

NATIONAL UNIVERSITY OF IRELAND, MAYNOOTH



NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

# Non-Contact Sleep Monitoring

Lorcan Walsh

A thesis submitted in partial fulfillment for the degree of

Doctor of Philosophy

in the

Faculty of Science and Engineering

Electronic Engineering Department

Supervisor and Head of Department: Dr. Seán McLoone

November 30, 2012

# Declaration of Authorship

I, Lorcan Walsh, declare that this thesis titled, ‘Non-Contact Sleep Monitoring’ and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

---

Date:

---

*“The real voyage of discovery consists not in seeking new landscapes but in having new eyes.”*

Marcel Proust

# *Abstract*

“The road ahead for preventive medicine seems clear. It is the delivery of high quality, personalised (as opposed to depersonalised) comprehensive medical care to all.”  
Burney, Steiger, and Georges (1964)

This world’s population is ageing, and this is set to intensify over the next forty years. This demographic shift will result in significant economic and societal burdens (particularly on healthcare systems). The instantiation of a proactive, preventative approach to delivering healthcare is long recognised, yet is still proving challenging. Recent work has focussed on enabling older adults to *age in place* in their own homes. This may be realised through the recent technological advancements of affordable healthcare sensors and systems which continuously support *independent living*, particularly through longitudinally monitoring deviations in behavioural and health metrics. Overall health status is contingent on multiple factors including, but not limited to, physical health, mental health, and social and emotional wellbeing; sleep is implicitly linked to each of these factors.

This thesis focusses on the investigation and development of an unobtrusive sleep monitoring system, particularly suited towards long-term placement in the homes of older adults. The Under Mattress Bed Sensor (UMBS) is an unobtrusive, pressure sensing grid designed to infer bed times and bed exits, and also for the detection of development of bedsores. This work extends the capacity of this sensor. Specifically, the novel contributions contained within this thesis focus on an in-depth review of the state-of-the-art advances in sleep monitoring, and the development and validation of algorithms which extract and quantify UMBS-derived sleep metrics.

Preliminary experimental and community deployments investigated the suitability of the sensor for long-term monitoring. Rigorous experimental development refined algorithms which extract respiration rate as well as motion metrics which outperform traditional forms of ambulatory sleep monitoring. Spatial, temporal, statistical and spatiotemporal features were derived from UMBS data as a means of describing movement during sleep. These features were compared across experimental, domestic and clinical data sets, and across multiple sleeping episodes. Lastly, the optimal classifier (built using a combination of the UMBS-derived features) was shown to infer sleep/wake state accurately and reliably across both younger and older cohorts.

Through long-term deployment, it is envisaged that the UMBS-derived features (including spatial, temporal, statistical and spatiotemporal features, respiration rate, and

sleep/wake state) may be used to provide unobtrusive, continuous insights into overall health status, the progression of the symptoms of chronic conditions, and allow the objective measurement of daily (sleep/wake) patterns and routines.

# *Acknowledgements*

I would like to express my sincerest gratitude to my PhD supervisor Dr. Seán McLoone for his constant guidance, and unbounded assistance. His in-depth, and sometimes mistifying, extensive knowledge of mathematical and engineering principles are only matched by his rich explanations of their intricacies and their application. Particular thanks are due to Prof. Charles Czeisler and Dr. Jeanne Duffy of the Division of Sleep Medicine, Harvard Medical School, whose open, inclusive nature of driving knowledge forward, particularly in sleep research, is an inspiration.

This work would not have been possible without support from the Irish Research Council for Science, Engineering and Technology, the Health Research and Innovation Group at Intel Ireland (including Michael McGrath, Terry O'Shea, Barry Greene, Julie Behan, Adrian Burns, Cliodhna Case, and Emer Doheny), the Technology Research for Independent Living Centre, and the Health Informatics Society of Ireland. I am particularly grateful to the FÁS Science Challenge Program and the Fulbright Commission whose support allowed me to create and maintain a collaboration with the Division of Sleep Medicine at Harvard Medical School. Furthermore I would like to thank the staff and postgrads of the Department of Electronic Engineering at N.U.I. Maynooth.

Most importantly, I would like to thank my family, including my parents, my siblings, Aidan, Ciara, Eilish, Gareth, Martin and my extended family, and my friends, including Aileen, Amy, Avril, Catherine, Ciarán, David, John, Kevin, Leslie, Mick, Niall, Paul, Shane, Tom, and especially Chris Gordon. Were it not for their support and encouragement, I would have lost any and all conviction I possessed.

# Contents

<b>Declaration of Authorship</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background and Motivation . . . . .	1
1.2 Sleep . . . . .	4
1.3 Aims and Scope of this Thesis . . . . .	6
1.4 Contributions of this Thesis . . . . .	7
1.5 List of Publications . . . . .	9
1.6 Thesis Layout . . . . .	10
<b>2 An Introduction to Sleep</b>	<b>12</b>
2.1 The Definition of Sleep . . . . .	12
2.2 Sleep Scoring . . . . .	13
2.2.1 Wake . . . . .	15
2.2.2 Rapid Eye Movement (REM) Sleep . . . . .	16
2.2.3 Non-Rapid Eye Movement (NREM) Sleep . . . . .	16
2.3 Other Physiological Changes During Sleep . . . . .	18
2.4 Commonly Used Sleep Metrics . . . . .	18
2.5 The Architecture of Sleep and Sleep Stages . . . . .	20
2.6 Age and Illness Related Changes in Sleep . . . . .	20
2.6.1 Sleep and Age . . . . .	20
2.6.1.1 Sleep and Older Adults . . . . .	21
2.6.2 Sleep Degradation from Sleep Disorders, Diseases and Illnesses . . . . .	22
2.6.2.1 Apnoea/Hypopnoea . . . . .	22
2.6.2.2 Excessive Daytime Sleepiness . . . . .	23
2.6.2.3 REM Sleep Behaviour Disorder . . . . .	24
2.6.2.4 Restless Leg Syndrome . . . . .	24
2.6.2.5 Periodic Limb Movement Disorder . . . . .	25
2.6.2.6 Nocturia . . . . .	25

2.6.2.7	Circadian Rhythm Disorders . . . . .	25
2.6.2.8	The Effect of Parkinson’s Disease on Sleep . . . . .	26
2.6.2.9	The Effect of Alzheimer’s Disease on Sleep . . . . .	27
2.6.2.10	Insomnia . . . . .	27
2.7	Conclusion . . . . .	28
<b>3</b>	<b>Classification Methods</b>	<b>29</b>
3.1	Data Sets . . . . .	29
3.2	Performance Metrics . . . . .	31
3.3	Classifiers . . . . .	33
3.3.1	Training, Validation, Testing Data and Cross Validation . . . . .	33
3.3.2	Binary Classification Test . . . . .	34
3.3.3	Discriminant Analysis . . . . .	36
3.3.3.1	Mathematical Derivation . . . . .	37
3.3.3.2	Application of LDA and QDA . . . . .	39
3.3.3.3	Discussion . . . . .	39
3.3.4	<i>k</i> -Nearest Neighbour . . . . .	40
3.3.4.1	Discussion . . . . .	41
3.3.5	Artificial Neural Networks . . . . .	41
3.3.5.1	The Mathematical Representation of a Neuron . . . . .	41
3.3.5.2	Multilayer Perceptrons . . . . .	42
3.3.5.3	Neural Networks: Learning and Storing Information . . . . .	43
3.3.5.4	Discussion . . . . .	46
3.3.6	Support Vector Machines . . . . .	48
3.3.7	Multiclass Classification . . . . .	53
3.3.8	Application and Performance of the Classifiers . . . . .	54
3.4	Conclusion . . . . .	55
<b>4</b>	<b>A Review of Sleep Measurement Technologies</b>	<b>57</b>
4.1	Polysomnography . . . . .	57
4.1.1	Polysomnography Signals . . . . .	58
4.1.2	Types of Polysomnography . . . . .	60
4.1.2.1	Clinical PSG . . . . .	60
4.1.2.2	Unattended PSG . . . . .	61
4.1.2.3	Ambulatory PSG . . . . .	61
4.1.3	PSG in Children . . . . .	61
4.1.4	Sleep Scoring . . . . .	62
4.2	Sleep Diaries . . . . .	62
4.3	Sleep Tests and Subjective Measures . . . . .	62
4.3.1	Common Sleep Tests and Measures . . . . .	64
4.4	Actigraphy . . . . .	66
4.4.1	Actigraphy Data Preprocessing . . . . .	66
4.4.2	Sleep/Wake Classification and Scoring Procedures for Humans . . . . .	67
4.4.3	Clinical Performance . . . . .	68
4.4.3.1	The Validity of Wrist Actigraphy Estimates . . . . .	69
4.4.4	Actigraphy - Deployment in Practice . . . . .	69
4.4.4.1	Actigraphy Device Placement Issues . . . . .	70

4.4.4.2	Circadian Rhythm Analysis . . . . .	70
4.4.4.3	Limitations of Actigraphic Monitoring . . . . .	70
4.4.4.4	Advances in Wrist-based Actigraphic Monitoring . . . . .	71
4.5	Advances in Sleep Monitoring . . . . .	72
4.5.1	Automated PSG Sleep Analysis . . . . .	72
4.5.1.1	Partial PSG Automated Sleep Scoring . . . . .	72
4.5.1.2	Full Complement Automated Sleep Scoring . . . . .	78
4.5.2	Contact-Based Sleep Monitoring . . . . .	81
4.5.2.1	Electrode-based Advances . . . . .	81
4.5.2.2	Embedded and Textile Electrodes . . . . .	82
4.5.2.3	Ambulatory Physiological Monitoring . . . . .	83
4.5.3	Non-Contact Sleep Monitoring . . . . .	84
4.5.3.1	Optical Monitoring . . . . .	84
4.5.3.2	Smart Phone Application - Sleep Cycle Alarm Clock- /Sleep On It . . . . .	85
4.5.3.3	Bed Position Monitoring (Bed Sores) . . . . .	86
4.5.4	Non-Contact Vital Signs Monitoring . . . . .	87
4.5.4.1	Ballistocardiography . . . . .	87
4.5.4.2	Radio Frequency-Based Solutions: . . . . .	96
4.5.5	Multi-Modal Solutions: . . . . .	97
4.5.6	Sleep Apnoea Monitoring . . . . .	98
4.5.7	Brain Imaging and Sleep Monitoring . . . . .	102
4.6	Conclusions . . . . .	102
<b>5</b>	<b>The Non-Obtrusive Non-Contact Sleep Sensor</b>	<b>105</b>
5.1	Under-Mattress Bed Sensor . . . . .	105
5.1.1	Tactels . . . . .	107
5.1.2	UMBS Sensor Types . . . . .	107
5.1.3	Communications Protocol . . . . .	107
5.1.4	Data Collection System . . . . .	108
5.1.5	Sampling Rate . . . . .	108
5.1.6	Pressure Measurement . . . . .	110
5.1.7	UMBS Notation . . . . .	110
5.2	Initial UMBS Deployment . . . . .	111
5.2.1	Example Data . . . . .	111
5.2.2	Entire Night's Data . . . . .	111
5.2.3	Turns Data Set . . . . .	113
5.3	UMBS Community Deployment in Older Adults . . . . .	115
5.3.1	Methods . . . . .	116
5.3.1.1	Detection of In-Bed Presence and Calculation of Bed Restlessness . . . . .	117
5.3.2	Nocturnal Routines and Bed Restlessness . . . . .	118
5.3.2.1	Methods . . . . .	119
5.3.2.2	Results . . . . .	119
5.3.2.3	Discussion . . . . .	119
5.3.3	UMBS vs Wrist Actigraphy . . . . .	121
5.3.3.1	Methods . . . . .	121

5.3.3.2	Results . . . . .	122
5.3.3.3	Discussion . . . . .	123
5.3.4	UMBS and Daily Activity . . . . .	124
5.3.4.1	Methods . . . . .	124
5.3.4.2	Results . . . . .	124
5.3.4.3	Discussion . . . . .	124
5.3.5	Overall Discussion . . . . .	125
5.4	Parallel UMBS Research . . . . .	125
5.5	Conclusions . . . . .	127
<b>6</b>	<b>Experimental Validation of the UMBS</b>	<b>129</b>
6.1	Experiment 1 - Detection of Physiological Signals . . . . .	130
6.1.1	Methods . . . . .	130
6.1.2	Results . . . . .	131
6.1.3	Discussion . . . . .	133
6.2	Experiment 2 - Automated Respiration and Heart Rate Estimation . . . . .	136
6.2.1	Methods . . . . .	136
6.2.2	Data Preparation . . . . .	139
6.2.2.1	UMBS Data Preparation . . . . .	139
6.2.3	Automated Physiological Signal Estimation . . . . .	143
6.2.3.1	Time-based Techniques . . . . .	143
6.2.3.2	Frequency Based Analysis . . . . .	146
6.2.4	Results . . . . .	147
6.2.4.1	Respiration Estimation . . . . .	147
6.2.4.2	Heart Rate Estimation . . . . .	147
6.2.5	Discussion . . . . .	149
6.3	Experiment 3 - Movement Detection Capacity . . . . .	153
6.3.1	Data Collection and Processing . . . . .	153
6.3.1.1	UMBS . . . . .	153
6.3.1.2	Video . . . . .	154
6.3.1.3	PIR SHIMMER . . . . .	155
6.3.1.4	Wrist Actigraphy . . . . .	156
6.3.2	Results . . . . .	156
6.3.2.1	Movement Binary Classification Test . . . . .	156
6.3.2.2	Movement Classification Using Multiple Inputs . . . . .	159
6.3.2.3	Alternative Technologies . . . . .	160
6.3.3	Discussion . . . . .	161
6.4	Summary . . . . .	161
<b>7</b>	<b>UMBS Deployment in Clinical and Domestic Environments</b>	<b>164</b>
7.1	Introduction . . . . .	165
7.2	UMBS Feature Extraction Algorithms . . . . .	166
7.2.1	Temporal Movement Features . . . . .	167
7.2.2	Spatial Movement Features . . . . .	170
7.2.3	Statistical Movement Features . . . . .	175
7.2.4	Statiotemporal Movement Features . . . . .	175
7.3	Methods . . . . .	177

7.3.1	Summerhill Data Set . . . . .	177
7.3.2	Maynooth Data Set . . . . .	178
7.3.3	Peamount Data Set . . . . .	179
7.3.4	Boston Data Set . . . . .	179
7.3.5	Pre-Processing . . . . .	179
7.4	Results . . . . .	180
7.4.1	Distribution Plots . . . . .	180
7.4.2	Comparison Across All Subjects and Cohorts . . . . .	180
7.4.2.1	Temporal, Spatial and Statistical Features . . . . .	180
7.4.2.2	Spatiotemporal Features . . . . .	181
7.4.3	Variance Over Multiple Days . . . . .	181
7.4.4	Cohort Classification of Spatiotemporal Movement Features . . . . .	186
7.5	Discussion . . . . .	188
7.5.1	Movement Features Across All Subjects and Cohorts . . . . .	188
7.5.1.1	Temporal Movement Feature - $TMF_2$ . . . . .	189
7.5.1.2	Spatial Movement Feature - SMF . . . . .	190
7.5.1.3	Statistical Movement Feature - UMBS 2 . . . . .	190
7.5.1.4	Spatiotemporal Movement Features . . . . .	190
7.5.2	Consistency of Measurements Over Multiple Days . . . . .	192
7.5.3	Spatiotemporal Movement Feature Derived Cohort Classification . . . . .	194
7.5.4	Non-Environmental Specific Features . . . . .	194
7.6	Conclusions . . . . .	195
<b>8</b>	<b>UMBS Sleep Classification</b>	<b>199</b>
8.1	Introduction . . . . .	199
8.2	Methods . . . . .	199
8.2.1	Study Protocol . . . . .	199
8.2.2	Subject Details . . . . .	202
8.2.3	Data Preparation . . . . .	202
8.2.4	Features . . . . .	205
8.2.5	Performance Measures . . . . .	207
8.2.6	Feature Reduction Mechanisms . . . . .	208
8.2.6.1	Principal Component Analysis . . . . .	208
8.2.6.2	Forward Selection Component Analysis . . . . .	208
8.2.6.3	Feature Subset Selection . . . . .	209
8.3	Sleep/Wake Classification . . . . .	211
8.3.1	Cohort Specific Classification . . . . .	212
8.3.1.1	Classifier Training and Validation Preprocessing Procedure	213
8.3.1.2	Single Feature Classification . . . . .	213
8.3.1.3	Multi-Feature Classification . . . . .	214
8.3.1.4	Sequential Feature Selection . . . . .	218
8.3.1.5	Optimal Classifier Selection . . . . .	220
8.3.1.6	Cohort Independent Classifier Testing . . . . .	220
8.3.2	Subject Specific Classification . . . . .	222
8.3.2.1	Pre-Processing . . . . .	222
8.3.2.2	Multi-Feature Classification . . . . .	222
8.3.2.3	Optimal Classifier Selection . . . . .	227

8.3.2.4	Subject Specific Classifier Validation . . . . .	227
8.3.3	Discussion . . . . .	228
8.4	Sleep Stage Classification . . . . .	233
8.4.1	Multi-Class Classification . . . . .	233
8.4.2	Hierarchical Binary Classification . . . . .	235
8.4.3	Discussion . . . . .	236
8.5	Overall Discussion . . . . .	237
8.6	Conclusions . . . . .	238
<b>9</b>	<b>Conclusions and Future Work</b>	<b>240</b>
9.1	Overall Conclusions . . . . .	240
9.1.1	Literature Review . . . . .	241
9.1.2	Experimental Validation of Movement and Respiration Extraction Algorithms . . . . .	243
9.1.3	Deployment of the UMBS in Clinical and Domestic Environments	244
9.1.4	UMBS Sleep Classification . . . . .	245
9.2	Further Work . . . . .	246
9.2.1	User-Acceptance Testing of Domestic Sleep Monitoring Technology	246
9.2.2	Addition of Markov Models in Predicting Sleep State Transitions for UMBS data . . . . .	247
9.2.3	Assessment of Ambient Methods for Long-Term State Transition Analysis . . . . .	247
9.2.4	An Investigation into Behavioural Change Technologies for Adopting Healthy Sleep Patterns . . . . .	248
9.2.5	An Investigation into the Electronic Delivery of Cognitive Behavioural Therapy for Insomnia . . . . .	248
9.3	Overall Summary . . . . .	249
<b>A</b>	<b>UMBS Spatiotemporal Features - Variance Between Multiple Cohorts</b>	<b>250</b>
<b>B</b>	<b>UMBS Spatiotemporal Features - Variance over Multiple days</b>	<b>260</b>
	<b>References</b>	<b>266</b>

# List of Figures

1.1	Demographic breakdown of the population of the EU 27 (diagram reproduced from European Union (2011) and data taken from EuroStat) . . . .	2
1.2	Old age dependency ratios (data from 2010-2060 are predicted values; data taken from Information Unit, An Roinn Slainte (Department of Health), Ireland (2010)) . . . . .	2
2.1	Positions of the electrodes for Electroencephalography (EEG) (brown), Electro-oculography (EOG) (purple) and EMG (blue). . . . .	14
2.2	Sleep spindle and K-Complex EEG patterns which characterise Stage 2 sleep. . . . .	17
2.3	Schematic representation of the trends of objectively recorded sleep parameters, such as Wake-time After Sleep Onset (WASO), Sleep Latency (SL), Time in Bed (TIB), Total Sleep Time (TST) and all stages, are shown. Data are best estimates taken from all the publications cited in (Miles and Dement, 1980, chap. 3). [taken from (Miles and Dement, 1980, chap. 3).] . . . . .	19
2.4	Typical progression of sleep stages throughout an eight hour sleeping period . .	20
2.5	The values shown in the legend are the mean age of each group. The fifth group, mean age 55.3, experienced only four cycles. Data taken from (Feinberg, 1974). .	21
2.6	Age-related trends for stage 1 sleep, stage 2 sleep, Slow Wave Sleep (SWS), Rapid Eye Movement (REM) sleep, WASO and SL (in minutes). Taken from Ohayon et al. (2004). . . . .	21
3.1	Five artificially generated data sets containing data from two classes (grey and brown respectively). . . . .	30
3.2	Application of BCT classifier on <i>separate sides</i> data set . . . . .	35
3.3	Distribution of data (shown using the Gaussian bell curves) from two classes on the original features and on a new axis which optimally separates the classes. The minimum overlap between the distributions occurs for this new axis. . . . .	36
3.4	Decision regions (marked by blue and red backgrounds) and a discriminating boundary (pink) of LDA and QDA classifiers on two data sets (containing two classes of data marked by brown square and light blue data diamond points). . . . .	39
3.5	Decision regions (marked by blue and red backgrounds) of kNN classifiers on two data sets (containing two classes of data marked by brown square and light blue data diamond points). . . . .	42
3.6	The McCullough-Pitts neuron. . . . .	43
3.7	Multilayer perceptron ANN. . . . .	44

3.8	Decision regions (red and blue backgrounds) for four Multilayer Perceptrons (MLP) neural network classifiers trained using the ‘partially surrounded’ data set and randomly assigned initial conditions. Different decision regions were found for each different set of initial conditions. . . .	46
3.9	Decision regions (red and blue backgrounds) for six MLP neural network classifiers trained using the ‘partially surrounded’ data set and different topologies. The decision regions become increasingly complex (in terms of the function that is being approximated) as the number of neurons in the hidden layer increases. . . . .	47
3.10	Optimal discriminating hyperplane (solid line) for two separate classes (red and blue points) with two hyperplanes either side of maximal distance away (dashed lines). . . . .	49
3.11	Decision regions of SVM applied to the ‘partially surrounded’ data set for the specified kernel, and, where applicable, internal parameters. . . . .	52
3.12	Decision regions of SVM with an RBF kernel applied to the ‘fully surrounded’ data set for the specified internal parameters. . . . .	52
3.13	Decision regions of SVM with an RBF kernel applied to the ‘partially surrounded’ data set for the specified internal parameters. . . . .	53
4.1	Approximate central, frontal and occipital regions of the brain . . . . .	58
4.2	Typical Sleep Diary . . . . .	63
4.3	Sample wrist actigraphy collected over nine days from an elderly subject with each day plotted on subsequent lines. The light blue patterns indicate periods of sleep and the pink patterns indicate periods when the watch was not worn. . . . .	67
5.1	Layout of the UMBS. Each pressure sensing location is marked by the blue circles. The signal from each tactel is sampled by a multiplexer. . . .	106
5.2	Typical placement of the UMBS . . . . .	106
5.3	UMBS open cell urethane foam substrate mechanism of transducing downward pressure into a signal. The foam is covered with a heavy duty flat plastic material (black). The amount of light permeating throughout the substrate is illustrated by the yellow colouring. . . . .	107
5.4	UMBS data collection algorithm . . . . .	109
5.5	Sampling rates attained by the data collection equipment. . . . .	109
5.6	Sample UMBS data. The signals from each tactel are plotted as separate lines. . . . .	112
5.7	2D map of pressure placed on the UMBS at one time instant. . . . .	112
5.8	UMBS data over an entire sleeping period for a healthy adult subject. . .	113
5.9	5 active tactels over an entire sleeping period for a healthy adult subject.	114
5.10	Sample UMBS Turns Data. The subject is lying on their back (B), left (L), right (R) and front (F) and where 1) represents rapid transition without lateral displacement, 2) represents rapid transition with lateral displacement and 3) represents slow rolling transitions . . . . .	115
5.11	UMBS data with movements times highlighted using the standard deviation of the windowed UMBS data . . . . .	118
5.12	Sleeping pattern data from subject DH701 over the two week trial. The magnitude of the blue data is given on the left axis, while the magnitude for the brown data is given on the right axis. . . . .	120

5.13	Comparison of UMBS data, standard deviation metric and sum of difference metric. . . . .	122
5.14	MCC and MCR for both movement decision functions over 8 window lengths. . . . .	123
6.1	Example UMBS data recorded from a female subject lying in a prone position . . . . .	132
6.2	Active UMBS signals recorded from a female subject lying in a prone position . . . . .	132
6.3	Scatterplots of UMBS derived physiological measures to actual and respiratory rates. Diagonal line represents point of perfect alignment. . . . .	135
6.4	Concomitant raw UMBS, strain gauge and PPG data . . . . .	137
6.5	UMBS and strain gauge placement . . . . .	138
6.6	A sample of the concomitant data recorded from one subject. The respiratory effort is captured using the strain gauge as shown in (a). . . . .	138
6.7	Concomitant data recorded from one subject. A correlation exists between the first principal component and the strain gauge (SG) data. . . . .	141
6.8	A correlation can be seen between the first principal component (PC1) and the strain gauge in both figures. The data was normalised to a range of [-1, +1]. . . . .	142
6.9	Simulated UMBS data showing the capacity of the ALC to compensate for the phase variance between UMBS signals. The frequency peak in the output signal is at the expected value of 0.2 Hz. . . . .	144
6.10	Actual UMBS data showing the application of the ALC in compensating for the phase variance between UMBS signals. In Figures (a) and (b), the UMBS signal with the highest variance is used as the desired signal. The output is compared to the actual strain gauge data in Figure (c). In Figures (d) and (e), the actual strain gauge data is set as the desired signal.	145
6.11	The absolute mean difference (solid line) and standard deviation (dotted lines) between the respiratory peaks manually observed and automatically detected using the various techniques. . . . .	146
6.12	Absolute difference between the frequency peak of actual heart rate (using Photoplethysmography (PPG)) and the estimated heart rate (dotted lines separate data belonging to different subjects). The corresponding derived quality metric (peak frequency height/mean frequency height) is also shown.	149
6.13	Comparison of the ability of the UMBS to correctly estimate respiration using direct contact (grey diamonds) and through a mattress (brown squares) against manual respiration counts using a strain gauge over a 5 minute epoch. The diagonal line represents perfect estimation. . . . .	151
6.14	Video motion metric histogram for the entire data set. The high occurrence of low motion metric values represent illustrates that subjects remained still over extended periods of time. . . . .	155
6.15	UMBS data and concomitant video and wrist actigraphy data over an extended period . . . . .	157
6.16	Direct comparison of 2 derived UMBS motion metrics, the PIR motion metric and the video motion metric over 45 sixty second epochs. . . . .	158
6.17	Example Mirrored ROC Curve (sensitivity vs. specificity) . . . . .	158
7.1	Extraction of movement features from UMBS data. . . . .	167

