targets the inter subject variability.

6.2.4 Feature classification

A total of five binary classification problems were assessed, namely APNEIC versus HEALTHY breathing, WAKE versus SLEEP at sleep onset (first hour of the night), WAKE versus SLEEP after sleep onset (all but first hour of the night), REM versus NREM and N1N2 versus N3. The results of the latter four were also combined to evaluate the overall classification problem of WAKE-REM-N1N2-N3. The problem of WAKE versus SLEEP was split up in two

Table 6.3: Overview of time domain feature definitions. The rr_i represent the time series values or signal values within a 60-second epoch. E represents the expected value. F represents the cumulative distribution function (CDF).

	Time domain features	**Feature definition**	#
1	Mean	$\mu_{rr} = \mathrm{E}\left[rr_i\right]$	1
2	Percentiles (perc)	$\mathrm{perc}_{rr} = \mathrm{F}^{-1}\left([0.10, 0.25, 0.50, 0.75, 0.90]\right)$ $\mathrm{F} \rightarrow$ CDF of rr_i	5
3	Inter Quartile Range (IQR) Inter Decile Range (IDR)	$IQR_{rr} = \mathrm{perc}_{rr}\left(0.75\right) - \mathrm{perc}_{rr}\left(0.25\right)$ $IDR_{rr} = \mathrm{perc}_{rr}\left(0.90\right) - \mathrm{perc}_{rr}\left(0.10\right)$	2
4	Standard deviation (STD)	$\sigma_{rr} = \sqrt{\mathrm{E}\left[\left(rr_i - \mu_{rr}\right)^2\right]}$	1
5	Standard deviation of inter-beat differentials (STDi)	$\sigma_{rd} = \sqrt{\mathrm{E}\left[\left(rd_i - \mu_{rd}\right)^2\right]}$ $rd_i = rr_{i+1} - rr_i$ and $\mu_{rd} = \mathrm{E}\left[rd_i\right]$	1
6	Mean absolute deviation (MAD1)	$MAD_1 = \mathrm{E}\left[\left\|rr_i - \mu_{rr}\right\|\right]$	1
7	Median absolute deviation (MAD2)	$MAD_2 = \mathrm{F}^{-1}\left(0.5\right)$ $\mathrm{F} \rightarrow$ CDF of $\left\|rr_i - \mathrm{perc}_{rr}\left(0.50\right)\right\|$	1
8	Root mean square of inter-beat differentials (RMSi)	$RMSi = \sqrt{\mathrm{E}\left[rd_i^2\right]}$	1
9	Serial correlation coefficients (k = 1, ..., 5) (SerCorrCoef)	$r_k = \frac{\sum_{i=1}^{m}(rr_i - \mu_{rr})(rr_{i+k} - \mu rr)}{\sum_{i=1}^{m}(rr_i - \mu_{rr})^2}$	5

Table 6.4: Overview of frequency domain feature definitions. The rr_i represent the time series values or signal values within a 60-second epoch. FFT represents the Fast Fourier Transform, with hann representing the Hanning windowing function.

	Frequency domain features	**Feature definition**	#
10	VLF, LF, HF and TF energy content (EC)	$EC_{rr}(f) = \left\|\text{FFT}\left(\text{hann}\left(rr_i\right)\right)\right\|^2$ $VLF_{rr} = \sum_{f=0.003}^{0.04} EC_{rr}$ and $LF_{rr} = \sum_{f=0.04}^{0.15} EC_{rr}$ $HF_{rr} = \sum_{f=0.15}^{0.40} EC_{rr}$ and $TF_{rr} = \sum_{f=0.003}^{0.40} EC_{rr}$	4
11	LF to HF energy content ratio (ECR)	$LFHR_{rr}^{ratio} = \frac{LF_{rr}}{HF_{rr}}$	1
12	Normalized VLF, LF and HF energy content ratio (ECR)	$VLF_{rr}^{norm} = \frac{VLF_{rr}}{TF_{rr}}$, $LF_{rr}^{norm} = \frac{LF_{rr}}{TF_{rr}}$, $HF_{rr}^{norm} = \frac{HF_{rr}}{TF_{rr}}$	3
13	Mean frequency	$MF_{rr} = \frac{\sum_{f=0.003}^{0.40} f\, EC_{rr}(f)}{TF_{rr}}$	1
14	Peak frequency	$PF_{rr} = f_{max}$ with $EC_{rr}(f_{max}) = \max(EC_{rr})$	1
15	Cumulative energy frequency percentile (CE freq perc)	$percF_{rr} = f_{[0.25,0.50,0.75]}$ with $\frac{ECcum_{rr}\left(f_{[0.25,0.50,0.75]}\right)}{max(ECcum_{rr})} = [0.25, 0.50, 0.75]$ and $ECcum_{rr}(k) = \sum_{f=0.003}^{k} EC_{rr}$	3
16	Cumulative energy frequency Inter Quartile Range (CE freq IQR)	$IQRF_{rr} = percF_{rr}(0.75) - percF_{rr}(0.25)$	1

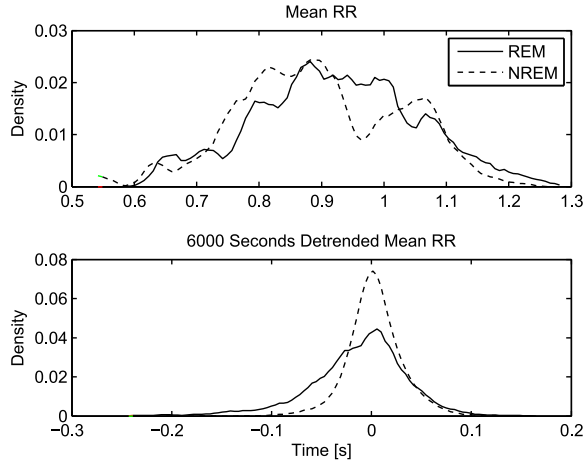


Figure 6.1: Example effect of detrending on the probability density function.

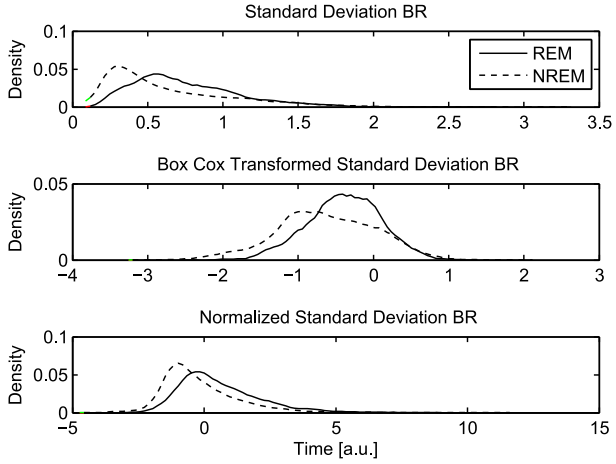


Figure 6.2: Example effect of Box Cox transform and normalization on the probability density function.

parts, since the physiological changes occurring around sleep onset are different from awakenings during the night.

Since the reference annotations are linked to 30-second epochs, every 60-second epoch (on which features were calculated) essentially contains two reference annotations. To determine the annotation of the 60-second epoch (to be used in the classifier training phase), both 30-second epoch reference annotations were

combined, prioritizing in the case of a tie APNEIC over HEALTHY breathing, WAKE over SLEEP, REM over NREM and N1N2 over N3. For example, if a 60-second epoch contains the annotations N2 and REM, it will be referred to as REM.

Classification was performed using Linear Discriminant Analysis (LDA) (Ripley 1996). The approach is outlined below for ease of reference. Variables are indicated in *italic*; vectors and matrices are also in ***bold***. For a given (to be classified) input feature vector $\boldsymbol{x}_i$, representing the features defined on a 60-second epoch, the trained LDA classifier will output two discriminant values y_1 and y_2 (one for each class k) calculated as

$$y_k(\boldsymbol{x}_i) = -\frac{1}{2}\boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_k + \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i + \log(\pi_k), \tag{6.3}$$

with π_k a chosen prior class probability parameter for class k, $\boldsymbol{\mu}_k$ the vector containing the feature means of the training vectors of class k and $\boldsymbol{\Sigma}$ the covariance matrix common for both classes. The latter two are calculated as

$$\boldsymbol{\Sigma} = \frac{1}{N-2}\sum_{k=1}^{2}\sum_{n=1}^{N_k}(\boldsymbol{x}_{nk} - \boldsymbol{\mu}_k)(\boldsymbol{x}_{nk} - \boldsymbol{\mu}_k)^T, \tag{6.4}$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N_k}\boldsymbol{x}_{nk}, \tag{6.5}$$

with $\boldsymbol{x}_{nk}$ the n-th training vector of class k, N_k the total amount of training vectors of class k, and N the total amount of training vectors from both classes combined. Regularization can be applied to the covariance matrix to alter the bias/variance trade-off of the classifier using

$$\boldsymbol{\Sigma} = (1-\alpha)\boldsymbol{\Sigma} + \alpha\boldsymbol{I}, \tag{6.6}$$

with α the covariance matrix regularization parameter, and $\boldsymbol{I}$ the identity matrix. From the two discriminant values y_1 and y_2, posterior class probability estimates can be derived using

$$P(k|\boldsymbol{x}_i) = \frac{\exp(y_k)}{\sum_{l=1}^{2}\exp(y_l)}. \tag{6.7}$$

The class of the input feature vector $\boldsymbol{x}_i$ is then determined as the class with the highest posterior probability estimate.

Table 6.5: Pseudo-code of the preprocessing step for the training feature vectors.

```
for k = 1 to number of classes do
    for i = 1 to number of features do
        perc_i(25) = 25th percentile over all n values x_nk(i)
        perc_i(75) = 75th percentile over all n values x_nk(i)
        IQR_i = perc_i(75) − perc_i(25)
        for n = 1 to number of feature vectors x_nk do
            if (x_nk(i) > perc_i(75) + 1.5 IQR_i) then
                x_nk(i) = perc_i(75) + 1.5 IQR_i
            elseif (x_nk(i) < perc_i(25) − 1.5 IQR_i) then
                x_nk(i) = perc_i(25) − 1.5 IQR_i
            end if
        end for
    end for
end for
for i = 1 to number of features do
    μ_i = combined mean over all n values x_n1(i) and x_n2(i)
    maxAbs_i = combined max over all n values |x_n1(i)| and |x_n2(i)|
    for k = 1 to number of classes do
        for n = 1 to number of feature vectors x_nk do
            x_nk(i) = (x_nk(i) − μ_i) / maxAbs_i
        end for
    end for
end for
```

Before calculating $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}$, the training vectors $\boldsymbol{x}_{nk}$ were first ridded from outliers within their feature values. Outlier values were replaced by their class-specific outlier bounds. Elimination would have led to the removal of too many training vectors (due to the large amount of features), while replacing outlier values by class-specific medians was not deemed ideal due to the sometimes long-tailed distributions of the original features (features without Box Cox transform). Next, training vectors were normalized to the range [-1,1]. The same normalization to the range [-1,1] was also applied on every input feature vector $\boldsymbol{x}_i$ before classification. The pseudo-code for the training vector preprocessing step is displayed in Table 6.5.

Due to the 30 second overlap of the 60-second epochs, every 30-second epoch linked to the reference annotations is essentially scored twice by the 60-second epoch classifier, as visualized in Fig. 6.3. In order to take both classifications into account, the posterior class probability estimates from both of the 60-second epochs were averaged, resulting in 30-second epoch posterior class probability

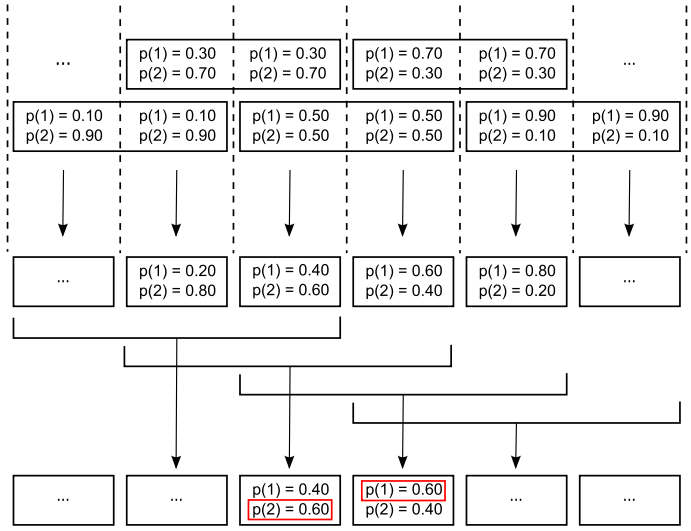


Figure 6.3: Visualization of the process to determine the final class output on 30-second epochs based on the classifier output on 60-second epochs.

estimates. It is known however that most periods of apneic breathing as well as most sleep stages have a time span of several epochs. Therefore, to filter out rapid transitions between APNEIC and HEALTHY breathing, or between different sleep stages, the 30-second epoch posterior class probability estimates were averaged over those from their two neighbouring epochs. The final class of the 30-second epoch classifier annotation was then determined by the highest average probability estimate.

To prevent overfitting of the data, a triple layer validation scheme was used, as visualized in Fig. 6.4. In the first layer, the 25 nights of the dataset were split up via stratified sampling into on average 2/3 training nights and 1/3 test nights. Stratified sampling makes sure that a certain aspect within the data is equally represented. The goal of this study is to assess the influence of apneic events on the ability to use cardiac and respiratory information for sleep staging. Therefore, a more or less equal ratio of APNEIC to HEALTHY epochs in each training and test set was taken as the stratified sampling constraint. This process was repeated ten times, in order to provide robust measures of classifier performance, which are less affected by choice of training and test nights. For all of the five binary classification problems, the same ten distributions of training and test nights were used.

In the second layer, each set of training nights from the first layer was similarly

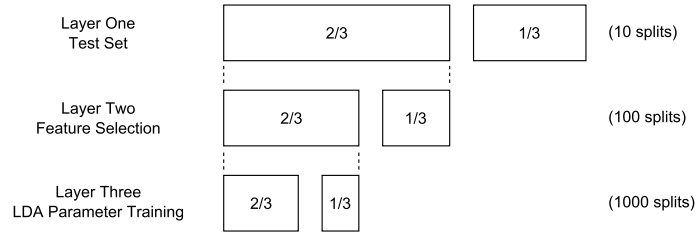


Figure 6.4: Visualization of the triple layer validation scheme.

split up ten times via stratified sampling into on average 2/3 training nights and 1/3 validation nights, resulting in a total of 100 splits. This second layer allowed for a robust selection of ten feature subsets, one for each of the ten folds of the first layer. A feature subset was constructed based on a maximum value of Cohen's kappa (Landis & Koch 1977) on the validation nights, averaged over the ten folds of the second layer. Feature selection will be further discussed in section 6.2.5. Again, the same distributions of training and validation nights were used for each of the five binary classification problems.

A similar third layer (resulting in a total of 1000 splits) allowed for a robust gridsearch optimization of the LDA classifier parameters (α, π_1, $\pi_2 \in [0,1]$) to be used in the respective fold of the second layer. Optimal parameters were selected based on a maximum value of kappa on the validation nights, averaged over the ten folds of the third layer. The values of the classifier parameters to be used in the respective fold of the first layer were derived from the selected parameters used within the second layer, more specifically by taking the median of their values.

Classifier performance was also assessed using a leave-one-out validation scheme (splitting up the dataset 25 times in 24 training nights, and 1 test night), resulting in a night-by-night evaluation of classifier performance. Posterior class probability estimates were averaged over the outputs of the ten constructed classifiers from the triple layer validation scheme (each with their own optimized feature subsets and classifier parameters).

6.2.5 Feature selection

In order to reduce complexity and noise on the classification result, unique subsets of features were selected as being most descriptive for their respective binary classification problems. To reduce computational cost, the feature selection process was performed in two phases. In phase one, classifier performance was assessed (on layer two) for each of the 510 features separately.

From the feature sets of each of the four time series or signals (RR, BR, IEratio and BS), only the top 25 features achieving the highest kappa value were kept, reducing the feature set to 100 features. In phase two, a sequential forward floating search (SFFS) algorithm (Pudil et al. 1994) extracted a subset of maximum ten features from this reduced set based on a maximal increase in kappa (on layer two) after every step. The SFFS algorithm alternates between forward selection and unselection of one of the previously selected features, as long it leads to an increase in kappa. If at a certain point neither the forward selection nor the unselection of a previously selected feature leads to an improvement in kappa, the algorithm stops and outputs the final feature subset. The algorithm also stops if the maximum of ten features is reached.

Kappa is defined as

$$\text{kappa} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \tag{6.8}$$

which corrects agreement,

$$\Pr(a) = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6.9}$$

for agreement occurring by chance,

$$\Pr(e) = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2}, \tag{6.10}$$

where a True Positive (TP) was defined as a correctly identified class one instance (i.e. APNEIC, WAKE, REM, N3), a True Negative (TN) as a correctly identified class two instance (i.e. HEALTHY, SLEEP, NREM, N1N2), a False Positive (FP) as a class two instance incorrectly identified as a class one instance, and a False Negative (FN) as a class one instance incorrectly identified as a class two instance.

During each forward selection step of the SFFS algorithm, in order to prevent selection of too similar features (and to prevent problems due to multicollinearity), selectable features from the remaining feature set were assessed based on an inter feature mutual information (MI) criterion. The inter feature mutual information measures the information shared between two features, and is defined as

$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{y\in\boldsymbol{Y}}\sum_{x\in\boldsymbol{X}} \boldsymbol{p}(x,y)\log\left(\frac{\boldsymbol{p}(x,y)}{\boldsymbol{p}(x)\,\boldsymbol{p}(y)}\right), \tag{6.11}$$

with x and y representing the elements of the feature vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, and $\boldsymbol{p}(x,y)$, $\boldsymbol{p}(x)$ and $\boldsymbol{p}(y)$ their probability density distributions. The criterion states: 'Whenever the inter feature mutual information of a selectable feature with any feature from the already selected feature set was higher than an heuristically determined threshold of 1.246, the selectable feature was disregarded from the remaining feature set.' The threshold of 1.246 was determined by calculating the inter feature mutual information for all possible combinations of the 510 features on the complete dataset (a total of 510 x 510 = 260100 values). The resulting values were sorted and the elbow criterion applied. This criterion selects the point in a curve of sorted values where a sharp transition in tangent is observed. The elbow point coincided with the 85 percentile, of which the corresponding mutual information value of 1.246 was thus taken. This means that every feature will have on average 15% features that are deemed too similar.

6.2.6 Visual representation

The gold standard technique of sleep stage and apneic event annotation is done manually, leading to a discrete set of decisions on the current sleep stage, and the presence or absence of apneic breathing. Several limitations to this gold standard approach are discussed by Himanen et al. (Himanen & Hasan 2000). It is well known that a significant disagreement ratio exists when PSG's are annotated by different sleep experts (Whitney et al. 1998). This can be due to ambiguous data, ambiguous interpretation of the AASM standard, or just due to simple human error. Opportunities of computer based sleep recording and analysis are extensively discussed by Penzel et al. (Penzel & Conradt 2000b) and Schulz (Schulz 2008). Advantages of a machine learning approach are not only the consistency in decisions, but also the possibility to extract class probability estimates. A machine learning approach can thus e.g. give an indication of how certain it is about its decision, based on the information it has gathered on other subjects and nights during its training phase.

An alternative, non-discrete visual representation of the classifier output is therefore proposed, referred to as the hypnocorrogram (for the distribution of sleep stages during the night) or the apneacorrogram (for the presence of apneic breathing). By extracting the posterior class probability estimates for each of the four sleep stage classification problems, the outputs can be combined using a probabilistic decision tree, as visualized in Fig. 6.5. In the example

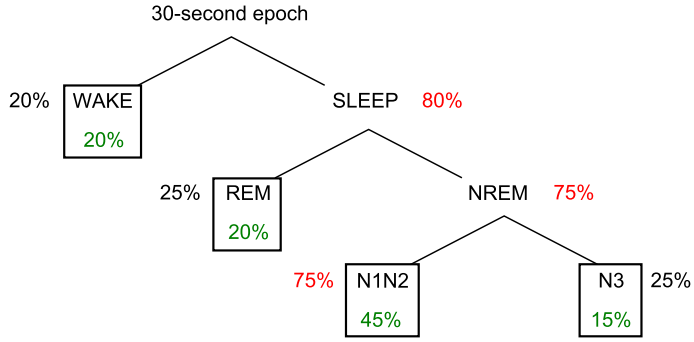


Figure 6.5: Combining binary classifier posterior probability estimates to attain multiclass posterior probability estimates.

given, a 30-second epoch has a binary posterior class probability of 20% with respect to WAKE and 80% with respect to SLEEP. Furthermore, it has a binary posterior class probability of 25% with respect to REM and 75% with respect to NREM. The latter values allow to split the 80% binary probability of SLEEP into multiclass probabilities of 20% with respect to REM and 60% with respect to NREM. The multiclass probabilities of 45% with respect to N1N2 and 15% with respect to N3 can be derived in a similar way.

This thus leads to four multiclass posterior probability estimates, which can be plotted simultaneously on a nightly basis as in Fig. 6.6. The different coloured areas represent the magnitude of the probability at every 30-second epoch during the night. The reference annotation (with only 0% or 100% posterior probability values) is displayed below for comparison. To improve visualization of the REM posterior probability estimate, the WAKE estimate can be removed from the hypnocorrogram, with the sum of the other three estimates scaled to one. As visualized in Fig. 6.7, it now allows to compare both the N3 as well as the REM posterior probability estimates to a horizontal baseline.

The WAKE vs SLEEP probabilistic representation can then be visualized separately by scaling the posterior probability estimate of WAKE ($\in$ [0,1]) on the range [-1,1], where the zero-line now represents the 50% posterior probability. The coloured area above the zero-line represents the magnitude of the probability above 50% (i.e. the classifier leaning more towards WAKE), while the coloured area below the zero-line represents the magnitude below 50% (i.e. the classifier leaning more towards SLEEP). Fig. 6.8 visualizes this binary hypnocorrogram, with the reference annotation similarly displayed below for comparison.

Fig. 6.9 gives a probabilistic representation on the presence or absence of apneic breathing, referred to as the apneacorrogram. The coloured area above the

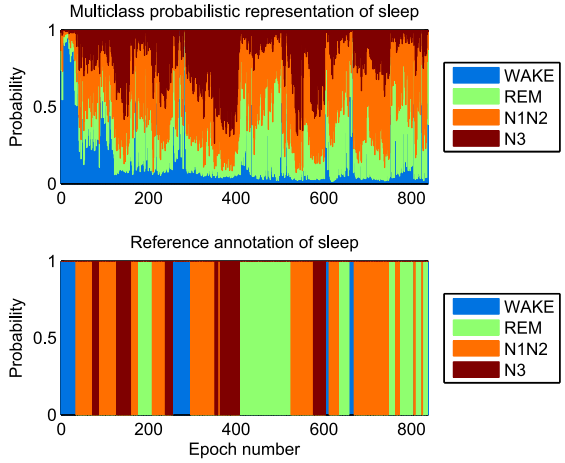


Figure 6.6: An example hypnocorrogram, visualizing the multiclass probabilistic representation of the different sleep stages.

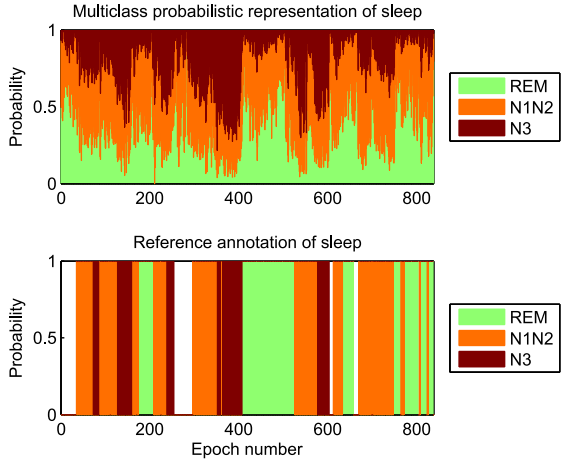


Figure 6.7: An example hypnocorrogram, visualizing the multiclass probabilistic representation of the different sleep stages, excluding WAKE.

zero-line represents again the magnitude of the probability above 50% (i.e. the classifier leaning more towards APNEIC breathing), while the coloured area below the zero-line represents the magnitude below 50% (i.e. the classifier leaning more towards HEALTHY breathing). The reference annotation is displayed below for comparison.

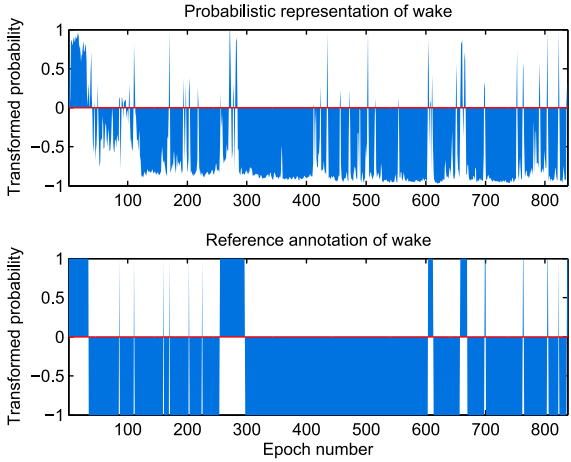


Figure 6.8: An example binary hypnocorrogram, visualizing the probabilistic representation of WAKE.

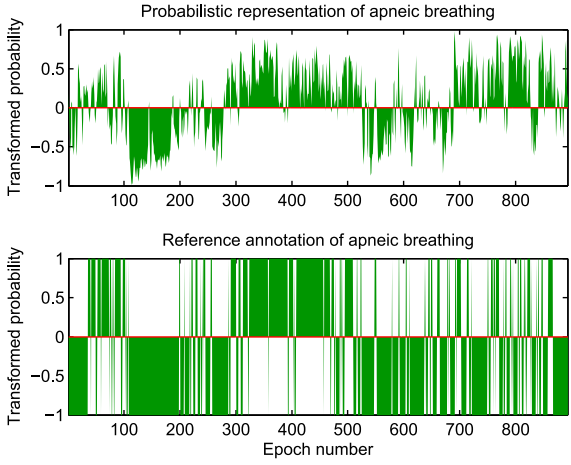


Figure 6.9: An example apneacorrogram, visualizing the probabilistic representation of apneic breathing.

The content of the figures will be further discussed in section 6.3.3.

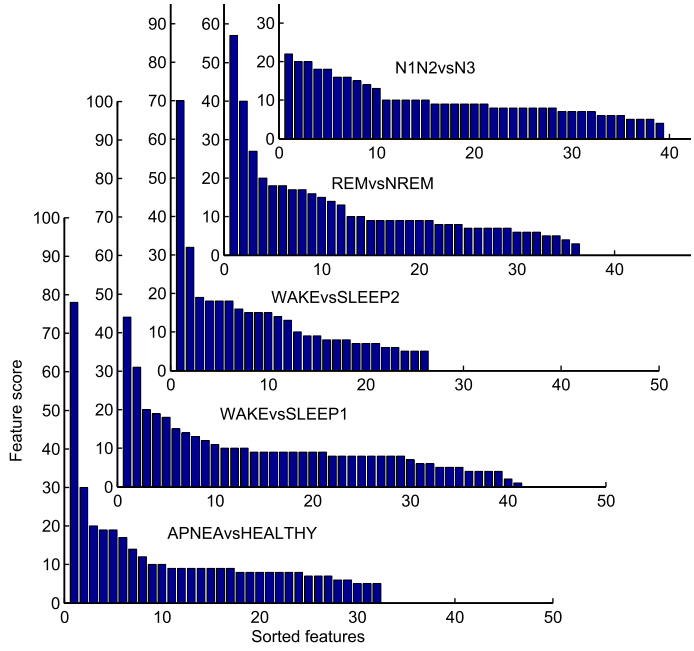


Figure 6.10: Sorted feature scores for all five binary classification problems.

6.3 Results

6.3.1 Feature selection

In order to determine the relevance of selected features present in one or more of the ten different feature subsets extracted by the triple layer validation scheme, selected features were given a feature score (separately for every binary classification problem). Whenever a feature was present in of the feature subsets, its feature score was increased by a number between 1 and 10 based on how early it was detected by the SFFS algorithm (10 for the first feature, 9 for the second feature, etc.). Since ten different feature subsets exist, the maximum feature score attainable was 100. Fig. 6.10 shows the sorted feature scores for all five binary classification problems.

Due to the large feature starting set of 510 features, a lot of different features were selected at least once. Most features were derived from the BS signal (33%), followed by the RR series (32%), IEratio series (20%) and lastly the BR series (15%). More than half of all selected features (56%) were normalized versions of features, proving the added value of this transformation. On the

Table 6.6: Overview of selected features with a feature score of at least 20. Full feature definitions can be found in Table 6.3.

Feature score	Feature description
	APNEIC vs HEALTHY
78	Box Cox VLF EC of BR
30	Box Cox MAD2 of BR
20	Box Cox MAD1 of BR
	WAKE vs SLEEP 1 (first hour)
44	Normalized LF EC norm of BS
31	LF EC norm of RR
20	Normalized Box Cox ECR of BS
	WAKE vs SLEEP 2 (after first hour)
70	HF EC norm of BS
32	Normalized ECR of BS
	REM vs NREM
57	Normalized 25th perc of 6000s detrended RR
40	Normalized VLF EC norm of RR
27	Normalized Box Cox HF EC norm of RR
20	Normalized mean of 6000s detrended BR
	N1N2 vs N3
22	Normalized Box Cox VLF EC of BR
20	Box Cox STD of BR
20	Box Cox ECR of BS

features where a Box Cox transform could have been defined, it was present in 44%. Since LDA makes the assumption that the probability density functions of features are normally distributed, applying the Box Cox transform should always be considered. Table 6.6 lists the few selected features with a feature score of at least 20.

6.3.2 Classifier performance

The classifier performance for each of the classification problems is listed in Table 6.7 for both the triple layer validation scheme as the leave-one-out validation scheme. Performance metrics given are the mean, standard deviation and range (over the different folds) of kappa, agreement, precision, recall (also known as sensitivity) and specificity. The latter three are defined as

$$\text{precision} = \frac{TP}{TP + FP}, \tag{6.12}$$

$$\text{recall} = \frac{TP}{TP + FN}, \tag{6.13}$$

$$\text{specificity} = \frac{TN}{TN + FP}. \tag{6.14}$$

Recall quantifies the amount of class one instances identified within the complete set of instances (usually the smaller class), while precision quantifies how precise the classifier was in this identification. Specificity quantifies the amount of class two instances identified within the complete set of instances.

Agreement values are around 75% to 80% for all binary classification problems, indicating reasonable performance at first sight. Sensitivity is however clearly lacking, especially for the REM vs NREM classification problem (<40%). Precision values are only around 40% to 45% for all binary classification problems. The low sensitivity and precision are clearly reflected in the low values of kappa, with values between 0.20 and 0.40 only considered as fair agreement, and values above 0.40 as moderate agreement (Landis & Koch 1977).

Although most performance metric values are similar between the two validation schemes, larger standard deviations are present in the leave-one-out scheme. This can be accounted to the increased amount of possible data distributions in its test sets, like class asymmetry amongst others. In the leave-one-out scheme, test sets existed only out of a single night. In the triple layer scheme however, test sets were always built up from on average 1/3 of the total amount of nights, thus averaging out differences in data distributions between different nights. Even then, the range of values in the triple layer scheme indicates the considerable effect the choice of training and test nights can have on measures of classifier performance.

To assess the combined sleep stage classification problem, Table 6.8 shows the complete confusion matrix, built up by adding the confusion matrices from the individual folds of the leave-one-out validation scheme. The total sum of elements adds up to the total amount of epochs present in the dataset (=20793). Table 6.9 also shows the kappa and agreement metrics for the WAKE-REM-NREM and WAKE-REM-N1N2-N3 classification problem. As seen from the confusion matrix, the classifier has a lot of difficulties in distinguishing Wake, REM and N3 from the most common state N1N2, leading to mediocre overall performance values.

Table 6.7: Classifier performance using the triple layer and leave-one-out validation scheme. Kappa has been scaled with a factor 100. Other results are given in percentages. The range of values is displayed between brackets.

	Kappa	**Agreement**	**Precision**	**Recall/ Sensitivity**	**Specificity**
Triple layer scheme					
APNEIC vs HEALTHY	30.24 ± 5.15 [24.70 .. 40.46]	73.58 ± 5.20 [63.47 .. 79.52]	42.93 ± 4.50 [36.48 .. 48.33]	59.52 ± 15.19 [34.66 .. 75.87]	76.97 ± 9.82 [60.46 .. 90.03]
WAKE vs SLEEP	37.07 ± 11.21 [13.66 .. 47.31]	74.22 ± 6.23 [60.89 .. 79.30]	46.07 ± 8.38 [31.04 .. 55.98]	71.66 ± 6.97 [62.18 .. 88.00]	75.24 ± 8.75 [58.20 .. 82.86]
REM vs NREM	23.50 ± 5.93 [16.46 .. 34.58]	77.28 ± 3.04 [70.86 .. 81.75]	38.37 ± 7.54 [27.95 .. 51.02]	38.52 ± 9.33 [27.34 .. 55.59]	85.89 ± 4.99 [75.42 .. 93.28]
N3 vs N1N2	29.86 ± 5.99 [18.89 .. 39.78]	73.27 ± 2.94 [68.00 .. 77.29]	39.14 ± 6.45 [26.69 .. 52.24]	62.92 ± 5.96 [53.96 .. 75.15]	75.72 ± 4.32 [68.10 .. 80.55]
Leave-one-out scheme					
APNEIC vs HEALTHY	27.21 ± 17.92 [-11.48 .. 55.90]	75.90 ± 8.75 [48.97 .. 90.63]	41.13 ± 22.75 [2.13 .. 88.03]	62.30 ± 21.18 [3.70 .. 100.00]	76.57 ± 14.62 [38.14 .. 93.79]
WAKE vs SLEEP	43.91 ± 19.30 [-7.55 .. 74.93]	77.95 ± 9.36 [46.41 .. 91.04]	51.98 ± 20.20 [24.01 .. 88.89]	79.46 ± 10.62 [54.55 .. 95.65]	78.68 ± 12.65 [33.94 .. 94.30]
REM vs NREM	29.80 ± 22.99 [-8.69 .. 70.66]	81.52 ± 7.46 [62.01 .. 93.48]	45.62 ± 23.20 [0.00 .. 85.53]	39.38 ± 20.45 [0.00 .. 78.65]	90.55 ± 4.25 [75.95 .. 96.88]
N3 vs N1N2	33.90 ± 23.15 [-27.10 .. 66.59]	76.92 ± 12.60 [38.34 .. 96.22]	44.01 ± 19.67 [0.00 .. 77.87]	69.99 ± 20.84 [1.92 .. 100.00]	78.58 ± 14.75 [32.01 .. 96.75]

Table 6.8: Confusion matrix of the combined sleep stage classification problem, more specifically the sum of the confusion matrices from each of the 25 nights/folds of the leave-one-out scheme. The columns represent the predicted annotations by the classifier (indicated as P). The rows represent the actual reference annotations (indicated as A).

A ↓ \ P →	**Wake**	**REM**	**N1N2**	**N3**
Wake	2870	276	1239	341
REM	381	956	1401	278
N1N2	1372	710	6173	2133
N3	83	41	694	1845

Table 6.9: Classifier performance of the combined sleep stage classification problem, for both the triple layer as the leave-one-out validation scheme. Kappa has been scaled with a factor 100. Agreement is given in percentages. The range is displayed between brackets.

	Kappa	**Agreement**
Triple layer scheme		
WAKE-REM-NREM	33.72 ± 5.52 [19.36 .. 38.00]	65.46 ± 3.71 [55.35 .. 68.05]
WAKE-REM-N1N2-N3	26.95 ± 5.41 [12.94 .. 32.88]	50.10 ± 4.17 [37.79 .. 53.57]
Leave-one-out scheme		
WAKE-REM-NREM	40.67 ± 12.17 [4.68 .. 55.85]	70.36 ± 6.92 [50.28 .. 79.21]
WAKE-REM-N1N2-N3	33.68 ± 11.55 [2.01 .. 50.15]	56.83 ± 8.22 [33.21 .. 71.32]

6.3.3 Visual representation

Fig. 6.6 and Fig. 6.7 show the hypnocorrogram of the combined sleep stage classification problem for the night of a subject with an AHI index of 14. In order to visualize the proposed representation in a sufficiently clear way, it is an example with a kappa performance value of the WAKE-REM-N1N2-N3 problem on the upper end of the spectrum (44.28 to be exact). Even then, it is quite clear from the figures that the classifier is in most cases far from certain about its decisions, which is exactly the point of this representation. These classifiers

were built to assess how unseen data corresponds to training data. They should therefore exactly output how likely this unseen data corresponds to each of the classes within the training data, without pretending to know which class it would fit in the end. The same is true for Fig. 6.8 and Fig. 6.9, representing the binary hypnocorrogram of WAKE vs SLEEP and the apneacorrogram of APNEIC vs HEALTHY breathing. The information present in the uncertainty of the classifier is valuable and worth visualizing.

6.4 Discussion

This study evaluated the use of cardiac and respiratory information for discriminating between the different sleep stages WAKE, REM, light (N1N2) and deep (N3) NREM sleep, as well as discriminating between APNEIC and HEALTHY breathing. Moreover, it presented an alternative visual representation of the classifier output, the hypnocorrogram and apneacorrogram, incorporating the uncertainty of the classifier. The authors believe that visualizing the correspondence of unseen data to training data allows to put the classifier output into perspective, and will increase the likelihood of bringing these kinds of classifiers into clinical practice.

Table 6.10 compares the obtained classification results to similar classification algorithms in literature. The last row of the table lists the presented results of this study. The top two rows present our previous results on cardiac and respiratory based classification of sleep stages in healthy subjects (Willemen et al. 2014a), and on cardiac and respiratory based classification of apneic breathing in subjects with mild, moderate and high levels of sleep apnea syndrome (Willemen et al. 2014c). During our literature research, the study of Redmond et al. (Redmond & Heneghan 2006) was the only one found on cardiac and respiratory based classification of sleep stages in apnea patients, and was therefore added to Table 6.10. Their used database was however not publicly available. A total of five studies were found to be using the same public UCD database as used in this study. Two of them however used the UCD database for evaluation of pulse oximetry based or EEG based classification, which was not part of this work. Two others studied the presence of apneic breathing using cardiac and respiratory information, but added the UCD database to a larger non-public database, making direct comparison impossible. The fifth study, from Xie et al. (Xie & Minn 2012), did use only the UCD database for their cardiac and respiratory based classification of apneic breathing, and was therefore also included in Table 6.10.

As shown in Table 6.10, our results on the APNEIC vs HEALTHY classification

Table 6.10: Overview of similar classification algorithms in literature. EDR = ECG derived respiration; RIP = respiratory inductance plethysmography (breathing belts); MOV = movement. Agr = agreement. Nights: h = healthy; a = apneic.

Author	**Signals**	**Classification**	**Agr**	**Kappa**	**Nights**	**Epoch**
(Willemen et al. 2014a)	RR, RIP, MOV	WAKE-REM-NREM WAKE-REM-N1N2-N3	81%[a] 69%[a]	0.62[a] 0.56[a]	85 h	60s
(Willemen et al. 2014c)	RR, EDR	APNEIC-HEALTHY	90%[a]	0.80[a]	70 a	60s
(Redmond & Heneghan 2006)	RR, EDR, RIP	WAKE-REM-NREM	67%[b]	0.32[b]	37 a	30s
(Xie & Minn 2012)	RR, EDR	APNEIC-HEALTHY	78%[c]	-	25 a	60s
This study	RR, RIP	APNEIC-HEALTHY WAKE-REM-NREM WAKE-REM-N1N2-N3	74%[d]/76%[b] 65%[d]/70%[b] 50%[d]/57%[b]	0.30[d]/0.27[b] 0.34[d]/0.41[b] 0.27[d]/0.34[b]	25 a	30s

[a] Result on separate test set of 35 nights
[b] Averaged result of leave-one-out cross-validation
[c] Averaged result of ten-fold cross-validation (possible night-specific training)
[d] Averaged result of triple layer validation scheme

problem are significantly worse in this study, compared to our previous study (Willemen et al. 2014c) which used the Physionet Apnea-ECG database (Penzel et al. 2000a). There were however two important differences in methodology between both studies, which are discussed below. Firstly, in our previous study (Willemen et al. 2014c), averaging of features over 300-second windows was performed to achieve the results listed in the table. Without this averaging, results dropped to an agreement of 86%. Feature averaging was skipped in the current study, since it contradicts with our visual representation methodology which shows how unseen data corresponds to training data at every epoch. Averaging over large interval widths only artificially enhances results without actually enhancing information.

Secondly, our previous study made use of 60-second reference intervals, while in the current study 30-second reference intervals were used. This was done to comply more easily with the 30-second expert reference annotations for the sleep stages. Features were still calculated however on 60-second intervals (with 30-second shifts), since this interval length was reported to deliver the best results for APNEIC vs HEALTHY breathing classification (de Chazal et al. 2004). Changing back to 60-second reference intervals led to a small increase in values of kappa (from 0.30 / 0.27 to 0.32 / 0.27), but an unexpected decrease in agreement (from 74% / 76% to 72% / 73%). The increase in kappa can be explained by the decrease in class asymmetry. With 60-second reference intervals, a higher total amount of the night will be annotated as APNEIC (which was the smaller class), thus reducing the agreement occurring by chance which kappa heavily corrects for. The decrease in agreement was however due to this same increase in APNEIC annotation range. Complete 60-second intervals containing only a small fraction of an apnea event are also annotated as APNEIC, which led to a small increase in false negative class estimations.

Nevertheless, the feature set and classifier algorithm of Bsoul et al. (Bsoul et al. 2011), which reached an agreement of 86% on the Apnea-ECG database, only achieved an agreement of 71% on the UCD database. The latter result was reported by Xie et al. (Xie & Minn 2012), who used exactly the same feature set of Bsoul et al. (Bsoul et al. 2011), but evaluated it on ten different classifier algorithms. The large differences achieved between the two databases are therefore not fully explainable by methodological differences. On a side note, reported agreements by Xie et al. (Xie & Minn 2012) ranged between 68% and 78%, showing a significant influence of the classifier algorithm on the overall agreement. Although it is known that supervised learning algorithms can have significant variability in their performance across problems (Caruana & Niculescu-Mizil 2006), the large differences in agreement found by Xie et al. (Xie & Minn 2012) could also be explained by their reported use of default parameter settings for all but one classifier.

The difference between the Apnea-ECG and UCD database might further be explained by differences in apnea severity of the subjects contained within the databases. The Apnea-ECG database contains a total of 13002 apnea events and 1337 hypopnea events. The UCD database however contains only 695 apnea events, but a total of 2623 hypopnea events. This means that the ratio of hypopnea to apnea events in the UCD database is 36.7 times larger than the same ratio in the Apnea-ECG database. Since hypopnea events are the less-severe variant of apnea events, they have less of an impact on the cardiac and respiratory system, and are thus more difficult to detect.

Another large difference between the two databases are the event based apneic annotations. While the latter normally only contain periods of reduced breathing volume (as in the UCD database), the annotations in the Apnea-ECG database also contain the associated recovery breaths (de Chazal et al. 2004). The use of 60-second epochs with 30-second overlap made sure that features from the recovery breaths were also linked to the apneic reference annotation epochs in the UCD database. However, this information will also have been linked to healthy reference annotation epochs, obscuring the decision boundary between apneic and healthy. A solution might thus be to artificially increase the length of each of the event based apneic annotations of the UCD database (for example by 10 seconds) as to include at least the first recovery breaths into the annotation.

The results of our combined WAKE-REM-NREM sleep stage classification are similar to those reported in the study by Redmond et al. (Redmond & Heneghan 2006). They are however significantly worse when compared with our previous study on sleep stage classification in healthy subjects (Willemen et al. 2014a). This can again be due to several reasons. Firstly, in our previous study (Willemen et al. 2014a), 60-second reference intervals were used, transformed from the 30-second sleep expert annotations. Transforming the reference intervals of the current study in the same way led to an unexpected decrease in accuracy on the WAKE-REM-NREM problem in both values of kappa (0.34 / 0.41 to 0.33 / 0.40), as well as agreement (65% / 70% to 64% /69 %). The decrease in agreement was again due to an increased amount of false negative class estimations. The effect on class asymmetry was however smaller, preventing an increase in values of kappa (as was the case in the APNEIC vs HEALTHY problem).

Secondly, our previous study also included features related to movement, as measured by indentation sensors inside of the bed system (Verhaert et al. 2011a). These movement features proved to be very selective, not only in classification of wake epochs, but also in distinguishing light sleep (N1N2) from deep sleep (N3) (longterm absence of movement activity during deep sleep), as well as REM sleep from NREM sleep (short twitches present in REM sleep). A signal related to movement was however not present in the UCD database.

Thirdly, the presence of apneic events in both REM as NREM sleep will have definitely obscured the decision boundary between both, since apneic events have a similar impact on the cardiac and respiratory system independently of the sleep stage. The frequent arousals from sleep due to apneic events will have also obscured the decision boundary between WAKE and SLEEP, since the effect of an arousal on the heart rhythm is quite similar to the effect of an awakening. Redmond et al. (Redmond & Heneghan 2006) found no significant difference in classification accuracy between patients with low AHI (< 10) and high AHI (> 10) index. In our results however, kappa of the WAKE-REM-NREM classification problem was significantly correlated with the AHI index ($r = -0.682$; $p < 0.001$). An increased amount of apneic events during the night thus has a negative impact on the obtainable accuracy of the sleep stage classification problem. The question rises however whether a reliable distribution of sleep stages is necessary when a significant amount of apneic events occur. Diagnosis of sleep apnea is in that case of most concern, since its negative impact on sleep continuity and depth is anyhow implied.