7.2	Data from one UMBS tactel over time and the derivative of that data. . .	168
7.3	UMBS data and derived temporal features from a participant shifting between postures on the UMBS. . . . .	169
7.4	UMBS pressure data with outline of original tactels registering values above 500. . . . .	170
7.5	Interpolated tactels values for one row. . . . .	171
7.6	UMBS terms of location . . . . .	171
7.7	UMBS data and derived spatial features from a participant shifting between postures on the UMBS. . . . .	172
7.8	UMBS data and spread of pressure from a participant shifting between postures on the UMBS. . . . .	174
7.9	UMBS and spatiotemporal data from a participant shifting between postures on the UMBS. . . . .	176
7.10	Pressure impinged on the UMBS at the beginning and end of a movement	176
7.11	Distribution plots for uniform random (left), sinusoidal (centre), and normally distributed random time series data. . . . .	181
7.12	Distribution of $TMF_2$ . . . . .	182
7.13	Distribution of $SMF$ . . . . .	183
7.14	Distribution of UMBS 2 . . . . .	184
7.15	Number of spatiotemporal movements . . . . .	186
7.16	$TMF_2$ CompOverMultNights . . . . .	187
7.17	$SMF$ CompOverMultNights . . . . .	187
7.18	UMBS 2 CompOverMultNights . . . . .	188
8.1	Double raster plot for protocol A (black represents scheduled periods of sleep). . . . .	201
8.2	Double raster plot for protocol B (black represents scheduled periods of sleep). . . . .	202
8.3	Distribution of sleep stages per subject over all sleep episodes. BN101-BN108 belong to the younger adult cohort, while BN109-116 are part of the older adult cohort. . . . .	204
8.4	Performance of each single feature in discriminating between sleep and wake over the three cohorts: younger adults (green), older adults (blue) and all adults (red). The standard deviation in the performance over 5 repetitions is shaded in. . . . .	214
8.5	Performance of the cohort specific classification over the three cohorts. . .	216
8.6	SVM Performance over all box constraint ( $C$ ) and sigma ( $\sigma$ ) values applied to the younger adult cohort (cohort specific classification). . . . .	218
8.7	Sequential feature selection performance per number of input features. . .	219
8.8	LDA classifier performance over six sets of input features. . . . .	223
8.9	QDA classifier performance over six sets of input features. . . . .	223
8.10	kNN1 classifier performance over six sets of input features. . . . .	224
8.11	kNN5 classifier performance over six sets of input features. . . . .	224
8.12	kNN9 classifier performance over six sets of input features. . . . .	224
8.13	NN20 classifier performance over six sets of input features. . . . .	225
8.14	NN40 classifier performance over six sets of input features. . . . .	225
8.15	SVM classifier performance over six sets of input features. . . . .	225
8.16	Mean subject specific classifier performance. . . . .	226

8.17 Hierarchical sleep classifier . . . . .	235
8.18 Distribution of sleep stages per cluster. . . . .	236
A.1 Movement Area . . . . .	251
A.2 Change in spread . . . . .	252
A.3 Movement duration . . . . .	253
A.4 Maximum movement magnitude . . . . .	254
A.5 Lateral change in position . . . . .	255
A.6 Magnitude of movement . . . . .	256
A.7 Percentage to peak movement . . . . .	257
A.8 Time to peak movement . . . . .	258
A.9 Spread movement index . . . . .	259
B.1 numMovements CompOverMultNights . . . . .	260
B.2 movementArea CompOverMultNights . . . . .	261
B.3 deltaSpread CompOverMultNights . . . . .	261
B.4 duration CompOverMultNights . . . . .	262
B.5 magMovement CompOverMultNights . . . . .	262
B.6 mlChange CompOverMultNights . . . . .	263
B.7 mx CompOverMultNights . . . . .	263
B.8 percentToePeak CompOverMultNights . . . . .	264
B.9 timePeak CompOverMultNights . . . . .	264
B.10 sumDiffSpread CompOverMultNights . . . . .	265

# List of Tables

2.1	Physiological rules for sleep scoring as defined by the AASM 2007 standard	15
3.1	Determination of TP, FP, FN and TN	31
3.2	Performance of the SVM, ANN, kNN, LDA and QDA on the <i>fully sur- rounded</i> data set.	55
3.3	The hyperparameters and parameters associated with each classifier.	56
4.1	A summary of an approaches to automated PSG scoring using EOG data	73
4.2	A summary of some approaches to automated PSG scoring using EEG data	77
4.3	A summary of some approaches to automated PSG scoring using EEG and EOG data	77
4.4	A summary of some approaches to automated PSG scoring using EEG, EOG and EMG data	81
4.5	Suitability of sleep monitoring modalities for long-term use.	104
5.1	UMBS Sensor Types	108
5.2	Subject details from UMBS community deployment	116
5.3	Determination of the optimal coefficients	123
5.4	Comparison of UMBS derived restlessness versus activity count (via pe- dometer).	125
6.1	Heart beat and respiration rates over 3 subjects for 3 positions each. The rates for each epoch are measured manually by an observer, by visually examining the UMBS time series data (UMBS time) and also by visually examining the UMBS frequency spectrum (UMBS freq). A manual de- termination of the heart rate was not possible by a visual examination of the UMBS time series data (-). For some epochs a visual examination of the frequency spectrum could not discern a clear frequency peak for both respiration and heart rate data (Not Applicable (N/A)).	134
6.2	The difference between respiration rates captured over all subjects, mean $\pm$ standard deviation breaths per minute. The heart rate difference could not be calculated using UMBS time series data and also excludes Subject 3 in the prone position, mean $\pm$ standard deviation beats per minute.	134
6.3	Pearson's correlation coefficient and mean percentage error for the esti- mated UMBS data against the actual breathing and heart rates over all subjects. Data is also shown for the heart rate estimate excluding Subject 3 Prone position.	135
6.4	Particulars of automated respiration estimation data set.	139

6.5	Accuracy of the respiration rate estimation methods when the UMBS is placed underneath the mattress (UM) and in direct contact with the subject (DC). . . . .	148
6.6	Performance of the UMBS compared to similar ambient sensors. Mean Absolute Difference (MAD), Mean Difference (MD), Mean Error (ME), Mean Percentage Error (MPE), Breaths Per Minute (BrPM) . . . . .	151
6.7	Results for the binary classification test at the optimal UMBS threshold (as defined by the ROC curve) applied to each UMBS derived motion metric and PIR defined motion against the gold standard video based movement metric. . . . .	159
6.8	Results (mean $\pm$ st dev; %) for the accuracy of movement detection when fusing the four UMBS derived motion metrics together and for the UMBS 2 metric. . . . .	160
6.9	Wrist actigraphy movement detection performance over various thresholds compared to the video based gold standard. The optimal BCT applied to the derived UMBS data is also given. . . . .	160
6.10	Performance of wrist actigraphy, PIR and UMBS metrics in detecting motion compared to a video-based gold standard. . . . .	162
7.1	Derived UMBS Features . . . . .	166
7.2	Spatiotemporal Features . . . . .	177
7.3	Data from the home of relatively healthy older adults. . . . .	178
7.4	Data from the homes of 3 healthy young adults. . . . .	178
7.5	Data from a Sleep Clinic from individuals with a suspected sleep disorder. . . . .	179
7.6	Data from a research clinic from healthy younger and older adults. . . . .	180
7.7	Comparison of temporal, spatial, statistical and spatiotemporal movement features for each metric averaged across the four cohorts. . . . .	185
7.8	Confusion Matrix discriminating cohorts using spatiotemporal features using linear discriminant analysis (percent) . . . . .	188
7.9	Confusion Matrix discriminating cohorts using spatiotemporal features using quadratic discriminant analysis (percent) . . . . .	189
8.1	Data Set A - young healthy adults used for classifier training and validation. . . . .	203
8.2	Data Set B - older healthy adults used for classifier training and validation. . . . .	203
8.3	Data Set C - young healthy adults used for classifier testing. * indicates subjects with less than twenty 60 second epochs of wake. . . . .	203
8.4	Data Set D - young healthy adults used for extracting coefficients used in pre-processing steps. No corresponding sleep stages associated. . . . .	204
8.5	Data Set E - older healthy adult used for classifier testing. . . . .	204
8.6	UMBS-Derived Features Used to Discriminate Sleep from Wake. . . . .	205
8.7	Correlation ( $r$ ) Between the UMBS Features Averaged over the 6 Subjects in Data Set D. . . . .	206
8.8	Confusion Matrix Showing Determination of the Number of True Positives, False Positives, True Negatives and False Negatives. . . . .	207
8.9	Top four FSCs. . . . .	209
8.10	Classifiers and samples sizes used for sleep/wake discrimination. . . . .	215
8.11	Performance (F-score) of the classifiers in discriminating sleep and wake over the younger adult cohort . . . . .	217

8.12	Performance (F-score) of the classifiers in discriminating sleep and wake over the older adult cohort . . . . .	217
8.13	Performance (F-score) of the classifiers in discriminating sleep and wake over the entire cohort . . . . .	217
8.14	Sequential feature selection chosen features per cohort. . . . .	219
8.15	Best group specific classifier. . . . .	220
8.16	Optimal group specific classifier. . . . .	220
8.17	Testing performance of the optimal cohort specific classifiers on all three cohorts. . . . .	222
8.18	Performance of the optimal cohort specific classifiers on their cohorts' independent unseen data (data sets C and E). . . . .	222
8.19	Performance of the optimal cohort specific classifiers on their cohorts' data during training, testing and validation. . . . .	222
8.20	Optimal subject specific classifier (results reported per cohort). . . . .	227
8.21	Validation performance of the optimal subject specific classifiers on both cohorts. . . . .	227
8.22	Performance of the optimal subject specific classifiers on their cohorts' data during training, testing and validation. . . . .	228
8.23	Performance of the optimal subject specific classifiers on their cohorts' testing data. . . . .	228
8.24	Confusion matrix of LDA classification of wake/sleep stages using the 13 UMBS features. . . . .	234
8.25	Confusion matrix of LDA classification of a reduced number of wake/sleep stages using the 13 UMBS features. . . . .	234
8.26	Confusion matrix of LDA classification of wake/NREM/REM stages using the 13 UMBS features. . . . .	234
8.27	Overall accuracy (total percentage of correct classifications) of LDA and QDA classification over the three sets of data. . . . .	235
8.28	Comparison of wrist actigraphy, BiancaMed SleepMinder and UMBS for sleep/wake discrimination . . . . .	239

# List of Acronyms

<b>AASM</b>	American Academy of Sleep Medicine
<b>ADL</b>	Activities of Daily Living
<b>ANN</b>	Artificial Neural Networks
<b>AHI</b>	Apnea-Hypopnea Index
<b>ALC</b>	Adaptive Linear Combiner
<b>A-PSG</b>	Ambulatory Polysomnography
<b>AR</b>	Autoregressive
<b>BCT</b>	Binary Classification Test
<b>BCG</b>	Ballistocardiography
<b>BFGS</b>	Broyden-Fletcher-Goldfarb-Shannon
<b>BP</b>	Blood Pressure
<b>BPM</b>	Beats Per Minute
<b>BrP5M</b>	Breaths per Five Minutes
<b>BrPM</b>	Breaths per Minutes
<b>CHF</b>	Congestive Heart Failure
<b>COPD</b>	Chronic Obstructive Pulmonary Disease
<b>CNS</b>	Central Nervous System
<b>CPU</b>	Central Processing Unit
<b>DC</b>	Direct Current
<b>ECG</b>	Electrocardiography
<b>EDR</b>	ECG derived respiration
<b>EDS</b>	Excessive Daytime Sleepiness
<b>EEG</b>	Electroencephalography
<b>EER</b>	Equal Error Rate
<b>EMFi</b>	Electromagnetic Film
<b>EMG</b>	Electromyography
<b>EOG</b>	Electro-oculography
<b>ESS</b>	Epworth Sleepiness Scale
<b>FBG</b>	Fibre Bragg Grating

<b>FFT</b>	Fast Fourier Transform
<b>fMRI</b>	functional Magnetic Resonance Imaging
<b>FN</b>	False negatives
<b>FP</b>	False positives
<b>FSCA</b>	Forward Selection Component Analysis
<b>FSC</b>	Forward Selection Component(s)
<b>HMM</b>	Hidden Markov Models
<b>HRV</b>	Heart Rate Variability
<b>Hz</b>	Hertz
<b>ICA</b>	Independent Component Analysis
<b>ISS</b>	Insomnia Severity Scale
<b>kNN</b>	<i>k</i> -Nearest Neighbour
<b>KSS</b>	Karolinska Sleepiness Scale
<b>LDA</b>	Linear Discriminant Analysis
<b>LM</b>	Levenberg-Marquardt
<b>LMS</b>	Least Mean Squares
<b>MCC</b>	Matthew's Correlation Coefficient
<b>MCR</b>	Misclassification Rate
<b>MLP</b>	Multilayer Perceptrons
<b>MPE</b>	Mean Percentage Error
<b>MSE</b>	Mean Squared Error
<b>MSLT</b>	Multiple Sleep Latency Test
<b>MSL</b>	Mean Sleep Latency
<b>MWT</b>	Maintenance of Wakefulness Test
<b>N/A</b>	Not Applicable
<b>NIH</b>	National Institute of Health
<b>NMCR</b>	Normalised Misclassification Rate
<b>NREM</b>	Non-Rapid Eye Movement
<b>OSAS</b>	Obstructive Sleep Apnoea Syndrome
<b>PAT</b>	Peripheral Arterial Tonometry
<b>PCA</b>	Principle Component Analysis
<b>PC</b>	Principle Component(s)
<b>PCC</b>	Pearson's Correlation Coefficient
<b>PDA</b>	Personal Digital Assistant
<b>PDR</b>	Posterior Dominant Rhythm
<b>PIR</b>	Passive Infra-red
<b>PLMS</b>	Periodic Limb Movement Disorder
<b>PPG</b>	Photoplethysmography

<b>PSD</b>	Power Spectral Density
<b>PSG</b>	Polysomnography
<b>PSQI</b>	Pittsburgh Sleep Quality Index
<b>QDA</b>	Quadratic Discriminant Analysis
<b>RBD</b>	REM Sleep Behaviour Disorder
<b>RBF</b>	Radial Basis Function
<b>REM</b>	Rapid Eye Movement
<b>RF</b>	Radio Frequency
<b>RIB</b>	Respiratory Inductance Band
<b>RIP</b>	Respiratory Inductance Plethysmography
<b>RLS</b>	Restless Leg Syndrome
<b>ROC</b>	Receiver Operator Characteristic
<b>SBD</b>	Sleep Behaviour Disorder
<b>SCN</b>	Supra-Chiasmatic Nucleus
<b>SCSB</b>	Static Charge Sensitive Bed
<b>SD</b>	Standard Deviation
<b>SDB</b>	Sleep Disordered Breathing
<b>SE</b>	Sleep Efficiency
<b>SFS</b>	Sequential Forward Selection
<b>SL</b>	Sleep Latency
<b>SMF</b>	Spatial Movement Feature
<b>SOM</b>	Self-organising Map
<b>SRI</b>	Sleep Related Injury
<b>SSE</b>	Sum of Squared Errors
<b>SSS</b>	Stanford Sleepiness Scale
<b>SVM</b>	Support Vector Machines
<b>SWS</b>	Slow Wave Sleep
<b>TIB</b>	Time in Bed
<b>TMF</b>	Temporal Movement Feature
<b>TN</b>	True negatives
<b>TP</b>	True positives
<b>TST</b>	Total Sleep Time
<b>UMBS</b>	Under Mattress Bed Sensor
<b>WASO</b>	Wake-time After Sleep Onset

# Chapter 1

## Introduction

### 1.1 Background and Motivation

The world's population is ageing as a result of changes in life expectancy, fertility and migration, and this demographic shift is set to intensify over the next forty years (McMorrow and Roeger, 2012). This will be most acute in North America, Europe and Japan and is set to place an enormous burden on healthcare systems (McMorrow and Roeger, 2012). In a global context, the proportion of the population over 60 years of age has risen from 8% in 1950 to 11% in 2009, and is expected to dramatically increase further to 22% in 2050 (Department of Economic and Social Affairs, Population Division, United Nations, 2009). In a European context (of the EU 27 countries as of May 2012), the proportion of adults aged 65 and above is set to increase from 17.4% in 2010 to 30% by 2060 (see Figure 1.1). The *Old Age Dependency Ratio*, the ratio of those over 65 over the working population (15-64 years old), is set to increase dramatically in both a European (from 25.9% in 2010 to 52.6% in 2060) and an Irish (from 16.8% in 2010 to 36.7% in 2060) context (see Figure 1.2). Additionally, there will be an increase in the prevalence of chronic diseases, such as diabetes, Chronic Obstructive Pulmonary Disease (COPD), and arthritis, as a result of changing diet and lifestyles. For example, the incidence of diabetes is set to double between 2005 and 2030 to an estimated number of 366 million people living with diabetes in 2030 (World Health Organisation (WHO), 2004). This will place further strain on healthcare systems as chronic conditions may exist for a long number of years, the resultant complications may be severe, and the means to control them costly (Agoulmine et al., 2011).

The current healthcare paradigm for individuals who have lost the ability to fully care for themselves, particularly amongst *older adults* (aged 60 and older), is centred around delivering institutional care (for example either through hospital or nursing home) when

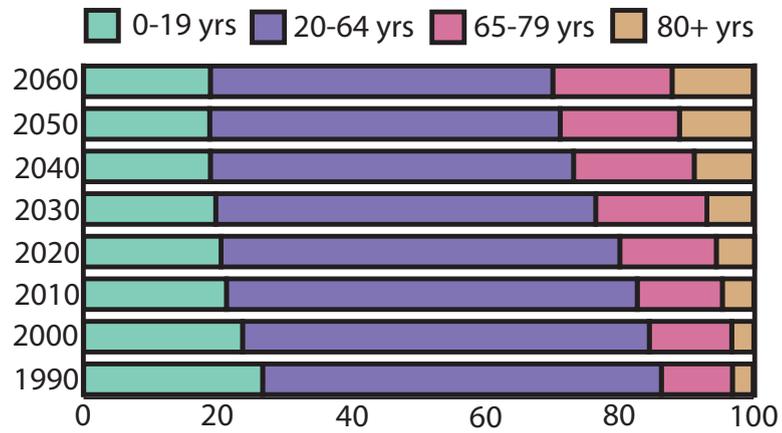


FIGURE 1.1: Demographic breakdown of the population of the EU 27 (diagram reproduced from European Union (2011) and data taken from EuroStat)

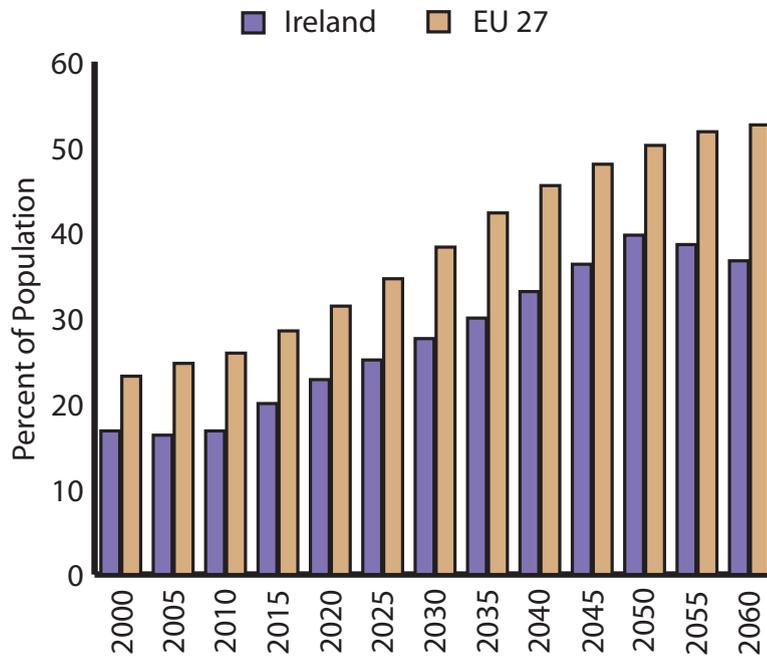


FIGURE 1.2: Old age dependency ratios (data from 2010-2060 are predicted values; data taken from Information Unit, An Roinn Slainte (Department of Health), Ireland (2010))

no caregivers are available or appropriately skilled. This model is costly and significantly reduces the independence and quality of life of those being cared for. Institutional care is suboptimal in cases where some additional support allows the older adults to remain in their home. Ongoing research proposes a move towards technologies which deliver and support *independent living* allowing the older adult to *age in place* for as long as possible.

Recent advances in health technology for older adults have focussed broadly on two areas: 1) clinic-based technology for health assessment resulting in a better delivery of care and 2) home-based technology to support independent living. Both of these areas are diverse. The former ranges from the instantiation and large-scale analysis of electronic health records, to the collection of large data sets using instrumented versions of standard clinical tests for the modelling and prediction of a deteriorating health status, and to the development of new technologies for the prevention, early diagnosis and management of chronic diseases (such as COPD, Congestive Heart Failure (CHF) and diabetes). The latter consists of body sensor networks (the continuous collection and analysis of physiological, kinematic and biomechanical data from miniaturised sensors attached or implanted onto the person), robotics (inclusive of mechatronic systems which may restore motor function or for limb recovery, robots to assist in the performance of Activities of Daily Living (ADL), or the monitoring of overall health status such as for cognition, or for companionship), and ambient intelligent smart homes (where the environment is sensorised using novel unobtrusive technologies which infer, monitor and report the overall health status of the resident). Jointly these approaches facilitate a more effective and efficient delivery of healthcare.

Clinical tests provide a deep insight into the health status of an individual, however due to their intrusive nature and costly overhead they are often performed infrequently. Simple and easily performed standardised tests have been developed to provide quantitative health status indicators (such as the quantitative *Timed Up and Go* test (Greene et al., 2010)). While these may provide a much richer source of data in more convenient settings (for example through primary care clinics), inter- and intra-daily variations are still not measured. This is particularly important for those conditions which cause a rapid deterioration in health (such as COPD and CHF). Body sensor networks and non-contact solutions have been proposed to address this issue (Bonato, 2010).

Ambient living aware homes capture data continuously from various sources throughout the participant's residence; these range from monitoring in-house movement levels (Walsh et al., 2011a), electricity and water usage (Froehlich et al., 2011), inferring a depression index (Dickerson et al., 2011) and many other topics. Through combining features extracted from such technologies valuable insights into the occupant's life may

be obtained. Central to this is the prediction of the overall health status of the resident, however this is contingent on multiple factors including, but not limited to, their physical health, mental health, and social and emotional wellbeing; sleep is implicitly linked to each of these factors.

The fundamental focus of this thesis is the development of an appropriate technology for the long-term monitoring of sleep, particularly suited for placement in the homes of older adults. The Under Mattress Bed Sensor (UMBS) is a pressure sensing grid placed beneath the mattress. In its original clinical instantiation (mainly in nursing homes), it provides measurements of *presence in bed* and quantifies the timing between *bed movements* for inferring the development of bedsores. Specifically, the contributions of this work are in the development of algorithms which extend the capacity of the sensor and are shown to accurately provide metrics of in-bed movement levels, respiration rates and sleep state. Through long-term deployment, it is envisaged that this system may be used to provide unobtrusive insights into overall health status, the progression of the symptoms of chronic conditions, and allow the objective measurement of daily (sleep/wake) patterns and routines.

## 1.2 Sleep

Sleep is a fundamental physiological process with important restorative functions. It occurs in all living mammals and generally over a significant portion of each day (Zepelin, 2000). Sleep problems have been shown to be detrimental to human health. In humans, short (seven hours or less) and long (nine hours or more) durations of sleep have been shown to be significant predictors of death in prospective population studies (Cappuccio et al., 2010). Sleep disturbances may be indicative of poor health and functional deficits, especially in older adults (Manabe et al., 2000; Miles and Dement, 1980). Total sleep time is reduced in the elderly and this is not due to a reduced need for sleep, but in a diminished ability to sleep (Ancoli-Israel, 1997). Sleep complaints are commonly reported by over 50% of those aged 65 and older (Miles and Dement, 1980). These complaints include getting less sleep, frequent awakenings, waking up too early, excessive daytime sleepiness, and napping during the day. Decreased quality of life, and higher rates of depression and anxiety are reported in patients with sleeping difficulties (Barbar et al., 2000). In direct comparisons against matched controls, aged patients with sleep difficulties have significant cognitive impairment and limited attention spans. Additionally, high incidences of balance, ambulatory and visual difficulties (after controlling for medication use) have been reported in older adults with sleep problems (Brassington et al., 2000). Furthermore, decreased TST, an increased SL, defined as the time taken to fall

asleep in bed, and a poor Sleep Efficiency (SE), defined as the percentage of TST over total TIB, are linked to a greater risk for mortality (even when controlling for related covariates) (Dew et al., 2003). Additionally, the symptoms of various chronic conditions continue into the night and result in a disturbed sleep; these include movement disorders, neuromuscular diseases, depression, dementia, epilepsy, obesity and circadian rhythm disorders (Happe, 2003).