6.5 Conclusions and future work

This study evaluated the use of cardiac and respiratory information for detection of apneic breathing, and for discrimination between sleep stages. Analysis was performed on the UCD sleep apnea database, available on Physionet. Detection of apneic breathing proved to be more difficult when compared with the Apnea-ECG database (also available on Physionet), related to a higher proportion of hypopneas and differences in expert apneic annotations. Future work could investigate whether improved agreement can be achieved when recovery breaths would be automatically included to the expert apneic annotations. Further on, the presence of apneic events proved to have a significant impact on the performance of a cardiac and respiratory based algorithm for sleep stage discrimination. While the probabilistic representation of the hypno- and apneacorrogram allows to put the classifier output in perspective, further improvements would be necessary to make the classifier reliable for use on patients with sleep apnea. Future work could investigate whether including movement can significantly improve results.

Acknowledgements

Research supported by Research Council KUL: GOA MaNet, PFV/10/002 (OPTEC), several PhD/postdoc & fellow grants; iMinds: SBO dotatie 2013,

ICON: NXT_Sleep, Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, 2012-2017), FWO postdoctoral fellows.

Chapter 7

Characterization of the respiratory and heart beat signal from an air pressure-based ballistocardiographic setup

Tim Willemen[1,2,3], Dorien Van Deun[1], Vincent Verhaert[1], Sabine Van Huffel [2,3], Bart Haex[1], Jos Vander Sloten[1]

[1] Biomechanics Section, Department of Mechanical Engineering, KU Leuven, 3001 Leuven, Belgium

[2] STADIUS, Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium

[3] Medical IT, iMinds, 3001 Leuven, Belgium

As adapted from *Proceedings of the 36th Annual International Conference of the IEEE EMBS*, Chicago, pp. 6298-6301, Aug. 2014.

Statement of relevance

This chapter is the first of two chapters addressing research question 3. It outlines the results of a study building and characterizing a ballistocardiographic setup, implemented inside of a bed. The results of this study have been presented with an oral presentation in the *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society - 2014*, and published online as a conference proceeding paper in the *IEEE Xplore Digital Library* (Willemen et al. 2014b). Readers who went through the general introduction of this thesis (chapter two) are able to skip the first paragraph of the introduction of this chapter. Readers with limited time can skip this chapter entirely and narrow down their reading to chapter eight, which holds most relevant information related to research question 3.

Abstract

Off-body detection of respiratory and cardiac activity presents an enormous opportunity for general health, stress and sleep quality monitoring. The presented setup detects the mechanical activity of both heart and lungs by measuring pressure difference fluctuations between two air volumes underneath the chest area of the subject. The registered signals were characterized over four different sleep postures, three different base air pressures within the air volumes and three different mattress top layer materials. Highest signal strength was detected in prone posture for both the respiratory and heart beat signal. Respiratory signal strength was the lowest in supine posture, while heart beat signal strength was lowest for right lateral. Heart beat cycle variability was highest in prone and lowest in supine posture. Increasing the base air pressure caused a reduction in signal amplitude for both the respiratory and the heart beat signal. A visco-elastic poly-urethane foam top layer had significantly higher respiration amplitude compared to high resilient poly-urethane foam and latex foam. For the heart beat signal, differences between the top layers were small. The authors conclude that, while the influence of the mattress top layer material is small, the base air pressure can be tuned for optimal mechanical transmission from heart and lungs towards the registration setup.

Keywords

Ballistocardiography, biomedical signal processing, health monitoring, smart bedding systems

7.1 Introduction

Since we spend almost a third of our lives in bed, it presents an enormous opportunity for general health monitoring. During the past decade, numerous methods and techniques have been developed to record vital body parameters such as respiratory and cardiac activity from within or around our bed. Some methods measure electrical activity of the heart from textile electrodes placed into the bed sheets (Peltokangas et al. 2012), while others focus on the mechanical activity of both heart and lungs. The registration of the latter is formerly called a ballistocardiogram (BCG) (Starr et al. 1939). Once heart and breathing rate are acquired, heart and breathing rate variability analysis (i.e. temporal changes in heart and breathing rate analyzed in both time and frequency domain) can lead to information regarding cardiac health (Malik et al. 1996), stress (Hall et al. 2004) and sleep quality (Willemen et al. 2014a, Stein & Pu 2012).

This paper presents an off-body registration methodology for respiratory and cardiac activity, more specifically a mechanical sensor pad based on pressure variations in between two air volumes. The setup was adapted from the one proposed by Shin et al. (Shin et al. 2010). The registered signals are characterized through a number of experiments, looking at four different sleep postures, as well as differences in base air pressure and material type of the mattress top layer.

7.2 Methodology

7.2.1 Setup of the measurement system

The used set-up is adapted from (Shin et al. 2010) and consists of two 0.85 m air-filled tubes combined with a differential pressure sensor. The tubes are placed laterally and in parallel under the thorax of the subject. The system is covered by a 4 cm mattress top layer to prevent any discomfort from the tubes, which are 3 cm in diameter when inflated. Fig. 7.1 shows a visualization of the measurement setup.

A certain flexibility of the tube material is required to ensure sufficient deformation of the tubes under the body contours of the subject and the small body motions of respiration and heart rate. Butyl tubes (cf. the inner tube of a standard bicycle tire) were found to have the desired properties. One of the tubes has a stem valve and gauge pressure sensor with digital display, which allows modifying the inner pressure of the tube system by inflation and deflation.

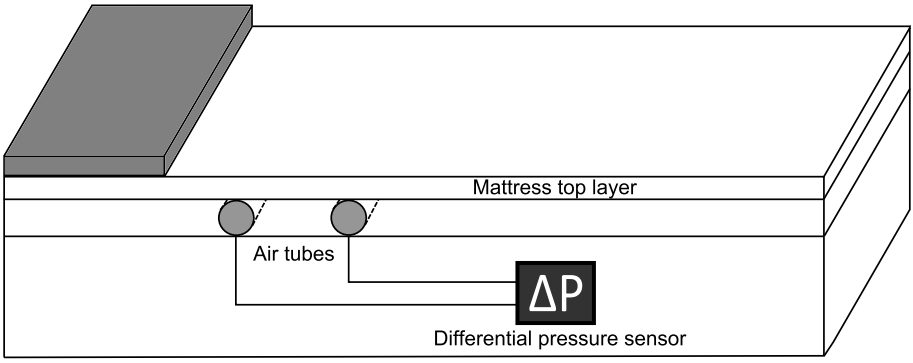


Figure 7.1: Visualisation of the measurement setup. The two circular tubes are placed underneath the thorax and connected to a differential pressure sensor. A mattress top layer prevents discomfort from the inflated tubes.

The observed gauge pressure variations due to respiration (< 50 Pa) and heart rate (< 5 Pa) are relatively small compared to the gauge pressure of a subject lying on the tubes. For this reason, a differential pressure sensor is able to track the signal of interest better than two separate gauge pressure sensors. For the current set-up, a Honeywell HSC Series - High Accuracy differential pressure sensor with a pressure range of -1 kPa to +1 kPa is used (Honeywell 2011).

The differential pressure sensor is connected to the computer by means of a microcontroller, in this case the Arduino Uno. The Arduino makes use of I2C communication to retrieve the data stream from the sensor, at a sample rate of 100 Hz. For validation of the heart beat registration, an analog pulse sensor (PulseSensor Amped 2011) was included in the measurement system, simultaneously sampled by the same Arduino Uno. The good signal-to-noise ratio of the systolic peak within this signal allows for an easy and accurate extraction of the sequential heart beat cycle locations using the standard Pan-Tompkins algorithm for ECG. Rare missed detections were manually corrected.

7.2.2 Experiments

The goal of the experiments was to quantify differences in the respiratory and heart beat signal in four basic sleep postures (supine, left lateral, right lateral, prone), as well as different base air pressures (5 kPa, 10 kPa, 15 kPa) and material types of the 4 cm mattress top layer (latex foam (Latex), high resilient poly-urethane foam (HR), visco-elastic poly-urethane foam (Visco)). Density values of the mattress top layers are respectively 65 kg/m^3, 40 kg/m^3 and 65

Table 7.1: Overview of the different experimental conditions.

#	Tube pressure	Topper material	Body postures	Time
1-4	15 kPa	HR	Supine, Left Lateral, Right Lateral, Prone	3 minutes
5-8	10 kPa	HR		
9-12	5 kPa	HR		
13-16	5 kPa	Latex		
17-20	5 kPa	Visco		

kg/m^3. Indentation hardness index values at 40% compression (as explained in (ISO 2439 2008)) are 70N, 100N and 70N. A healthy test person (male, age 27) went through the different experimental conditions as summarized in Table 7.1. In order to be able to compare in a correct way, the test person was instructed to always use the same breathing rhythm and depth (5 seconds per breath, silently counting 1-2-3 during both inspiration and expiration), and the shoulders were always positioned on the same area of the mattress.

7.2.3 Signal characterization

The acquired pressure sensor signal was first decomposed into a respiration and heart beat signal. Both periodic signals have a different, non-overlapping frequency range: 0.05-0.5 Hz for the respiration and 2-15 Hz for the heart beat signal (Shin et al. 2010). Decomposition was performed by means of a second-order zero-phase forward and reverse digital Butterworth band pass filter with the mentioned cut-off frequencies. Fig. 7.2 shows that for the respiration only the low-frequency component of the original signal is preserved, whereas for the heart beat signal a high-frequency signal is obtained. Next, the heart beat signal was segmented into individual heart beat cycles using the annotated systolic peak locations from the pulse sensor. Every heart beat cycle segment consists of a one second interval, centered on the systolic peak location. Also for the respiration signal, peaks and valleys, corresponding to initiation of respectively expiration and inspiration, were manually annotated.

In order to quantify the experimental differences between the registered signals, four different parameters were calculated for each of the 20 experimental conditions of Table 7.1: three parameters for the heart beat signal and one for the respiration signal. First, to estimate the power of the heart beat cycle (and thus its detectability), median absolute deviation values were calculated for every segment. Higher values correspond to larger oscillations, thus to

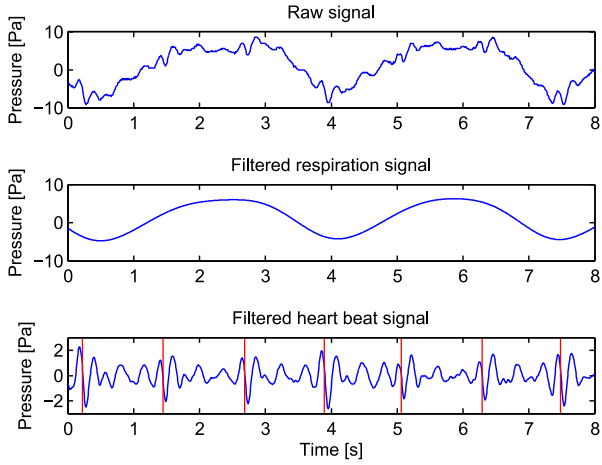


Figure 7.2: Eight second fragment of the raw signal from the pressure sensor decomposed via band pass filtering into a respiration signal (0.05-0.5 Hz) and heart beat signal (2-15 Hz). Locations of systolic peaks from reference pulse sensor are indicated by vertical lines in the bottom plot.

higher detectability. Second, to estimate the variability between the different heart beat cycle segments of the same experimental condition, a mean template of the heart beat cycle was constructed by averaging over the different heart beat cycle segments of that condition ($\approx$ 180 segments). Next, this mean template was subtracted from each of the heart beat cycle segments, after which the power of these difference segments was again estimated using the median absolute deviation. Higher values correspond to larger deviations from the mean template, thus to higher variability. Third, since larger heart beat cycle power would normally also lead to larger variability in absolute terms, the ratio of both is taken in order to have a relative comparison of power versus variability. A high ratio will indicate an easy to detect heart beat cycle with low variability. Fourth, to estimate the detectability of the inspiration and expiration periods of the respiration signal, the amplitude difference between the peaks and preceding valleys was calculated. Higher values correspond to larger respiration signal oscillations, thus to higher detectability. Statistical differences between the different base air pressures and mattress top layers were evaluated using the Kruskal-Wallis one-way analysis of variance by ranks; values of $p < 0.001$ are deemed significant.

7.3 Results

The values of the heart beat and respiration signal parameters are displayed in Table 7.2. From left to right, they list the power of the heart beat cycle (Power heart), the variability of the heart beat cycle (Variability heart), the ratio of heart beat cycle power and variability (Power/Variability ratio) and the detectability of the inspiration and expiration periods in the respiration signal (Detectability respiration). Values given are the median and inter-quartile range (IQR).

Looking at Table 7.2, there is a significant increase in detectability for both the heart beat cycles (Power Heart) as well as the respiration cycles (Detectability Resp) when going from a larger to a smaller base air pressure. For the heart beat cycle power to variability ratio, differences between base air pressures are significant for supine, left lateral and prone. For supine and prone, a lower base air pressure leads to a higher power to variability ratio; for left lateral and right lateral the ratio increases from 15 kPa to 10 kPa, but drops again from 10 kPa to 5 kPa.

There is also a significant increase in detectability for the respiration cycles going from a HR or Latex mattress top layer to a Visco top layer. For the heart beat power, differences are smaller but still significant for supine, left lateral and right lateral. Values seem to be the lowest for the Latex mattress top layer and mostly equal for the HR and Visco top layers. For the power to variability ratio, differences between the different mattress top layers are significant for supine, left lateral and prone, with the largest values for the Latex top layer; except for supine, which has its largest value in the Visco top layer.

Between the different postures, respiration detectability is the lowest for supine and by far the highest for prone. Heart beat cycle detectability is the lowest for right lateral, and again the highest for prone. The latter also has the highest variability of the heart beat cycle though, giving supine and left lateral the highest heart beat cycle power to variability ratio.

The results on the heart beat signal parameters of Table 7.2 are visualized in Fig. 7.3, 7.4 and 7.5. Fig. 7.3 visualizes the variability of the heart beat cycle for the case of a HR mattress top layer with 5 kPa base air pressure. It shows the individual heart beat cycles plotted together with the mean template in bold. Fig. 7.4 and 7.5 show only the mean heart beat cycle templates, respectively for the different base air pressures and mattress top layers.

Table 7.2: Heart beat and respiration signal parameters for all experimental conditions.

	Pressure - Top layer	**Body posture**	**Power heart [Pa]**		**Variability heart [Pa]**		**Power/Var. ratio [-]**		**Detectability respiration [Pa]**	
			Median	*IQR*	*Median*	*IQR*	*Median*	*IQR*	*Median*	*IQR*
1		S	0.15	0.04	0.09	0.03	1.62	0.58	2.82	1.82
2	15 kPa -	LL	0.17	0.05	0.09	0.04	2.01	0.68	7.56	2.52
3	HR	RL	0.15	0.04	0.09	0.03	1.56	0.54	6.90	2.96
4		P	0.17	0.06	0.12	0.05	1.41	0.49	22.49	6.71
5		S	0.19	0.05	0.09	0.04	2.10	0.85	7.90	2.07
6	10 kPa -	LL	0.27	0.06	0.12	0.05	2.41	0.90	15.32	4.00
7	HR	RL	0.17	0.06	0.10	0.05	1.61	0.52	6.94	2.84
8		P	0.25	0.09	0.16	0.08	1.61	0.70	27.80	6.72
9		S	0.39	0.12	0.17	0.07	2.16	1.00	12.10	6.31
10	5 kPa -	LL	0.41	0.12	0.21	0.08	1.88	0.95	21.37	4.01
11	HR	RL	0.28	0.09	0.18	0.08	1.50	0.59	18.01	5.51
12		P	0.56	0.15	0.32	0.12	1.77	0.81	49.30	7.54
13		S	0.33	0.08	0.16	0.06	1.94	0.91	19.05	3.38
14	5 kPa -	LL	0.40	0.10	0.19	0.07	2.12	0.94	22.36	6.80
15	Latex	RL	0.22	0.06	0.15	0.05	1.59	0.51	19.42	2.85
16		P	0.57	0.16	0.30	0.12	1.82	0.67	64.79	11.91
17		S	0.39	0.08	0.17	0.06	2.22	1.14	18.73	5.71
18	5 kPa -	LL	0.46	0.12	0.26	0.11	1.78	0.71	42.53	6.94
19	Visco	RL	0.27	0.10	0.18	0.09	1.50	0.51	29.87	4.76
20		P	0.57	0.20	0.43	0.17	1.35	0.44	116.02	23.07

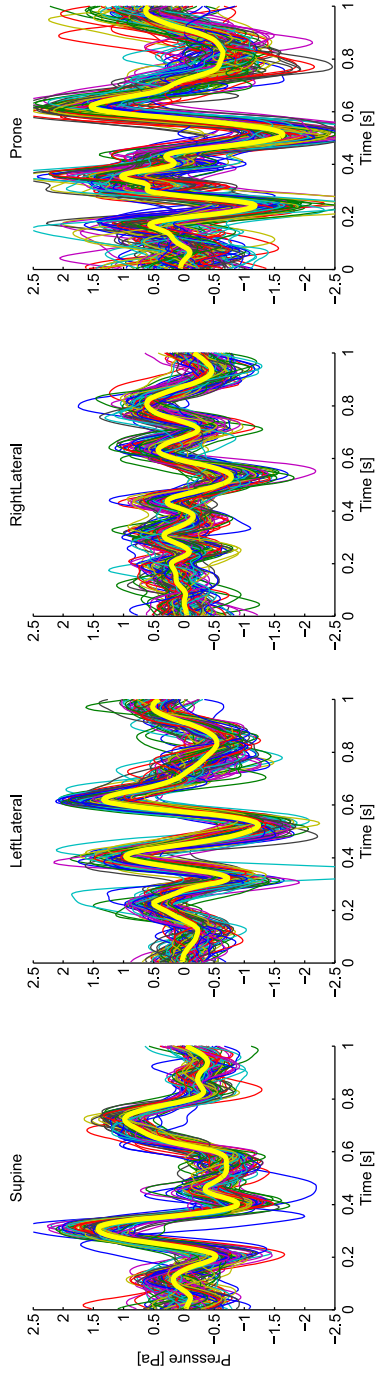


Figure 7.3: Visual representation of the variability in the heart beat signal (HR mattress top layer, 5 kPa base air pressure). The individual heart beat cycles are plotted together with the template in bold.

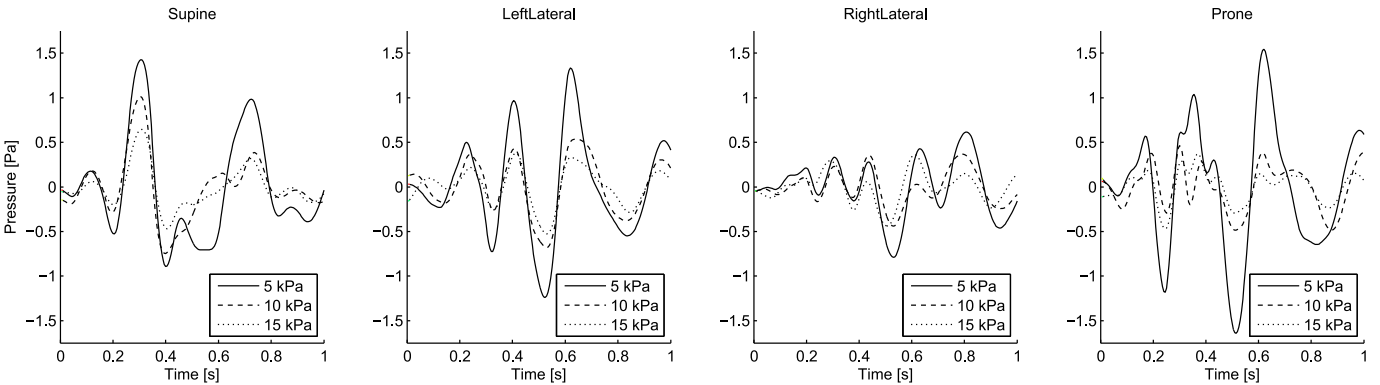


Figure 7.4: Heart beat templates for the three different base air pressures (5 kPa, 10 kPa, 15 kPa) on a HR mattress top layer.

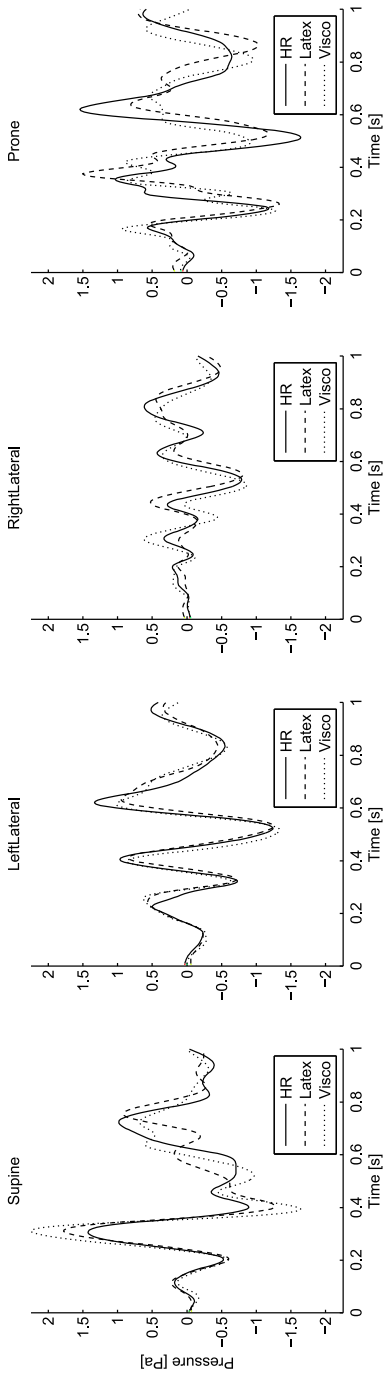


Figure 7.5: Heart beat templates for the three different mattress top layers (HR, Latex, Visco) with 5 kPa base air pressure.

7.4 Discussion

This paper described a bed-implemented off-body registration setup for respiratory and cardiac activity, and characterized the quality of the registered signals over different sleeping postures, base air pressures and mattress top layers. In (Watanabe et al. 2005), heart beat signal strength is characterized by the signal to noise ratio of the first heart beat harmonic in the FFT spectrum, an approach which should normally lead to similar results. Our approach however does not limit the signal strength to only one harmonic.

Increasing the base air pressure caused a reduction in signal strength and amplitude, which can be explained by a reduction in flexibility of the tubes. Although a more sensitive pressure sensor could cope with this reduced signal amplitude, it would not only be more costly but also more fragile due to possible overpressure from large and sudden body movements. Furthermore, the reduced flexibility of the tubes could require an increased thickness of the mattress top layer to prevent discomfort.

The significantly higher respiration amplitude for the Visco top layer could possibly be explained by a higher compressibility compared to the other top layers. A higher compressibility would not only bring the body closer to the air tubes but would also improve contact, thus enabling better mechanical transmission. More in-depth research should point out whether others factors play a role. For the heart beat signal parameters, differences between the mattress top layers were small. The highest signal amplitude for prone in both the respiratory and heart beat signal can easily be explained by the smallest distance from respectively the chest and heart to the air tubes. The lowest amplitude for the respiratory signal in supine and for the heart beat signal in right lateral is explained similarly by the larger distance of chest and heart (asymmetric position towards left side) to the registration setup. The highest variability of the heart beat cycle in prone and the lowest in supine can probably be explained by interference with the respiratory movement of chest and abdomen.

The authors conclude that, while the influence of the mattress top layer material is small, the base air pressure can be tuned for optimal mechanical transmission from heart and lungs towards the registration setup. Future work will investigate whether this tuning could be done automatically using a feedback loop, taking into account interpersonal body characteristics and sleep posture changes during the night. Furthermore, algorithms will be developed for extracting heart beat locations and respiration dynamics.

Acknowledgements

Research supported by Research Council KUL: GOA MaNet, PFV/10/002 (OPTEC), several PhD/postdoc & fellow grants; iMinds: SBO dotatie 2013, ICON: NXT_Sleep, Belgian Federal Science Policy Office: IUAP P719/ (DYSCO, 2012-2017), IWT SBO EASI-sleep.

Chapter 8

A Pan-Tompkins based algorithm for heart beat detection in the ballistocardiogram

Tim Willemen[1,2,3], Dorien Van Deun[1], Vincent Verhaert[1], Bart Haex[1], Jos Vander Sloten[1], Sabine Van Huffel [2,3]

[1] Biomechanics Section, Department of Mechanical Engineering, KU Leuven, 3001 Leuven, Belgium

[2] STADIUS, Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium

[3] Medical IT, iMinds, 3001 Leuven, Belgium

[4] IRISI, 3001 Leuven, Belgium

As adapted from a submission to *IEEE Journal of Biomedical and Health Informatics*, Nov. 2014.

Statement of relevance

This chapter is the second of two chapters addressing research question 3. It outlines the results of a study presenting and benchmarking an algorithm for the accurate detection of inter beat intervals in the ballistocardiogram. The results of this study have been submitted for publication to the journal *IEEE Journal of Biomedical and Health Informatics* (Willemen et al. 2014d). Readers who went through the general introduction of this thesis (chapter two) are able to skip the first two paragraphs of the introduction of this chapter.

Abstract

Smart bedding systems, able to record vital body parameters such as respiration and heart rate in an unobtrusive way, present an enormous opportunity for health monitoring. This paper presents a comparison between three algorithmic approaches for the localization of individual heart beats in the ballistocardiogram, as evaluated on a dataset of fifteen subjects lying in the four standard sleeping postures. Their common basis is the well-known Pan-Tompkins algorithm for electrocardio-graphy, now adapted for use on the ballistocardiogram. Although the basic algorithm is able to correctly segment almost all individual heart beat cycles, it failed to deliver a precise estimation of the individual heart beat locations due to the varying shape of sequential heart beat cycle segments. Therefore, two additional algorithms were defined for feature-based and template-based retrospective alignment of the individual heart beat location estimates, first detected by the basic algorithm. The template-based approach delivered the best results, mainly because of its possibility to adapt to the varying shape of the heart beat cycle segments over different postures and persons. The feature-based approach, looking for both the highest and steepest peak within the cycle, missed accuracy especially for the right lateral and prone posture. Future work will consists in implementing and validating an online version of the proposed algorithm.

Keywords

Ballistocardiography, biomedical signal processing, health monitoring, smart bedding systems

8.1 Introduction

Smart bedding systems able to record vital body parameters present an enormous opportunity for health monitoring. In the literature, several methods and techniques are described to detect respiration and heart rate from within or around the bed system. They are based on either measuring the electrical or mechanical activity of heart and lungs, without attaching sensors directly to the subject's body. Via heart and breathing rate variability analysis (HRV and BRV), information regarding cardiac health (Malik et al. 1996), stress (Hall et al. 2004) and sleep quality (Willemen et al. 2014a) can be obtained. The latter two are based on research that links the variability within respiration and heart rate to the sympathetic and parasympathetic variations within the autonomic nervous system (Somers et al. 1993).

As described in the literature, electrical activity of the heart can be registered by embedding textile electrodes in the bed sheets (Peltokangas et al. 2012, Devot et al. 2007). Similarly to how a standard electrocardiogram (ECG) is obtained, it registers the voltage difference between the different electrodes in contact with the subject's body, providing information on both heart rate and respiration. Mechanical activity of heart and lungs leads to subtle body motions. Respiration will cause thorax movement, whereas the ejection of blood with every heart beat gives a small but detectable oscillation of the body in accordance with the principle of conservation of momentum. The registration of the latter is formerly called a ballistocardiogram (BCG) (Braunstein 1953).

Existing methodologies to register this BCG use e.g. Doppler radar directed at the chest area (Guohua et al. 2007), place vibration sensors at the bed posts (Brink et al. 2006, Nukaya et al. 2012) or sensor pads underneath the chest area that work either electronically (Alihanka et al. 1981), mechanically (Shin et al. 2010, Willemen et al. 2014b) or optically (Brüser et al. 2012). The latter sensor pads are also placed underneath or within the pillow (Zhu et al. 2006). For increased reliability or detection coverage, multiple techniques could be combined.

For extracting the individual heart beat locations from the BCG, numerous algorithms of different complexity have been developed, of which a non-exhaustive overview is given below. A first category uses characteristic features of the general BCG heart beat cycle shape to determine the heart beat locations, either using thresholds or using machine-learning mechanisms such as clustering (Brüser et al. 2011, Rosales et al. 2012, Heise et al. 2013). Sometimes these features are improved by first applying a differentiation or squaring filter (Mack et al. 2009, Kortelainen & Virkkala 2007). A second category tries to determine the heart beat locations via correlation to either a BCG heart beat cycle

template constructed from a training set (Jansen et al. 1991), or via correlation to the previously detected heart beat cycle segment (Brüser et al. 2011, Friedrich et al. 2010). Heart beat cycle modeling is sometimes used to improve correlation, either using a respiratory latent variable model (Paalasmaa 2010, Paalasmaa et al. 2014) or via dynamic time warping (Wang et al. 2013). Multiple algorithms have also been combined for improved robustness (Brüser et al. 2011, Friedrich et al. 2010, Sprager & Zazula 2012, Brüser et al. 2013).

This paper discusses a comparison between three algorithmic approaches for the localization of the individual heart beats in the BCG. Their basis is the well-known Pan-Tompkins algorithm for ECG, now adapted for use on the BCG. In addition to the basic algorithm, both a feature-based and template-based approach are evaluated for retrospective alignment of the individual heart beat location estimates, first detected by the basic algorithm. An algorithm for the correct identification of respiration cycles was not part of this study, but can for example be found in (Paalasmaa et al. 2011). Data was collected with an air pressure-based BCG setup, improved from the one proposed by Shin et al. (Shin et al. 2010). Our adaptations to the setup were introduced and characterized within (Willemen et al. 2014b). The algorithmic approach described in this paper, although developed using the air pressure-based registration method, is more broadly applicable to most BCG methods, since all originate from the same BCG principle.

8.2 Materials and methods

8.2.1 Set-up of the measurement system

The used set-up was introduced and characterized in (Willemen et al. 2014b). It consists of two 0.85 m air-filled tubes combined with a differential pressure sensor, more specifically a Honeywell HSC Series - High Accuracy differential pressure sensor with a pressure range of -1 kPa to +1 kPa (Honeywell 2011). The Butyl tubes, about 3 cm in diameter when inflated to an unloaded base pressure of 5 kPa, are placed laterally and in parallel under the thorax of the subject within an opening carved within the top of the mattress. Next, they are covered by a four cm mattress top layer, more specifically a visco-elastic poly-urethane foam (Visco) with a density of 65 kg/m^3. This top layer serves to prevent any discomfort from the inflated tubes, as well as to prevent the tubes from being pinched off due to the subject lying on top of them. Fig. 8.1 shows a visualization of the measurement setup.

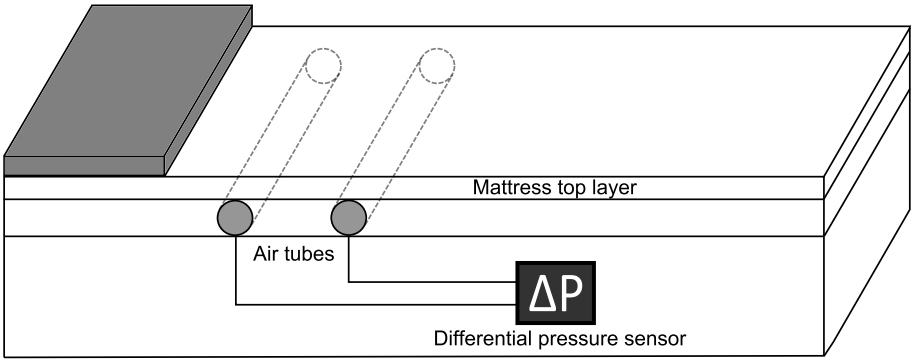


Figure 8.1: Visualisation of the measurement setup. The two circular tubes are placed underneath the thorax and connected to a differential pressure sensor. A mattress top layer prevents discomfort from the inflated tubes.

The differential pressure sensor is connected to the computer by means of a microcontroller, in this case the Arduino Uno. The Arduino makes use of I2C (inter-integrated circuits) communication to retrieve the data stream from the sensor, within a loop time of 10 ms (sample rate $\approx$ 100 Hz). Since the expected frequencies of relevance for respiration and heart rate are within a frequency range of 0.1-15 Hz (Shin et al. 2010), this sample rate is sufficient. Because the sampling of the sensor signal is not uniform, it is resampled at a uniform sample rate of 100 Hz, using timestamp information coming from the sensor with each datapoint. For validation purposes, a NexFin ECG monitor was included in the measurement system, of which the lead II output was simultaneously sampled by the same Arduino Uno for convenient synchronization. The R-peaks within this ECG were extracted using the standard Pan-Tompkins algorithm for ECG. Rare misdetections were corrected manually.

8.2.2 Experiments

Fifteen healthy volunteers (10 males & 5 females; age 26.5 $\pm$ 2 years) were asked to lie on the bed with the implemented BCG setup, while being monitored by both the BCG setup as well as the ECG monitor. They were asked to first lie down in the supine posture, after which posture changes to respectively left lateral, prone and right lateral posture were requested with an interval of 2.5 minutes. They were also instructed to lie still in between these posture changes to simulate a state of sleep. No further instructions were given as to how exactly to position themselves on top of the mattress, in order to simulate a realistic environment as good as possible. This led to approximately 10 minute

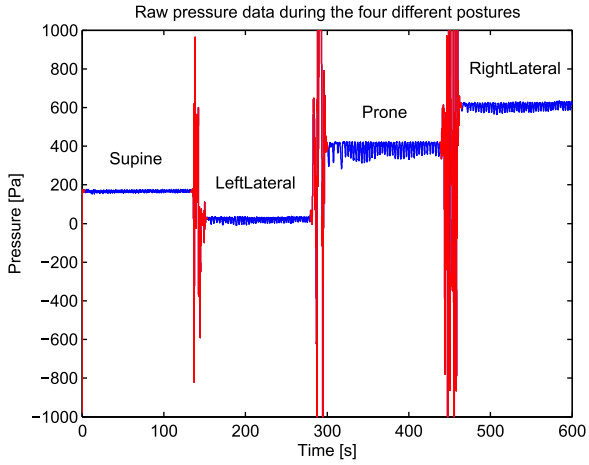


Figure 8.2: Visualization of the result of the movement detection algorithm applied to the 10-minute raw pressure data excerpt from subject 1. The three periods of detected movement activity are indicated in red.

recordings from every subject, distributed over four standard sleeping postures, as visualized in Fig. 8.2.

8.2.3 Movement detection algorithm

In order to split up the 10 minute recordings in movement-free parts, a simple off-line movement detection algorithm was developed. The raw pressure data from the differential pressure sensor was first differentiated in order to detrend the data and indicate significant pressure changes. Next, it was squared for increased emphasis on the periods of movement activity. This signal was finally thresholded with a heuristically determined threshold, equal to the mean plus mean absolute deviation (MAD) of this 10 minute signal. Movements with less than 5 seconds in between are classified as a single movement. The three seconds before and after every movement were also included within the movement activity to prevent transient behavior to be included within the movement-free parts. As an example, Fig. 8.2 visualizes the three periods of detected movement activity within the 10-minute raw pressure data excerpt of subject 1.

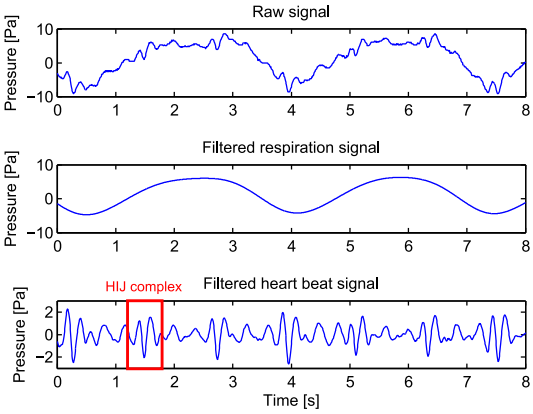


Figure 8.3: Decomposition of the original raw signal into a low-frequency component (respiration signal) and high-frequency component (heart beat or BCG). The characteristic HIJ complex is highlighted with a rectangle.

8.2.4 Heart beat localization: basic Pan-Tompkins (PT) algorithm

The aim of the heart beat localization algorithms is to automatically detect the exact locations of heart beat events in the movement-free parts of the pressure signal. First, the resampled pressure signal is decomposed into a respiration signal and heart beat signal (i.e. the BCG). Both periodic signals have a different, non-overlapping frequency range: 0.05-0.5 Hz for the respiration signal and 2-15 Hz for the BCG (Shin et al. 2010). Decomposition was done by means of a second-order zero-phase forward and reverse digital Butterworth bandpass filter with the mentioned cut-off frequencies. Fig. 8.3 shows that for the respiration only the low-frequency component of the original signal is preserved, whereas for the BCG a high-frequency signal is obtained. In the next step, both resulting signals can be processed separately. As mentioned in the introduction, an algorithm for the correct identification of respiration cycles was not part of this study, but can for example be found in (Paalasmaa et al. 2011).

According to (Braunstein 1953), every heart beat cycle segment in the BCG should contain a characteristic HIJ complex, i.e. a sequence of two positive and one intermediate negative peak, as visualized in Fig. 8.3 and in more detail in Fig. 8.5. Although the QRS complex in the ECG is more prominent than the HIJ complex in the BCG, similar algorithms can be applied. For R-peak

extraction from the QRS complex in the ECG, the Pan-Tompkins algorithm is commonly used (Pan & Tompkins 1985). Here, a novel adaptation for J-peak extraction from the HIJ complex in the BCG is presented.

First, since the steep slopes of the HIJ complex are an important feature to identify the different heart beat locations (similarly to the slopes of the QRS complex in the ECG signal), a derivative is taken,

$$\boldsymbol{y}(n) = \left(\frac{T}{8}\right)\left[\boldsymbol{x}(n-2) + 2\boldsymbol{x}(n-1) - 2\boldsymbol{x}(n+1) - \boldsymbol{x}(n+2)\right]. \tag{8.1}$$

Next, a point-wise squaring of the signal is performed,

$$\boldsymbol{z}(n) = \left[\boldsymbol{y}(n)\right]^2, \tag{8.2}$$

which makes all data points positive and emphasizes the higher frequencies by non-linear amplification (Pan & Tompkins 1985). The resulting adjacent positive peaks coming from the HIJ complex are then integrated to obtain a smoothed single peak for each corresponding heart beat cycle segment. Since the HIJ complex is much wider than the QRS complex, the number of samples in the moving integration window needs to be larger too. A wider rectangular window could however output integrated peaks which include information of the HIJ complexes from the two adjacent heart beat cycle segments, especially for an elevated heart beat rhythm. Therefore a triangular window is preferred,

$$\boldsymbol{w}(n) = \left(\frac{2}{N-1}\right)^2 \left(\frac{N-1}{2} - \left|n - \frac{N-1}{2}\right|\right), \tag{8.3}$$

where N is equal to the window width, empirically chosen to be 75 samples for a sample rate of 100Hz, which equals a width of 750 ms. Fig. 8.4 shows a BCG of 8 seconds on which the differentiation, squaring, and integration step are performed.

The peaks in the resulting signal after the integration step can be found by a straight-forward derivative-based peak detection method. However, due to noise on the BCG, not all peaks truly correspond to an HIJ complex. Selecting all of them would lead to numerous false positive peak detections. In the Pan-Tompkins algorithm for ECG, an adaptive thresholding routine is run to identify the correct integrated peak locations corresponding to the QRS complexes of the ECG. However, since the BCG is much more susceptible to noise, the integrated peaks corresponding to the HIJ complexes of the BCG are

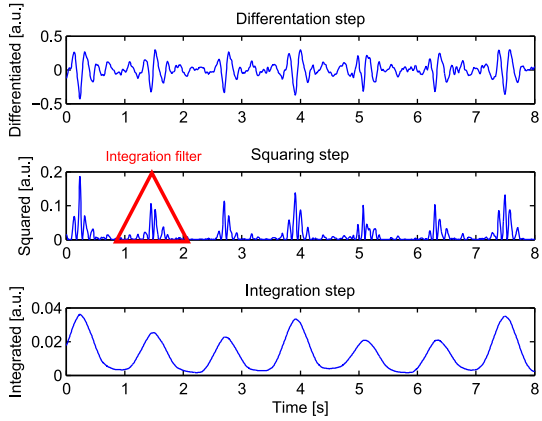


Figure 8.4: The differentiation, squaring and integration step of the heart beat localization algorithm applied on 8 seconds of example BCG data. The triangular integration filter is shown in the squaring step.

much more variable in amplitude. Therefore, the use of a thresholding routine for the peak height does not meet the required accuracy.

Instead, a signal-to-noise (SNR) ratio is calculated for every detected peak, by dividing its peak height by the average depth of its two neighboring valleys. If two peaks are located less than 0.5 seconds from each other, only the peak with the highest SNR ratio is kept. The other one is disregarded as being a false positive detection. Making this 0.5 seconds distance adaptive to the average heart rate rhythm of the subject at that time did not improve results on this dataset, but might be necessary on datasets including events of bradycardia and tachycardia. The remaining integrated peaks are regarded as the heart beat location estimates of the basic algorithm.

8.2.5 Heart beat localization: feature-based approach (Height-Steepness)

A step commonly performed after the basic Pan-Tompkins algorithm for ECG, is looking around the detected peak locations from the integrated signal (after filter delay correction) for the maximum value within the original ECG, those being the R-peaks of the corresponding QRS complexes. Similarly, the J-peaks from the HIJ complexes in the original BCG can be located by searching for a characteristic feature around the heart beat location estimate of the

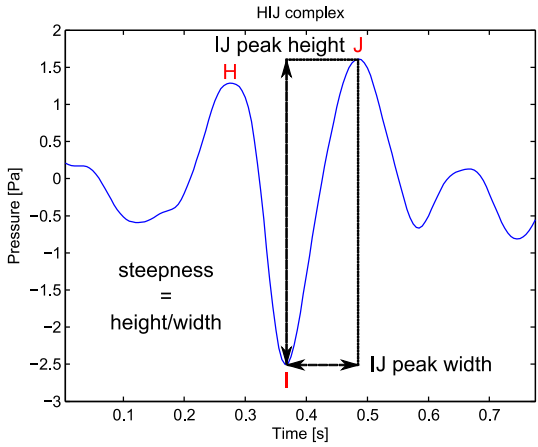


Figure 8.5: Individual heart beat cycle segment with indicated peak height, width and steepness parameters.

basic algorithm (also after filter delay correction), more specifically within an empirically determined symmetric range of 600 ms. For the BCG, only looking at one extreme value (the maximum) proved not to be robust. Therefore, a combination of two different characteristic features was used. Firstly the height difference between the negative I-peak and the positive J-peak, and secondly the average slope steepness of the J-peak, both indicated in Fig. 8.5. A formula was defined to determine the peak most likely to be the J-peak, namely

$$\boldsymbol{peakScore} = \frac{\boldsymbol{peakH}}{\max\left(\boldsymbol{peakH}\right)} + \frac{\boldsymbol{peakS}}{\max\left(\boldsymbol{peakS}\right)}, \tag{8.4}$$

where $\boldsymbol{peakH}$ and $\boldsymbol{peakS}$ are respectively the height and steepness of every peak in the BCG within the 600 ms symmetric range around the heart beat location estimate of the basic algorithm. The peak with the highest score is selected as the J-peak.

8.2.6 Heart beat localization: template-based approach

Instead of focusing on one specific part or feature of the heart beat cycle segment, another approach would be to align the complete heart beat cycle segment with a template. This template is constructed independently for every movement-free part, starting from the heart beat location estimates of the basic

algorithm. Individual heart beat cycle segments were derived by extracting BCG segments centered on each of these location estimates. The width of these segments equals 1.25 times the average inter beat interval of the subject in the corresponding movement-free part, as calculated from the heart beat location estimates of the basic algorithm. Segment width was made adaptive to prevent inclusion of neighboring heart beat cycles, while still ensuring the presence of the complete heart beat cycle. The individual heart beat cycle segments are stored as rows of a heart beat cycle segment matrix $\boldsymbol{M}$. Each row has a corresponding center location, representing the heart beat location estimate. A heart beat cycle template $\boldsymbol{T}$ can now be constructed by taking the median over the first dimension of this matrix $\boldsymbol{M}$.

In order to align the individual heart beat cycle segments with this template $\boldsymbol{T}$, an iteration is performed over all rows of the matrix $\boldsymbol{M}$. For every row, the cross-correlation is calculated between the individual segment and the constructed template $\boldsymbol{T}$. The location of the maximum cross-correlation value indicates over which distance d (either positive (right) or negative (left)) the individual segment needs to be shifted in order to align with the template. This distance d is then subtracted from the segment's original corresponding center location, i.e. relocating the heart beat location estimate to a more optimized position. The heart beat cycle segment of that row is then updated based on this new heart beat location estimate. The template $\boldsymbol{T}$ is updated after every relocation step (by taking again the median over the first dimension of the updated matrix $\boldsymbol{M}$) to improve its quality over time. The iteration over all rows of the matrix $\boldsymbol{M}$ is performed twice to ensure that also the first rows of the matrix $\boldsymbol{M}$ can be aligned to an improved version of the template $\boldsymbol{T}$. Begin and end result of the procedure is visualized in Fig 8.6, for the BCG heart beat cycle segments of subject 2 during prone posture.