The gold standard sleep assessment technology is polysomnography (PSG) which records multiple physiological signals (including brain activity, muscle tone, eye movements, heart rate and respiration) during sleep. This is generally performed in a sleep clinic and a trained sleep scorer uses a strictly defined set of rules to classify each 30 second epoch into either wake or a variety of sleep stages (Iber et al., 2007). However the application of these rules is subjective and an inter-rater agreement rate of 82% has been reported using data from multiple subjects and across separate sleep laboratories (Danker-Hopfe et al., 2009). Additionally, Polysomnography (PSG) is intrusive, costly, time consuming and often alienates the patient. Wrist actigraphy is the current ambulatory gold standard sleep monitoring device. It consists of a two axis accelerometer which records the rest/activity patterns of the wearer and converts this to sleep/wake estimates (using slightly modified thresholding algorithms). Wrist actigraphy has been shown to estimate nocturnal sleep duration and sleep-wake patterns reliably where PSG is not a suitable alternative (Kushida et al., 2001; Sadeh et al., 1994). However, a low wake detection capacity is often reported with this device as the device cannot discriminate between quiescent wake and sleep (Paquet et al., 2007). Wrist actigraphy is dependent on the adherence (and conscious participation) of the wearer. Sleep diaries are also used (often concomitantly with wrist actigraphy) to estimate sleep duration in normal, institutionalised and pediatric populations. However their validity relies upon the attentiveness of the individual filling the diaries out (in cases where the diaries are filled out by the individual). A trade-off exists for these technologies between accuracy and suitability for long term deployment.

An ambient living technology approach offers a more practical solution for long term sleep monitoring and will ideally avoid any conscious interaction with the subject. Video-based sleep monitoring solutions have been proposed Nakajima et al. (2001); Okada et al. (2008) and while their utility is impressive participants are often uncomfortable with the presence of video recording equipment in the home and especially in the bedroom (Townsend et al., 2011c). Privacy concerns were diminished only when unfavourable alternatives (such as nursing homes) were considered (Townsend et al., 2011c). However, other technologies may provide sufficient utility while retaining privacy. Passive Infra-red (PIR) based monitoring systems (Choi et al., 2006; Shin et al., 2003) have been developed and a high accuracy reported, but a number of potential usability issues

remain including the varying location of heaters between environments and the presence and types of bed sheets shielding sensor from the minute movements of the individual. Radar based technologies (de Chazal et al., 2008) report high accuracies in detecting movement compared to wrist actigraphy although the range between the sensor and the individual may be an issue in some cases (particularly when the user is in a double bed). Load-cell movement detection sensors (Brink et al., 2006; Choi et al., 2009) have been shown to have a high capacity in detecting respiration and movement although this may depend on the orientation of the individual. Sleep monitoring systems using pressure pads placed on top of the mattress, underneath the bed sheets, have been developed which detect the ballistocardiogram and detect the heart rate and respiratory rate accurately (Mack et al., 2009a). Under mattress sensors (Watanabe and Watanabe, 2004; Carlson et al., 1999; Shin et al., 2010; Kortelainen et al., 2010; Brser et al., 2011) have been shown to measure respiration or both respiration and heart rate effectively. Pillow based sensors Zhu et al. (2006) have been proposed as solutions to domestic non-contact sleep monitoring by measuring heart and respiration rate. These systems are particularly suited to non-contact long term sleep monitoring as they do not require specialist expertise to install. Overall, some systems have usability constraints ranging from limitations in a practical deployment to a rejection of the technology by the participants due to their design (for example, the noticeable presence of the sensor when in bed or the stigma attached to the requirement for an assisted living technology if the technology is within view of visitors to the environment). In order for any such technology to be adopted successfully over extended durations, they must meet the needs of the user, and, in the case of smart-home/telehealth technology, being unobtrusive is central to this (both in terms of aesthetic design and comfort when applied to long-term sleep monitoring).

### **1.3 Aims and Scope of this Thesis**

Reliable non-invasive long-term monitoring of sleep in a non-clinical setting remains a challenging problem and significant amounts of similar research is ongoing (as discussed above). Such pervasive technologies offer the ability to monitor both the slow progression of an illness as well as sudden physiological changes due to serious life events. The main objective of this thesis is in the development of an appropriate system for long-term in-bed monitoring, particularly for placement in the homes of older adults. This objective is achieved in four main sections.

Firstly, a brief introduction to sleep is given including the definition of sleep, the stages of sleep, physiological changes entering and during sleep, and commonly used sleep metrics.

Additionally, architecture of sleep stages throughout the night, as well as age and illness related changes in sleep are described. A comprehensive overview of traditional clinical and non-clinical sleep monitoring is also provided. An in-depth discussion of the state-of-the-art recent (clinical, contact-based and non-contact) advances in sleep monitoring is also given.

Secondly, the non-contact under-mattress sleep monitoring modality (the UMBS) and the various multiple data collection systems are described in detail. Results from an initial experimental deployment of the UMBS show the ability of the sensor to discern motion. Another initial deployment in a pilot study in the domestic residences of older adults provides a comparison to other sleep monitoring technologies as well as daily activity metrics.

Thirdly, a thorough validation of custom developed algorithms details the reliable and accurate detection of respiration rate and the movement detection capacity of the sensor. Three experiments provide the justification for this: 1) an initial investigation into the capacity of the sensor to detect respiration and heart rate, 2) the development of automated algorithms which extract these physiological signals, and 3) the development of algorithms to extract movement metrics and a comparison against alternative technologies.

Fourthly, multiple feature extraction algorithms which measure temporal, spatial, statistical, and spatiotemporal movement are proposed. These features are applied to and compared across clinical (research and hospital based) and domestic data sets from older and younger adults. A comparison across multiple nights is also made in order to assess inter-daily variations.

Finally, all of the features derived throughout this thesis are used to generate systems which discriminate between sleep and wake. This is performed using a large data set and multiple classifiers on both a younger and older adult data set and is shown to outperform wrist actigraphy (the current ambulatory gold standard for sleep monitoring).

## 1.4 Contributions of this Thesis

In its entirety, this thesis extends the original purpose of the sensor (intended for use in long-term care institutions) into a system which unobtrusively gathers objective respiration, movement and sleep state metrics reliably and accurately. The novel contributions of the work presented in the thesis are:

1. A comprehensive literature review of the state of the art in contact and non-contact sleep monitoring is conducted. This includes advances in automated PSG scoring, electrodes based advances, video, smartphone and Radio Frequency (RF) based applications, and pressure based sensors. The review also provides details on the methodologies used in validating sleep monitoring systems.
2. As part of preliminary investigations of the UMBS, the system was deployed in an experimental setting as well as in a pilot study in the real-world. This real-world deployment was part of a larger study investigating the routines of community dwelling older adults and was carried out by the Digital Health Group in Intel Ireland Ltd. The creation of the data collection platform and the analysis of the associated data was performed in collaboration with the Digital Health Group.
3. The preliminary detection of physiological signals using the system was performed using a rigorously developed experimental setup. Using the resultant data, algorithms were developed which automatically derived respiration rate from the UMBS. These results were optimised against gold standard measurements of respiration. Additionally, movement detection algorithms were developed which were found to outperform wrist actigraphy when compared to a custom-built video-based motion detection system.
4. Methods extracting quantitative measures of temporal, spatial, statistical and spatiotemporal motion in bed were developed as a means of describing movement during sleep. These techniques were validated using a custom-collected data set. Additionally, the system was deployed in complicated clinical (research and hospital-based) and non-clinical (domestic) settings amongst older and younger adults. Features were extracted from the resultant data, and compared across the cohorts. Furthermore, inter-daily comparisons of the metrics were also investigated. Consistencies in some features existed between subjects and cohorts as well as across days, while other features were found to vary largely. This provides an unobtrusive framework for the collection of objective sleep data (in the form of temporal, spatial, statistical and spatiotemporal movement descriptions) which may be representative of transient and long-lasting sleep quality. The hospital-based clinical data set is to be made available to the wider research community.
5. Data were collected in an intensive physiological monitoring clinical research setting from a large cohort of older and younger adults as part of wider research studies. The UMBS data was linked to sleep stages and advanced classification techniques were applied to this data in order to discriminate between sleep and wake. The large data set collected facilitated the application of many classification algorithms, and ensured the validity of the results. The algorithms with the highest accuracy

were shown to outperform wrist actigraphy in detecting sleep and wake using an unobtrusive and non-contact means. An investigation into discriminating between the constituent stages of sleep was also performed, however its performance was low and the system was deemed unsuitable for sleep stage detection. Other similar systems report greater accuracy, however they also use features derived from the heart rate which provides extra discriminative capacity.

## 1.5 List of Publications

1. L. Walsh and S. McLoone, "Validation of a Non-Contact Under-Mattress Sleep Monitoring System", Submitted to *The Journal of Ambient Intelligence and Smart Environments*.
2. L. Walsh and S. McLoone, "A review of recent advances in Non-Contact Sleep Monitoring", In Preparation, *IEEE Reviews in Biomedical Engineering*.
3. L. Walsh, S. McLoone, J.F. Duffy, A.M. Chang, E. Evans, and C.A. Czeisler, "Non-Contact Sleep/Wake Classifications using an Under-Mattress Pressure-Based Sensor", In Preparation, *IEEE Transactions in Biomedical Engineering*.
4. J. Behan, D. Pendergast, L. Walsh and B. Quigley, "Social Rhythms and Nocturnal Routines in Community Dwelling Older Adults." *International Journal of Assistive Robotics and Mechatronics (IJARM)*. Special edition entitled "Ambient Intelligence in Smart Homes: Eldercare with Interventions and Activity Modeling, Reasoning & Recognition", Vol.9, No.4. 2008.
5. J. Behan, D. Pendergast, B. Quigley and L. Walsh, "Determining the relationship between sleep and social activity in community-dwelling older people." *Gerontechnology*, 7(2):70, May 2008.
6. L. Walsh, E. Moloney, S. McLoone, "Identification of Nocturnal Movements During Sleep Using the Non-Contact under Mattress Bed Sensor," *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC11)*, Boston, USA, Sept 2011.
7. L. Walsh, S. McLoone, J. Behan, T. Dishongh, D. Prendergast, and B. Quigley, "Determining Sleep Metrics and Nocturnal Activity Using an Unobtrusive Pressure Sensing Grid." *Proceedings of the IET Assisted Living Conference 2009*, London, UK.
8. L. Walsh, S. McLoone, J. Behan and T. Dishongh, "The Deployment of a Non-Intrusive Alternative to Sleep/WakeWrist Actigraphy in a Home-Based Study of

the Elderly.” *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC08)*, Vancouver, Canada, 20-24 August 2008.

9. J. Behan, D. Pendergast, L. Walsh and B. Quigley, “Social Rhythms and Nocturnal Routines in Community Dwelling Older Adults.” *Lecture Notes in Computer Science: Smart Homes and Health Telematics (ICOST)*, Vol. 5120/2008, pp. 73-80, Springer-Verlag, 2008.

## 1.6 Thesis Layout

This thesis begins in Chapter 2 by providing an introduction to sleep and sleep medicine for the novice reader. It provides a definition of each of the constituent stages sleep is generally broken into and other physiological changes exhibited during sleep. Commonly reported sleep metrics are also defined. The typical architecture of sleep and sleep stages throughout the night are also described. Additionally, the changes to sleep caused by natural ageing, and by illness and disease are discussed.

In Chapter 3 an overview of the performance metrics and classification methods discussed in this thesis are presented. These classification methods include the binary classification test, discriminant analysis, k-nearest neighbour, artificial neural networks and support vector machines. An overview of how to rigorously implement these techniques using large data sets and how to implement multiclass classification is provided.

Chapter 4 provides a comprehensive literature review of sleep measurement technologies. It ranges from PSG (the traditional clinical based approach) and manual scoring, sleep diaries, sleep tests and subjective measures, to wrist actigraphy, automated PSG scoring techniques, and contact and non-contact advances in sleep monitoring. Additionally, multi-modal techniques, modalities for sleep apnea monitoring and sleep monitoring using brain imaging are briefly discussed.

Chapter 5 presents the candidate hardware and software used to unobtrusively monitor sleep. Results from initial experimental deployments investigating the detection of physiological signals, examining sensor data from an entire night and showing sensor data during typical nightly movements are presented. Additionally, results from the deployment of the sensor in a pilot study are presented. Data extracted from the sleep sensor during this pilot study are compared to wrist actigraphy, measures of daily activity and daily routines. A discussion on parallel research being carried out using this sensor is also reported.

Chapter 6 describes the thorough and rigorous validation of automated algorithms extracting respiration rates and in-bed movement metrics. These physiological results are optimised against clinical gold standards and a custom-built motion detection system. This chapter provides the evidence underpinning the suitability of the sensor for reporting data related to overall health and sleep status.

Chapter 7 discusses the extraction of temporal, spatial, statistical and spatiotemporal movement features from the proposed sensor. Data were collected from clinical (in research and hospital based settings) and non-clinical (domestic) settings from both older and younger adults and these features were extracted from each of these populations. A comparison of the movement features across the subjects over one night is made, as well as an investigation of the consistency of these metrics across multiple nights (mainly in the cohort of the older adults in their domestic residences). A comparison of the spatiotemporal movement features between cohorts is made in order to compare whether the quantitative components of each movement differs between the populations.

Chapter 8 applies various classification techniques to the features derived in the previous chapters in order to discriminate between sleep and wake, and between the sleep stages. A large data set was collected for this purpose in a cohort of older and younger adults in a clinical research setting. Cohort and subject specific classifiers are built for the discrimination of sleep and wake. Standard multiclass classification techniques and hierarchical binary classifiers are applied to predict the sleep stages from a mixture of the individual features.

Finally, Chapter 9 provides a summary of the novel contributions of this thesis, and discusses potential further avenues of research made available by this body of work.

## Chapter 2

# An Introduction to Sleep

Sleep is a complex and essential physiological process with an important restorative function. Unsurprisingly, it has been, and remains to be, the subject of intense scientific study. This chapter introduces the field of sleep science and begins by providing the scientific definition of sleep and the various sleep stages, each with their own distinct characteristics. Commonly used sleep metrics, derived from the pattern of sleep stages that occur during a sleeping period, are described. Sleep quality declines and sleep quantity shortens as a person gets older and this can be a result of ageing, the progression of a disease or illness, and/or sleep disorders. These changes (for example a reduced TST, lower SE or an increased SL) are often multifactorial. A description of how poor sleep affects overall health, and how various illnesses affect sleep is provided. Sleep science and sleep medicine is an ever-expanding area and its entire scope is outside the bounds of this research; for a thorough review the reader is directed to Lee-Chiong (2006).

### 2.1 The Definition of Sleep

Humans spend approximately one third of each day asleep (Feinberg, 1974; Miles and Dement, 1980). Before the development of technologies which measure internal biological processes, sleep was defined through: a) a species specific body posture, b) maintained behavioural quiescence, c) an elevated arousal threshold, and d) state reversibility with stimulation (Flanigan, 1972). However, in 1935, Loomis et al. conducted extensive comparisons investigating changes in cortical patterns discriminating patterns sleep and wake (Loomis et al., 1935a,b). Further work in 1937 examined brain waves at the point of sleep onset (Davis et al., 1937). This work was further expanded up as eye movements were found to also provide discriminative capacity in identifying when dreams occur (Dement and Kleitman, 1957). Rechtschaffen and Kales published their seminal sleep

scoring manual in 1968 which provided a strict and well-defined set of rules for trained scorers to identify periods of sleep and the constituent stages of sleep.

The exact purpose of sleep is unknown; however adaptive, restorative and homeostatic theories have been proposed to explain a need for sleep (Carlson et al., 2000; Malim and Birch, 2000). The adaptive theory discusses that sleep evolved to keep animals safe from predators as this was when they were most susceptible to attack. The restorative theory for sleep states that sleep is needed for recuperative purposes. The homeostatic theory suggests that sleep evolved to conserve energy, the antithesis of the waking day.

An endogenous drive for sleep exists in most animals (Lockley, 2009). In fact, chronic sleep restriction has been found to be fatal in rats, drosophila and possibly humans (Frank, 2006). This supports a life-sustaining function for sleep. The existence of an internal body clock that regulates the *circadian* (Latin for 'about a day') rhythms of many physiological processes, including sleep, has been proven to exist (Lockley, 2009). In humans, the circadian rhythm of sleep and wake drives a high desire for sleep at night and a smaller drive in the afternoon (Mitler, 1996). This is the opposite for nocturnal animals. The driving mechanism behind this rhythm has been found to exist in the Supra-Chiasmatic Nucleus (SCN) in the hypothalamus (Lockley, 2009). Light is a direct input to this system and the retino-hypothalamic tract has been shown to carry projections from the retina to the SCN. If a human is left to self regulate without any external influences their body has been shown to adopt an internal endogenous cycle lasting nearly twenty five hours (Lockley, 2009). Light, environmental cues (such as social demands) and hormones can be used to reset and regulate this internal clock to adhere to societal norms (Lockley, 2009). Disturbances in circadian rhythms have been associated with cognitive decline, mood, behavioural and sleep disturbances, and limitations of activities of daily living in elderly patients with dementia and their caregivers (Riemersma-van der Lek et al., 2008). A randomised control study on elderly patients with dementia found that the long-term treatment using whole-day bright light ( $\approx 1000$  lux) attenuated cognitive deterioration, ameliorated depressive symptoms and reduced the limitations in functional limitations over time (Riemersma-van der Lek et al., 2008).

## 2.2 Sleep Scoring

Upon initial investigation human sleep was divided into two distinct states known as REM sleep and Non-Rapid Eye Movement (NREM) sleep each characterised by the presence or absence of the fluttering, or rotation, of the eye (Aserinsky and Kleitman, 1953). In 1968, Rechtschaffen and Kales developed an analytical basis for scoring human sleep

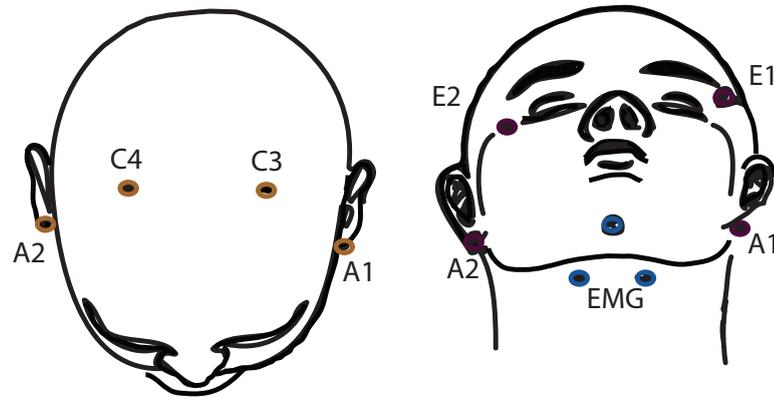


FIGURE 2.1: Positions of the electrodes for EEG (brown), EOG (purple) and EMG (blue).

recordings into separate stages using brain waves and physiological patterns (Rechtschaffen and Kales, 1968). Specific, well defined rules subdivided NREM sleep into 4 stages (stage 1, stage 2, stage 3, stage 4) (Rechtschaffen and Kales, 1968). Rechtschaffen and Kales also defined 'wake' and 'movement time' stages. Recently these standards have been replaced by the American Academy of Sleep Medicine (AASM) Manual for Scoring Sleep Stages and Associated Events (Iber et al., 2007) as outlined in Table 2.1. One major change between the AASM and the Rechtschaffen and Kales guidelines for scoring sleep is that sleep stages 3 and 4 are combined. REM sleep is also divided into phasic and tonic REM.

Sleep scoring mainly examines (Iber et al., 2007) (and the uses the electrode positions defined in Figure 2.1):

- EEG: the measurement and patterns of brain waves
- EOG: the measurement of eye movements
- Electromyography (EMG): the measurement of muscle tone

EEG provides essential information for sleep scoring, while EOG and EMG provide additional information (Iber et al., 2007). Other signals which can aid scoring are the Electrocardiography (ECG), the breathing signal and a movement signal. Further information on the measurement of these signals is provided in Chapter 4.

A brief overview of wake and each of the sleep stages is given in the following sections.

TABLE 2.1: Physiological rules for sleep scoring as defined by the AASM 2007 standard

Stage	EEG Findings	Eye Movements (EOG)	EMG submental
W	<b>Over 50 % of an epoch has an alpha rhythm over occipital rhythm</b>	<i>Typically, no eye movements seen</i>	<b>Normal to high muscle tone</b>
N1	<b>Attenuation of alpha rhythm for over 50 % of the epoch replaced with mixed frequency low-amplitude rhythm or a slowing of PDR from waking of greater than or equal to 1 Hz if no alpha rhythm was noted.</b> <i>Vertex sharp waves.</i> N1 stage continues until beginning of N2 stage or an arousal	Slow, rolling eye movements typically	Variable, typically less than wake
N2	<b>K complexes and/or sleep spindles (as presented in Figure 2.2) occurring in the first half of the epoch;</b> <i>Low-amplitude, mixed frequency EEG.</i> N2 stage persists until transition to N3 stage, R stage or an arousal.	<i>Typically, no eye movements, but slow eye movements may persist</i>	<i>Variable amplitude, typically lower than W and higher than R</i>
N3 <sup>1</sup>	<b>Slow-wave activity (0.5-2 Hz, greater than 75 microVolts) for greater than 20 % of an epoch. Sleep spindles may persist.</b> N3 persists until transition to N1, transition to N2, between K complexes without eye movements, or an arousal.	<i>Typically, no eye movements seen</i>	<i>Variable amplitude, typically lower than W and higher than R</i>
R	<b>Low-amplitude, mixed frequency EEG.</b> <i>Sawtooth waves.</i> R persists until transition to N1, transition to N2, between K complexes without eye movements, or an arousal	<b>REMs</b>	<b>Low muscle tone</b>

W=wakefulness; N1 = NREM stage 1 sleep; N2 = NREM stage 2 sleep; N3 = NREM stage 3 sleep; R = REM stage sleep. Bolded items are requirements for staging. Italicised items are nonrequired associated findings that may be present in that sleep stage. Table adopted from AASM Manual for the Scoring of Sleep and Associated Events (Iber et al., 2007).

<sup>1</sup> Previously known as NREM stage 3 and NREM stage 4 sleep.

### 2.2.1 Wake

Wake is characterised by an alpha (8-13 Hz) EEG rhythm over the occipital region (rearmost position of the brain) and this is commonly referred to as a Posterior Dominant Rhythm (PDR). It is also defined by the presence of rapid eye movements and continued muscle tone. Sleep is scored as wake when greater than half of the epoch has an alpha rhythm (Iber et al., 2007). Eye blinks, reading eye movements, irregular conjugate rapid

eye movements associated with normal or high chin muscle tone also define the wake state (Iber et al., 2007).

### 2.2.2 Rapid Eye Movement (REM) Sleep

REM sleep occurs for approximately twenty five percent of TST (Ehlers and Kupfer, 1997). REM sleep is characterised in EEG by low amplitude, mixed frequency EEG activity with slow alpha (8-12 Hz) and theta (4-8 Hz) waves. Saw-tooth waves (resembling in appearance the blade of a saw) which are brief bursts of 3-7 Hz EEG activity with an amplitude less than  $50\mu\text{Volts}$  and a duration of about 5 seconds can also be present (Stern, 2004). REM sleep can be divided up into tonic and phasic REM sleep. Tonic REM sleep is characterised by a desynchronised EEG, atonia of skeletal muscle groups and suppression of monosynaptic and polysynaptic reflexes. Rapid eye movements, transient swings in blood pressure, heart rate changes, irregular respiration, tongue movements and myoclonic twitching of chin and limb muscles are evident in Phasic REM sleep (Baust et al., 1972; Chokroverty, 1980; Iber et al., 2007; Oksenberg et al., 2001; Orem, 1980; Rama et al., 2006).

### 2.2.3 Non-Rapid Eye Movement (NREM) Sleep

Non-rapid eye movement sleep occupies approximately seventy-five percent of total sleep time in normal healthy adults. Traditionally it was broken into four sleep stages, although in the recent review of sleep scoring (Iber et al., 2007) it has been broken up into 3 stages. Stage 1 (also known as N1 in the latest revision of sleep scoring by Iber et al.) is often considered a transitional stage of sleep occurring as the individual enters sleep. NREM sleep occurs cyclically throughout the night interspersed with REM sleep (Feinberg, 1974). The following section describes the different stages of NREM sleep.

Stage NREM 1 (N1) sleep occurs mostly in the transition from wakefulness to sleep (including after arousals). Approximately five percent of sleep is spent in this stage (Ehlers and Kupfer, 1997). Some subjects if woken after stage 1 sleep will have no recollection of being asleep, while others will report being asleep (Feinberg, 1974). During stage 1 sleep the PDR is replaced by a low-voltage, mixed-frequency activity pattern, typically to a theta frequency (4-8 Hz). A decreased muscle tone and slow rolling eye movements are also evident. Vertex sharp waves (50-200 ms) are noted towards the end of stage 1 sleep (Iber et al., 2007; Rama et al., 2006; Vaughn and Gaiallanza, 2008).

Stage NREM 2 sleep (N2) is often also referred to as quiet sleep and occurs for over forty five percent of total sleep time (Ehlers and Kupfer, 1997). Two distinct EEG patterns,

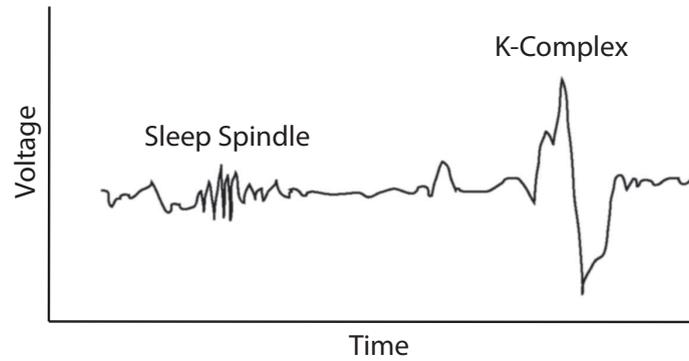


FIGURE 2.2: Sleep spindle and K-Complex EEG patterns which characterise Stage 2 sleep.

sleep spindles and K-complexes (as presented in Figure 2.2), which must occur in the first half of the scoring period, define the start of this sleep stage (Iber et al., 2007). A sleep spindle is a train of distinct waves of 11-16 Hz activity lasting at least 0.5 seconds (Iber et al., 2007). K-complexes are EEG activity consisting of a downward trend followed by a gradual upward trend lasting more than 0.5 seconds (Iber et al., 2007). Delta waves (0.5-4 Hz) may also be present in the EEG but occur in small amounts. Muscle tone and eye movements also diminish during this stage (Rama et al., 2006; Iber et al., 2007). This stage will continue to be scored until an arousal, major body movement or transition to another sleep stage occurs (Iber et al., 2007).

Stage NREM 3 and 4 sleep (N3) lasts for approximately fifteen to twenty percent of the total sleep time. High-amplitude ( $> 75\mu\text{Volts}$  peak to peak), slow wave delta (0.5-2.0 Hz) EEG activity characterises these stages. Stage NREM 3 and 4 are defined when this EEG pattern occurs for over twenty percent of the epoch (Iber et al., 2007). In the previous Rechtschaffen and Kales standard (Rechtschaffen and Kales, 1968), stages NREM 3 sleep and stage NREM 4 sleep were commonly referred to as SWS. Stage NREM 3 sleep contained over twenty percent delta activity and stage NREM 4 sleep contained over fifty percent delta activity. When a person is woken from this stage they appear groggy and disorientated. EOG does not register eye movements and muscle tone is decreased compared to wakefulness or stage NREM 1 sleep, typically lower than NREM 2 sleep and can be as low as that seen in REM sleep (Rama et al., 2006; Iber et al., 2007).

## 2.3 Other Physiological Changes During Sleep

**Autonomic Nervous System Variation** An increase in parasympathetic tone and a decrease in sympathetic tone is seen at the onset of sleep and NREM (Choudhary and Choudhary, 2009). These features are more exaggerated during tonic REM, however in phasic REM an increase in sympathetic activity is observed. For further information regarding the autonomic nervous system, sympathetic and parasympathetic tone, the reader is referred to Choudhary and Choudhary (2009).

**Body temperature Variation** The body temperature decreases during sleep by 1-2 degrees Celcius and rises during the waking day in a sine wave pattern (Choudhary and Choudhary, 2009). The regulation of temperature is maintained during NREM sleep, but this is attenuated during REM sleep (Choudhary and Choudhary, 2009).

**Hormonal Variations** A circadian variation in hormone secretion is common and this often peaks between 4am and 8am (Choudhary and Choudhary, 2009).

**Cardiovascular Variation** The cardiovascular system is influenced by parasympathetic activity (Choudhary and Choudhary, 2009). As a result, heart rate, blood pressure, stroke volume, cardiac output and systemic vascular resistance decreases during sleep (Choudhary and Choudhary, 2009). A 35% increase in heart rate occurs during phasic REM sleep (Choudhary and Choudhary, 2009).

**Respiratory Variation** A small decrease in ventilation occurs during the transition from wakefulness to NREM sleep and a reduction in respiratory drive occurs (Choudhary and Choudhary, 2009). The breathing pattern is regular during NREM sleep (Choudhary and Choudhary, 2009). During REM sleep the respiration becomes irregular (Choudhary and Choudhary, 2009).

## 2.4 Commonly Used Sleep Metrics

Various general statistics describing sleep are commonly reported across studies and individuals (Iber et al., 2007; Hori et al., 2001); some of which are detailed below. A comparison of some of these metrics across age groups and gender (see Figure 2.3) was performed by Miles and Dement (Miles and Dement, 1980). These metrics can be reported objectively from sleep recordings and also through subjective reports.

**Wake-time After Sleep Onset (WASO)** The length of time spent awake after sleep onset until the final awakening.

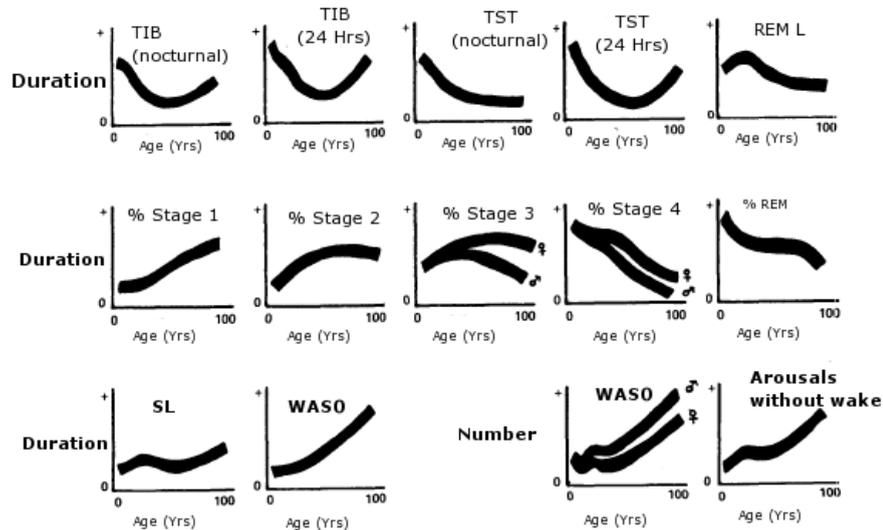


FIGURE 2.3: Schematic representation of the trends of objectively recorded sleep parameters, such as WASO, SL, TIB, TST and all stages, are shown. Data are best estimates taken from all the publications cited in (Miles and Dement, 1980, chap. 3). [taken from (Miles and Dement, 1980, chap. 3).]

**Sleep Latency (SL)** The length of time between lights out (the decision to sleep) and sleep onset.

**Rapid Eye Movement (REM) latency** The length of time between lights out and the first occurrence of REM sleep.

**Total Sleep Time (TST)** The total length of time spent asleep over the entire sleep episode.

**Time in Bed (TIB)** The total time spent in bed during the entire sleep episode.

**Sleep Efficiency (SE)** Sleep efficiency is defined as the TST divided by the TIB.

**Number of bed exits** The number of times that the participant leaves the bed (for example, to use the toilet) during the entire sleep episode.

**Length of bed exits** The total length of times that the participant leaves the bed for during the entire sleep episode.

**Number of arousals** This is defined as the number of individual arousals during the entire sleep episode.

**Length of arousals** The total length of arousals during the entire sleeping episodes.

**Percent of sleep stage** This is a percentage of the time spent in each stage divided by the TST.

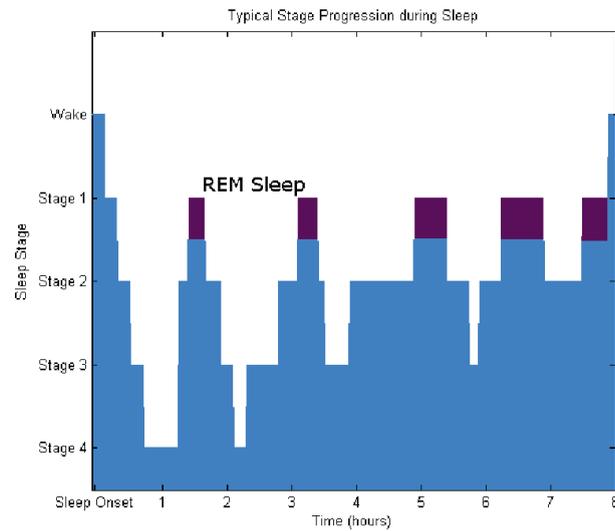


FIGURE 2.4: Typical progression of sleep stages throughout an eight hour sleeping period

## 2.5 The Architecture of Sleep and Sleep Stages

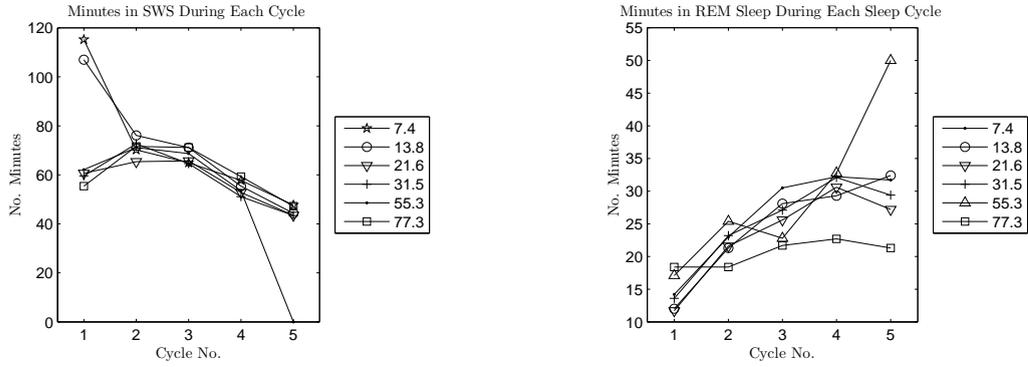
The distribution of the various sleep stages occurs in NREM-REM cycles throughout the night. Each cycle lasts approximately ninety minutes with approximately four to six cycles per major sleeping episode (Feinberg, 1974). As older adults tend to have a reduced TST, a lower number of cycles is generally seen (Feinberg, 1974). SWS sleep is more common at the beginning of the sleeping period with REM sleep more common towards the end. An example of the typical distribution of sleep stages throughout a night can be seen in Figure 2.4 generated with simulated data using typified data (Feinberg, 1974).

Changes in the amount of time spent in each sleep stage over a range of subjects stratified by age is shown in Figure 2.5. The data are reported were averaged results categorised by the mean age of each group. A general downward trend in the amount of time spent in NREM sleep can be seen over successive cycles. The time spent in REM sleep generally increases over successive cycles.

## 2.6 Age and Illness Related Changes in Sleep

### 2.6.1 Sleep and Age

Natural age-related changes in sleeping patterns occurs throughout the life cycle (see Figure 2.6). New born infants spend a significant portion of their day asleep (commonly



(a) Typical Minutes of SWS Sleep during Progressive Sleep Cycles.

(b) Typical Minutes of REM Sleep during Progressive Sleep Cycles.

FIGURE 2.5: The values shown in the legend are the mean age of each group. The fifth group, mean age 55.3, experienced only four cycles. Data taken from (Feinberg, 1974).

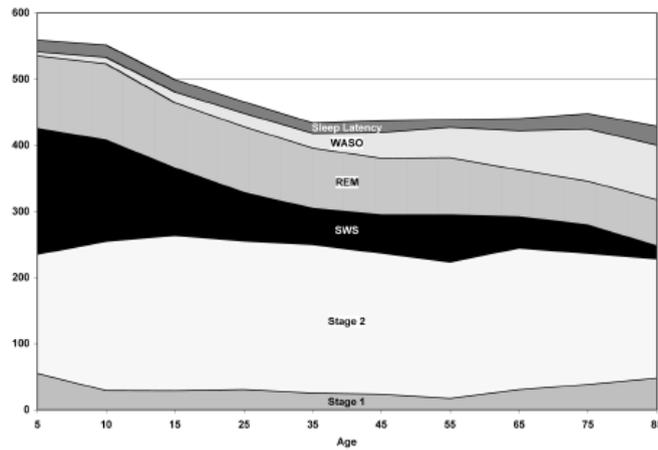


FIGURE 2.6: Age-related trends for stage 1 sleep, stage 2 sleep, SWS, REM sleep, WASO and SL (in minutes). Taken from Ohayon et al. (2004).

over sixteen hours asleep; however this is not consolidated) and over fifty percent of sleep is spent in the REM stage (Rosenthal, 2006). A consolidated nocturnal sleep should be possible from three months with addition of daytime naps. Older children and adults typically have consolidated nocturnal sleep without any daytime naps. However, older adults commonly report sleep disturbances (Ayalon et al., 2004; Ayalon and Ancoli-Israel, 2006; Bliwise, 2000). However, a distinction must be made between natural changes in sleep patterns with normal aging and pathological changes in sleep patterns.

### 2.6.1.1 Sleep and Older Adults

Total sleep time is reduced in the elderly (Miles and Dement, 1980) and this is not due to a reduced need for sleep, but in a diminished ability to sleep (Ancoli-Israel, 1997;

Foley et al., 2004). It has also been shown that sleep disturbances may be indicative of poor health and functional deficits, especially in older adults (Miles and Dement, 1980; Manabe et al., 2000). Over 50 % of older adults frequently complain about their sleep and these complaints include getting less sleep, difficulty initiating or maintaining sleep, frequent awakenings, waking up too early, excessive daytime sleepiness, and napping during the day (Foley et al., 1995; Vitiello et al., 2004; Bliwise et al., 1992)

A meta-analysis of older adults with a strict selection criteria found consistent TST, SL, WASO, percent stage NREM 1 sleep, percent stage NREM 2 sleep, percent stage SWS, percent stage REM sleep, REM latency and also found a decrease in sleep efficiency between the ages of sixty, and one-hundred-and-two years old (Ohayon et al., 2004). A second study by Vitiello (2006) with a less stringent selection criteria found that men showed evidence of poorer sleep with aging compared to women. A possible explanation for this postulated by Redline et al. (2004) is the excess subclinical morbidity in men. Correspondingly, the studies by Vitiello and Ohayon et al. are complimentary when the reader is mindful of the selection criteria for participants. The Ohayon et al. study uses a strict selection criteria and reports on older adults aging healthily. However, Vitiello uses a less strict selection criteria, which is more representative of the population, and reports poorer sleep as age increases. This suggests that monitoring sleep may be used as a proxy for assessing the health status (including subclinical degradations) of older adults.

## **2.6.2 Sleep Degradation from Sleep Disorders, Diseases and Illnesses**

The effects of many diseases extend nocturnally, negatively affecting sleep, while some manifest themselves solely during the waking day. Movement disorders, neuromuscular diseases, depression, dementia, epilepsy, obesity and circadian rhythm disorders directly affect the sleep of many people (Happe, 2003). Sleep disorders, occurring in the absence of a causative factor, also degrade sleep and sleep quality. This effect is often multifactorial. Some common sleep disorders are described below and the effects of some diseases and illnesses on sleep is also described.

### **2.6.2.1 Apnoea/Hypopnoea**

Apnoeas are characterised as a cessation of breathing during sleep for at least ten seconds (Lawati et al., 2009). This can be further broken down into central (lack of respiratory effort) or obstructive (caused by a blockage in the breathing passage) sleep apnoea. Sleep apnoea and its diagnosing criteria have been well defined and an outline is given below (The Report of an American Academy of Sleep Medicine Task Force, 1999). Hypopnoeas

are a milder form of sleep apnoea where a complete suspension of breathing does not fully occur. Instead, either 1) a reduction of over 30% of airflow for ten seconds and at least a 4% drop in oxygen desaturation or 2) a 50% reduction in airflow for ten seconds and at least a 3% drop in oxygen desaturation or an arousal must occur in order to score a hypopnoea.

**Diagnostic criteria for Central Sleep Apnoea:** Insomnia or excessive sleepiness, shallow or absent breathing during sleep; gasps or grunts or choking during sleep; frequent body movements; cyanosis; polysomnogram demonstrations of central apnoeas with arousals; bradycardia (prolonged occurrence of a slowing of the heart rate), or oxygen desaturation.

**Diagnostic Criteria for Obstructive Sleep Apnoea:** Excessive Daytime Sleepiness or insomnia; frequent episodes of obstructive breathing during sleep; associated features such as loud snoring; morning headaches; dry mouth upon waking, and chest retraction in children during sleep; PSG showing obstructive apnoeas, arousals, bradycardia, and arterial oxygen desaturation; and Multiple Sleep Latency Test (MSLT) showing increased daytime sleepiness. Not all of these need to be present.

The Apnoea-Hypopnoea Index (AHI) is used to quantify the severity of apnoeas/hypopnoeas. It is measured clinically as the average number of apnoeas and hypopnoeas which occur per hour. High rates of moderate (AHI of 15 to 30) to severe (AHI of over 30) sleep apnoea have been reported in approximately 9% in middle aged men and 4% of women (Lawati et al., 2009). This is greater in some patient populations such as those who are elderly, have hypertension, and those with coronary disease (Lawati et al., 2009). Obstructive sleep apnoea impairs quality of life, and is associated with cardiovascular disease and motor vehicle crashes (Lawati et al., 2009).

### 2.6.2.2 Excessive Daytime Sleepiness

Excessive Daytime Sleepiness (EDS) is defined by the need to sleep at abnormal, often inappropriate, times and places (American Sleep Disorders Association, 1997). Twelve percent of the general population are affected by it (Happe, 2003; Roth and Roehrs, 1996). EDS is thought to have four major causes:

1. Quantitative and qualitative sleep deficiencies
2. Central Nervous System (CNS) pathological abnormalities (eg. neurological disorders)

3. Circadian rhythm disorders (eg. jet lag or shift work)
4. Drugs (Roth and Roehrs, 1996)

### **2.6.2.3 REM Sleep Behaviour Disorder**

REM sleep is partially characterised by muscle atonia, except in the extraocular and diaphragm muscles (Rechtschaffen and Kales, 1968). This atonic state is thought to avoid the 'acting out' of dreams. REM Sleep Behaviour Disorder (RBD) occurs when this suppression of bodily activity does not function. PSG-coordinated video recordings often show dream mentation during these RBD episodes. It often presents itself in 60-80 year olds and can be preceded by limb movements, talking and yelling (American Sleep Disorders Association, 1997).

RBD is thought to be a precursor for many neurological disorders and this theory is supported by many studies (Happe, 2003). One such study has indicated that 38% of males diagnosed with idiopathic RBD developed Parkinson's disease over a mean of 3.7 years (Schenck et al., 1996). A further follow-up study showed that 65% developed delayed Parkinsonism and/or dementia seven years after the original article was published (Schenck et al., 2003). Another recent study showed 45% of a cohort of forty four patients with assumed idiopathic RBD developed a neurodegenerative disease, including Parkinson's disease, dementia with Lewey bodies (DLB), multiple system atrophy (MSA) and mild cognitive impairment, 10.7 years after reported RBD onset (Iranzo et al., 2006). Early clinical manifestations of neurodegenerative disorders, such as Parkinson's disease or MSA might be possible by screening for idiopathic RBD (Stiasny-Kolster et al., 2007).

Sleep Related Injury (SRI) often occurs jointly with the sudden aggressive movement associated with RBD. It can be caused to either the subject or to their caregiver. SRI was reported to be a lot higher in an RBD group, thirty three percent, than in a non-RBD group, six percent (Comella et al., 1998).

### **2.6.2.4 Restless Leg Syndrome**

Restless Leg Syndrome (RLS) is prevalent in adult populations (5%-15%) and increases with age (Happe, 2003; Allen et al., 2003). Individuals with RLS have a longer SL, a shorter TST, less SWS and a decreased SE (Happe, 2003). A National Institute of Health (NIH) workshop in collaboration with the International Restless Legs Syndrome Study Group defined the criteria for RLS (Allen et al., 2003):

1. Desire to move the limbs usually associated with paresthesias or dysesthesias

2. Symptoms begin or worsen during inactivity or rest (ie. lying or sitting)
3. Symptoms partially or totally relieved by movement (walking or stretching)
4. Symptoms are worse in the evening or night. (Allen et al., 2003)

#### **2.6.2.5 Periodic Limb Movement Disorder**

Periodic Limb Movement Disorder (PLMS) is classified by abnormal uncontrolled involuntary movements during sleep and have been reported to be associated in up to eighty eight percent of patients with RLS (Happe, 2003). PLMS is more common with increasing age and it is thought that thirty three percent of people with Parkinson's disease have PLMS (Happe, 2003). The American Sleep Disorders Association (ASDA) classified a diagnostic criteria for PLMS (American Sleep Disorders Association, 1997):

1. Insomnia or excessive sleepiness, occasionally patients are asymptotic and movements are noted by an observer
2. Repetitive, highly stereotyped muscle movements
3. PSG demonstrates repetitive episodes of muscle contraction, and arousal or awakenings may occur.

#### **2.6.2.6 Nocturia**

Nocturia, the process of waking during sleep in order to use the bathroom, has been found to be independently associated with sleep disordered breathing and also independently associated with cardiovascular co-morbidity in a community based study of older adults (Parthasarathy et al., 2012).

#### **2.6.2.7 Circadian Rhythm Disorders**

The most common circadian phase disorder is shift work sleep disorder; experienced by many of the 15 million American Shift workers (Lockley, 2009). Shift workers experience fatigue, sleep problems, poor performance, poor memory, gastrointestinal problems and have an increased risk for cardiovascular disease, diabetes and cancer. This is due to the desynchronisation between the light-dark cycle of the natural day and the shift worker's schedule (Lockley, 2009). It is common that the shift worker cannot adapt to inconsistent working times (imposed by varying times at which their working schedule

begins) which results in working and sleeping at biologically inappropriate times of the day.

Jet lag is another common circadian rhythm disorder caused by the insertion of an individual into a new circadian phase (Lockley, 2009). However, unlike shift work, jet lag does not occur that frequently (as most individuals do not travel across timezones very regularly).

Advanced and delayed sleep phase syndrome cause sleep to naturally occur very early or very late in the day (Lockley, 2009). It is common that older adults have advanced sleep phase syndrome which results in very early awakenings and consequently early sleep start times. However, this imposes bed times earlier than societally accepted norms. Often the older adults goes to bed at a more traditional time and their TST is reduced (Miles and Dement, 1980; Lockley, 2009).

#### **2.6.2.8 The Effect of Parkinson's Disease on Sleep**

A comprehensive review of Parkinson's disease and sleep has been given in various studies (Freedom, 2007; Happe, 2003; Clarenbach, 2000). Patients with Parkinson's disease are prone to sleep disturbances (Thorpy, 2004). A positive correlation between severity of Parkinson's disease and sleep disruption has been found (Comella, 2006). Another sleep related complaint which occurs in Parkinson's disease patients is nocturia and urinary incontinence (Chaudhuri et al., 2001). Two studies report high incidents of sleep disturbance or sleepiness of 82% and 98% in Parkinson's disease patients (Oerlemans and de Weerd, 2002; Factor et al., 1990). It has also been reported that there is a increased number of arousals, a lower sleep duration and efficiency as well as an increased sleep-related breathing disorder and sleep behaviour disorder in Parkinson's disease patients (Clarenbach, 2000). Such sleep disturbances may lead to an explanation for the increased daytime sleepiness (EDS) in Parkinson's Disease patients (as discussed in section 2.6.2.2).

EDS has been shown to increase in severity with increasing degradation due to Parkinson's disease (Gjerstad et al., 2002). An increase from 7.7% to 28.9% of sleepiness in patients with Parkinson's disease over four years has been reported (Gjerstad et al., 2002). EDS was reported to be prominent in over 15% of subjects compared to 1% of healthy controls (Tandberg et al., 1999). Another study reported a smaller difference in age matched controls, 19.9% EDS in Parkinson's disease patients compared to 9.8% of control subjects (Tan et al., 2002), however there is still a marked increase in EDS in those with Parkinson's disease. Sleep attacks, similar to narcolepsy, have also been reported in 20% of patients with Parkinson's disease (Roth et al., 2003). The occurrence

of both obstructive and central sleep apnoea increases with the prevalence of Parkinson's disease (Basta et al., 2003).

Many studies have highlighted the occurrence of RBD in parkinson's disease; RBD is even thought to be an indicator of parkinson's disease and other neurodegenerative diseases (briefly discussed previously in section 2.6.2.3).

Some studies note a possibility for an association between RLS and Parkinson's disease, while other don't show any association. PLMS has been shown to be more common in patients with neurodegeneration than in controls (Freedom, 2007). Medication treating neurodegenerative disorders, particularly Parkinson's disease, has been shown to lessen the effects of sleep degradation most notably caused by RLS (Rye, 2004).

### **2.6.2.9 The Effect of Alzheimer's Disease on Sleep**

Individuals with Alzheimer's disease (a neurodegenerative disease) often show a decreased SE, more frequent awakenings and arousals, and a decreased TST. Other features are nocturnal insomnia, nocturnal wandering (and an increase in nocturnal wandering), increased number of daytime naps, increased TIB and increased time spent awake in bed (Crowley, 2011; Pollak and Perlick, 1991). Patients with Alzheimer's disease often exhibit circadian rhythm disorders where a consolidated nocturnal sleep no longer occurs (Lockley, 2009). Often this results in the institutionalisation of the individual as a single caregiver cannot provide adequate care 24 hours per day (Lockley, 2009). Restoration of light/dark cycles (possibly by an intervention technology) results in more consolidated rest/activity patterns and can help alleviate these symptoms (Lockley, 2009).