8.2.7 Evaluation of the HRV accuracy

The goal of the experiments was to compare the heart beat localization capabilities of the three algorithms described in section 8.2.4, 8.2.5 and 8.2.6. Good estimates of heart beat locations are important for accurate Heart Rate Variability (HRV) analysis, which is based on time and frequency domain measures calculated on the time intervals between individual heart beat locations. In the ECG, these interbeat time intervals are usually defined based on the R-peak locations (the so-called RR intervals). In the BCG, similar BB intervals can be defined based on the heart beat location estimates of the three algorithms. For the basic algorithm, this corresponds to the locations of the integrated peaks. For the feature-based approach, it corresponds to the locations of the

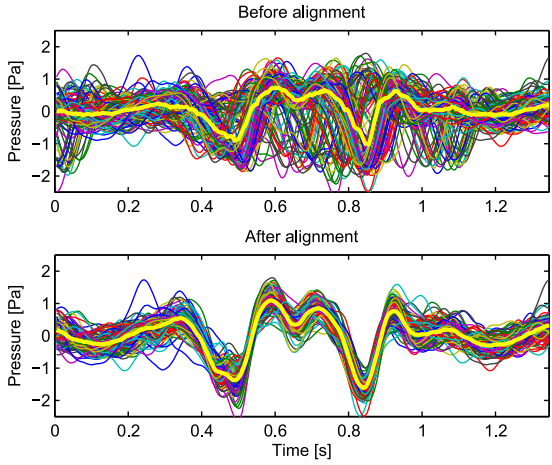


Figure 8.6: Visualization of the result of the template alignment algorithm on the band-pass filtered pressure data from subject 2 during prone posture. The heart beat cycle template T is indicated in bold yellow, before and after alignment of the individual heart beat cycle segments.

J-peaks. For the template-based approach, it corresponds to the relocated center locations of each row of the updated heart beat cycle segment matrix $\boldsymbol{M}$.

Since the gold standard for HRV analysis makes use of RR intervals extracted from the ECG, the HRV accuracy was defined to reflect the similarity between the detected BB intervals from the BCG and their corresponding reference RR intervals from the ECG. A BB interval is defined to be similar to an RR interval if the difference in length between them is smaller than an allowed difference threshold T. Results of HRV accuracy will be plotted as a function of this allowed difference threshold T (varying between 0.01 and 0.40 seconds), and Tables will include results for T equal to 0.05 and 0.10 seconds.

To be precise, the HRV accuracy was defined as a combination of precision and recall,

$$HRV\ accuracy = \frac{precision + recall}{2}. \tag{8.5}$$

The precision is a measure that indicates how well an algorithm correctly distinguishes true positives over false positives,

$$precision = \frac{TP}{TP + FP}, \tag{8.6}$$

where in this case a true positive (TP) is found when for a detected RR interval from the ECG, the corresponding BB interval from the BCG has similar length, whereas for a false positive (FP) the corresponding BB interval is shorter due to an extra invalid detected beat within the BCG. Similarly, the recall is a measure that indicates how well an algorithm is able to pick up all true positives,

$$recall = \frac{TP}{TP + FN}, \tag{8.7}$$

where in this case a true positive (TP) is found when for a detected BB interval from the BCG, the corresponding RR interval from the ECG has similar length, whereas for a false negative (FN) the corresponding RR interval is shorter due to a missed beat within the BCG.

8.2.8 Optimization of the bandwidth from the BCG

The frequency spectrum of the raw pressure sensor signal contains (next to the respiration) both the fundamental frequency as well as the harmonics of the heart beat events. Depending on the chosen bandwidth and the average heart rate rhythm of the subject, the frequency spectrum of the BCG contains a selection of this fundamental frequency and harmonics. Changing the cut-off frequencies of the Butterworth bandpass filter used in the basic algorithm of section 8.2.4 can significantly alter the BCG, and thus the heart beat localization capabilities of the three algorithms investigated.

In order to check whether the standard bandwidth of 2-15 Hz (as proposed in (Shin et al. 2010)) could be further improved, an optimal low-pass and high-pass cut-off frequency for the bandpass filter of the BCG was looked for. All low-pass and high-pass cut-off frequency combinations were possible between 1 and 20 Hz, with a minimum step size of 1 Hz. For the high-pass cut-off frequency, a minimum value of 0.66 Hz was also allowed. This cut-off frequency still filters out the respiration, but does allow for the presence of the heart beat fundamental frequency within the possible bandwidths. Optimal bandwidths were specific for each of the three algorithms and each of the four postures. They were determined based on a maximal HRV accuracy criterion averaged over all subjects, in particular for an allowed length difference threshold T equal to 0.05 and 0.10 seconds. An optimal bandwidth common for all postures was also determined.

8.3 Results

A comparison between the HRV accuracies and corresponding optimal bandwidths of the three algorithms are displayed in Table 8.1, for each of the four postures separately, as well as for an optimized bandwidth over all postures combined. Accuracy values given are averaged over all subjects, and are displayed for the allowed length difference threshold values T equal to 0.05 and 0.10 seconds. HRV accuracy values are the highest for the template-based approach, followed by (depending on the chosen threshold T) either the feature-based approach or the basic Pan-Tompkins algorithm. Values in supine posture are comparable for all algorithms, whereas primarily in right lateral and prone posture differences between the algorithms are significant ($p < 0.001$). The optimal bandwidths include the lower frequencies for supine and prone posture, whereas in right lateral posture they are mostly excluded. Table 8.2 displays a comparison between the standard bandwidth of 2-15 Hz (proposed by (Shin et al. 2010)) and the optimal bandwidth for all postures combined (as displayed on the right side in Table 8.1). Differences are rather small, because other than the optimal bandwidth for the template-based approach at the threshold T = 0.10 s, the optimal bandwidths do not differ that much from the standard bandwidth of 2-15 Hz.

The results of Table 8.2, for the optimal bandwidths corresponding to the threshold T equal to 0.10 seconds, are also plotted in Fig. 8.7 as a function of the allowed length difference threshold T. It displays a clear comparison in the HRV accuracy between the three algorithms described within this paper. While the basic algorithm shows a steady increase in HRV accuracy with increasing threshold T, both the feature-based and template-based approach show a step-like increase, primarily visible for the feature-based approach in left lateral and prone posture, and for the template-based approach in right lateral and prone posture. Fig. 8.8 displays an inter subject comparison for the template-based approach. It shows the same step-like behavior in HRV accuracy for certain subjects. It also shows that in most cases, HRV accuracy values are comparable between a threshold T equal to 0.05 and 0.02 seconds. The smaller the length difference between the RR and BB intervals, the more reliable time and frequency domain HRV measures could be extracted from the BB intervals. With the current sample rate (100 Hz) and the intrinsic error from the limited frequency content in the BCG (compared to the ECG), smaller length differences are difficult to obtain.

Table 8.1: Bandwidth-optimized HRV accuracy values averaged over all subjects for threshold values T=0.05 seconds and T=0.10 seconds. [Acc = accuracy; BW = bandwidth]

	T	**Supine**		**LeftLateral**		**RightLateral**		**Prone**		**All postures**	
		Acc	*BW*	*Acc*	*BW*	*Acc*	*BW*	*Acc*	*BW*	*Acc*	*BW*
Basic PT	0.05 s	0.9509	1-13 Hz	0.8843	4-20 Hz	0.8208	3-20 Hz	0.8207	3-10 Hz	0.8549	1-19 Hz
Feature		0.9834	1-6 Hz	0.9468	4-18 Hz	0.7803	2-7 Hz	0.8686	2-13 Hz	0.8745	2-19 Hz
Template		0.9946	2-10 Hz	0.9822	5-18 Hz	0.9713	14-19 Hz	0.9844	2-15 Hz	0.9666	3-20 Hz
Basic PT	0.10 s	0.9721	1-19 Hz	0.9679	4-19 Hz	0.8950	7-20 Hz	0.8973	5-8 Hz	0.9232	3-20 Hz
Feature		0.9922	1-20 Hz	0.9517	4-20 Hz	0.8657	16-19 Hz	0.8894	1-15 Hz	0.9030	2-20 Hz
Template		0.9965	2-10 Hz	0.9847	6-19 Hz	0.9768	12-20 Hz	0.9882	2-15 Hz	0.9728	6-20 Hz

Table 8.2: HRV accuracy values averaged over all subjects for threshold values T=0.05 seconds and T=0.10 seconds. Values given are for both the standard 2-15 Hz Bandwidth, as well as the method-specific optimal bandwidth for all postures as derived from Table 8.1. [Stnd = standard; Opt = optimal]

	T	**Supine**		**LeftLateral**		**RightLateral**		**Prone**		**All postures**	
		Stnd	*Opt*	*Stnd*	*Opt*	*Stnd*	*Opt*	*Stnd*	*Opt*	*Stnd*	*Opt*
Basic PT	0.05 s	0.9195	0.9445	0.8573	0.8614	0.7988	0.7957	0.8163	0.8179	0.8480	0.8549
Feature		0.9369	0.9382	0.9188	0.9277	0.7643	0.7721	0.8623	0.8599	0.8707	0.8745
Template		0.9939	0.9913	0.9536	0.9666	0.8959	0.9324	0.9844	0.9760	0.9569	0.9666
		Stnd	*Opt*	*Stnd*	*Opt*	*Stnd*	*Opt*	*Stnd*	*Opt*	*Stnd*	*Opt*
Basic PT	0.10 s	0.9566	0.9484	0.9309	0.9579	0.8723	0.8937	0.8970	0.8929	0.9142	0.9232
Feature		0.9767	0.9820	0.9253	0.9343	0.8007	0.8156	0.8783	0.8799	0.8952	0.9030
Template		0.9956	0.9946	0.9605	0.9843	0.9144	0.9601	0.9882	0.9521	0.9647	0.9728

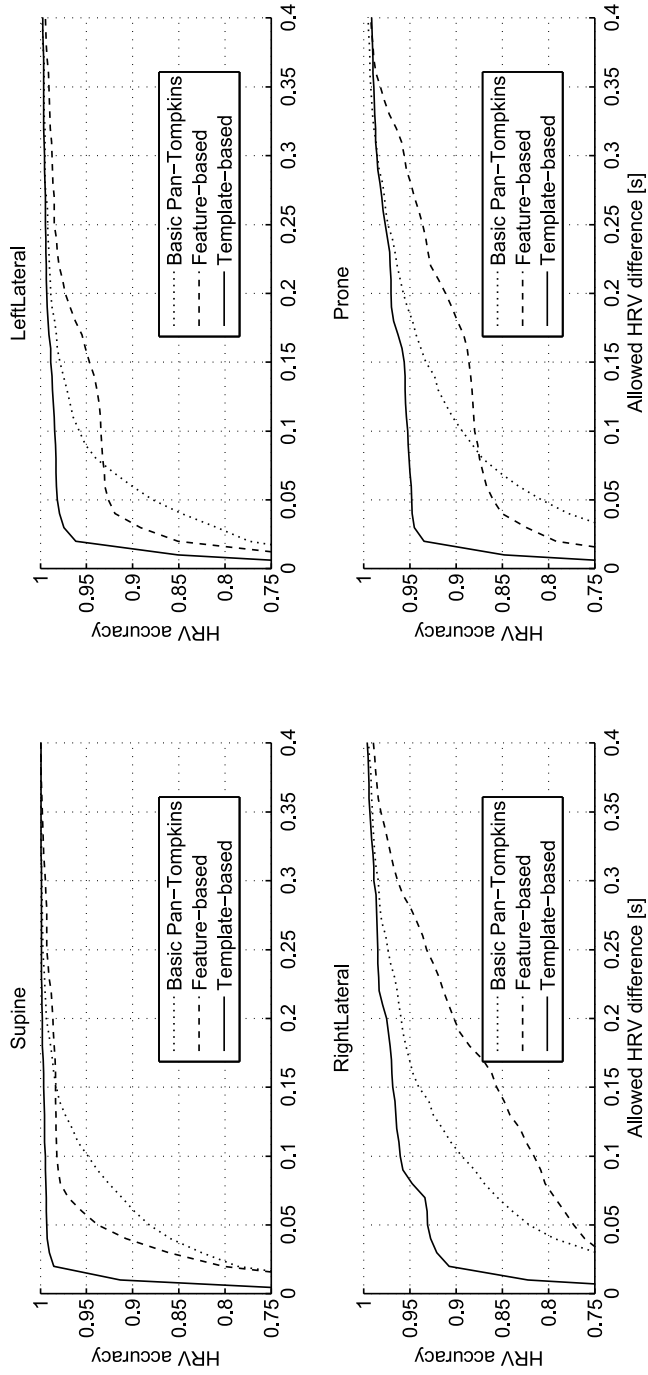


Figure 8.7: Comparison of the HRV accuracy (averaged over all subjects) between the four different postures and three different BCG beat detection algorithms. HRV accuracy is plotted as a function of the allowed HRV length difference between BCG and ECG inter-beat intervals.

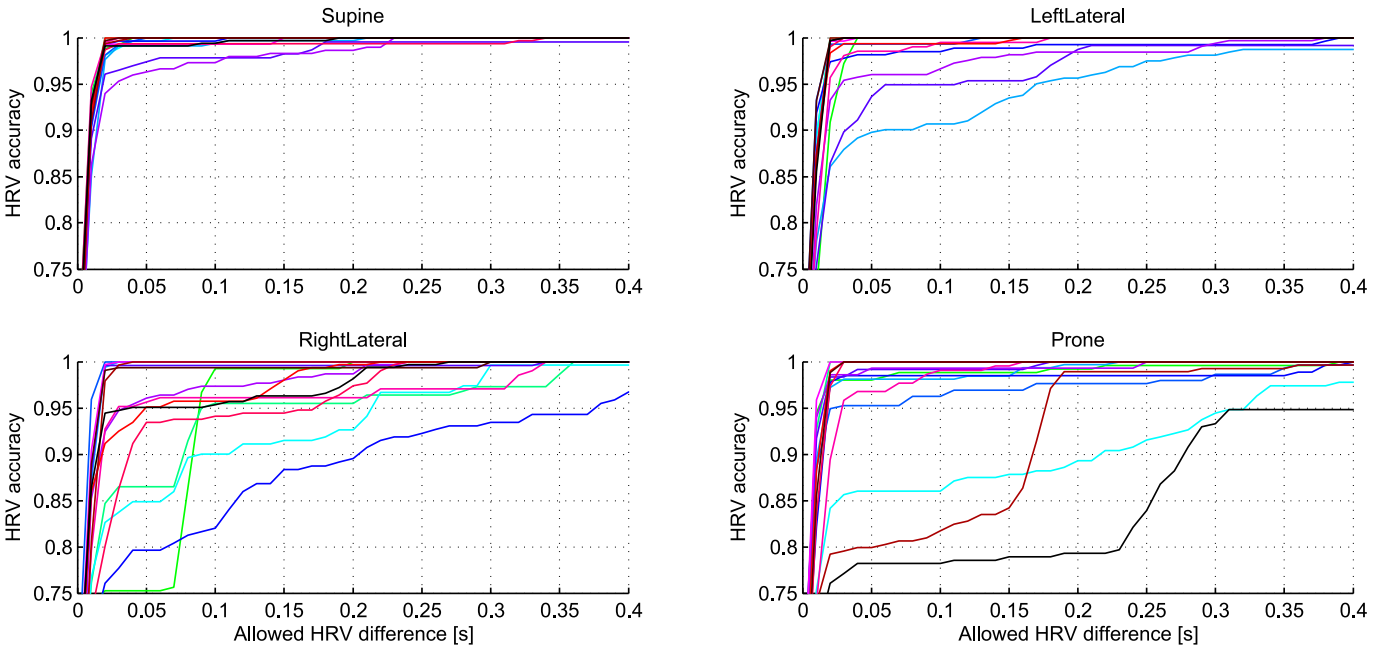


Figure 8.8: Comparison of the 'Template-based' HRV accuracy between the four different postures and fifteen different subjects. HRV accuracy is plotted as a function of the allowed HRV difference between BCG and ECG inter-beat intervals.

8.4 Discussion

This paper described and evaluated three algorithmic approaches for the localization of individual heart beats in the BCG. Their basis is the well-known Pan-Tompkins algorithm for ECG, now adapted for use on the BCG. In addition to the basic algorithm, both a feature-based (Height-Steepness) and template-based approach were evaluated for retrospective alignment of the individual heart beat location estimates, first detected by the basic algorithm. Although the basic algorithm was able to correctly segment almost all individual heart beat cycles, it failed to deliver a precise estimation of the individual heart beat locations due to the varying shape of sequential heart beat cycle segments. These variations can be attributed to respiratory influences and other body-function and environmental noise.

The origin of the difference in HIJ complex shape over respiration is discussed in (Starr & Friedland 1946): evidenced by experiments, they attribute this variation to changes in cardiac filling, with a delayed but smaller influence to the left ventricle compared to the right one. The fact that the BCG is modulated upon the respiration signal might however be the largest influence, as proposed in our previous work (Willemen et al. 2014b). The rising and declining of the respiration signal alter the filtered BCG differently during inspiration compared to expiration. This could explain why HRV accuracy values of the basic algorithm are the lowest in prone posture, because of the highest breathing signal amplitudes and thus the highest possibility of distortion within that posture, as shown in (Willemen et al. 2014b).

The SNR of the integrated peaks from the basic algorithm can give a first indication on the signal power (median of the SNR) and variability (median absolute deviation of the SNR) for a certain subject and posture. Signal power was overall the lowest for the right lateral posture (largest distance between heart and air tubes), while signal variability was the highest in prone posture (largest breathing amplitudes). The ratio of this power and variability measure showed a significant positive correlation with the HRV accuracy ($T = 0.05$ seconds) of the basic algorithm ($r = 0.6062$, $p < 0.001$). Power to variability ratios were (as expected) the lowest for the right lateral and prone posture.

In order to account for the varying shape of the heart beat cycle segments, a comparison was made between a feature-based and a template-based method for retrospective alignment of the individual heart beat location estimates, first detected by the basic algorithm. The template-based method delivered the best results, mainly because of its possibility to adapt to the varying shape of the heart beat cycle segments over different postures and persons. The feature-based method, looking for both the highest and steepest peak within the cycle, lacked

accuracy especially within the right lateral and prone posture.

The step-like increase in HRV accuracy with increasing threshold T, as visible in both Fig. 8.7 and Fig. 8.8, is caused by the retrospective alignment used by both algorithms. In the feature-based approach, for some heart beat cycle segments, instead of the correct J-peak, a neighboring peak of the heart beat cycle segment will have been selected. The prone posture heart beat data in Fig. 8.6 are a good example of a difficult to distinguish J-peak, and clearly shows the limitation of the feature-based approach due to the possible high similarity in peak height and steepness between two peaks within the same heart beat cycle segment. This second peak will always be located at about the same distance from the correct J-peak. It thus causes the HRV accuracy to stabilize for a while on the same accuracy value until the threshold T becomes higher than this inter-peak distance.

Although the template-based approach was chosen to overcome this limitation, the effect is thus still there, albeit smaller. The template-based approach might be further improved by modeling the influence of the respiration on the heart beat template, or by using different templates for different moments within the respiratory cycle. A simple amplitude model (normalizing the amplitude of the rows in the heart beat cycle segment matrix $\boldsymbol{M}$) did not improve results on this dataset; more sophisticated techniques however might be succesfull, for example a respiratory latent variable model (Paalasmaa 2010) or dynamic time warping (Wang et al. 2013), which are able to morph the template in a non-linear way.

The overall HRV accuracy might also be improved by combining the heart beat localization algorithms with a posture detection technique, such as described in (Verhaert et al. 2011a). This way, the lower frequencies can be filtered out for the lateral postures, and included for the prone posture. Another possibility might be to disregard the lower frequencies if the heart beat signal has an overall low median SNR of the integrated peaks from the basic algorithm, indicating right lateral posture. Determining the correct threshold might be difficult however due to large inter-subject variations. A final possibility might be to use a multi-modal approach consisting of multiple BCG's with different bandwidths, originating from the same pressure signal. The heart beat location estimates detected on these different BCG's might then be combined similarly as in (Brüser et al. 2011), where combinations were made based on different detection algorithms. Alternatively, the most reliable sequence of BB intervals might be selected based on the calculation of Hjorth's mobility parameter on this sequence (Hjorth 1970), where a lower value would indicate a more smooth sequence and therefore a more reliable result. The BCG from multiple detection setups could be combined in a similar way.

The inter-subject variations within Fig. 8.8 can partly be explained by differences

in average heart beat rhythm. When heart beat cycle segments are further apart, the oscillations have more time to diminish, causing less interference between neighboring segments. Also, a slower heart beat rhythm can be linked to a larger amount of blood ejected by the heart at every heart beat. This leads to an increased momentum of the ejected blood, thus (according to the principle of conservation of momentum) to an increased momentum of the body. Whether or not this momentum leads to a small or large body oscillation depends amongst others on the body weight of the person. The effect of the average heart beat rhythm is proven by a significant negative correlation found between average heart beat rhythm and median SNR of the integrated peaks from the basic algorithm ($r = -0.530$, $p < 0.001$).

8.5 Conclusion

Since differences between subjects, sleeping postures and location within the breathing cycle all lead to changes in heart beat cycle shape, an adaptive algorithm is required which does not solely focus on parts of the highly variable heart beat cycle characteristics. The template-approach presented in this paper shows promising results, and due to its rather simple approach, has possibilities for online monitoring. In future work, the automatic adaptation of bandwidths or the multi-modal approach will be investigated, as well as an online implementation of the template-based approach. Overnight testing of the system and algorithm will be performed to further investigate their reliability for use in HRV monitoring.

Acknowledgements

This research was supported by Research Council KUL: GOA MaNet, PFV/10/002 (OPTEC), several PhD/postdoc & fellow grants; iMinds: SBO dotatie 2013, ICON: NXT_Sleep, Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, 2012-2017), IWT SBO EASI-sleep.

Chapter 9

General discussion

9.1 Summary of chapters

Chapter two outlined the societal aspects of sleep, after which the physiology of sleep was illustrated, making use of the two-process and flip-flop models of sleep. Next an overview was given of the current standards of sleep assessment, the most common sleep disorders and some of their existing treatment options. Observed limitations to the current standards were highlighted, and possible solutions and opportunities revealed. The current gap in clinical and research practice towards assessment of sleep and its pathologies was addressed. Multiple night PSG measurements are found too costly, impractical and intrusive for longterm screening, while the currently available unobtrusive measurements (actigraphy, PVT's, subjective questionnaires) might lack sufficient information and/or accuracy towards certain sleep disorders. This thesis investigated whether automated cardiac, respiratory and movement-based analysis could be able to bridge this gap, especially when all signals are monitored off-body in a mechanical way. Sleep and its disorders have a significant influence on the autonomic nervous system, which controls cardiac, respiratory and movement activity. The principle of ballistocardiography allows the mechanical registration of these signals in an off-body way.