### **2.6.2.10 Insomnia**

A difficulty in falling asleep or maintaining sleep for at least one month, with a resulting impairment in daytime functioning is defined as insomnia (Crowley, 2011). This can either be secondary to another medical condition, or exist alone (primary insomnia). Insomnia is estimated to be as prevalent as 40 % in those over 65 years of age (Foley et al., 2004), although this varies between 4 and 11 % for the general population (Crowley, 2011). Insomnia patients report problems with memory, concentration, slower reaction times, poor attention, higher risk of falling (resultant from the use of hypnotics) and impaired cognition (Foley et al., 1995; Ancoli-Israel, 2005; Ancoli-Israel et al., 2005; Roth and Ancoli-Israel, 1999).

## 2.7 Conclusion

An introduction and foundation to sleep and sleep research are provided in this chapter. A working definition of sleep is given and an outline of the constituent stages of sleep is described (further details can be found in Iber et al. (2007); Malim and Birch (2000)). The rules for scoring sleep into its constituent stages, and the general patterns in which they occur, are also given. Other physiological changes which occur during sleep, and various sleep stages, are outlined. Commonly used metrics describing sleep (derived on a nightly basis from sleeping patterns and sleep stage information) are also introduced.

Sleep is an important physiological process and any degradation in its quality or quantity can result in poor health. This chapter discussed the negative effects various sleep disorders, common illnesses and diseases (particularly in older adults), and ageing has on sleep, and as a result their effects on health.

In Chapter 4, an in-depth description of the traditional clinical and ambulatory technologies used to monitor sleep will be presented and a review of the recent advances in sleep monitoring will also be given. Prior to this, Chapter 3 provides a description of discriminating algorithms used in this thesis and introduces commonly used performance metrics. These are illustrated using a simulated data set.

## Chapter 3

# Classification Methods

This chapter provides an outline of the methods used in this thesis to discriminate between sleep and wake, and between sleep stages. It also details strategies which are employed to ensure the generation of valid results. Such a consistent approach is particularly warranted due to the multiple sleep stage inferencing methods and sleep measurement technologies currently under investigation. Additionally, multiple performance metrics are used to report the accuracy of a classifier and many are reported inconsistently across the literature making comparisons difficult. The common performance metrics reported in the literature are presented in this chapter.

### 3.1 Data Sets

Five data sets were artificially generated using a randomised procedure (as shown in Figure 3.1) to illustrate the differences between the various classification algorithms described later. In the first *separate sides* data set (Figure 3.1(a)), data are separated by a relatively large distance. In the second *straight split* data set (Figure 3.1(b)), the horizontal distance between the data sets was reduced dramatically until a straight line could no longer be used to separate both classes. The classes in the third data set (*diagonal split*) could also be discriminated using a straight line (see Figure 3.1(c)). However, in previous examples one feature could discriminate between the classes whereas a linear combination of both features is required to separate both classes for these data. The remaining two data sets were designed to be increasingly difficult to classify accurately. In data set 4 *fully surrounded* (see Figure 3.1(d)), data from one class was completely surrounded by the other class. In the last data set *partially surrounded* (see Figure 3.1(e)), data from one class exists at the edge of the bounds of the data set and surrounded on all other sides by the other class.

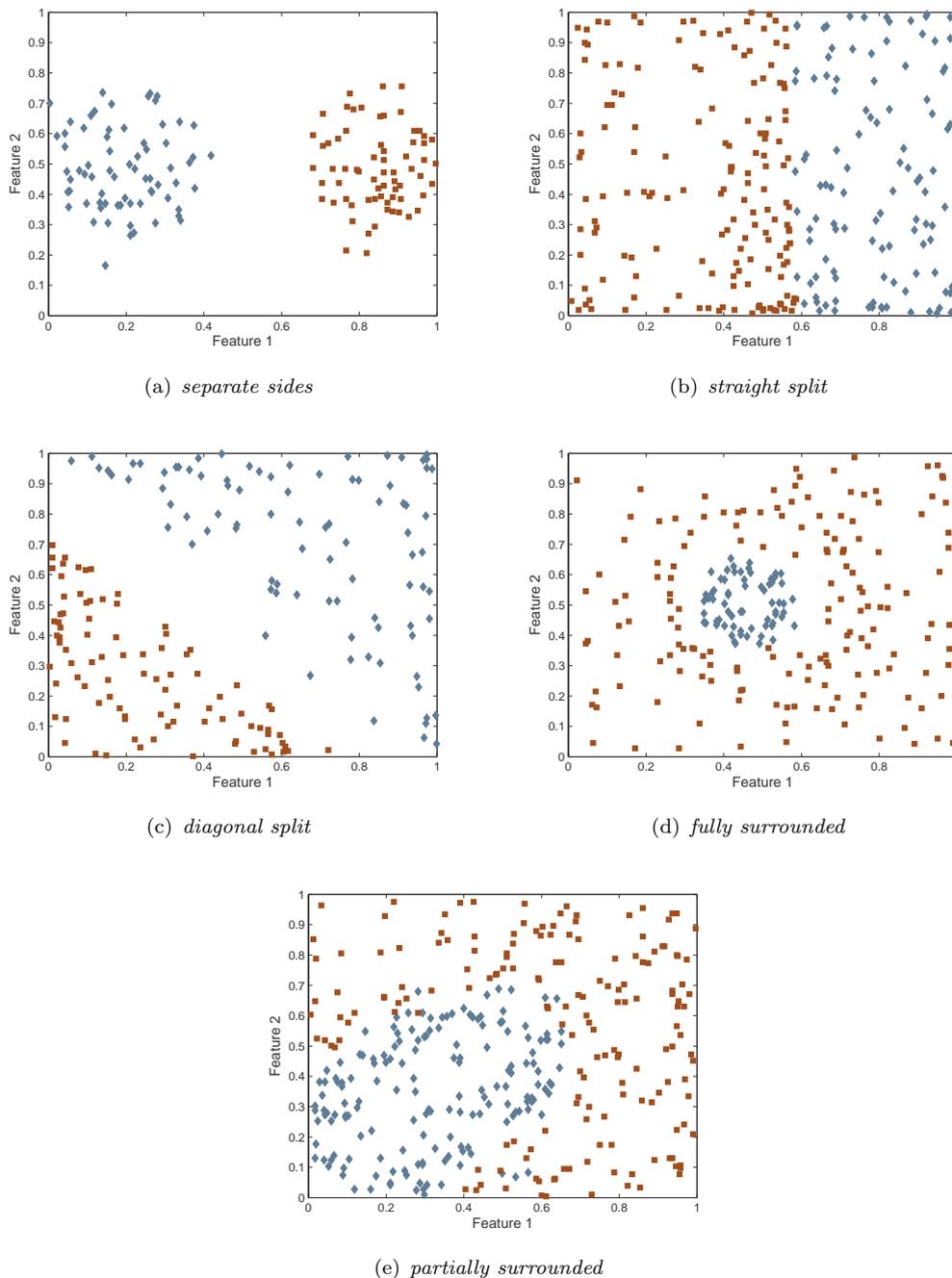


FIGURE 3.1: Five artificially generated data sets containing data from two classes (grey and brown respectively).

TABLE 3.1: Determination of TP, FP, FN and TN

		Actual Class	
		1	0
Predicted Class	1	TP	FP
	0	FN	TN

### 3.2 Performance Metrics

Many metrics may be used to report the performance of a classification algorithm including accuracy, Misclassification Rate (MCR), sensitivity, specificity, precision, recall, Matthew’s Correlation Coefficient (MCC), F Score and Cohen’s Kappa. Accuracy (defined as the number of correctly classified samples over the total number of samples) provides a good indication of the overall performance of a system, however it is unsuitable when the data set is biased towards a particular class. For example, a biased data set is often present during sleep recordings as the subject can be asleep for over 90% of the period under investigation. Thus, if a test, using such data, states that the person is always asleep, without interrogating that data, an high accuracy of 90% would be reported. However, the system would never report when the subject is awake. As such, other metrics, which report the accuracy of the system in detecting both a positive and negative case, are often used to report the system performance.

In order to fully explain each of these metrics, the reader must first become familiar with the concept of True positives (TP), False positives (FP), False negatives (FN), and True negatives (TN). TP occur when a sample is correctly predicted (in this case by a classifier) as belonging to the positive class, while TN occur when a sample is correctly predicted as belonging to a negative class. FP are when a sample is incorrectly predicted to belong to the positive class, while FN occur when a sample is incorrectly predicted as belonging to the negative class. This is further illustrated in Table 3.1.

The mathematical formulation of each of the performance metrics is given below.

**Accuracy:** Accuracy, as defined above is the number of correctly classified samples ( $TP + TN$ ) over the total number of samples, is a valid measure of performance when the data set is evenly distributed between both states (sleep and wake in this instance). However the performance of the accuracy metric suffers in cases where the data set is biased towards a particular state. For example when 90% of the data set is sleep, if a

classifier states that the entire data set is sleep an accuracy of 90% will be reported.

$$Accuracy = \frac{(TP + TN)}{(TP + FN + FP + TN)} \quad (3.1)$$

**Misclassification Rate:** MCR is a measure of the number of incorrectly classified samples and is directly related to accuracy as  $(1 - Accuracy)$ . Its suitability also suffers when the data sets is biased towards one particular state.

$$Misclassification Rate (MCR) = \frac{(FP + FN)}{(TP + FN + FP + TN)} \quad (3.2)$$

**Sensitivity, Specificity and Precision:** Sensitivity and specificity are measures of the accuracy of the system in detecting true positives and true negatives. For this application, sensitivity refers to the number of true sleep epochs which are correctly classified as sleep, while specificity refers to the number of true wake epochs correctly labelled as wake. Sensitivity is also known as recall and hit rate. Precision measures the number of samples the classifier correctly predicted as a proportion of the total number of samples the classifier predicted to be positive. Putting this into context, it refers to the number of samples the system correctly identified as sleep over the total number of sleep samples the system predicted. Each of these measures are not affected whether the data set is biased or not.

$$Sensitivity(recall) = \frac{TP}{(TP + FN)} \quad (3.3)$$

$$Specificity = \frac{TN}{(FP + TN)} \quad (3.4)$$

$$Precision = \frac{(TP)}{(TP + FP)} \quad (3.5)$$

**Cohen's Kappa ( $\kappa$ ):** Cohen's Kappa is a statistic used to assess inter-judge agreement and is used for nominally coded data Cohen (1960). In the context of sleep/wake discrimination only two output cases are allowed (although it is still applicable when more output cases exist). If both raters were in complete agreement Cohen's Kappa ( $\kappa$ ) would be one, however if there was no agreement Cohen's Kappa ( $\kappa$ ) would be zero. This test is insensitive to a biased data set.

$$\text{Probability of observed agreement } (Po) = \frac{(TP + FP)}{(TP + FN + FP + TN)} \quad (3.6)$$

$$\text{Probability expected by chance } (Pc) = \frac{(TP + FN)}{(TP + FN + FP + TN)} \times \frac{(TP + FP)}{(TP + FN + FP + TN)} \quad (3.7)$$

$$\text{Cohen's Kappa } (\kappa) = \frac{Po - Pc}{1 - Pc} \quad (3.8)$$

**F1-Score** The F1-Score van Rijsbergen (1979) is a measure of a tests accuracy and uses a weighted average of (the previously defined values of) precision and recall, specifically the harmonic mean,

$$F \text{ Score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.9)$$

This metric is also insensitive to a biased data set.

**Matthew's Correlation Coefficient** MCC is another metric of the quality of agreement between two classes which is not sensitive to a biased data set. Values may range from -1 (complete disagreement) to +1 (complete agreement) with 0 representing no better prediction than by chance.

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FN)(TP + FP)(FP + TN)(TN + FN)}} \quad (3.10)$$

### 3.3 Classifiers

A brief review of the theory behind the classification algorithms employed in this thesis is given below. Additionally, artificially generated data (described in Section 3.1) are used to provide a more intuitive understanding of the inner workings of each classifier as well as some of the advantages and disadvantages of applying each classifier. Firstly however, a common approach used to ensure valid classification results is briefly described.

#### 3.3.1 Training, Validation, Testing Data and Cross Validation

*Training data* (a data set of inputs with known outputs, or states) is used to create a function which can discriminate between two or more classes of data. During that classification process the *hyperparameters* (the pre-configured set-up of the classifiers, such as the structure of a neural network) and the *parameters* (the internal parameters tuned during the classification process) are optimised. An optimally tuned classifier may report excellent performance on the data it has been trained upon. However the

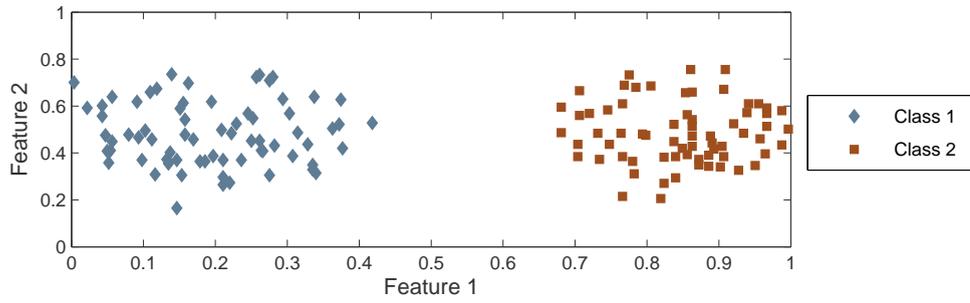
discriminating function developed may be too specialised upon the training data and, as such, report significantly poorer performance when applied to unseen data (a data set external to the training data commonly referred to as *validation data*). The process of ensuring that the classifier does not become too *overfitted* on the specific variations of the training data is referred to as *generalisation*. The *generalisability* of a classifier is optimised during the training procedure. This is performed using unseen data during or after the training process. For computationally inefficient algorithms, the training procedure is often stopped early when the performance of the classifier, on the validation data, decreases. Often multiple classifiers are tested and the performance curves (from both the training and testing data sets) are analysed in order to find the optimal choice of classifier (and choice of internal classification parameters, where appropriate).

*Cross validation* is an additional methodology used to select the optimal classifier (and internal parameters of a classifier) which performs well, on both training and validation data. This process consists of training the classifier multiple times over different combinations of samples in the training and validation data sets which produces multiple sets of results. Through re-running the classification process on different combinations of samples in the training and validation sets, the bias inherent in a particular set of training samples or in a specific set of random chosen initial variables can be controlled for. The overall training and validation performance is averaged for all classification iterations in order to cater for any such bias. However a computational overhead is associated with each cross-validation, and for certain classifiers (and configuration of classifier) this limits the number of runs.

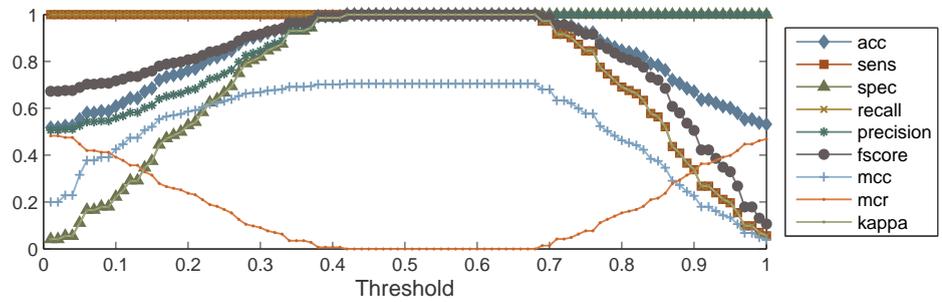
Often multiple classifiers (with various combinations of internal parameters) are trained and validated using this process. However, this also introduces a bias, as one data set has been continually split used to generate this optimal classifier (and internal parameter) selection. As such, the results are not completely independent of the training procedure. Accordingly, additional and completely separate *testing data* is sometimes used to generate independent results which quantifies the performance of the optimal classifier (found during the training/validation procedure). The testing data set is only applied to the optimal classifier. This final test step is often not employed in practice as this requires a very large data set which may not be practical.

### 3.3.2 Binary Classification Test

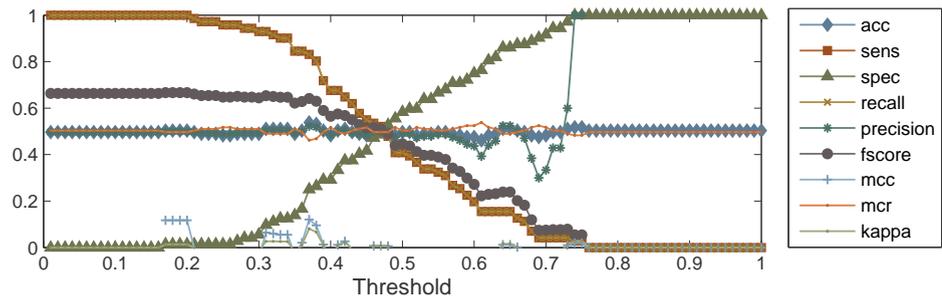
A Binary Classification Test (BCT) is one of the simplest methods for discriminating between two classes. This algorithm finds the optimal threshold value (of a feature) which best separates two classes. This discriminator works in situations where classes



(a) *separate sides* data



(b) Results from BCT on Feature 1



(c) Results from BCT on Feature 2

FIGURE 3.2: Application of BCT classifier on *separate sides* data set

can be separated by applying a threshold to a single feature. It is ideal for the *straight split* and *separate sides* data sets (ie. where a thresholded value of Feature 1 can split the data into the two separate classes). Figure 3.2 presents results from the application of a BCT on both features in the ‘*separate sides*’ data set. Maximum performance is found for a range of values over Feature 1 (see figure 3.2(b)) and poor performance is reported for Feature 2 (see figure 3.2(c)). The optimum threshold may be chosen as the point which maximises either the overall accuracy, sensitivity, specificity, recall, precision, f-score, MCC, MCR or Cohen’s Kappa ( $\kappa$ ). For the *separate sides* data set, this occurs at many places.

However for cases where more than one feature is required to separate the classes (for example, *diagonal split*, see Figure 3.1(c)), BCT is not suitable (as a linear combination

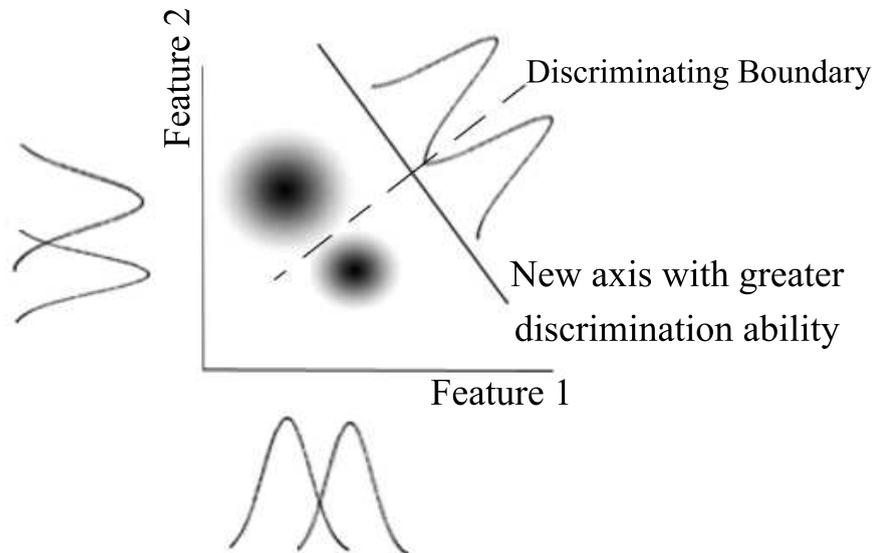


FIGURE 3.3: Distribution of data (shown using the Gaussian bell curves) from two classes on the original features and on a new axis which optimally separates the classes. The minimum overlap between the distributions occurs for this new axis.

of more than one features is required to accurately split the data into the separate classes). Additionally, for cases where a straight line will not discriminate between classes, this type of classifier will perform very poorly (for example the *fully surrounded* and *partially surrounded* cases, Figures 3.1(d) and 3.1(e)).

### 3.3.3 Discriminant Analysis

Discriminant Analysis predicts data as belonging to between two or more classes using distributions of that data. Two side-by-side normal distributions show the projection of simulated data (in this case modelled using a Gaussian distribution) from two classes onto the two original features (see Figure 3.3). Due to the overlap between the distributions, we cannot say with certainty to which class every sample belongs when investigating the two original features separately. However, the two classes may be discriminated when projected onto a new axis (containing a combination of Feature 1 and Feature 2) which minimises this overlap (see Figure 3.3). As a result, we are much more confident in predicting a class to which a sample belongs (including an unseen sample).

In order to achieve this, discriminant analysis models the data (producing a model similar to the Gaussian distributions shown above) using a training set in order to find the optimal axis (and discriminating boundary) which separates the classes. In order to find this axis, discriminant analysis maximises the ratio of the between-class variance to within-class variance. If the data from each class were normally distributed (with an

equal number of samples and spread), the discriminating boundary would occur along a line which runs equidistant from both clusters of data (for the two class case). However in practice, the distribution of both classes is often non-normal and accordingly the spread of the data must be taken into account to find the optimal discriminating boundary (which in such a case would not be equidistant from each cluster centroid).

### 3.3.3.1 Mathematical Derivation

Given a single measurement of data,  $\mathbf{x}$ , the probability of the current class being  $R_\kappa$  is  $P(R_\kappa | \mathbf{x})$ . However, the range of  $\mathbf{x}$  is vast, and calculating  $P(R_\kappa | \mathbf{x})$  for each  $\mathbf{x}$  will be a lengthy process; Bayes' rule (see Equation 3.11) may be used to simplify this calculation:

$$P(R_\kappa | \mathbf{x}) = \frac{P(\mathbf{x} | R_\kappa) P(R_\kappa)}{P(\mathbf{x})} \quad (3.11)$$

where  $P(R_\kappa)$  is the probability of class  $R_\kappa$ ,  $P(\mathbf{x})$  is the probability of the sample  $\mathbf{x}$  occurring and  $P(\mathbf{x} | R_\kappa)$  is the probability of the data occurring given the class.  $P(R_\kappa)$  is usually assumed to be equal across all classes (where no data exists to suggest otherwise).  $P(\mathbf{x})$  may be calculated using

$$P(\mathbf{x}) = \sum_{\kappa=1}^M P(\mathbf{x} | R_\kappa) P(R_\kappa) \quad (3.12)$$

where  $M$  is the total number of classes.

The distribution of data for each class may be modelled using a Gaussian function as:

$$P(\mathbf{x} | R_\kappa) = \frac{1}{(2\pi)^{\frac{k}{2}} |\boldsymbol{\Sigma}_\kappa|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\kappa)^T \boldsymbol{\Sigma}_\kappa^{-1} (\mathbf{x} - \boldsymbol{\mu}_\kappa)\right) \quad (3.13)$$

where  $\mathbf{x}$  is the data,  $R_\kappa$  is the current class,  $P(\mathbf{x} | R_\kappa)$  is the probability of the data for a given class  $\kappa$ ,  $\boldsymbol{\mu}_\kappa$  is the mean of the data for each class,  $\boldsymbol{\Sigma}_\kappa$  is the covariance of the data for a class  $\kappa$ ,  $|\boldsymbol{\Sigma}_\kappa|$  is the determinant of the covariance for a particular class.

The relative probability of class  $\kappa$  occurring ( $P(R_\kappa)P(\mathbf{x} | R_\kappa)$ ) can be expressed as

$$\delta_\kappa(\mathbf{x}) = P(R_\kappa) \frac{1}{(2\pi)^{\frac{k}{2}} |\boldsymbol{\Sigma}_\kappa|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\kappa)^T \boldsymbol{\Sigma}_\kappa^{-1} (\mathbf{x} - \boldsymbol{\mu}_\kappa)\right) \quad (3.14)$$

or using the log likelihood,

$$\delta_\kappa(\mathbf{x}) = \log \left[ P(R_\kappa) \frac{1}{(2\pi)^{\frac{k}{2}} |\boldsymbol{\Sigma}_\kappa|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\kappa)^T \boldsymbol{\Sigma}_\kappa^{-1} (\mathbf{x} - \boldsymbol{\mu}_\kappa)\right) \right] \quad (3.15)$$

This may be simplified further to,

$$\delta_\kappa(\mathbf{x}) = \log(P(R_\kappa)) - \frac{1}{2} \log(|\Sigma_K|) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\kappa)^T \Sigma_\kappa^{-1}(\mathbf{x} - \boldsymbol{\mu}_\kappa) \quad (3.16)$$

The optimal discriminant function occurs for,

$$D(\mathbf{x}) = \arg \max_{\kappa} (\delta_\kappa(\mathbf{x})) \quad (3.17)$$

or,

$$D(\mathbf{x}) = \arg \max_{\kappa} \left( \log(P(R_\kappa)) - \frac{1}{2} \log(|\Sigma_K|) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\kappa)^T \Sigma_\kappa^{-1}(\mathbf{x} - \boldsymbol{\mu}_\kappa) \right) \quad (3.18)$$

**Linear Discriminant Analysis:** Linear Discriminant Analysis (LDA) assumes that the distribution of the data for all classes (the covariance,  $\Sigma$ ) is the same which simplifies the discriminating function as  $\frac{1}{2} \log(|\Sigma_K|)$  is constant across all classes (producing the decision function in equation 3.19). The point  $\mathbf{x}$  is estimated as belonging to the class,  $\kappa$ , which results in the largest  $D_{LDA}(\mathbf{x})$  value. The discriminating boundary may be calculated as the points which result in equal  $D_{LDA}(\mathbf{x})$  values for both classes.

$$D_{LDA}(\mathbf{x}) = \arg \max_{\kappa} \left( \log(P(R_\kappa)) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\kappa)^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_\kappa) \right) \quad (3.19)$$

**Quadratic Discriminant Analysis:** Quadratic Discriminant Analysis (QDA) calculates the covariance of the data for each class and uses these to calculate the discriminant function (as given in Equation 3.20) using the method as described for the linear case. In this case the discriminating function will be non-linear.

$$D_{QDA}(\mathbf{x}) = \arg \max_{\kappa} \left( \log(P(R_\kappa)) - \frac{1}{2} \log(|\Sigma_K|) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_\kappa)^T \Sigma_\kappa^{-1}(\mathbf{x} - \boldsymbol{\mu}_\kappa) \right) \quad (3.20)$$

For both LDA and QDA, the mean ( $\mu$ ) and covariance ( $\Sigma$ ) are calculated using,

$$\mu = \sum_{g_i=\kappa} x_i / N_\kappa \quad (3.21)$$

$$\Sigma = \sum_{\kappa=1}^K \sum_{g_i=\kappa} (x_i - \mu_k)(x_i - \mu_k)^T / (N - K) \quad (3.22)$$

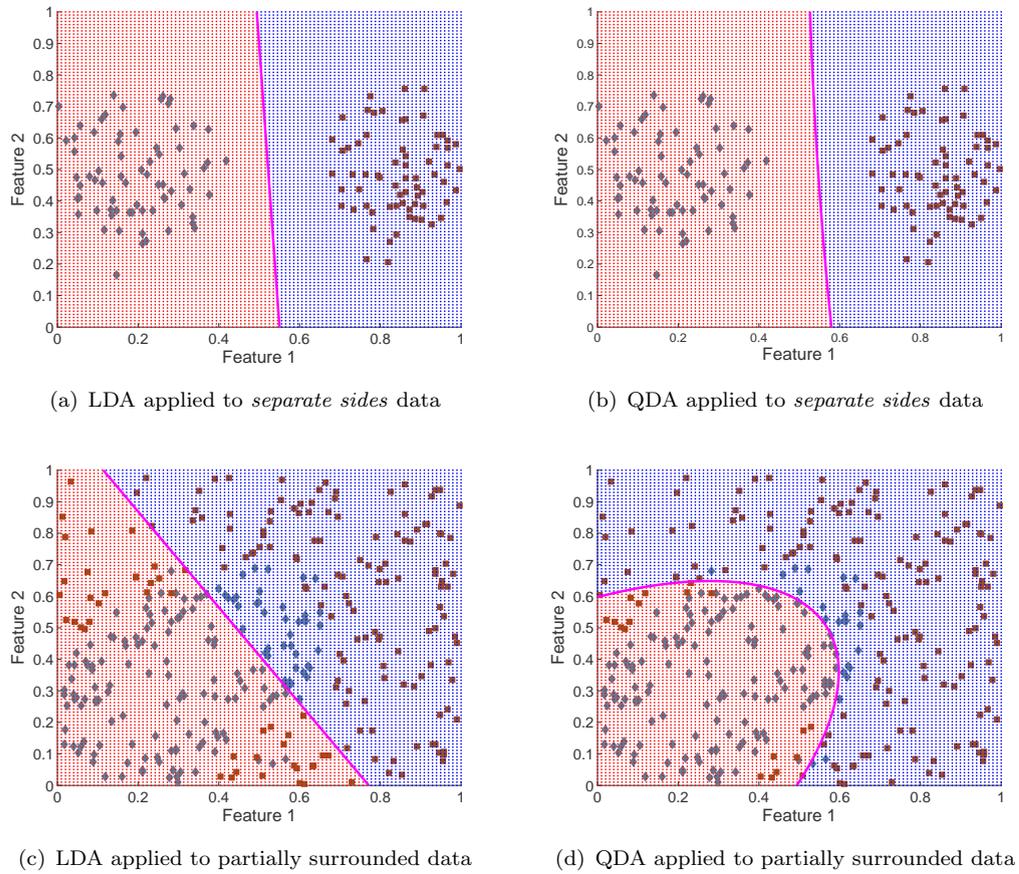


FIGURE 3.4: Decision regions (marked by blue and red backgrounds) and a discriminating boundary (pink) of LDA and QDA classifiers on two data sets (containing two classes of data marked by brown square and light blue data diamond points).

### 3.3.3.2 Application of LDA and QDA

LDA and QDA were applied to two of the sample data sets (as shown in Figure 3.4). For the separate side data set, a straight line will easily discriminate between both classes of data. There is no added advantage in using a quadratic discriminator (as demonstrated in Figure 3.4(b)). The partially surrounded data set is more difficult to separate and its performance is poor using LDA (see Figure 3.4(c)). However, a quadratic line provides a greater discriminative capacity as shown in Figure 3.4(d).

### 3.3.3.3 Discussion

Discriminant analysis estimates the covariance matrix from the training samples. For LDA the same covariance matrix is estimated and used to calculate a discriminating function which is linear. For QDA, the covariance matrices are calculated separately for each class of data. This results in a quadratic discriminating hyperplane which may

provide better performance for non-Gaussian and closely inter-mingled data sets. As such, QDA may fit the data better than LDA, however more parameters are estimated for QDA than for LDA (which may reduce the performance of the classifier). QDA is more computationally complex than LDA. Higher dimensional hyperplanes can be also be generated however they are outside the bounds of this work. The order of the hyperplane (where 1 is linear, 2 is quadratic, etc.) is referred to as a hyperparameter and is set prior to training.

Another method of performing non-linear discriminant analysis involves transforming the data to a higher dimensional hyperspace and then discriminates the data using a linear hyperplane (performing LDA in the higher dimensional feature space). This has been found to provide similar classification flexibility to the Gaussian QDA method (Hastie et al., 2001). The reader is directed to Hastie et al. (2001) for a more detailed discussion on this and on discriminant analysis in general.

### 3.3.4 $k$ -Nearest Neighbour

The  $k$ -Nearest Neighbour (kNN) classifier is a simple, non-linear classification method which generally results in a relatively high accuracy. It is a non-parametric, memory-based algorithm which classifies new samples based on how closely they occur (often using a Euclidian distance metric) relative to samples in a training set. Generally, a large training set is used as this captures a distribution representative of the occurrence of classes in the feature space. The Euclidian distance is calculated between each test sample that is to be classified ( $\check{x}$ ) and each sample in the training data set (as given in Equation 3.23).

$$e_n = \sqrt{(x_n^1 - \check{x}^1)^2 + \dots + (x_n^d - \check{x}^d)^2 + \dots + (x_n^D - \check{x}^D)^2} \quad (3.23)$$

where the training set consists of  $N$  sample vectors with  $D$  features,  $n \in \mathfrak{R}, 1 \leq n \leq N$ ,  $1 \leq d \leq D$  and  $e_n$  is the root mean squared distance between the test sample and each point in the training set.

The vector of the distance metric is arranged in ascending order (from small to large) and each sample has a corresponding class ( $R_\kappa$ ). This effectively arranges the training samples in order of their proximity to the test sample. The nearest  $k$  samples (belonging to one class) define which class the test samples is predicted as belonging to. The value of  $k$  is a hyperparameter and can be optimised over multiple iterations.

### 3.3.4.1 Discussion

$k$  is a user-defined value which controls the flexibility and generalisability of the decision regions (the area in the feature space which are related to a particular class). For example, if  $k$  is too low the accuracy of the classifier becomes very dependent on the training data. However, if  $k$  is too large the classifier may become too generalised. In practice,  $k$  is often optimised using a cross-validation approach. The results of applying kNN is illustrated in Figure 3.5. The non-linear decision boundaries are particularly evident for the closely intermingled data. With increasing  $k$ , the decision boundary can be seen to become less complex and more generalisable (and more representative of the underlying structure of the data).

The kNN algorithm has a very high computational complexity (as the distance to each sample in the training matrix must be calculated for every test sample). This is particularly problematic for large data sets. Many other classifiers rely on representations of the data sets which reduces the complexity of the algorithms.

### 3.3.5 Artificial Neural Networks

Artificial Neural Networks (ANN) are a method of processing information which mimics the technique used by the human brain. It consists of an interconnected network of nodes, or neurons, in which a learning strategy developed over time allows information to transfer throughout the network. Signals are allowed to proliferate across the network based on the previous amount of information transfer between specific neurons. This is achieved biologically, in the brain, by lowering the resistance of a pathway between two neurons iteratively over time as signals are sent across that pathway. A mathematical implementation achieves this by weighting each interconnection between each neuron resulting in a process which either allows the signal to pass or not (McCulloch and Pitts, 1943; Rosenblatt, 1962). By passing signals through multiple interconnecting internal layers of neurons with non-linear activation functions (for example a sigmoid activation function), non-linear relationships between the inputs and outputs to the network can be generated.

#### 3.3.5.1 The Mathematical Representation of a Neuron

The McCulloch-Pitts implementation produces an output using the summation of a weighted version ( $w_i$ ) of all inputs ( $x_i$ ) as well as a bias ( $b$ ) offset (as shown in Figure 3.6) (McCulloch and Pitts, 1943). The inputs may be external to the network or from

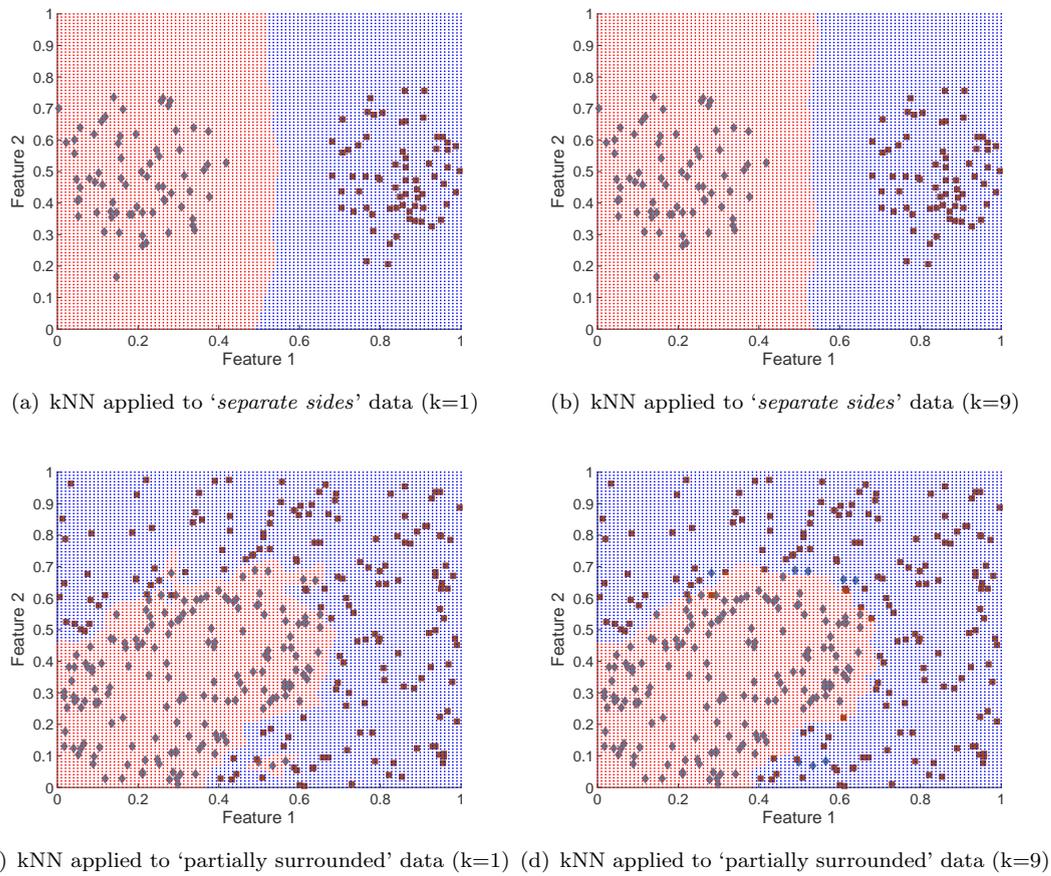


FIGURE 3.5: Decision regions (marked by blue and red backgrounds) of kNN classifiers on two data sets (containing two classes of data marked by brown square and light blue data diamond points).

other neurons. The output is generated using an activation function  $f(\cdot)$ . This activation function may be either linear or non-linear (such as a step function).

### 3.3.5.2 Multilayer Perceptrons

Multiple neurons may be arranged in a network of interconnecting nodes in order to approximate complex functions similar to the neuronal structure in the brain. Multiple layer neural networks with linear activation functions can be simplified to single layer networks (although these networks remain incapable of representing certain functions such as the XOR function) (Minsky and Papert, 1969). However, the use of non-linear activation functions overcomes this limitation and allows the approximation of an extended set of functions.

MLP neural networks consist of three separate layers of neurons: 1) the input layer (where the inputs to the network are directly connected), 2) the hidden layer (which

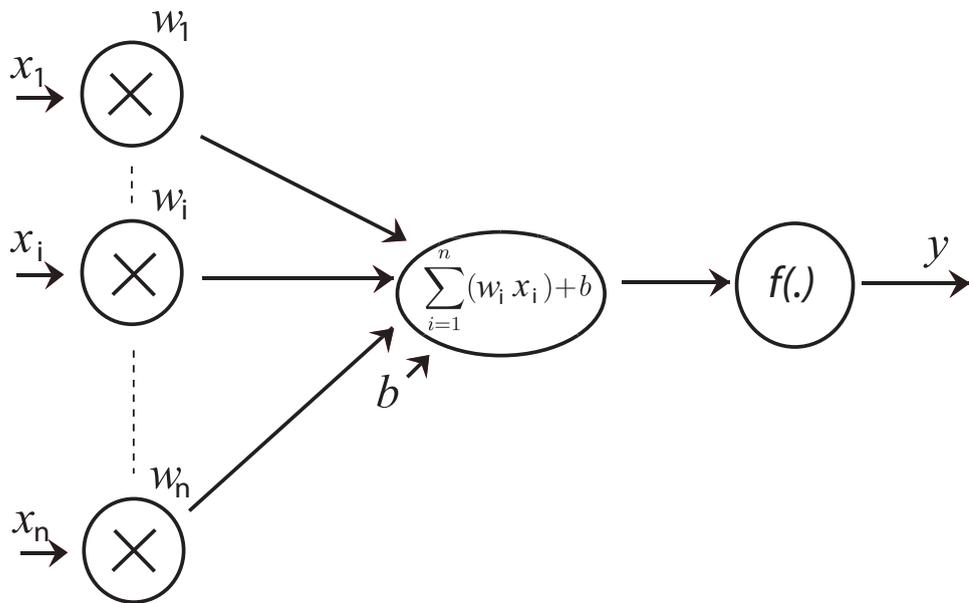


FIGURE 3.6: The McCullough-Pitts neuron.

consists of one or more layers), and 3) the output layer (connected directly to the outputs). The hidden layer may itself consist of multiple layers. Each layer consists of multiple neurons and the neurons in each layer are interconnected to each neuron in the previous layer (as shown in Figure 3.7), and this is commonly referred to as a feedforward neural network. Hyperparameters define the overall structure of the neural network (including the number of neurons in each layer and the number of hidden layers). These are set prior to training the ANN.

Although MLP neural networks have been shown to be able to approximate almost any continuous function (Irie and Miyake, 1988; Funahashi, 1989), in practice it may be prohibitive to approximate some functions due to computational constraints such as processing power and time. Additionally, this may require the collection of large amounts of data which might be unrealistic in practice (Zhang et al., 2001).

Multiple MLP are commonly configured in networks forming ANN. Other implementations of ANN include radial basis function networks, Hopfield networks and self-organising feature maps, however the MLP is the most popular and is considered in this work.

### 3.3.5.3 Neural Networks: Learning and Storing Information

Information is stored in a neural network using the internal parameters: the weights and the biases. Through a training procedure, the biases to each neuron, and the weights

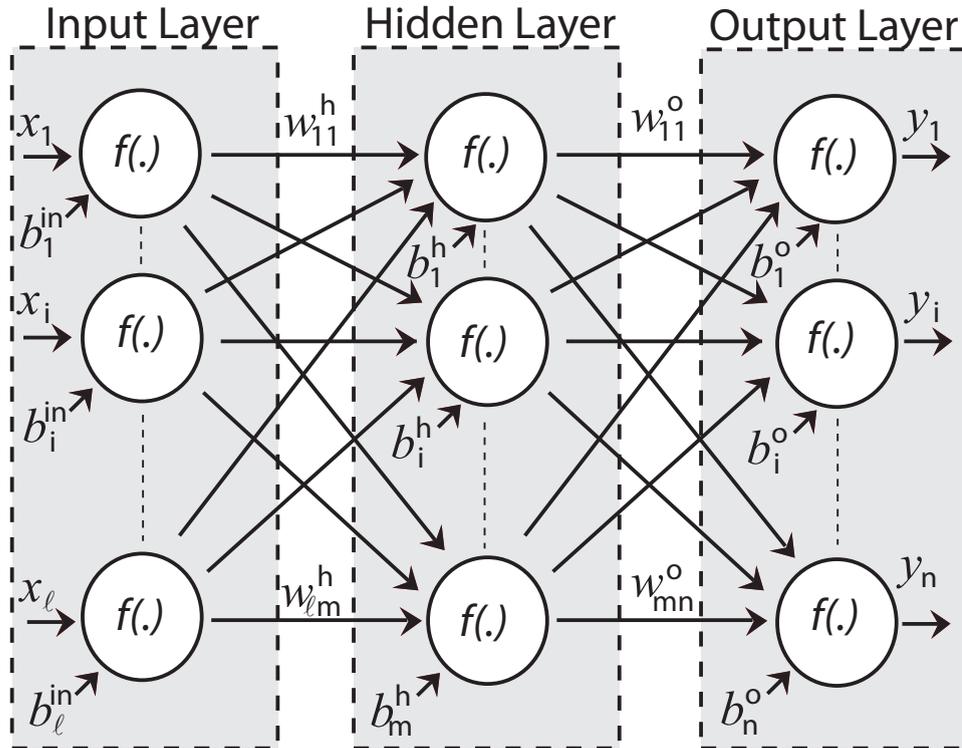


FIGURE 3.7: Multilayer perceptron ANN.

applied to each input and to each connection between each neuron are tuned. The optimal parameters are found when the error between the output of the neural network ( $\hat{y}$ ) and the desired output ( $y$ ) is minimised. In order to achieve this, a training data set is used where the output of the system is known for particular inputs. After the training procedure, the neural network should be able to correctly approximate the underlying relationship between the inputs and outputs for new unseen data.

The *backpropagation* algorithm was devised to tune the weights and biases for MLP neural networks (Rumelhart et al., 1986). As a first step, the weights are initialised randomly (although intelligent techniques may be used to select better initial parameters (Hernandez-Espinosa and Fernandez-Redondo, 2001)). Subsequently, the output of the network is calculated using the input data from the training data set for the initial weights ( $\hat{y}$ ). An error metric is generated using the difference between the desired (or correct) output ( $y$ ) and ( $\hat{y}$ ). The Sum of Squared Errors (SSE) is often used as this performance metric (or cost function), that is,

$$SSE = \sum_{i=1}^N \sum_{j=1}^M (\hat{y}_{ij} - y_{ij})^2 \quad (3.24)$$

where  $\hat{y}_{ij}$  represents the estimated value for the  $j$ th output at the  $i$ th sample instant,  $y_{ij}$  represents the desired value for the  $j$ th output at the  $i$ th sample instant,  $N$  is the

number of sample instants, and  $M$  is the number of outputs.

Learning occurs through modifying the internal parameters of the neural network such that the SSE is reduced. A *gradient descent* algorithm is often used to iteratively tune the weights until the SSE is minimised. The general form of the gradient descent algorithm for MLP neural network weight optimisation is,

$$\mathbf{w}(m+1) = \mathbf{w}(m) - v\Delta(m) \quad (3.25)$$

where  $m$  is the training iteration number,  $\mathbf{w}$  is a vector containing the weight and bias values,  $v$  is the learning rate.  $\Delta(m)$  is the gradient of the error function with respect to the weight vector,

$$\Delta(m) = \frac{\delta(SSE)}{\delta\mathbf{w}(m)} \quad (3.26)$$

This process is repeated until the minimal SSE has been found. This process can be applied to either all data in the training set at once (a batch process), or iteratively one sample at a time.

More computationally efficient techniques of finding the optimal weights have been proposed, such as the 2nd order gradient descent Broyden-Fletcher-Goldfarb-Shannon (BFGS) method (Barttiti and Masulli, 1990) or the Levenberg-Marquardt (LM) method (Hagan and Menhaj, 1994; Marquardt, 1963).

Often in practice, this training procedure is run multiple times per training structure in order to ensure the optimal weights are found. For some cases, the randomly assigned initial weights may produce a sub-optimal set of weights. In some cases, these may occur locally in the feature space as opposed to an optimal set of weights which would occur globally in the feature space. Randomly assigned initial conditions may result in slightly different weights when optimised using multiple runs. Additionally, many different MLP neural network structures may be tested in order to find the optimal configuration. For each test, the error metric, network configuration and internal parameters are recorded and compared. The network with the best performance is then retained for future use. When MLP neural networks are applied to classification, they may result in different decision regions for different initial configurations (as shown in Figure 3.8) and for different topologies (as shown in Figure 3.9).

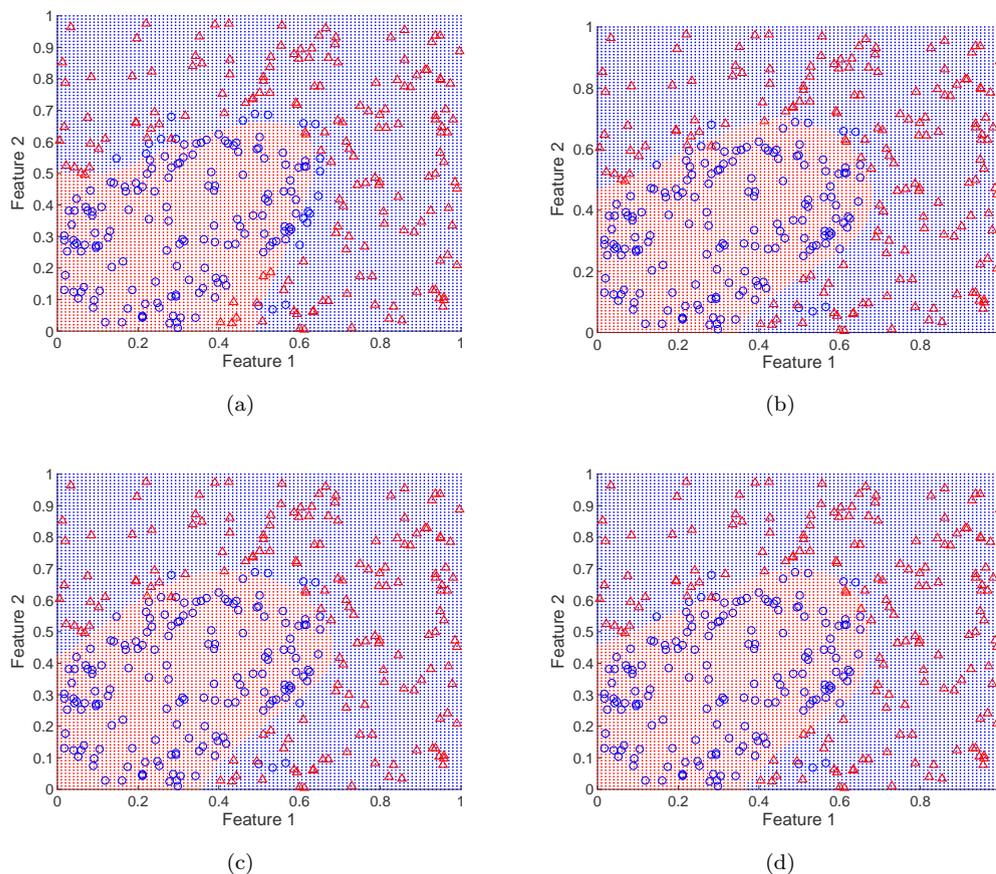


FIGURE 3.8: Decision regions (red and blue backgrounds) for four MLP neural network classifiers trained using the ‘partially surrounded’ data set and randomly assigned initial conditions. Different decision regions were found for each different set of initial conditions.

### 3.3.5.4 Discussion

ANN provide a black box system which models an input-output relationship of variables (capable of approximating non-linear functions) and no understanding of this relationship is required. While this may be extremely useful, many caveats exist. These include making an intelligent choice of the model topology for a system which is slightly non-linear, versus a system which is completely non-linear. An in-depth understanding of the system often proves useful, especially when large sets of unseen data are passed through the system. In order to reduce errors, the system is built and tested using training, testing and validation data. ANN are computationally intensive and typically require large training data sets to achieve optimal results (Cerny, 2001).