Chapter three outlined the results of a proof-of-concept study distinguishing the combination of wake and active sleep (Wake-REM-N1) from passive sleep (N2-N3) using easy to register signals, i.e. cardiac activity from electrocardiogram, respiratory activity from a thorax breathing belt, and movement activity from a continuously measured mattress indentation. The study was performed

on a small 16-night subset of a private database of polysomnographic nights from healthy subjects, as collected during a 4-year IWT-SBO project entitled 'EASI-sleep'. Two different training and test set compositions were investigated, varying with respect to the presence of person-specific information in the training nights. Features were selected using a filter method, based on independent two-sample t-tests (to extract feature relevance) and Pearson's correlation coefficient (to determine inter-feature correlation). Classification was performed using Least-Squares Support Vector machines. Results illustrated the ability to use cardiac, respiratory and movement features for distinguishing the combination of wake and active sleep from passive sleep in healthy subjects, but suggested the importance of person-specificity in the extracted features.

Chapter four outlined the results of a larger follow-up study, extending the sleep classification towards a full hypnogram (distinguishing between Wake, REM, light sleep (N1-N2) and deep sleep (N3)). The study was performed on a larger 85-night subset of the same private EASI-sleep database, using the same signals. Features were selected using a combination of a wrapper method (single-feature training set accuracy) and filter method (a minimum Redundancy Maximum Relevance criterion based on Mutual Information). Classification was performed using Support Vector Machines. Results demonstrated that subject-independent sleep stage classification using cardiac, respiratory and movement activity is feasible in healthy subjects, with agreement values around 80% and kappa values around 0.60. Features should however be sufficiently transformed to cope with the large amount of variability in heart and lung functioning among different subjects. The intermittent absence of sympathetic activation in REM sleep, and the close correspondence between REM and N1 could however lead to difficult-to-avoid misclassifications.

Chapter five outlined the results of a study distinguishing between apneic events and healthy breathing. The study was performed on a public 70-night database with apnea annotations, only containing an electrocardiogram from which both cardiac as well as respiratory activity was derived. Features were selected using a combination of two wrapper methods: one for feature preselection (single-feature training set accuracy) and one for final feature subset construction (greedy forward selection). Classification was performed using Linear Discriminant Analysis. Results indicated a positive influence on classifier performance when information from neighboring segments was incorporated. Averaging of feature values over neighboring segments led to classifier agreement values around 90% and kappa values up to 0.80. Extracting features over larger window segments led however to a reduction in classifier agreement due to its susceptibility to noise.

Chapter six outlined the results of a study distinguishing between apneic events and healthy breathing, as well as between the different sleep stages Wake, REM,

light sleep (N1-N2) and deep sleep (N3). The study was performed on a public 25-night polysomnography database of apneic patients. Cardiac activity was derived from electrocardiogram, while respiratory activity was derived from a thorax breathing belt. Features were selected using a combination of two wrapper methods: one for feature preselection (single-feature training set accuracy) and one for final feature subset construction (sequential forward floating search). Classification was performed using Linear Discriminant Analysis. Results showed a significant impact of the presence of apneic events on sleep stage classifier performance, since apneic events have similar sympathetic activation of the cardiac and respiratory system, independent of the sleep stage. Agreement values were only around 70% and kappa values mostly below 0.40. A high proportion of hypopneic events and the absence of recovery breaths in the expert apneic annotations probably led to mediocre performance values in detection of apnea, with agreement values around 75% and kappa values around 0.30. A probabilistic representation (referred to as the hypnocorrogram and apneacorrogram) did allow to put classifier output in perspective, visualizing classifier uncertainty in correspondence to training data. Further improvements might be necessary however to make the classifier reliable for use on patients with sleep apnea.

Chapter seven outlined the results of a study building and characterizing a ballistocardiographic setup, implemented inside of a bed. The setup proved able to detect movement as well as the mechanical activity of both heart and lungs by measuring pressure difference fluctuations between two air volumes underneath the chest area of the subject. Signal strength of both cardiac and respiratory activity was inversely correlated with the base air pressure inside of the two air volumes. The influence of different mattress top layer materials was small. Cardiac and respiratory signal strength varied between postures, inversely correlated with the distance between heart/chest and air volumes. Respiratory movement of chest and abdomen influenced the variability within the cardiac signal, with higher respiratory signal strength leading to higher cardiac signal variability.

Chapter eight outlined the results of a study presenting and benchmarking an algorithm for the accurate detection of inter beat intervals in the ballistocardiogram. The basis of the algorithm is the well-known Pan-Tompkins algorithm for electrocardiography, now adapted for use on the ballistocardiogram. Two additional algorithms were defined for feature-based and template-based retrospective alignment of the individual heart beat location estimates, first detected by the basic Pan-Tompkins algorithm. The three algorithms were evaluated on a dataset of fifteen subjects lying in the four standard sleeping postures. Although the basic algorithm was able to correctly segment almost all individual heart beat cycles, it failed to deliver a precise estimation of the

individual heart beat locations due to the varying shape of sequential heart beat cycle segments. This varying shape proved also detrimental to the feature-based method, which focused too much on only one part of the heart beat cycle. The template-based method surmounted this limitation, displaying an average correspondence of 97% with respect to inter beat intervals detected by electrocardiogram.

9.2 Discussion of chapters

9.2.1 Choice of signals

This PhD thesis hypothesized that the current gap in clinical and research standards towards assessment of sleep and its pathologies could be bridged by supplementing the movement-based analysis of actigraphy with a cardiac and respiratory-based analysis of sleep and its disorders; especially when all of these signals can be monitored off-body without attaching sensors directly to the skin.

However, for the development of the sleep classification models of chapters three through six, off-body registered signals were not available (except for the movement signal used in chapters three and four). Instead, cardiac activity was registered by ECG, while respiratory activity was registered by both oro-nasal airflow (thermistor) as well as ribcage and abdomen movements (breathing belts). Although the thermistor signal is usually of much better quality and significance compared to the breathing belt signals, respiratory activity was always extracted only from these breathing belts. This was done to simulate the idea of measuring respiratory activity in an off-body way, where respiratory activity can only be monitored via chest or abdomen movements. Similarly for the cardiac activity, only the interbeat RR intervals were used, since those interbeat intervals can be detected off-body using the BCG.

The reliability of off-body detected cardiac and respiratory activity was assessed in chapters seven and eight. An algorithm for the reliable detection of the breathing signal from the BCG setup was however not presented. An estimation of the detectability of the breathing signal was provided by calculating the average breathing signal amplitude (height difference between the peaks and preceding valleys), of which results were shown in Table 7.2. Detectability was the lowest for supine and the highest for prone, explained by the difference in distance and contact area between chest and measurement setup. Breathing signal amplitude seemed sufficient however to warrant focus of chapter eight on the reliable detection of the heart beat signal.

A higher breathing signal amplitude could have been established by positioning one of the BCG setup's air tubes more towards the abdominal area. The latter usually displays a larger amount of breathing movement, especially during REM sleep (Tabachnik et al. 1981). This placement would have however negatively affected the signal quality of the heart beat signal. First, the larger distance between the heart and the air tubes would have decreased the amplitude of the heart beat signal. Second, the increased amplitude of the breathing signal would have caused an increased amount of interference, leading to a larger amount of heart beat signal variability.

In order to simulate close-correspondence to the signals coming from the BCG setup, it was for this reason that in chapters three, four and six, the thorax breathing belt signal was preferred. Use of the abdominal breathing belt signal would probably have led to an improvement of the sleep and apneic classification models. A future adaptation of the BCG setup might benefit from decoupling the measurement systems for heart beat and breathing signal, by placing the former underneath the chest area and the latter underneath the abdominal area.

9.2.2 Characterization of apneic events

The apneic breathing classification models of chapters five and six have been mostly presented in a way which is regarded as a black box, i.e. without much knowledge of its internal workings. A list of top selected features was always provided, but not extensively discussed as was done for the sleep classification model in chapter four. In order to better understand the apneic detection capabilities of a heart rate and thorax breathing signal, the following section will go over some representative visual examples which better illustrate both the possibilities and limitations of the method.

First, it's important to revisit the difference between an obstructive and a central apneic event (Quan et al. 1999). While during a central apnea, the thorax breathing signal will show no breathing due to absence of breathing effort, during an obstructive apnea, a phenomenon known as paradoxical breathing is observed. Even though air flow is absent, the patient's breathing effort will still cause breathing movement in the chest and abdominal area, albeit of less amplitude. An example thorax breathing signal of a mixed apnea, starting as a central apnea but evolving into an obstructive one, is given in Fig. 9.1, with the period of absent air flow indicated with horizontal lines at the top and bottom of the figure.

Second, while breathing movement might still be present during the apneic event due to the patient's breathing effort, "breathing" rate is usually decreased below

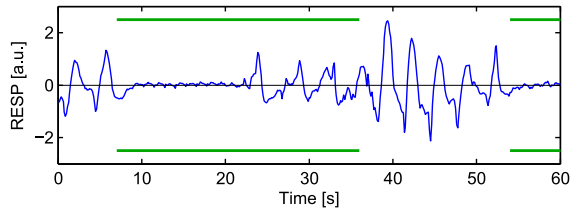


Figure 9.1: An example thorax breathing signal of a mixed apnea, starting as a central apnea but evolving into an obstructive one. The period of absent air flow is indicated with horizontal lines at the top and bottom of the figure.

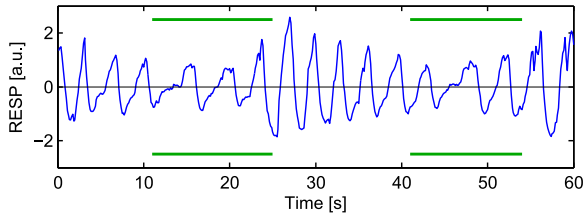


Figure 9.2: An example thorax breathing signal of an obstructive apnea, visualizing the below and above normal levels of breathing rate and amplitude. The period of absent air flow is indicated with horizontal lines at the top and bottom of the figure.

normal levels. Furthermore, after the apneic event, when air flow is reinitiated, both breathing rate as well as amplitude levels of breathing movement usually increase above normal levels. After these so-called recovery breaths, both return again to normal. An example thorax breathing signal of an obstructive apnea, visualizing the below and above normal levels of breathing rate and amplitude, is given in Fig. 9.2.

Third, the heart rate signal usually shows a characteristic reduction in heart rate during an apneic event (i.e. a period of bradycardia). The level of reduction varies with the degree of hypoxemia (oxygen deficiency in arterial blood) resulting from the absence of oxygenated air flow to the lungs (Zwillich et al. 1982). During the recovery breaths, when air flow is reinitiated, a rapid increase in heart rate above normal levels is sometimes perceived (i.e. a period of tachycardia), usually accompanied with an arousal. An example heart rate signal of an obstructive apnea, visualizing the periods of bradycardia, is given in Fig. 9.3.

Fourth, in the case of a hypopnea (the less severe variant of an apnea),

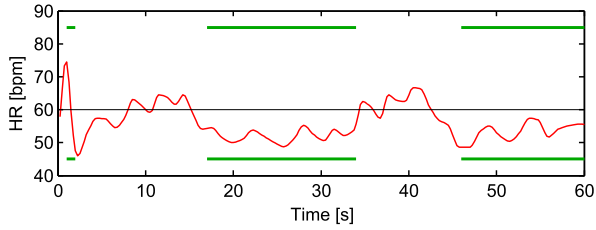


Figure 9.3: An example heart rate signal of an obstructive apnea, visualizing the periods of bradycardia. The period of absent air flow is indicated with horizontal lines at the top and bottom of the figure.

reduced breathing effort (central hypopnea) and partial obstructions (obstructive hypopnea) only lead to a reduction in air flow; by definition of at least 30% for at least 10 seconds (Berry et al. 2012). The effect this air flow reduction has on the amplitude of the breathing movement in the chest and abdominal area is sometimes difficult to conceive. The same holds true for the deceleration of both heart rate and breathing rate during the apneic event, and the acceleration of both during the recovery breaths. The manifestation of these characteristics depends on the severity of the hypopnea. An example of a more difficult to conceive hypopnea is given in Fig. 9.4. A reliable detection of hypopneic events could therefore require the use of extra signals, such as direct air flow measurements, or pulse oxymetry to monitor for significant oxyhemoglobin desaturations (Berry et al. 2012).

9.2.3 Choice of feature selection and classifier algorithms

Within chapters three to six of this thesis, several different classification models were built. For this, a multiple of different feature selection and classifier algorithms were utilized. Although not presented in the respective chapters itself, for both the REM vs NREM classification problem of chapter four, as well as the APNEIC vs HEALTHY breathing classification problem of chapter five, several of these feature selection and classifier algorithms were compared in order to make a founded choice of algorithm, as well as to assess the influence of this decision on the classification problem outcome.

For the REM vs NREM classification problem of chapter four, a comparison was made between the classifiers Naïve Bayes, J48tree, Random Forests, SVM and LSSVM. The results of this comparison on the 35 night healthy subject test set are displayed in Table 9.1. The highest value of kappa was reached for the SVM classifier, which was for this reason used in chapter four. Differences

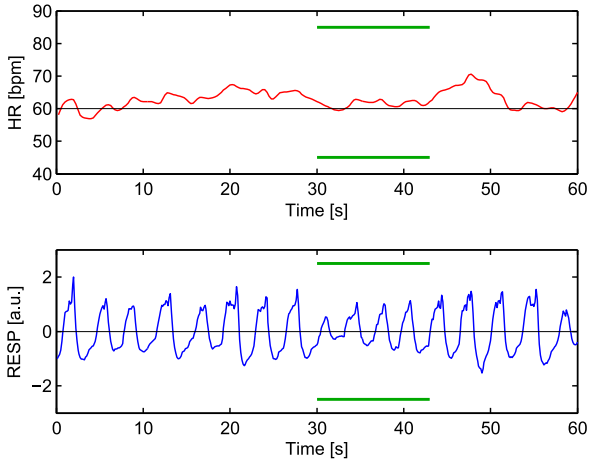


Figure 9.4: An example of a more difficult to conceive hypopnea. The period of absent air flow is indicated with horizontal lines at the top and bottom of the figure.

Table 9.1: Classifier comparison for the REM vs NREM classification problem of chapter four. Values given are the results on the test set of 35 healthy subject nights.

	Naïve Bayes	**J48tree**	**Random Forests**	**SVM**	**LSSVM**
Agreement	82.3	84.5	87.5	86.2	86.2
Kappa	48.9	46.7	55.6	55.8	52.2
Precision	49.2	54.9	64.5	55.8	59.8
Recall	75.6	57.3	61.8	75.3	61.4

between the classifier accuracies were however small, limiting the influence of this decision on the classification problem outcome.

For the APNEIC vs HEALTHY breathing classification problem of chapter five, a comparison was made between the classifiers LDA, SVM, LSSVM and Random Forests. The results of this comparison on the 35 night apnea test set are displayed in Fig. 9.5, as a function of the amount of features included in the classifier algorithm. Although this time the LSSVM classifier achieved the highest agreement values, the LDA classifier was preferred because of its much lower computational cost (at least an order of magnitude) while still achieving almost identical accuracy to the other three classifier algorithms.

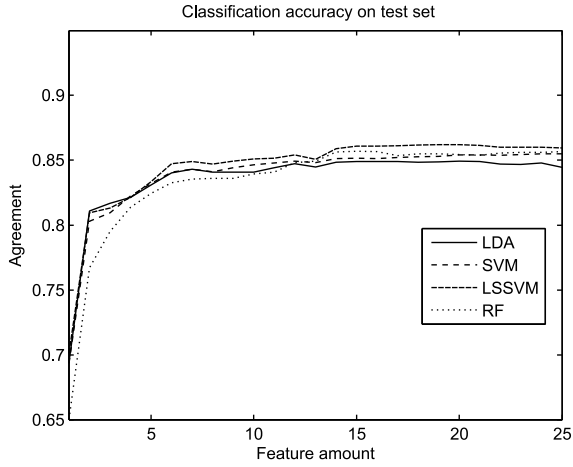


Figure 9.5: Classifier comparison for the APNEIC vs HEALTHY classification problem of chapter five. Results are displayed for the 35 night apnea test set, as a function of the amount of features included in the classifier algorithm.

Again, influence of this decision on the classification problem outcome was limited.

For the APNEIC vs HEALTHY breathing classification problem of chapter five, a comparison was also made between six feature selection algorithms, namely Kappa, MIMI, KappaSpearman, KappaKappa, GreedyKappa and GeneticAlgorithm, which are now shortly introduced. The Kappa algorithm selected out the top features with highest single-feature training set accuracy (i.e. kappa), without taking into account inter-feature correlation. This algorithm was used for feature preselection in chapters four, five and six, in order to reduce overall computational cost of the feature selection process. The MIMI, KappaSpearman and KappaKappa algorithms were all based on a minimum Redundancy Maximum Relevance algorithm, as used in chapter four. Relevance and redundancy can be determined using different measures, namely Mutual Information (MI; as used in chapter four), single-feature training set accuracy (Kappa), Spearman's correlation coefficient (Spearman; as used in chapter three) and double-feature training set accuracy (KappaKappa; were redundancy was based on the added or reduced accuracy (i.e. kappa) of adding a certain feature to an already selected feature). The GreedyKappa algorithm performed a greedy forward selection, as used in chapter five. The GeneticAlgorithm existed out of a combination of tournament parent selection, stochastic universal sampling, mutation and elitism to determine the optimal feature set with highest fitness (i.e. training set accuracy kappa). The GreedyKappa and GeneticAlgorithm

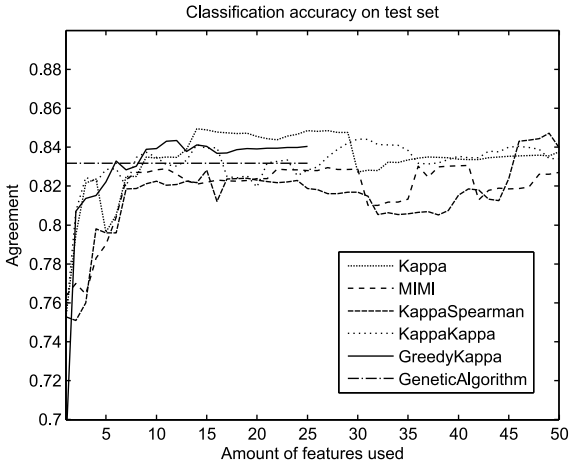


Figure 9.6: Feature selection comparison for the APNEIC vs HEALTHY classification problem of chapter five. Results are displayed for the 35 night apnea test set, as a function of the amount of features included in the LDA classifier algorithm.

were limited to a set of 25 features, while the other methods selected out 50 features. For the GeneticAlgorithm, only the single result at 25 features is displayed (horizontal line).

The results of the feature selection comparison on the 35 night apnea test set are displayed in Fig. 9.6, again as a function of the amount of features included in the classifier algorithm. Features were always classified with the LDA classifier. The GreedyKappa algorithm was preferred because it showed a quite consistent highest accuracy as a function of amount of features used. Still, since differences between the resulting accuracies were small, influence of this decision on the classification problem outcome was again limited.

Although the list of compared feature selection and classifier algorithms was not exhaustive, the above results stress the importance of well defined features above the choice of the right feature selection and classifier algorithm. While in this work, the starting list of features was mainly built up from literature, future work should focus on deriving new features which are able to surpass the results of the classification models proposed in this work. Improvements are deemed possible by e.g. including features with respect to cardiorespiratory coupling (Bartsch et al. 2014).

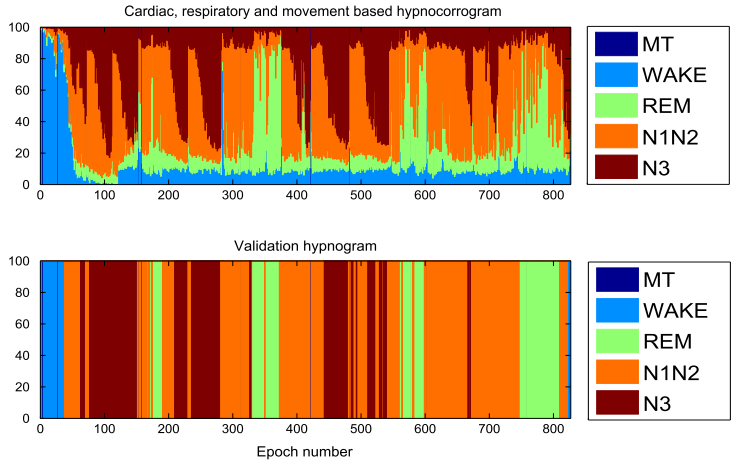


Figure 9.7: An example hypnocorrogram from the sleep classification model of chapter four (healthy sleep).

9.2.4 The hypnocorrogram in healthy sleep

In chapter six of this thesis, the hypnocorrogram and apneacorrogram were introduced as an alternative, non-discrete visual representation of respectively sleep stages and presence of apneic breathing. Due to the lacking accuracy of the sleep classification model within that chapter, the example hypnocorrogram shown in Fig. 6.6 displays posterior class probability estimates with often little bias towards one class. This lack of bias was explained by the effect apneic events have on the cardiac and respiratory system. Elevated amounts of SNS activity are introduced not only in close proximity to apneic events, but also as a more long-term effect during both sleep as well as wake (Somers et al. 1995). Both obscure the decision boundaries within the respective classification models. In order to demonstrate the full potential of the hypnocorrogram, an example from the sleep classification model of chapter four (healthy sleep) is shown in Fig. 9.7, which displays much better bias.

9.3 Evaluation of research questions

This thesis addressed the current gap in clinical and research practice towards assessment of sleep and its pathologies. Multiple night PSG measurements are found too costly, impractical and intrusive, while the currently available unobtrusive measurements (actigraphy, PVT's, subjective questionnaires) might

lack sufficient information and/or accuracy towards certain sleep disorders. This thesis investigated whether automated cardiac, respiratory and movement-based analysis could be able to bridge this gap, especially when all signals are monitored off-body in a mechanical way.

9.3.1 Research question 1

Can sleep be monitored reliably in healthy subjects using cardiac, respiratory and movement activity?

Chapters three and four of this work investigated the ability to use cardiac, respiratory and movement activity for sleep monitoring in healthy subjects. Using a dataset of 85 nights, several classification models were built to distinguish between Wake, REM, light sleep (N1-N2) and deep sleep (N3). The models were trained with and validated against gold standard polysomnography annotations, derived by sleep experts. Agreement values around 80% and kappa values around 0.60 confirmed the potential of a subject-independent method, with available improvement when going for subject-specific features.

While accuracy values are not on the same level as gold standard manual polysomnography, they are sufficient to bridge the gap between polysomnography and the currently available unobtrusive measurement techniques. Nevertheless, in order to fully answer this research question, future work should evaluate whether overall accuracy is sufficient for reliable detection of significant changes or deficiencies to a subject's hypnogram.

9.3.2 Research question 2

Can sleep be monitored reliably in sleep-disordered subjects using cardiac, respiratory and movement activity? More specifically in this thesis: can sleep and apneic events be monitored reliably in subjects with sleep apnea using cardiac and respiratory activity?