MLP neural networks can be designed as classifiers by specifying one or two outputs. In the case of one output, a sigmoid output function changes the probability of the output to be more likely to be 0 or 1. In practice, both of these states are set to represent

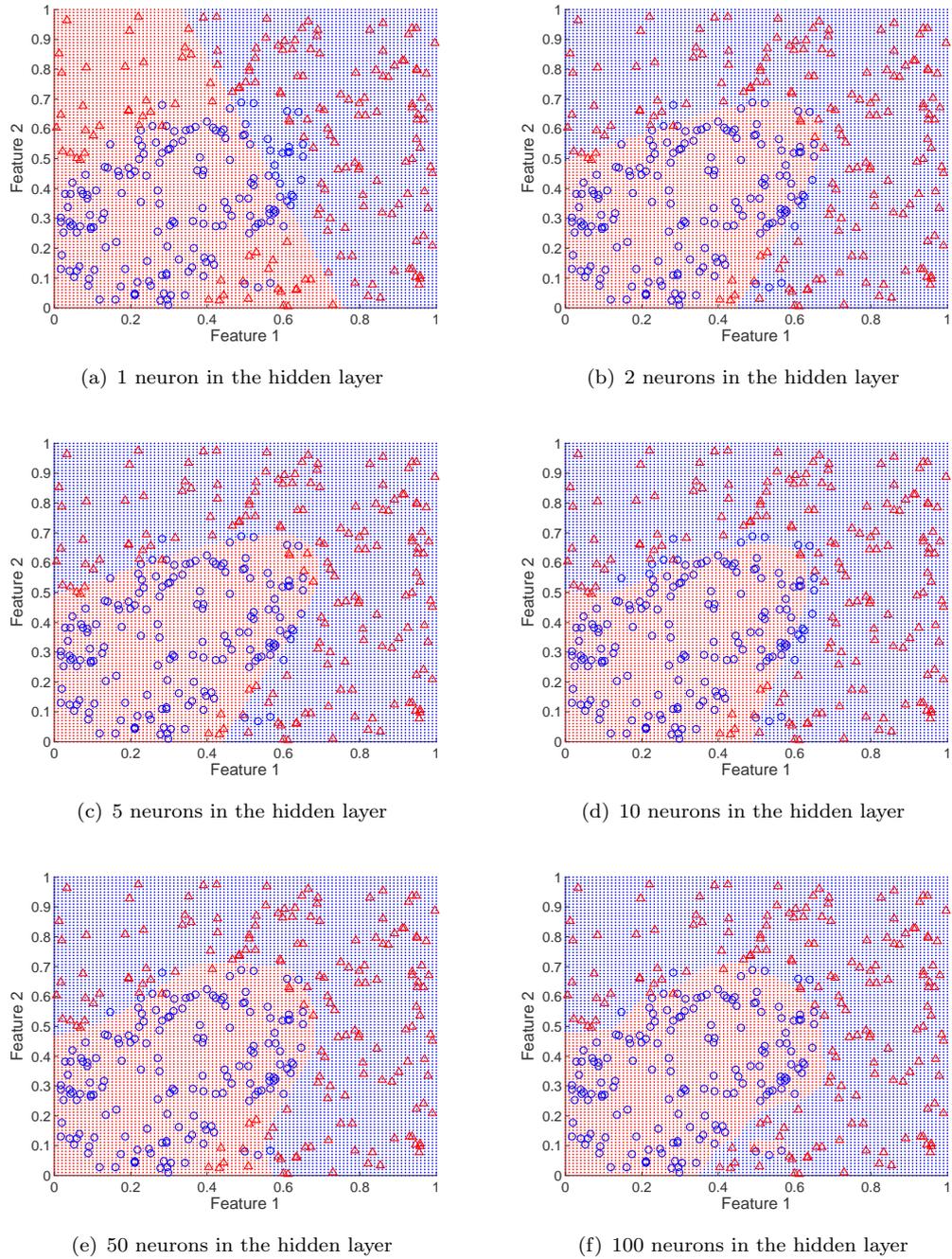


FIGURE 3.9: Decision regions (red and blue backgrounds) for six MLP neural network classifiers trained using the ‘partially surrounded’ data set and different topologies. The decision regions become increasingly complex (in terms of the function that is being approximated) as the number of neurons in the hidden layer increases.

both classes and training of the neural network is performed in the standard manner. The accuracy of this system is then tested using a separate data set. This approach was used when using an ANN classifier in this thesis. In the case of two separate outputs, one predicts the first class and the other output predicts the second class. Training and testing is performed as before, however four possible output may exist. The input may be classified as belonging to the first class, the second class, both classes, or neither class. No issues arise where only one class is selected. However for the two remaining two combinations, the score from each output neuron is a measure of the probability that the current sample belongs to the relevant class and, as such, the output with the highest probability is selected.

### 3.3.6 Support Vector Machines

Support Vector Machines (SVM) are another classification method which can separate classes using a non-linear boundary. However, SVM achieve this by increasing the dimensionality of the input feature space, and subsequently, a (possibly linear) hyperplane is used to maximally separate the classes in this higher dimensional feature space. Depending on the function used to increase the dimensionality, this may be equivalent to applying a non-linear hyperplane in the original feature space. A kernel function,  $K(x, y)$ , is used to translate the input to the higher dimensional space.

Initially, SVM can be explained using the linear case in the original feature space. Suppose, a set of data containing two features may be separated using a linear decision boundary (as defined by the hyperplane  $\mathbf{w}^T \cdot \mathbf{x} + b = 0$ , where  $\mathbf{w}^T$  is the slope of the hyperplane,  $\mathbf{w}$  is the normal vector, and  $b$  is the bias), two hyperplanes may be defined on either side of this decision boundary of maximal distance away from the decision boundary (Equations 3.27 and 3.28) as shown in Figure 3.10. These two hyperplanes pass through points (or vectors) closest to the decision boundary, and the distance from each hyperplane to the decision boundary is known as the margin. SVM maximise this margin in order to define the optimal decision boundary.

$$\mathbf{w}^T \cdot \mathbf{x} + b = 1 \tag{3.27}$$

$$\mathbf{w}^T \cdot \mathbf{x} + b = -1 \tag{3.28}$$

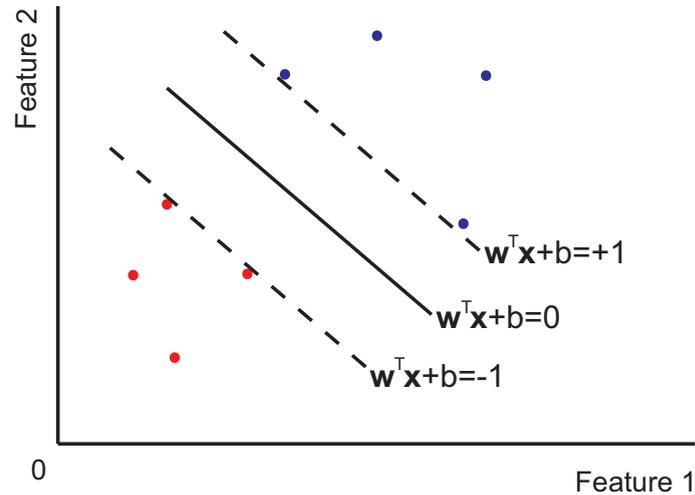


FIGURE 3.10: Optimal discriminating hyperplane (solid line) for two separate classes (red and blue points) with two hyperplanes either side of maximal distance away (dashed lines).

The distance between the hyperplanes may be calculated by first subtracting the equations of each hyperplane (Equations 3.27 and 3.27) from each other,

$$\mathbf{w} \cdot (\mathbf{x}_1 - \mathbf{x}_2) = 2 \quad (3.29)$$

Projecting this onto the normal to the hyperplane gives,

$$\frac{\mathbf{w}}{|\mathbf{w}|} \cdot (\mathbf{x}_1 - \mathbf{x}_2) \quad (3.30)$$

This can be used to calculate the distance between the hyperplanes (by substituting 3.30 into 3.29) resulting in a distance of  $\frac{2}{|\mathbf{w}|}$ . The optimal discriminating boundary is found when this margin is maximised (minimising  $|\mathbf{w}|$ ), however this is computationally complex (due to the square root). A computationally less intensive, yet equivalent, alternative is to minimise  $\frac{|\mathbf{w}|^2}{2}$ . Thus, defining the margin, such that no points lie within it, as,

$$\min_{\mathbf{w}, \mathbf{b}} \frac{1}{2} |\mathbf{w}|^2 \quad (3.31)$$

subject to

$$\mathbf{y}_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1, 1 \leq i \leq N \quad (3.32)$$

where  $\mathbf{y}_i$  takes on the values  $+1, -1$  for either class.

Lagrangian theory can be applied to reformulate the optimisation problem, where each Lagrange multiplier ( $\alpha_i > 0$ ) is subject to the constraint in Equation 3.32, as per,

$$\min_{\mathbf{w}, b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} |\mathbf{w}|^2 - \sum_{i=1}^N \alpha_i \{ \mathbf{y}_i (\mathbf{w}^T \mathbf{x}_i - b) - 1 \} \right\} \quad (3.33)$$

subject to the constraint,

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (3.34)$$

All points (or vectors) which are not support vectors, that is the points which do not lie on the margin, ( $\mathbf{y}_i (\mathbf{w}^T \mathbf{x}_i - b) - 1 > 0$ ) do not matter (as they do not influence the decision boundary). Correspondingly their  $\alpha_i$  is set to zero, and as such do not have to be considered in the optimisation process.

This problem may be written in the dual form:

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (3.35)$$

subject to  $\alpha_i \geq 0$  for all  $0 \leq i \leq n$  and the constraint

$$\sum_{i=1}^n \alpha_i \mathbf{y}_i \quad (3.36)$$

This facilitates the calculation of the weight vector,

$$\mathbf{w} = \sum_i^N \alpha_i \mathbf{y}_i \mathbf{x}_i \quad (3.37)$$

Currently we have only catered for linear decision hyperplanes. However, specific kernel functions,  $K(\mathbf{x}_i, \mathbf{x}_j)$  in Equation 3.35, can translate the input features into a higher dimensional space and linear decision hyperplanes in this higher dimensional space may be equivalent to non-linear decision regions when translated back into the original feature space. The translation of features from the original feature space into the higher dimensional space may be defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) \rightarrow \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \quad (3.38)$$

The selected Kernel must fit Mercer's condition:

$$\int K(\mathbf{x}_i, \mathbf{x}) \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) d\mathbf{x}_i d\mathbf{x} \geq 0 \quad \forall \varphi \neq 0, \quad \int \varphi^2(\mathbf{x}) d\mathbf{x} < \infty \quad (3.39)$$

Examples of Kernels which fit this condition include the dot product kernel, where  $K(\mathbf{x}_i, \mathbf{x})$  is the dot product of both inputs (this is a linear Kernel which produces a linear decision boundary), polynomial kernels, where  $K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i \cdot \mathbf{x})^d \forall d \geq 1$  (that is, the dot product of the two input vectors is raised to a power greater than one), a Gaussian Radial Basis Function (RBF), where  $K(\mathbf{x}_i, \mathbf{x}) = \exp(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2\sigma^2})$  (that is, the distance of each point from a centroid determines the magnitude of projection into a new dimension), and many others. The internal parameters of the Kernel being implemented are optimised as part of the classifier training procedure.

The process of finding the perfect hyperplane which separates the data in both classes may be exhaustive when data is closely intermingled, and may be impossible where samples from both classes overlap upon each other (possibly due to the nature of the data or to noise). Allowing some misclassifications to occur efficiently caters for these conditions and may be realised by making the decision function less strict (through introducing a slack term ( $\epsilon$ ))

$$\mathbf{y}_i(\mathbf{w}^T \mathbf{x} + b) \geq 1 - \epsilon_i \quad (3.40)$$

Thus, the original optimisation condition for the selection of the optimal margin (Equation 3.30) is modified to

$$\min_{\mathbf{w}, \mathbf{b}} \frac{1}{2} |\mathbf{w}|^2 + C \sum_{i=1}^N \epsilon_i \quad (3.41)$$

where  $C$  may be thought of as a penalty on errors and is referred to as the *box constant*. Greater box constant values will increase the cost on the optimisation process. In practice, this means that larger box constant values result in a more complicated decision region while smaller values result in a smoother decision region (and an increased number of errors). Large box constant values may be suboptimal as the generalisability of the classifier can become reduced as the classifier will become more specialised on the training data.

Examples of the decision regions found with the ‘partially surrounded’ data set using SVM with linear, quadratic and polynomial kernels are given in Figure 3.11. The complexity of the decision region increases as the order of the polynomial kernel increases. RBF offer a more non-linear approach by extending the feature space into another dimension centred around a centroid with a Gaussian distribution of width  $\sigma$  (Figure 3.12 illustrates the effect of different  $\sigma$  values). The box constant value may also be optimised while training the classifier. As shown in Figure 3.13, smaller box constant values result in a smoother decision region (and an increased number of errors), while larger values result in more complex decision regions (and less misclassifications).

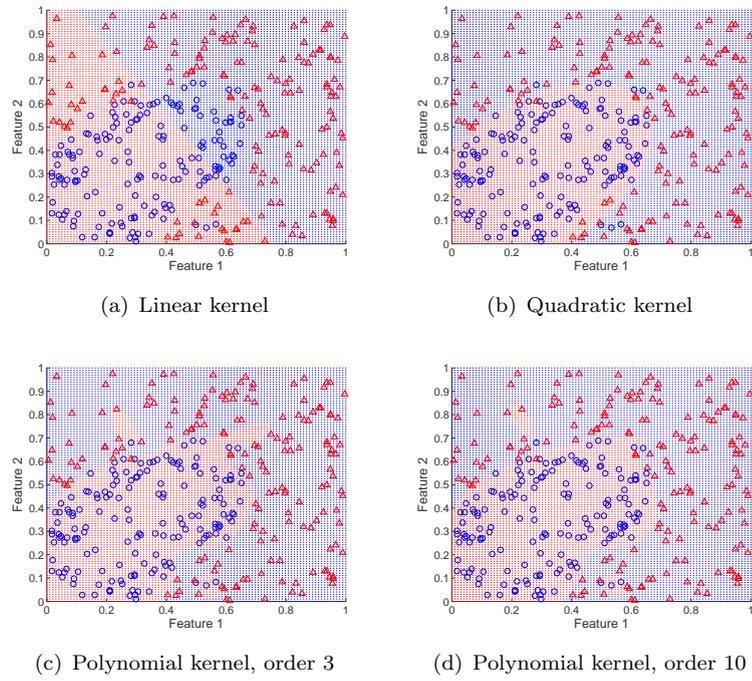


FIGURE 3.11: Decision regions of SVM applied to the ‘partially surrounded’ data set for the specified kernel, and, where applicable, internal parameters.

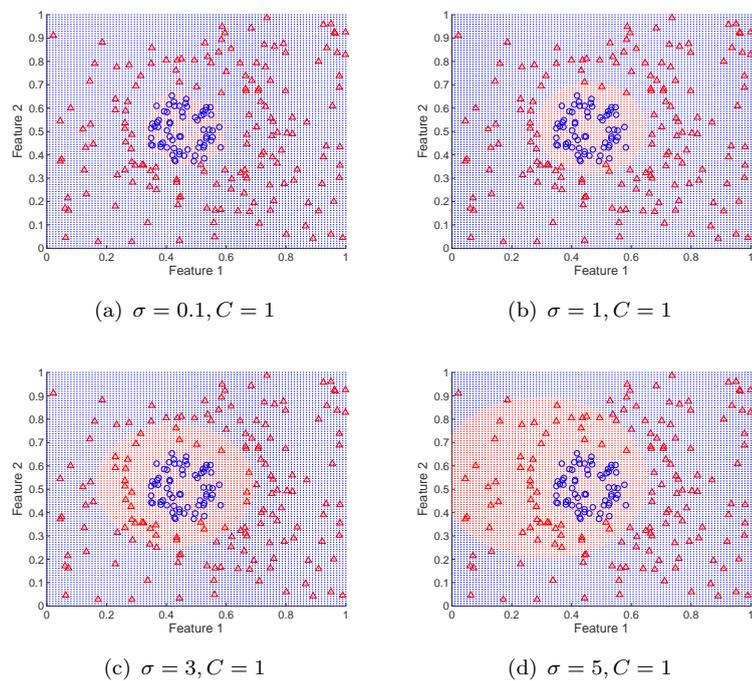


FIGURE 3.12: Decision regions of SVM with an RBF kernel applied to the ‘fully surrounded’ data set for the specified internal parameters.

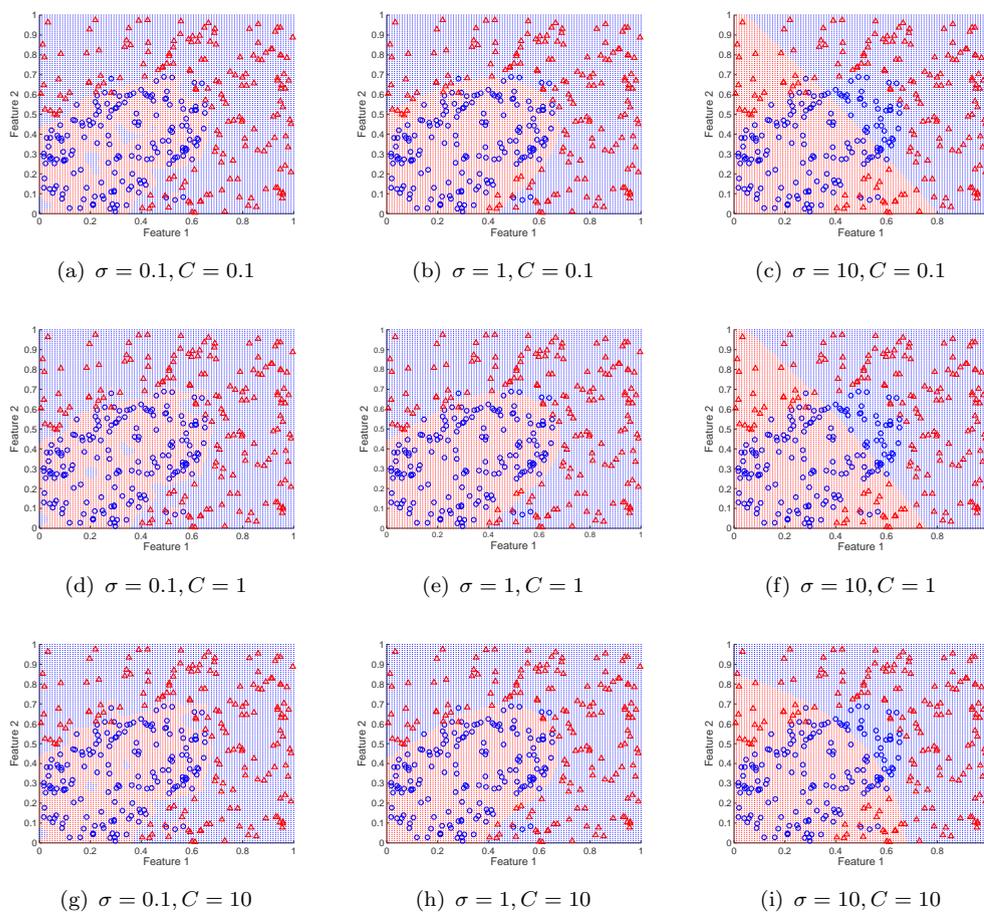


FIGURE 3.13: Decision regions of SVM with an RBF kernel applied to the ‘partially surrounded’ data set for the specified internal parameters.

For a more detailed description of SVM the reader may refer to Shawe-Taylor and Christiani (2000), and Burges (1998).

### 3.3.7 Multiclass Classification

Although some classifiers are naturally able to discriminate between more than two classes, many classifiers reduce the multi-class problem into a series of binary classifications. Two common approaches are 1) the *one versus all* strategy, and 2) the *one versus one* strategy.

**One versus all:** For this strategy, a classifier is trained to discriminate whether a new sample belongs to a specific class or to the remaining classes. Ideally, this results in the new sample being assigned to only one class. However, in practice the sample may be classified as belonging to two or more classes. In this case, the class which the sample

has the highest probability of belonging to is often chosen (known as the *winner takes all* strategy).

**One versus one:** Multiple classifiers are trained to discriminate between every possible pair of classes. In practice, each new sample is assigned to one of two classes for every possible pair of classes. The new sample is assigned to an overall class based on which class it has been assigned to the most.

### 3.3.8 Application and Performance of the Classifiers

Each of the classifiers described above were applied to the *fully surrounded* data set. This is a difficult set of classes to discriminate as no linear combination of the features may distinguish the features. As non-linear methods are employed (QDA, kNN, NN and SVM), significant improvements in performance were found (see Table 3.2). The ANN were trained using sixty percent of the data (and tested on the remaining forty percent), while the remaining classifiers were trained using two-thirds of the data and tested using the remaining one-third.

Increasing the  $k$  parameter in the kNN classifier (increasing the non-linearity of the discriminating hyperplane) was shown to reduce performance. Similarly, using a large number of neurons in the hidden layer of the ANN classifier was also shown to decrease performance. The SVM was optimised using a grid search method on the hyperparameters (the box constant and sigma) which resulted in the best performance (along with

QDA and kNN when  $k=1$ ).

TABLE 3.2: Performance of the SVM, ANN<sup>1</sup>, kNN<sup>2</sup>, LDA and QDA on the *fully surrounded* data set.

	LDA	QDA	kNN 1	kNN 5	kNN 9	ANN 1	ANN 5	ANN 10	ANN 100	SVM
TP	30	53	53	52	51	53	69	62	65	53
FP	8	0	0	0	0	4	0	0	0	0
TN	13	21	21	21	21	23	22	30	25	21
FN	23	0	0	1	2	12	1	0	2	0
Accuracy	58	100	100	99	97	83	99	100	98	100
MCR	42	0	0	1	3	17	1	0	2	0
Sensitivity	57	100	100	98	96	82	99	100	97	100
Specificity	62	100	100	100	100	85	100	100	100	100
Precision	79	100	100	100	100	93	100	100	100	100
Recall	57	100	100	98	96	82	99	100	97	100
PPV	79	100	100	100	100	93	100	100	100	100
NPV	36	100	100	95	91	66	96	100	93	100
MCC	16	85	85	82	79	54	85	82	81	85
F Score	0.66	1	1	0.99	0.98	0.87	0.99	1	0.98	1
Kappa	0.15	1	1	0.97	0.94	0.61	0.97	1	0.95	1

<sup>1</sup> The number of neurons in the hidden layer is specified in the column header.

<sup>2</sup> The value of  $k$  is specified in the column header.

### 3.4 Conclusion

In this chapter, the classification algorithms commonly used in recently developed sleep technologies and in the later chapters of this thesis were briefly described. Additionally, commonly used performance metrics were introduced. Artificial data sets were generated in order to provide the reader with a more intuitive understanding of these classifiers, the hyperparameters and the parameters (as given in Table 3.3). The procedure of splitting the data into training, validation and test data in order to produce valid results, select the optimal classifier and report its unbiased performance was also described.

As previously discussed, each classifier has their own advantages and disadvantages. During implementation, these mainly focus around time and processing power constraints. The BCT, LDA and QDA classifiers have relatively little computational overhead during training and implementation and their simplicity makes them ideal for low cost systems. The increased complexity of kNN, ANN and SVM provide increased discriminatory power where highly non-linear hyperplanes are required. However in cases

where such systems provide marginal benefit simpler and more low cost discriminating systems may be preferential.

TABLE 3.3: The hyperparameters and parameters associated with each classifier.

<b>Classifier</b>	<b>Hyperparameters</b>	<b>Parameters</b>
DA <sup>1</sup>	Order of the discriminating hyperplane (where 1 is linear, etc.)	Equation of the optimal separating hyperplane
kNN	k	distance between new test point and each training point
ANN	number of hidden layers, number of neurons in each layer	weights and bias applied to each neuron
SVM	Box constant, sigma and the Kernel type	Equation of the optimal separating hyperplane

<sup>1</sup> Discriminant Analysis

## Chapter 4

# A Review of Sleep Measurement Technologies

This chapter provides a comprehensive review of the traditional clinical, non-clinical and research technologies used in sleep monitoring, namely PSG, wrist actigraphy, subjective sleep scales, sleep indices, and sleep diaries. Various advances in sleep technology have been made and this chapter provides an overview of the new approaches to contact and non-contact based sleep monitoring. Such advances include advancements reducing the number of electrodes required, embedding electrodes into bed sheets and textiles, the use of smart phones, video monitoring, and advancements in recording vital signs (respiration and heart rate) via a non-contact (off-body) sensor. Finally, a comparative evaluation of these technologies investigating their suitability for long term use (particularly amongst sensitive populations) is made.

### 4.1 Polysomnography

PSG is the process of recording multiple physiological signals during sleep. This includes EEG, EOG and EMG and often can extend to ECG, Blood Pressure (BP), respiration (including air flow and respiratory effort) and less often renal function. A set of recommended recording guidelines (concerning sampling rate, etc.) was established by the AASM (Iber et al., 2007).