Chapters five and six of this work investigated the ability to use cardiac and respiratory activity for sleep monitoring in subjects with sleep apnea. Using a dataset of 70 nights, a classification model was built to distinguish between apneic and healthy breathing episodes. This model was again trained with and validated against gold standard polysomnography. Agreement values around 90% and kappa values around 0.80 showed great potential of the method. Using a second dataset of 25 nights, containing a high proportion of hypopneic events, a similar classification model led however to agreement values around 75% and

kappa values around 0.30, stressing the possible need of pulse oxymetry for reliable detection of hypopneic events.

The presence of apneic events also proved to have a significant impact on sleep stage classification model performance, with agreement values dropping to around 70% and kappa values to around 0.40 (although the influence of a missing movement signal should be investigated). The question rises however whether a reliable distribution of sleep stages is necessary when a significant amount of apneic events occur. Diagnosis of sleep apnea is in that case of most concern, since its negative impact on sleep continuity and depth is anyhow implied.

While the proposed probabilistic representations of the classification model outputs (referred to as the hypnocorrogram and apneacorrogram) are deemed to be a valuable addition by visualizing uncertainty of correspondence, further improvements will be necessary to make cardiac and respiratory-based sleep monitoring in subjects with sleep apnea reliable. Nevertheless, in order to fully answer this research question, future work should again evaluate whether overall accuracy is sufficient for reliable night-based diagnosis of potential (early) sleep apnea. In addition, more specific analysis should be performed towards other common sleep disorders such as insomnia and periodic limb movement disorder, which was not included in this thesis.

9.3.3 Research question 3

Can cardiac, respiratory and movement activity be monitored reliably in an off-body mechanical way?

Chapters seven and eight of this work investigated the ability to monitor cardiac, respiratory and movement activity in an off-body mechanical way. A pressure-based ballistocardiographic setup was implemented inside of a bed, measuring pressure difference fluctuations between two air volumes underneath the chest area of the subject. Further on, an adaptation of the Pan-Tompkins algorithm was proposed for accurate detection of cardiac inter beat intervals. The setup and algorithm were evaluated on a dataset of fifteen subjects lying in the four standard sleeping postures.

While respiratory and movement activity could be easily detected, cardiac signal quality varied between postures as well as persons. Still, in between episodes of movement activity, an average correspondence of 97% was achieved with respect to inter beat intervals detected by electrocardiogram, confirming the possibility for reliable detection. Although the validation experiments simulated a realistic sleeping environment, overnight testing will offer a more complete verification

of the system and algorithm, and will allow to check whether correspondence is sufficient to achieve comparable outcomes in detection of both sleep stages and apneic events.

9.4 General conclusion

In conclusion, this PhD thesis supports the opportunity of a biomechanics-based approach in sleep analysis, although its accuracy might be insufficient towards detection of minor changes or problems (such as small deficiencies in a subject's hypnogram or the presence of hypopneas). Future work should evaluate whether overall accuracy of sleep stage analysis and apneic breathing detection is sufficient for night-based diagnosis of potential (early) sleep disorders or disturbances, based on information from multiple consecutive nights. In addition, more specific analysis should be performed towards other common sleep disorders such as insomnia and periodic limb movement disorder, which was not included in this thesis.

Appendix A

Encompassing interdisciplinary research projects

A.1 EASI-SLEEP

The 4-year EASI-SLEEP project (2008 - 2011) focused on the improvement of prevention, diagnosis and treatment strategies for the different aspects of sleep disturbances. One of the most important targets of the project was preventing the evolution of situational insomnia into chronic insomnia, specifically into populations at risk such as shift workers, people living in 'at risk' zones with difficult to avoid noise and light disturbances, and frequent travelers.

The effect of different behavioral and environmental aspects was evaluated, in order to formulate specific recommendations towards these populations. Research was also performed towards low-cost and easy-to-use monitoring tools for the home environment, combining both quantitative as well as qualitative aspects of sleep, to be distributed by general practitioners to patients showing up with sleep complaints. A final important part of the project was the transition of going from monitoring to also guiding and manipulating these behavioral and environmental aspects, towards the realization of the intelligent bedroom of the future.

A.2 NXT_SLEEP

The 2-year NXT_SLEEP project (2013 - 2014) focused on the development of a next-generation sleep monitoring platform that is less obtrusive and more comfortable than the traditional polysomnography approach. The platform not only allows for sleep monitoring (for several nights) at people's homes, but also uses a new (beyond state-of-the-art) multi-sensor approach in combination with complex algorithms to accurately detect, process and deliver complete and useful information regarding the physiological parameters that come with sleep-related breathing disorders.

The multi-sensor approach was realized by integrating several beyond state-of-the-art technologies from different companies into one platform. Firstly, intelligent mattress technologies were explored to capture on the one hand movements that are related to heart rate and respiration, and on the other hand humidity and temperature. Secondly, an innovative body patch was integrated, recording wireless and long term electrocardiogram. Lastly, a novel microsensor was incorporated, which directly senses CO_2 in the exhaled air. Data was automatically collected by a cloud-based platform, ensuring its availability at all times. A specifically designed GUI handled the automatic data preprocessing, analysis and report generation.

Not only the medical relevance of the registered data was investigated, but also the usability and the user experience of the platform in order to satisfy the patients' and doctors' needs in the best way. In parallel, an appropriate business model for penetrating the sleep monitoring market was worked out in depth, including financial and organizational design and the development of a health care model.

Bibliography

Adam, K. & Oswald, I. (1984). Sleep helps healing., *British medical journal (Clinical research ed.)* **289**(6456): 1400–1401.

Åkerstedt, T. & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual, *International Journal of Neuroscience* **52**(1-2): 29–37.

Alihanka, J., Vaahtoranta, K. & Saarikivi, I. (1981). A new method for long-term monitoring of the ballistocardiogram, heart rate, and respiration, *Am J Physiol* **240**(5): R384–92.

Anderer, P., Moreau, A., Woertz, M., Ross, M., Gruber, G., Parapatics, S., Loretz, E., Heller, E., Schmidt, A., Boeck, M. et al. (2010). Computer-assisted sleep classification according to the standard of the american academy of sleep medicine: validation study of the aasm version of the somnolyzer 24× 7, *Neuropsychobiology* **62**(4): 250–264.

Aurora, R. N., Kristo, D. A., Bista, S. R., Rowley, J. A., Zak, R. S., Casey, K. R., Lamm, C. I., Tracy, S. L. & Rosenberg, R. S. (2012). The treatment of restless legs syndrome and periodic limb movement disorder in adults – an update for 2012: practice parameters with an evidence-based systematic review and meta-analyses: an american academy of sleep medicine clinical practice guideline, *Sleep* **35**(8): 1039–1062.

Bader, G. G. & Engdal, S. (2000). The influence of bed firmness on sleep quality, *Applied ergonomics* **31**(5): 487–497.

Bartsch, R. P., Liu, K. K., Ma, Q. D. & Ivanov, P. C. (2014). Three independent forms of cardio-respiratory coupling: Transitions across sleep stages, *Proc. of the 41st Computers in Cardiology Conference*, pp. 1–4.

Basner, M. & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test to sleep loss, *Sleep* **34**(5): 581.

Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., Marcus, C. L., Mehra, R., Parthasarathy, S., Quan, S. F. et al. (2012). Rules for scoring respiratory events in sleep: Update of the 2007 aasm manual for the scoring of sleep and associated events, *Journal of Clinical Sleep Medicine* **8**(5): 597–619.

Better Sleep Council (2008). Poor sleep affecting accuracy and attitude on the web. `http://www.bettersleep.org/OnBetterSleep/sleep_work.asp` [Last accessed: 2014-11-04].

Bixler, E., Vgontzas, A., Ten Have, T., Tyson, K. & Kales, A. (1998). Effects of age on sleep apnea in men: I. prevalence and severity., *American Journal of Respiratory and Critical Care Medicine* **157**(1): 144–148.

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

Bonnet, M. & Arand, D. (1998). Heart rate variability in insomniacs and matched normal sleepers, *Psychosomatic Medicine* **60**(5): 610–615.

Bonnet, M. H. & Arand, D. L. (2003). Situational insomnia: Consistency, predictors, and outcomes, *Sleep* **26**(8): 1029–1036.

Borbély, A. A. & Achermann, P. (1992). Concepts and models of sleep regulation: an overview, *Journal of sleep research* **1**(2): 63–79.

Botev, Z., Grotowski, J., Kroese, D. et al. (2010). Kernel density estimation via diffusion, *The Annals of Statistics* **38**(5): 2916–2957.

Boudewyns, A., Willemen, M., De Cock, W., Verbraecken, J., Coen, E., Wagemans, M., De Backer, W. & Van de Heyning, P. (1997). Does socially disturbing snoring and/or excessive daytime sleepiness warrant polysomnography?, *Clinical Otolaryngology & Allied Sciences* **22**(5): 403–407.

Braunstein, J. R. (1953). *The ballistocardiogram; a dynamic record of the heart beat*, Charles C. Thomas.

Brink, M., Müller, C. H. & Schierz, C. (2006). Contact-free measurement of heart rate, respiration rate, and body movements during sleep, *Behavior research methods* **38**(3): 511–521.

Brooks, C., Coleman, K., Hoffman, B., Kleyntjens, F., Koenig, E., Suckling, E. & Treumann, H. (1956). Sleep and variations in certain functional activities accompanying cyclic changes in depth of sleep., *Journal of applied physiology* **9**(1): 97–104.

Brüser, C., Kerekes, A., Winter, S. & Leonhardt, S. (2012). Multi-channel optical sensor-array for measuring ballistocardiograms and respiratory activity in bed, *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 5042–5045.

Brüser, C., Stadlthanner, K., de Waele, S. & Leonhardt, S. (2011). Adaptive beat-to-beat heart rate estimation in ballistocardiograms, *Information Technology in Biomedicine, IEEE Transactions on* **15**(5): 778–786.

Brüser, C., Winter, S. & Leonhardt, S. (2013). Robust inter-beat interval estimation in cardiac vibration signals, *Physiological measurement* **34**(2): 123.

Bsoul, M., Minn, H. & Tamil, L. (2011). Apnea medassist: real-time sleep apnea monitor using single-lead ecg, *Information Technology in Biomedicine, IEEE Transactions on* **15**(3): 416–427.

Bulckaert, A., Exadaktylos, V., De Bruyne, G., Haex, B., De Valck, E., Wuyts, J., Verbraecken, J. & Berckmans, D. (2010). Heart rate-based nighttime awakening detection, *European journal of applied physiology* **109**(2): 317–322.

Burgess, H. J., Trinder, J. & Kim, Y. (1999). Cardiac autonomic nervous system activity during presleep wakefulness and stage 2 nrem sleep, *Journal of sleep research* **8**(2): 113–122.

Burgess, H. J., Trinder, J., Kim, Y. & Luke, D. (1997). Sleep and circadian influences on cardiac autonomic nervous system activity, *American Journal of Physiology-Heart and Circulatory Physiology* **273**(4): H1761–H1768.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R. & Kupfer, D. J. (1989). The pittsburgh sleep quality index: a new instrument for psychiatric practice and research, *Psychiatry research* **28**(2): 193–213.

Canisius, S., Ploch, T., Gross, V., Jerrentrup, A., Penzel, T. & Kesper, K. (2008). Detection of sleep disordered breathing by automated ecg analysis, *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 2602–2605.

Canisius, S., Ploch, T., Penzel, T., A., J., Krefting, D. & Kesper, K. (2010). Classifying sleep stages using the heart rate spectrum – comparison of feasibility in healthy subjects and sleep apnea patients, *Proceedings of the international biosignal processing conference*, pp. 1–4.

Carlile, J. & Carlile, N. (2008). Repeat study of 149 patients suspected of having sleep apnea but with an ahi < 5, *Sleep*, Vol. 31, American Academy of Sleep Medicine, Westchester, IL 60154 USA, pp. A153–A153.

Caruana, R. & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms, *Proceedings of the 23rd international conference on Machine learning*, ACM, pp. 161–168.

Chang, C.-C. & Lin, C.-J. (2011). Libsvm: a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3): 1–27.

Chesson, A., Littner, M., Davila, D., MacDowell Anderson, W., Grigg-Damberger, M., Hartse, K., Johnson, S. & Wise, M. (1999). Practice parameters for the use of light therapy in the treatment of sleep disorders, *Sleep* **22**: 641–660.

Chesson Jr, A. L., Berry, R. B. & Pack, A. (2003). Practice parameters for the use of portable monitoring devices in the investigation of suspected obstructive sleep apnea in adults., *Sleep* **26**(7): 907–913.

Cristianini, N. & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press.

Curcio, G., Ferrara, M. & De Gennaro, L. (2006). Sleep loss, learning capacity and academic performance, *Sleep medicine reviews* **10**(5): 323–337.

Danker-hopfe, H., Anderer, P., Zeitlhofer, J., Boeck, M., Dorn, H., Gruber, G., Heller, E., Loretz, E., Moser, D., Parapatics, S. et al. (2009). Interrater reliability for sleep scoring according to the rechtschaffen & kales and the new aasm standard, *Journal of sleep research* **18**(1): 74–84.

Daubechies, I. et al. (1992). *Ten lectures on wavelets*, Vol. 61, SIAM.

De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., De Moor, B., Vandewalle, J. & Suykens, J. (2010). Ls-svmlab toolbox user's guide version 1.8, *Internal Report* **10-146**.

de Chazal, P., Penzel, T. & Heneghan, C. (2004). Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram, *Physiological measurement* **25**(4): 967–983.

Devot, S., Bianchi, A. M., Naujokat, E., Mendez, M. O., Brauers, A. & Cerutti, S. (2007). Sleep monitoring through a textile recording system, *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 2560–2563.

Devot, S., Dratwa, R. & Naujokat, E. (2010). Sleep/wake detection based on cardiorespiratory signals and actigraphy, *32nd Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pp. 5089–5092.

Epstein, L., Kristo, D., Strollo, P., Friedman, N., Malhotra, A., Patil, S., Ramar, K., Rogers, R., Schwab, R., Weaver, E. et al. (2009). Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults, *Journal of clinical sleep medicine* **5**(3): 263–276.

Espie, C. A., Kyle, S. D., Williams, C., Ong, J. C., Douglas, N. J., Hames, P. & Brown, J. S. (2012). A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application, *Sleep* **35**(6): 769.

Feinberg, I. (1974). Changes in sleep cycle patterns with age, *Journal of psychiatric research* **10**(3): 283–306.

Friedrich, D., Aubert, X. L., Fuhr, H. & Brauers, A. (2010). Heart rate estimation on a beat-to-beat basis via ballistocardiography-a hybrid approach, *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, IEEE, pp. 4048–4051.

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

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K. & Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals, *Circulation* **101**(23): e215–e220.

Guohua, L., Jianqi, W., Yu, Y. & Xijing, J. (2007). Study of the ballistocardiogram signal in life detection system based on radar, *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 2191–2194.

Haex, B. (2004). *Back and bed: ergonomic aspects of sleeping*, CRC press.

Hall, M., Vasko, R., Buysse, D., Ombao, H., Chen, Q., Cashmere, J. D., Kupfer, D. & Thayer, J. F. (2004). Acute stress affects heart rate variability during sleep, *Psychosomatic medicine* **66**(1): 56–62.

Harper, R. M., Schechtman, V. L. & Kluge, K. A. (1987). Machine classification of infant sleep state using cardiorespiratory measures, *Electroencephalography and clinical neurophysiology* **67**(4): 379–387.

Heise, D., Rosales, L., Sheahen, M., Su, B.-Y. & Skubic, M. (2013). Non-invasive measurement of heartbeat with a hydraulic bed sensor progress, challenges, and opportunities, *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 397–402.

Herscovici, S., Pe'er, A., Papyan, S. & Lavie, P. (2006). Detecting rem sleep from the finger: automatic rem sleep algorithm based on peripheral arterial tone (pat) and actigraphy, *Physiological measurement* **27**: 1–12.

Himanen, S.-L. & Hasan, J. (2000). Limitations of rechtschaffen and kales, *Sleep medicine reviews* **4**(2): 149–167.

Hjorth, B. (1970). Eeg analysis based on time domain properties, *Electroencephalography and clinical neurophysiology* **29**(3): 306–310.

Hobson, J. A., Pace-Schott, E. F. & Stickgold, R. (2000). Dreaming and the brain: toward a cognitive neuroscience of conscious states, *Behavioral and brain sciences* **23**(06): 793–842.

Hoddes, E. (1972). The development and use of the stanford sleepiness scale (sss), *Psychophysiology* **9**: 150–152.

Honeywell (2011). TruStability Board Mount Pressure Sensors: HSC Series – High Accuracy, Honeywell – Sensing and Control.

Inan, O., Etemadi, M., Wiard, R., Giovangrandi, L. & Kovacs, G. (2009). Robust ballistocardiogram acquisition for home monitoring, *Physiological measurement* **30**(2): 169.

ISO 2439 (2008). Flexible cellular polymeric materials – determination of hardness indentation technique.

Jansen, B. H., Larson, B. H. & Shankar, K. (1991). Monitoring of the ballistocardiogram with the static charge sensitive bed, *Biomedical Engineering, IEEE Transactions on* **38**(8): 748–751.

Johns, M. W. et al. (1991). A new method for measuring daytime sleepiness: the epworth sleepiness scale, *sleep* **14**(6): 540–545.

Kishi, A., Yasuda, H., Matsumoto, T., Inami, Y., Horiguchi, J., Tamaki, M., Struzik, Z. R. & Yamamoto, Y. (2011). Nrem sleep stage transitions control ultradian rem sleep rhythm, *Sleep* **34**(10): 1423.

Kortelainen, J. M., Mendez, M. O., Bianchi, A. M., Matteucci, M. & Cerutti, S. (2010). Sleep staging based on signals acquired through bed sensor, *Information Technology in Biomedicine, IEEE Transactions on* **14**(3): 776–785.

Kortelainen, J. M. & Virkkala, J. (2007). Fft averaging of multichannel bcg signals from bed mattress sensor to improve estimation of heart beat interval, *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 6685–6688.

Kräuchi, K., Cajochen, C., Werth, E. & Wirz-Justice, A. (1999). Physiology: Warm feet promote the rapid onset of sleep, *Nature* **401**(6748): 36–37.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data, *biometrics* pp. 159–174.

Le Bon, O., Staner, L., Hoffmann, G., Dramaix, M., San Sebastian, I., Murphy, J. R., Kentos, M., Pelc, I. & Linkowski, P. (2001). The first-night effect may last more than one night, *Journal of psychiatric research* **35**(3): 165–172.

Leger, D. & Poursain, B. (2005). An international survey of insomnia: under-recognition and under-treatment of a polysymptomatic condition, *Current Medical Research and Opinion®* **21**(11): 1785–1792.

Leung, R. S. & Douglas Bradley, T. (2001). Sleep apnea and cardiovascular disease, *American journal of respiratory and critical care medicine* **164**(12): 2147–2165.

Levendowski, D. J., Zack, N., Rao, S., Wong, K., Gendreau, M., Kranzler, J., Zavora, T. & Westbrook, P. R. (2009). Assessment of the test–retest reliability of laboratory polysomnography, *Sleep and Breathing* **13**(2): 163–167.

Li, J., Kapania, R. K. & Spillman, W. B. (2008). Placement optimization of distributed-sensing fiber optic sensors using genetic algorithms, *AIAA journal* **46**(4): 824–836.

Lisenby, M., Richardson, P. & Welch, A. (1976). Detection of cyclic sleep phenomena using instantaneous heart rate, *Electroencephalography and clinical neurophysiology* **40**(2): 169–177.

Littner, M., Kushida, C., Wise, M., Davila, D., Morgenthaler, T., Lee-Chiong, T., Hirshkowitz, M., Daniel, L., Bailey, D., Berry, R. et al. (2005). Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test, *Sleep* **28**(1): 113–121.

Loh, S., Lamond, N., Dorrian, J., Roach, G. & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration, *Behavior Research Methods, Instruments, & Computers* **36**(2): 339–346.

Mack, D. C., Patrie, J. T., Suratt, P. M., Felder, R. A. & Alwan, M. (2009). Development and preliminary validation of heart rate and breathing rate detection using a passive, ballistocardiography-based sleep monitoring system, *Information Technology in Biomedicine, IEEE Transactions on* **13**(1): 111–120.

MacWilliam, J. (1923). Some applications of physiology to medicine: Blood pressure and heart action in sleep and dreams: Their relation to haemorrhages, angina, and sudden death, *British medical journal* **2**(3286): 1196–1200.

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J. & Schwartz, P. J. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use, *European heart journal* **17**(3): 354–381.

McCall, W. V. (2004). Sleep in the elderly: burden, diagnosis, and treatment, *Primary care companion to the Journal of clinical psychiatry* **6**(1): 9–20.

Meerlo, P., Mistlberger, R. E., Jacobs, B. L., Craig Heller, H. & McGinty, D. (2009). New neurons in the adult brain: the role of sleep and consequences of sleep loss, *Sleep medicine reviews* **13**(3): 187–194.

Mendez, M., Corthout, J., Van Huffel, S., Matteucci, M., Penzel, T., Cerutti, S. & Bianchi, A. (2010b). Automatic screening of obstructive sleep apnea from the ecg based on empirical mode decomposition and wavelet analysis, *Physiological measurement* **31**(3): 273.

Mendez, M. O., Matteucci, M., Castronovo, V., Ferini-Strambi, L., Cerutti, S. & Bianchi, A. (2010a). Sleep staging from heart rate variability: time-varying spectral features and hidden markov models, *International Journal of Biomedical Engineering and Technology* **3**(3): 246–263.

Metersky, M. & Castriotta, R. (1996). The effect of polysomnography on sleep position: possible implications on the diagnosis of positional obstructive sleep apnea, *Respiration* **63**(5): 283–287.

Migliorini, M., Bianchi, A. M., Nisticò, D., Kortelainen, J., Arce-Santana, E., Cerutti, S. & Mendez, M. O. (2010). Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors, *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, IEEE, pp. 3273–3276.

Mitler, M. M., Carskadon, M. A., Czeisler, C. A., Dement, W. C., Dinges, D. F. & Graeber, R. C. (1988). Catastrophes, sleep, and public policy: consensus report, *Sleep* **11**(1): 100–109.

Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., Brown, T., Chesson Jr, A., Coleman, J., Lee-Chiong, T. et al. (2007). Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007, *Sleep* **30**(4): 519–529.

Morgenthaler, T., Kramer, M., Alessi, C., Friedman, L., Boehlecke, B., Brown, T., Coleman, J., Kapur, V., Lee-Chiong, T., Owens, J. et al. (2006). Practice parameters for the psychological and behavioral treatment of insomnia: an update, *Sleep* **29**(11): 1415–1419.

Muzet, A. (2007). Environmental noise, sleep and health, *Sleep medicine reviews* **11**(2): 135–142.

Noppe, J., Vanderleyden, L. & Callens, M. (2012). *De sociale staat van Vlaanderen 2011*, Studiedienst Vlaamse Regering.

Nukaya, S., Shino, T., Kurihara, Y., Watanabe, K. & Tanaka, H. (2012). Noninvasive bed sensing of human biosignals via piezoceramic devices sandwiched between the floor and bed, *IEEE Sensors Journal* **12**(3): 431–438.

Ohayon, M. M. & Roth, T. (2002). Prevalence of restless legs syndrome and periodic limb movement disorder in the general population, *Journal of psychosomatic research* **53**(1): 547–554.

Paalasmaa, J. (2010). A respiratory latent variable model for mechanically measured heartbeats, *Physiological measurement* **31**(10): 1331.

Paalasmaa, J., Leppakorpi, L. & Partinen, M. (2011). Quantifying respiratory variation with force sensor measurements, *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 3812–3815.

Paalasmaa, J., Toivonen, H. & Partinen, M. (2014). Adaptive heartbeat modelling for beat-to-beat heart rate measurement in ballistocardiograms, *IEEE journal of biomedical and health informatics* **PP**(99). Epub ahead of print.

Pan, J. & Tompkins, W. J. (1985). A real-time qrs detection algorithm, *Biomedical Engineering, IEEE Transactions on* **32**(3): 230–236.

Paquet, J., Kawinska, A. & Carrier, J. (2007). Wake detection capacity of actigraphy during sleep, *Sleep* **30**(10): 1362.

Peltokangas, M., Verho, J. & Vehkaoja, A. (2012). Night-time ekg and hrv monitoring with bed sheet integrated textile electrodes, *Information Technology in Biomedicine, IEEE Transactions on* **16**(5): 935–942.

Peng, H., Long, F. & Ding, C. (2005). 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.

Penzel, T. & Conradt, R. (2000b). Computer based sleep recording and analysis, *Sleep medicine reviews* **4**(2): 131–148.

Penzel, T., McNames, J., De Chazal, P., Raymond, B., Murray, A. & Moody, G. (2002). Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings, *Medical and Biological Engineering and Computing* **40**(4): 402–407.

Penzel, T., Moody, G., Mark, R., Goldberger, A. & Peter, J. (2000a). The apnea-ecg database, *Computers in Cardiology 2000*, IEEE, pp. 255–258.

Peppard, P. E., Young, T., Barnet, J. H., Palta, M., Hagen, E. W. & Hla, K. M. (2013). Increased prevalence of sleep-disordered breathing in adults, *American journal of epidemiology* **177**(9): 1006–1014.

Piéron, H. (1913). *Le problème physiologique du sommeil*, Masson et Cie.

Pudil, P., Novovičová, J. & Kittler, J. (1994). Floating search methods in feature selection, *Pattern recognition letters* **15**(11): 1119–1125.

PulseSensor Amped (2011). An open-source heart–rate sensor for Arduino. `http://pulsesensor.myshopify.com/` [Last accessed at: 2014-11-08].

Quan, S., Gillin, J. C., Littner, M. & Shepard, J. (1999). Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. editorials, *Sleep* **22**(5): 667–689.

Raymann, R. J., Swaab, D. F. & Van Someren, E. J. (2008). Skin deep: enhanced sleep depth by cutaneous temperature manipulation, *Brain* **131**(2): 500–513.

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

Redmond, S. J., de Chazal, P., O'Brien, C., Ryan, S., McNicholas, W. T. & Heneghan, C. (2007). Sleep staging using cardiorespiratory signals, *Somnologie-Schlafforschung und Schlafmedizin* **11**(4): 245–256.

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

Ripley, B. D. (1996). *Pattern recognition and neural networks*, Cambridge university press.

Rosales, L., Skubic, M., Heise, D., Devaney, M. J. & Schaumburg, M. (2012). Heartbeat detection from a hydraulic bed sensor using a clustering approach, *34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 2383–2387.

Roth, T. (2005). Prevalence, associated risks, and treatment patterns of insomnia, *The Journal of clinical psychiatry. Supplement* **66**(9): 10–13.

Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: an update, *Sleep medicine reviews* **15**(4): 259–267.

Saper, C. B., Chou, T. C. & Scammell, T. E. (2001). The sleep switch: hypothalamic control of sleep and wakefulness, *Trends in neurosciences* **24**(12): 726–731.

Schulz, H. (2008). Rethinking sleep analysis, *J Clin Sleep Med* **4**(2): 99–103.

Schutte-Rodin, S., Broch, L., Buysse, D., Dorsey, C. & Sateia, M. (2008). Clinical guideline for the evaluation and management of chronic insomnia in adults, *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine* **4**(5): 487–504.

Schwaibold, M., Harms, R., Schöller, B., Pinnow, I., Cassel, W., Penzel, T., F Becker, H. & Bolz, A. (2003). Knowledge-based automatic sleep-stage recognition–reduction in the interpretation variability, *Somnologie* **7**(2): 59–65.

Shin, J. H., Chee, Y. J., Jeong, D.-U. & Park, K. S. (2010). Nonconstrained sleep monitoring system and algorithms using air-mattress with balancing tube method, *Information Technology in Biomedicine, IEEE Transactions on* **14**(1): 147–156.

Snyder, F., Hobson, J. A., Morrison, D. F. & Goldfrank, F. (1964). Changes in respiration, heart rate, and systolic blood pressure in human sleep, *Journal of Applied Physiology* **19**(3): 417–422.

Somers, V. K., Dyken, M. E., Clary, M. P. & Abboud, F. M. (1995). Sympathetic neural mechanisms in obstructive sleep apnea., *Journal of Clinical Investigation* **96**(4): 1897–1904.

Somers, V. K., Dyken, M. E., Mark, A. L. & Abboud, F. M. (1993). Sympathetic-nerve activity during sleep in normal subjects, *New England Journal of Medicine* **328**(5): 303–307.

Somnio (2010). Helpt om uw slapeloosheid onder controle te krijgen. `http://www.somnio.eu` [Last accessed at: 2014-11-09].

Spencer, J., Moran, D., Lee, A. & Talbert, D. (1990). White noise and sleep induction., *Archives of disease in childhood* **65**(1): 135–137.

Sprager, S. & Zazula, D. (2012). Heartbeat and respiration detection from optical interferometric signals by using a multimethod approach, *Biomedical Engineering, IEEE Transactions on* **59**(10): 2922–2929.

St. Vincent's University Hospital/University College Dublin (2008). Sleep Apnea Database. `http://www.physionet.org/pn3/ucddb/` [Last accessed: 2014-11-07].

Stanchina, M. L., Abu-Hijleh, M., Chaudhry, B. K., Carlisle, C. C. & Millman, R. P. (2005). The influence of white noise on sleep in subjects exposed to icu noise, *Sleep medicine* **6**(5): 423–428.

Starr, I. & Friedland, C. K. (1946). On the cause of the respiratory variation of the ballistocardiogram, with a note on sinus arrhythmia, *Journal of Clinical Investigation* **25**(1): 53.

Starr, I., Rawson, A., Schroeder, H. & Joseph, N. (1939). Studies on the estimation of cardiac ouptut in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts; the ballistocardiogram, *American Journal of Physiology–Legacy Content* **127**(1): 1–28.

Stein, P. K. & Pu, Y. (2012). Heart rate variability, sleep and sleep disorders, *Sleep medicine reviews* **16**(1): 47–66.

Suykens, J. A., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., Suykens, J. & Van Gestel, T. (2002). *Least squares support vector machines*, Vol. 4, World Scientific.

Tabachnik, E., Muller, N., Bryan, C. & Levison, H. (1981). Changes in ventilation and chest wall mechanics during sleep in normal adolescents, *J Appl Physiol* **51**(3): 557–564.

The National Sleep Foundation (2012). What makes a good night's sleep. `http://www.sleepfoundation.org/healthy-sleep-tips` [Last accessed at: 2014-11-04].

van de Water, A., Holmes, A. & Hurley, D. A. (2011). Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography–a systematic review, *Journal of Sleep Research* **20**(1pt2): 183–200.

Van Deun, D., Verhaert, V., Willemen, T., Wuyts, J., Verbraecken, J., Exadaktylos, V., Haex, B. & Vander Sloten, J. (2012). Biomechanics-based active control of bedding support properties and its influence on sleep, *Work: A Journal of Prevention, Assessment and Rehabilitation* **41**: 1274–1280.

Van Wormer, J. J., Linde, J. A., Harnack, L. J., Stovitz, S. D. & Jeffery, R. W. (2012). Self-weighing frequency is associated with weight gain prevention over 2 years among working adults, *International journal of behavioral medicine* **19**(3): 351–358.

Verboven, S. & Hubert, M. (2005). Libra: a matlab library for robust analysis, *Chemometrics and intelligent laboratory systems* **75**(2): 127–136.

Verhaert, V., Haex, B., De Wilde, T., Berckmans, D., Vandekerckhove, M., Verbraecken, J. & Sloten, J. V. (2011a). Unobtrusive assessment of motor patterns during sleep based on mattress indentation measurements, *Information Technology in Biomedicine, IEEE Transactions on* **15**(5): 787–794.

Verhaert, V., Haex, B., De Wilde, T., Cluydts, R., Berckmans, D., Verbraecken, J. & Vander Sloten, J. (2009). Automatic posture recognition during sleep, *Proceedings of the 17th world congress on ergonomics*, pp. 9–14.

Verhaert, V., Haex, B., Wilde, T. D., Berckmans, D., Verbraecken, J., Valck, E. d. & Sloten, J. V. (2011b). Ergonomics in bed design: the effect of spinal alignment on sleep parameters, *Ergonomics* **54**(2): 169–178.

Wang, C., Wang, X., Xiong, X. & Wang, R. (2013). The segmented dynamic time warping algorithm for beat-to-beat heart rate estimation based on ballistocardiogram signals, *Journal of Fiber Bioengineering and Informatics* **6**(4): 415–425.

Watanabe, K., Watanabe, T., Watanabe, H., Ando, H., Ishikawa, T. & Kobayashi, K. (2005). Noninvasive measurement of heartbeat, respiration, snoring and body movements of a subject in bed via a pneumatic method, *Biomedical Engineering, IEEE Transactions on* **52**(12): 2100–2107.

Welch, A. J. & Richardson, P. C. (1973). Computer sleep stage classification using heart rate data, *Electroencephalography and clinical neurophysiology* **34**(2): 145–152.

Whitney, C. W., Gottlieb, D. J., Redline, S., Norman, R. G., Dodge, R. R., Shahar, E., Surovec, S. & Nieto, F. J. (1998). Reliability of scoring respiratory disturbance indices and sleep staging., *Sleep* **21**(7): 749–757.

Widjaja, D., Varon, C., Dorado, A. C., Suykens, J. A. & Van Huffel, S. (2012). Application of kernel principal component analysis for single-lead-ecg-derived respiration, *Biomedical Engineering, IEEE Transactions on* **59**(4): 1169–1176.

Willemen, T., Van Deun, D., Verhaert, V., Haex, B., Vander Sloten, J. & Van Huffel, S. (2014d). A pan-tompkins based algorithm for heart beat detection in the ballistocardiogram, *IEEE journal of biomedical and health informatics* . Submitted for review.

Willemen, T., Van Deun, D., Verhaert, V., Pirrera, S., Exadaktylos, V., Verbraecken, J., Haex, B. & Sloten, J. V. (2012). Automatic sleep stage classification based on easy to register signals as a validation tool for ergonomic steering in smart bedding systems, *Work: A Journal of Prevention, Assessment and Rehabilitation* **41**: 1985–1989.

Willemen, T., Van Deun, D., Verhaert, V., Van Huffel, S., Haex, B. & Vander Sloten, J. (2014b). Characterization of the respiratory and heart beat signal from an air pressure-based ballistocardiographic setup, *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 6298–6301.

Willemen, T., Van Deun, D., Verhaert, V., Vandekerckhove, M., Exadaktylos, V., Verbraecken, J., Huffel, S., Haex, B. & Vander Sloten, J. (2014a). An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification, *IEEE journal of biomedical and health informatics* **18**(2): 661–669.

Willemen, T., Varon, C., Caicedo, A. D., Haex, B., Vander Sloten, J. & Van Huffel, S. (2014e). Probabilistic cardiac and respiratory based classification of sleep and apneic events in subjects with sleep apnea, *Physiological measurement* . Submitted for review.

Willemen, T., Varon, C., Haex, B., Vander Sloten, J. & Van Huffel, S. (2014c). Assessment of different methodologies to include temporal information in classifying episodes of sleep apnea based on single-lead electrocardiogram, *Proc. of the 41st Computers in Cardiology Conference*, pp. 1–4.

Xie, B. & Minn, H. (2012). Real-time sleep apnea detection by classifier combination, *Information Technology in Biomedicine, IEEE Transactions on* **16**(3): 469–477.

Young, T., Palta, M., Dempsey, J., Skatrud, J., Weber, S. & Badr, S. (1993). The occurrence of sleep-disordered breathing among middle-aged adults, *New England Journal of Medicine* **328**(17): 1230–1235.

Zhu, X., Chen, W., Nemoto, T., Kanemitsu, Y., Kitamura, K.-i., Yamakoshi, K.-i. & Wei, D. (2006). Real-time monitoring of respiration rhythm and pulse rate during sleep, *Biomedical Engineering, IEEE Transactions on* **53**(12): 2553–2563.

Zwillich, C., Devlin, T., White, D., Douglas, N., Weil, J. & Martin, R. (1982). Bradycardia during sleep apnea. characteristics and mechanism., *Journal of Clinical Investigation* **69**(6): 1286.

Curriculum vitae

Personalia

Name:	Tim Willemen
Date of birth:	January 8, 1987
Place of birth:	Wilrijk, Belgium

Contact information

Biomechanics section
Department of Mechanical Engineering, KU Leuven
Celestijnenlaan 300C, B-3001 Heverlee, Belgium

STADIUS
Department of Electrical Engineering, KU Leuven
Arenbergpark 10, B-3001 Heverlee, Belgium

Tel: +32 16 37 27 99
Email: tim.willemen@kuleuven.be

Work experience

2010 - Present

PhD student, KU Leuven, Belgium
Keywords: Sleep (disorders), Biomedical Signal Processing, Machine Learning, Unobtrusive Monitoring, MATLAB
Projects: iMinds-ICON NXT_SLEEP (functioning as project coördinator;

partners: KUL, VUB, UZA, NXP, IMEC, Custom8, Sensotiss, Fifthplay), IWT-SBO EASI-SLEEP (partners: KUL, VUB, UZA)

2009 - 2010

Summer Jobs, Antwerp University Hospital, Belgium
Keywords: Breathing (disorders), Data Processing, Java
Projects: NAVA, Inspire

2003 - 2008

Various Summer Jobs, Antwerp University Hospital & RVT Christine, Belgium
Keywords: Internal Transport, Global Maintenance & Prevention, Cleaning Service

Education

2008 - 2010

Erasmus Mundus Master in Nanotechnology and Nanoscience, Magna cum laude
KU Leuven, Belgium (1st year) & Chalmers University of Technology, Sweden (2nd year)
Projects: Wireless Body Area Networks, Clean Room Fabrication, Superconductivity, Quantum Bits

2005 - 2008

Bachelor with major in Electrical Engineering, minor in Material Sciences, Magna cum laude
KU Leuven, Belgium

Publications

Articles in internationally reviewed academic journals

Willemen, T., Varon, C., Caicedo, A. D., Haex, B., Vander Sloten, J., Van Huffel, S. (2014) Probabilistic cardiac and respiratory based classification of sleep and apneic events in subjects with sleep apnea. *Physiological measurement*, Submitted for review.

Willemen, T., Van Deun, D., Verhaert, V., Haex, B., Vander Sloten, J., Van Huffel, S. (2014). A pan-tompkins based algorithm for heart beat detection in the ballistocardiogram. *IEEE Journal of Biomedical and Health Informatics*, Submitted for review.

Maes, J., Verbraecken, J., Willemen, M., De Volder, I., van Gastel, A., Michiels, N., Verbeek, I., Vandekerckhove, M., Wuyts, J., Haex, B., Willemen, T., Exadaktylos, V., Bulckaert, A., Cluydts, R. (2014). Sleep misperception, EEG characteristics and Autonomic Nervous System activity in primary insomnia: A retrospective study on polysomnographic data. *International Journal of Psychophysiology*, 91 (3), 163-171.

Willemen, T., Van Deun, D., Verhaert, V., Vandekerckhove, M., Exadaktylos, V., Verbraecken, J., Van Huffel, S., Haex, B., Vander Sloten, J. (2013). An evaluation of cardio-respiratory and movement features with respect to sleep stage classification. *IEEE Journal of Biomedical and Health Informatic*s, 18 (2), 661-669.

Van Deun, D., Verhaert, V., Willemen, T., Wuyts, J., Verbraecken, J., Exadaktylos, V., Haex, B., Vander Sloten, J. (2012). Biomechanics-based active control of bedding support properties and its influence on sleep. *Work*, 41 (Suppl. 1), 1274-1280.

Willemen, T., Van Deun, D., Verhaert, V., Pirrera, S., Exadaktylos, V., Verbraecken, J., Haex, B., Vander Sloten, J. (2012). Automatic sleep stage classification based on easy to register signals as a validation tool for ergonomic steering in smart bedding systems. *Work*, 41 (Suppl. 1), 1985-1989.

Articles in other academic journals

Van Deun, D., Verhaert, V., Willemen, T., Haex, B., Vander Sloten, J. (2012). Uw bed beïnvloedt uw slaap. *Tijdschrift voor Ergonomie*, 37 (4), 5-8.

Articles in academic books

Van Deun, D., Willemen, T., Verhaert, V., Van Huffel, S., Haex, B., Vander Sloten, J. (2014). Ambient intelligence in the bedroom. In: Curran K. (Eds.), *Recent advances in ambient intelligence and context-aware computing.* Hershey, Pennsylvania US: IGI Global.

Papers at international scientific conferences

Willemen, T., Van Deun, D., Verhaert, V., Van Huffel, S., Haex, B., Vander Sloten, J. (2014). Characterization of the respiratory and heart beat signal from an air pressure-based ballistocardiographic setup. *Proc. of the 36th Annual International Conference of the IEEE EMBS*. IEEE EMBS 2014. Chicago, Illinois USA), Aug. 2014 (pp. 6298-6301).

Willemen, T., Varon, C., Haex, B., Vander Sloten, J., Van Huffel, S. (2014). Assessment of Different Methodologies to Include Temporal Information in Classifying Episodes of Sleep Apnea Based on Single-Lead Electrocardiogram. *Proc. of the 41st Computers in Cardiology Conference*. CinC2014. Cambridge, USA, Sep. 2014 (pp. 1-4).

Abstracts presented at international scientific conferences

Willemen, T., Haex, B., Vander Sloten, J., Van Huffel, S. (2015). Biomechanics based analysis of sleep. *Proceedings of the 5th Dutch Bio-Medical Engineering Conference*. Dutch Bio-Medical Engineering Conference. Egmond aan Zee, 22-23 January 2015.

Willemen, T., Van Deun, D., Verhaert, V., Wuyts, J., Exadaktylos, V., Verbraecken, J., Van Huffel, S., Haex, B., Vander Sloten, J. (2012). A user-friendly approach to sleep stage classification using heart rate, breathing rate and movement. *Proceedings of the European Sleep Research Society*. Congress of the European Sleep Research Society. Paris, September 2012.

Abstracts presented at other scientific conferences

Willemen, T., Haex, B., Vander Sloten, J., Van Huffel, S. (2014). Biomechanics based analysis of sleep. *Proceedings of the Contact Forum 13th Belgian Day on Biomedical Engineering*. Belgian Day on Biomedical Engineering. Brussels, 28 November 2014.

Willemen, T., Van Deun, D., Verhaert, V., Van Huffel, S., Haex, B., Vander Sloten, J. (2013). Off-body measurement of respiration and heart rate for use in smart bedding systems. *Proceedings of the Contact Forum 12th Belgian Day on Biomedical Engineering*. Belgian Day on Biomedical Engineering. Brussels, 06 December 2013.

Willemen, T., Van Deun, D., Verhaert, V., Exadaktylos, V., Vandekerckhove, M., Verbraecken, J., Van Huffel, S., Haex, B., Vander Sloten, J. (2012). Heart

rate-, breathing rate- and movement-based sleep classification. *Proceedings of the Contact Forum 11th Belgian Day on Biomedical Engineering.* Belgian Day on Biomedical Engineering. Brussels, 07 December 2012.

Willemen, T., Van Deun, D., Verhaert, V., Exadaktylos, V., Vandekerckhove, M., Verbraecken, J., Van Huffel, S., Haex, B., Vander Sloten, J. (2012). Heart rate-, breathing rate- and movement-based sleep classification. Autumn Meeting Belgian Association for Sleep Research and Sleep Medicine. Brussels, 05 October 2012.

Van Deun, D., Verhaert, V., Willemen, T., Haex, B., Vander Sloten, J. (2012). Generation of customized human avatars using contour information. *Proceedings of the Contact Forum 11th Belgian Day on Biomedical Engineering.* Belgian Day on Biomedical Engineering. Brussels, 07 December 2012.

Van Deun, D., Verhaert, V., Willemen, T., Wuyts, J., Verbraecken, J., Exadaktylos, V., Haex, B., Vander Sloten, J. (2011). The effect of intelligent bedding support on sleep. *Proceedings of the Contact Forum 10th Belgian Day on Biomedical Engineering.* Tenth Belgian Day on Biomedical Engineering. Brussels, 2 December 2011.

Willemen, T., Van Deun, D., Verhaert, V., Pirrera, S., Exadaktylos, V., Verbraecken, J., Haex, B., Vander Sloten, J. (2011). Automatic sleep stage classification based on easy to register signals. *Proceedings of the Contact Forum 10th Belgian Day on Biomedical Engineering.* Tenth Belgian Day on Biomedical Engineering. Brussels, 2 December 2011.

Willemen, T., Van Deun, D., Verhaert, V., Pirrera, S., Exadaktylos, V., Verbraecken, J., Haex, B., Vander Sloten, J. (2011). Prospects of heart rate-, breathing rate- and movement-based sleep classification. Autumn Meeting Belgian Association for Sleep Research and Sleep Medicine. Gent, 28 October 2011.

Willemen, T., Haex, B., Vander Sloten, J. (2011). Investigation of sleep - past and future. Autumn Meeting Belgian Society of Sleep Technologists. Gent, 28 October 2011.

Van Deun, D., Verhaert, V., Willemen, T., Wuyts, J., Verbraecken, J., Exadaktylos, V., Haex, B., Vander Sloten, J. (2011). Detailed monitoring of movement in bed: current challenges - future perspectives. Autumn meeting Belgian Association for Sleep Research and Sleep Medicine. Gent, 28 October 2011.



FACULTY OF ENGINEERING SCIENCE
DEPARTMENTS OF MECHANICAL & ELECTRICAL ENGINEERING
BIOMECHANICS SECTION - STADIUS
Celestijnenlaan 200C box 2419 - Kasteelpark Arenberg 10 box 2446
B-3001 Heverlee
tim.willemen@kuleuven.be
http://www.mech.kuleuven.be/en/bme - http://www.esat.kuleuven.be/stadius/