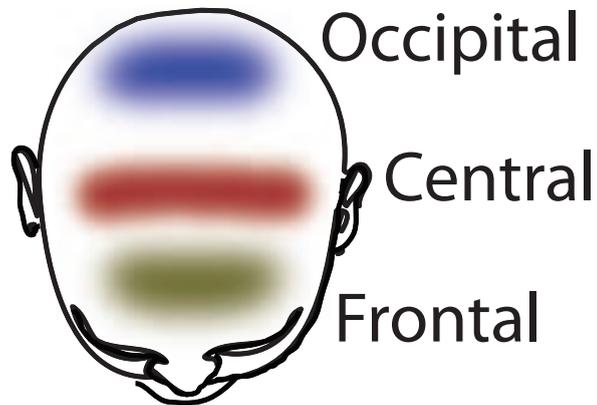


FIGURE 4.1: Approximate central, frontal and occipital regions of the brain

#### 4.1.1 Polysomnography Signals

**EEG** EEG records synaptic potentials from pyramidal cells or neurons in the brain. The EEG records the potential difference between electrodes placed on the skull (Benbadis, 2006). The new AASM guidelines state that the frontal, central and occipital regions (as given in Figure 4.1) are to be monitored with a minimum of 3 EEG recordings during sleep (Iber et al., 2007), whereas the older Rechtschaffen and Kales standard used EEG configurations recording from the central region of the brain. These electrodes are referenced to a lead placed on the mastoid (part of the bone located behind the ear).

**EOG** A natural dipole exists between the retina and the cornea (Benbadis, 2006). As the eye moves around its axis the potential between the retina and the cornea changes (Benbadis, 2006). Eye movements can be captured by recording this potential difference

over time (Benbadis, 2006; Collop, 2006; Iber et al., 2007). Electrodes are placed slightly above and below the outer canthus of each eye and are referenced to the same or opposite mastoid lead.

**EMG** Muscle tone is measured as a function of the electric fields generated by the membrane depolarisation of submental muscle fibers and is typically recognised as activity of greater than thirty hertz (Vaughn and Gaiallanza, 2008). Submental muscle tone is used in sleep studies to distinguish between different stages of sleep (for example REM sleep is noted to have low muscle tone) (Iber et al., 2007). The surface electrodes are generally placed 2 cm below the inferior edge of the mandible (the lower jaw) and 2 cm to the right and left of the midline. This electrode is referenced to an electrode placed at the midline about 1 cm above the inferior edge of the mandible and grounded to common electrode placed at the mastoid (Iber et al., 2007; Vaughn and Gaiallanza, 2008).

**ECG** A modified Lead II of a standard ECG is used to record the heart signal (Iber et al., 2007). Electrodes are placed below the right clavicle and below the left rib cage. A reference electrode is also used and is often also taken from the mastoid.

**Respiration** For an accurate representation of respiration both respiratory flow and effort should be recorded. Respiratory flow incorporates the movement of air in and out of the chest cavity. Respiratory effort must be measured as a surrogate to chest movement. A pneumotachometer can directly measure this, however it is both intrusive and cumbersome. Air temperature and nasal pressure are common techniques for flow measurement. A thermistor can be used to record breathing by detecting the flow of air at near body temperature being exhaled and the inhalation of warmer air which is at an ambient room temperature. A nasal pressure device can detect the air flow using a piezoelectric pressure sensor placed inside a small tube at the nostrils. Respiratory effort can be directly measured by intrathoracic pressure monitoring, however this is intrusive and will disturb sleep. Less invasive techniques include Respiratory Inductance Plethysmography (RIP), intercostal EMG and strain gauges. RIP measures changes in a magnetic field generated by a current passing through coil bands that are placed around the chest and abdomen. As such, deformations in either the diameter or the shape of the RIP band is used as a proxy measure of respiratory effort. Intercostal (between the ribs) EMG is the placement of surface electrodes on the fifth to eighth lateral intercostal space and has been shown to qualitatively assess respiratory effort (Stoohs et al., 2004; Vaughn and Gaiallanza, 2008). This can be of particular use during REM sleep where diaphragmatic activity exists but intercostal muscles are atonic

(Vaughn and Gaiallanza, 2008). Strain gauges, or belts, are piezoelectric devices which measure change in circumference of the chest. However, unlike RIP bands, the change in circumference is not measured throughout the device but at a fixed location along the belt. As the belt's circumference increases the strain on the piezoelectric sensor also changes. This strain is transduced into a recordable electrical signal. Strain gauges are not currently recommended by the AASM due to a lack of validation studies (Iber et al., 2007; Vaughn and Gaiallanza, 2008).

An oronasal thermal sensor is recommended for the detection of sleep apnoeas (discussed in Section 2.6.2.1) (Iber et al., 2007). A hypopnoea can be identified using a nasal air pressure transducer (Iber et al., 2007). Respiratory effort is often assessed using esophageal manometry or calibrated or uncalibrated inductance plethysmography (Iber et al., 2007). However, diaphragmatic or intercostal EMG can be used if none of these sensors are available Iber et al. (2007).

**Oxygen Saturation** The sensor recommended to record blood oxygenation is pulse oximetry with a maximum signal averaging time of 3 seconds (Iber et al., 2007). Details on how oxygen saturation and respiration can be used to assess sleep apnoea and hypopnoeas can be found in Section 2.6.2.1 and in (Iber et al., 2007).

#### 4.1.2 Types of Polysomnography

PSG can be performed in a clinical setting attended by clinicians, in an unattended setting (may be in a clinical or domestic setting) or in an ambulatory fashion (Iber et al., 2007).

##### 4.1.2.1 Clinical PSG

Clinical PSG should be attended by a trained health care clinician. A minimum of seven channels must be used (including EEG, EOG, EMG, ECG, air flow, respiratory effort and oxygen saturation). More comprehensive studies will use a full compliment PSG and be recorded in a specialised sleep laboratory. This facilitates the diagnosis of Obstructive Sleep Apnoea Syndrome (OSAS) and as well as various sleep disorders. It is also sometimes recommended that patients spend two nights in the sleep lab and data from the second night be used for analysis. This is due to 'the first-night effect' where increased arousals, reduced TST and a decreased REM percentage are seen in the first night of data collection. However, recording PSG over two nights has a high overhead and has not been proven to be cost and time beneficial.

#### **4.1.2.2 Unattended PSG**

Unattended studies range from recording data from one or two sensors (for example, air flow or oxygen saturation) to the seven channels described above. While recording using two channels does not yield the quantity of data collected in a clinical study, it does provide important information and can be used to screen for OSAS and other sleep disorders.

#### **4.1.2.3 Ambulatory PSG**

The ambulatory collection of PSG allows a participant to continue their daily routine without having to be kept in a fixed location. It can be cumbersome, intrusive and expensive. However it provides important information about sleep and sleeping patterns (for example, that of severely sleep restricted health care professionals (Barger et al., 2005; Lockley et al., 2004)). Results have shown an increased risk of motor vehicles crashes and attentional failures occurs in sleep restricted individuals (Barger et al., 2005; Lockley et al., 2004).

#### **4.1.3 PSG in Children**

Sleeping patterns are different in children to that seen in adults. There are many obstacles to pediatric sleep scoring including the developmental processes of pediatric EEG and sleep, changes in other physiological patterns with age, the changing etiologies of various disorders with age and the challenge of dealing with children in an intrusive clinical environment (Griebel and Moyer, 2006). However, there is still considerable benefit in pediatric PSG as an untreated sleep disorder can have long ranging effects such as growth failure, pulmonary hypertension, cor pulmonale (a heart condition) as well as academic and behavioural concerns (Griebel and Moyer, 2006). The recording of PSG during a nap session has also been suggested as a screening tool, however this can prove inconclusive as it is possible not all sleep stages will be seen (Griebel and Moyer, 2006). Consequently, a full overnight PSG could also be required, nullifying the cost and time benefits of a nap study. The ‘first-night effect’ is often called into question in pediatric sleep studies. However some have reported no difference between sleep metrics recorded on the first and second night spent in a sleep laboratory (Griebel and Moyer, 2006; Katz et al., 2002). More thorough information on pediatric PSG can be found in Griebel and Moyer (2006); Iber et al. (2007).

#### 4.1.4 Sleep Scoring

Sleep scoring is based on rules established by Rechtschaffen and Kales (Rechtschaffen and Kales, 1968) which were subsequently updated recently by the AASM (Iber et al., 2007). The scoring process is complicated and subjective, and often results in two trained and experienced sleep scorers having an imperfect agreement rate. Inter-rater agreement rates have been reported to be 80% for the R&K guidelines and 82% for the AASM guidelines over different subjects and sleep laboratories (Danker-Hopfe et al., 2009). EEG, EOG and EMG data are reviewed generally in 30 second epochs and assigned a sleep stage. Certain characteristic patterns define each sleep stage including sleep spindles, K-complexes, alpha and PDR activity, slow wave (also known as delta) activity, saw-tooth waves, differing levels of muscle tone and different types of eye movements. The reader is referred to Chapter 2.2 where these rules are specifically discussed.

## 4.2 Sleep Diaries

The sleep diary (see Figure 4.2) is an instrument consisting of multiple questions which record multiple sleep metrics. It records subjective information about sleep/wake rhythms, such as when 'lights out' time, 'lights on' time, sleep latency, number of and reasons for awakenings, time spent awake in bed until arising and visual acuity scales (VAS) for sleep quality, mood on wakening and alertness on final wakening (Monk et al., 1994). It is widely used in clinical and research studies. In comparison to a single survey report, a daily sleep diary relies less on memory and provides a quantitative measurement of sleep/wake schedules (Wolfson et al., 2003). A comparison of total sleep times between ambulatory PSG and sleep diaries have shown acceptable agreement rates (92.3% sensitivity and 95.6% specificity) (Rogers et al., 1993). Sleep diaries are often used as a screening tool for clinicians and researchers for sleep disorders, especially for circadian rhythm sleep disorders.

## 4.3 Sleep Tests and Subjective Measures

Subjective questionnaire-based data has also been used to compare both between and across study populations. Some of these are described below:

### Sleep Diary Template

ID: \_\_\_\_\_

Current Date/Time: \_\_\_\_\_ Lights Out: \_\_\_\_\_ Lights On: \_\_\_\_\_

1. How long did you take to fall asleep last night? \_\_\_\_ (hrs) \_\_\_\_ (mins)
2. How long did you sleep last night? \_\_\_\_ (hrs) \_\_\_\_ (mins)
3. How many times did you awaken during the night? \_\_\_\_
4. After the end of your sleep period, how long did you remain in bed before getting up? \_\_\_\_ (hrs) \_\_\_\_ (mins)
5. Was your sleep disturbed: [Yes / No] (If Yes, check all that apply)
  - noise     work duties     physical discomfort     voids, # of voids: \_\_\_\_
  - too hot     too cold     other
6. How did you sleep last night?
 

poorly |-----| great
7. How do you feel right now?
 

sleepy |-----| alert
8. Did you have any caffeine yesterday? [Yes / No] (If Yes, indicate how much)
  - coffee \_\_\_\_ (cups)                      caffeine pills \_\_\_\_ (100mg) \_\_\_\_ (200mg)
  - tea \_\_\_\_ (cups)                              caffeinated soft drinks \_\_\_\_ (glasses)

Indicate how long before bed your last caffeine intake was: \_\_\_\_ (hrs) \_\_\_\_ (mins)
9. Did you nap yesterday? [Yes / No] If Yes, indicate time of nap(s) \_\_\_\_\_
10. Did you take any medications yesterday? [Yes / No / Decline] (If Yes, list all) \_\_\_\_\_

Comments:

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

FIGURE 4.2: Typical Sleep Diary

### 4.3.1 Common Sleep Tests and Measures

**Multiple Sleep Latency Test (MSLT):** The MSLT is currently considered the ‘gold standard’ for the quantitative assessment of sleepiness (Carskadon et al., 1986). The MSLT examines the likelihood of falling asleep at multiple intervals throughout the day in an environment conducive to sleep. If, during a test, the user does fall asleep (as measured by PSG) the time taken to fall asleep is recorded. The average from the beginning of the test to sleep onset is averaged over four to five naps to obtain the Mean Sleep Latency (MSL). The MSLT is used to assess sleepiness with a lower score indicating a more severe case (Carskadon et al., 1986; Chervin and Guilleminault, 1995). MSLT is commonly often used as a primary indicator of sleepiness in a clinical setting. An MSL shorter than five minutes has been found to be associated with performance decrements and unintentional episodes of sleep in sleepy patients (Carskadon et al., 1986; Chervin and Guilleminault, 1995). The link between MSL and sleepiness has been questioned due to a lack of correlation with other factors including the amount of sleep during the previous night, the efficiency of that sleep, the number of awakenings and the amount of time spent in a particular sleep stage (Chervin, 2003; Chervin and Guilleminault, 1995). A thorough physiological construct for the mediation of sleepiness has not been established and as a result the optimal measurement scale for sleepiness remains elusive (Chervin, 2003; Chervin and Guilleminault, 1995).

**Epworth Sleepiness Scale (ESS)** The ESS is intended to assess excessive daytime sleepiness by measuring sleep propensity (Johns 1991, Johns 1992). It consists of eight questions asking whether or not the subject is likely to fall asleep in common situations of daily living. The four possible answers are weighted according to likelihood and subsequently summed to a single number. The final score has been shown to correlate well with the AHI in a large study sample (Gottlieb et al., 1999). A score of greater than ten is considered to indicate sleepiness (Buysse et al., 2008). The scale is widely used in clinical and research studies to assess sleepiness although there has been some dispute of its validity (Miletin and Hanly, 2003). Miletin and Hanly (2003) conducted a review and discussed the relationship between ESS and MSLT. A moderate but statistically significant relationship between ESS and MSL, as reported by MSLT, was cited, however conflicting results in other studies were also reported. It was noted that the ESS and MSLT do not measure the same type of sleepiness. The MSLT is an objective measure of sleepiness at the point of testing, but the ESS measures the tendency of sleepiness during certain recent activities (Miletin and Hanly, 2003).

**Karolinska Sleepiness Scale (KSS)** The KSS is a 9 point verbally anchored scale with the following steps: "Extremely alert" (score = 1), "Alert" (3), "Neither alert nor sleepy" (5), "Sleepy - but no difficulty remaining awake" (7), "Extremely sleepy - fighting sleep" (9) (Kaida et al., 2006). The steps in between have a scale value but no verbal value. Median reaction time, number of lapses, alpha and theta (as defined in Section 2.2) power density and the alpha attenuation coefficient (the ratio of mean eyes closed to mean eyes open alpha power) have been shown to have a highly significant increase with increasing KSS (Kaida et al., 2006).

**Stanford Sleepiness Scale (SSS)** The SSS is an analog scale used to assess sleepiness (Carskadon and Dement, 1981; Hoddes et al., 1973). It is a 7 point scale consisting of numbered statements which describe alertness/sleepiness levels. These range from "Feeling active, and vital" (score = 1) to "No longer fighting sleep, sleep onset soon" (score = 7) (Carskadon and Dement, 1981; Hoddes et al., 1973).

**Maintenance of Wakefulness Test (MWT)** The MWT assesses wake tendency by measuring the ability to remain awake in a sleep conducive setting (Doghranji et al., 1997). The MWT has been shown to be able to identify those with an impaired wake tendency and to identify those with a pathologic inability to remain awake while in a setting conducive to sleep (Doghranji et al., 1997).

**Other sleep scales and tests** Other sleep scales have been devised for specific conditions, for example the Parkinson's Disease sleepiness scale (PDSS) (Chaudhuri et al., 2002). The psychomotor vigilance test (PVT) is also commonly used to test reaction times in clinical and research studies (Wright et al., 2002).

**Pittsburgh Sleep Quality Index (PSQI)** The PSQI is a self rated questionnaire which assesses sleep quality over a period of one month (Buysse et al., 1989). Nineteen individual items are used to generate seven equally weighted "component" scores including subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication and daytime dysfunction. These component scores are summed to generate a global metric which ranges from 0-21, with higher scores indicating a worse sleep quality. The PSQI has been shown to identify good and poor sleepers and has been shown to have a high sensitivity and specificity for detecting a sleep disorder, using a cut-off of 5 (Buysse et al., 2008). An additional five questions are asked for clinical information only. Longitudinal applications of the PSQI have included its use in examining the course and natural history of sleep/wake

disorders, the progression of sleep disturbances and the detection of relapses detected by the re-emergence of a sleep disorder (Buysse et al., 1989).

**Insomnia Severity Scale (ISS)** The ISS comprises of seven items and is used to assess insomnia severity (Bastien et al., 2001). It generates a score varying from 0-28 with higher scores indicative of more severe insomnia. A total score of 0-7 indicates 'no clinically significant insomnia', 8-14 'sub-threshold insomnia', 15-21 'clinical insomnia (moderate severity)' and 22-28 'clinical insomnia (severe)' (O'Donoghue et al., 2009). A cut off of 14 is used to discriminate between individuals diagnosed with primary insomnia and those without (reporting a sensitivity of 94% and a specificity of 94%) (Smith and Trinder, 2001).

## 4.4 Actigraphy

Actigraphy refers to the collection of rest/activity profiles of humans and other animals, generally over extended periods of time in the animals normal habitat (Sadeh et al., 1994, 1995). These patterns have been widely used to document and present sleep/wake habits due to the inherent lack of physical activity during sleep. Mechanical actigraphs originally transduced physical movement into an electrical signal which used movement and inactivity as markers for wake and sleep. Subsequent manual scoring reported highly accurate sleep/wake classification (Mullaney et al., 1980). Modern actigraphs have developed into compact, reliable, single or multiple axis accelerometry based devices, often placed on the wrist, with the ability to record for several days, even weeks, without interruption.

### 4.4.1 Actigraphy Data Preprocessing

A bandpass filter is applied to the raw accelerometry data allowing signals in the general range of 0.25-4 Hz to pass. The selection of this range of frequencies removes the gravitational acceleration component and records only voluntary human movement (Ancoli-Israel et al., 2003). Redmond and Hegge (1987) reported that voluntary movement rarely exceeds 3-4 Hz and that involuntary movement generally exceeds 5 Hz. It has elsewhere been suggested to record faster movement in the 0.5-11 HZ range in order to capture movement in younger subjects (Van Someren et al., 1995). Various metrics, often reported as '*activity counts*', have been derived to represent motion over a user

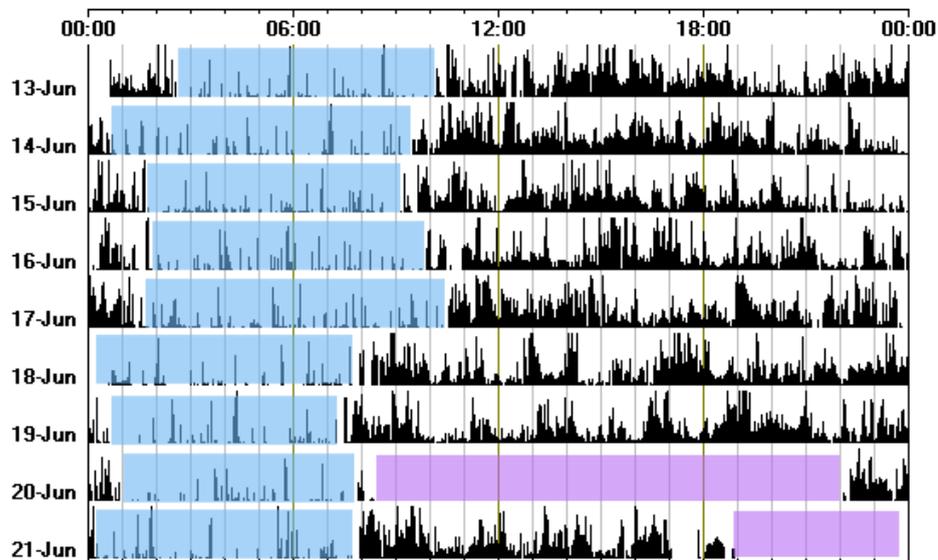


FIGURE 4.3: Sample wrist actigraphy collected over nine days from an elderly subject with each day plotted on subsequent lines. The light blue patterns indicate periods of sleep and the pink patterns indicate periods when the watch was not worn.

specified period of time, or epoch. Activity counts are calculated by integrating, or summing, the bandpassed accelerometry data over each epoch. Example wrist actigraphy data is shown in Figure 4.3 collected over a nine day period.

Unlike PSG, actigraphy generally does not suffer a ‘first night effect’ as the participant does not feel overly constricted (as they do in PSG). Compliance issues are generally less likely than PSG as the actigraph is considered unobtrusive. Although some participants, particularly older adults, may consider the device a hindrance (Behan et al., 2008a; Walsh et al., 2008). Actigraphy allows for an unconstrained, long-term and continuous monitoring period which PSG does not. This facilitates the examination of variations in sleeping patterns across multiple days in a community setting.

#### 4.4.2 Sleep/Wake Classification and Scoring Procedures for Humans

The manual scoring of actigraphic records for sleep/wake discrimination reported highly accurate results using wrist based devices against PSG in humans (Mullaney et al., 1980). This process was automated using a thresholding algorithm on activity count data to report active and inactive periods as sleep and wake, respectively, and was also found to report a high level of accuracy (Webster et al., 1982). Many of these algorithms generate a metric by weighting past, current and future activity count data together and a thresholding algorithm scores the current epoch as either sleep or wake (Cole et al., 1992; Sadeh et al., 1994). However, inaccurate short sleep latency times were found using

this approach. To improve the accuracy in detecting the correct sleep start time the scoring algorithms were modified to require a minimum consecutive period of inactivity (some allow for one epoch of activity) before sleep start could be defined. For example, the initial sleep epoch is scored as sleep when nineteen of the twenty subsequent epochs also report a value below the sleep/wake threshold (indicating sleep). This algorithm is also applied (although implemented in reverse) in order to find the last epoch of sleep before wake. This added complexity has addressed the issue of the inactive state of wakefulness prior to sleep start. Short sleep latency times were reported prior to the application of this algorithm (Ancoli-Israel et al., 2003; Cole et al., 1992).

#### 4.4.3 Clinical Performance

Comparisons of actigraphy and PSG have been reported for multiple populations including healthy adults, children, adolescents, older adults, psychiatric patients as well as people suffering from various sleep disorders including sleep apnoea and insomnia (Ancoli-Israel et al., 2003; Cole et al., 1992; Hauri and Wisbey, 1992; Hedner et al., 2004; Jean-Louis et al., 2001; Kushida et al., 2001; Mullaney et al., 1980; Pollak et al., 2001; Sadeh et al., 1994; Souza et al., 2003; Webster et al., 1982). Comparisons between actigraphy and PSG-defined sleep have yielded high levels of agreement with epoch-by-epoch concordance of up to 91-93% (Ancoli-Israel et al., 2003). A more recent review reported high agreement rates of over 85% in healthy adults of different age groups (Acebo and LeBourgeois, 2006).

Van De Water et al. (2011) presented a review of objective sleep measurements for non-laboratory settings and this included a large variety of wrist actigraphy sleep/wake systems deployed in hospital intensive care units, sleep laboratories and nursing homes. This review presents a number of metrics (including TST, SE, agreement rate, sensitivity, specificity, accuracy, total wake time, SL, WASO) which were used to compare the proposed technology to PSG. For most systems, specificity rates were much lower than sensitivity rates, although high accuracy rates were still reported as the data is generally highly biased towards the sleep state. No comparisons of wrist actigraphy against PSG were reported in home-based environments, however the optimal results for a sleep laboratory based setting had a sensitivity of 91%, a specificity of 65% and an accuracy of 88% ((Paquet et al., 2007) in (Van De Water et al., 2011)). Night to night TST (and other variables) correlations between PSG and actigraphic methods were commonly reported, however these can provide misleading results as they are a measure of the commonality of the trend between two signals.

#### 4.4.3.1 The Validity of Wrist Actigraphy Estimates

Actigraphic sleep monitoring uses movement information to discriminate sleep from wakefulness. As such, actigraphic estimates of sleep do not contain the ability to distinguish inactive, or quiescent, wake from sleep or conversely active sleep (such as sleep walking) from wake. Actigraphy, at its optimal performance, can only approach the accuracy of PSG, which uses multiple physiological variables to determine the sleep-/wakefulness state, and as such actigraphy can only be viewed as an estimate of sleep or wakefulness.

Commonly reported measures of comparison between actigraphy and PSG in sleep monitoring are Pearson's Correlation Coefficient (PCC) for total sleep time TST, accuracy, sensitivity and specificity of a binary classifier, Cohen's Kappa and Bland-Altman analysis (Tryon, 1991). The suitability of PCC for comparing sleep metrics, particularly TST, against the gold standard of PSG has been questioned as this comparison is relative not absolute (Ancoli-Israel et al., 2003; Souza et al., 2003). For example, a high correlation would remain if the trend of TST was similar for both measures despite a possible order of magnitude difference between the actual values (Ancoli-Israel et al., 2003; Souza et al., 2003). A direct epoch-by-epoch comparison of actigraphic and PSG measures of sleep state would be more suitable to determine the efficacy of the actigraph.

However, such a direct comparison presents a major methodological challenge known as 'time-locking' (ensuring all data across different devices starts and ends at the same time) (Ancoli-Israel et al., 2003; Pollak et al., 2001). Pollak et al. (2001) described the application of sensitivity, specificity and accuracy measures, as performance indicators of a standard binary classification test, in an epoch-by-epoch comparison. Positive results were reported, however a low wake detection capacity (specificity) was found. The Bland-Altman technique has also been used to compare two techniques of sleep monitoring (Souza et al., 2003). This technique compares two measures of the same metric recorded using two different devices. It is based on a plot of the difference between the two techniques against the mean of the two techniques for each subject in a sample (Bland and Altman, 1999).

#### 4.4.4 Actigraphy - Deployment in Practice

There are many situations where it is prohibitive to monitor sleep by PSG (for example in sensitive populations or in challenging environments such as space, or over multiple nights) and for these situations wrist actigraphy is a suitable alternative (Walsh et al., 2009a). Actigraphic values for sleep onset and offset, sleep duration and sleep efficiency

were found to be superior to sleep diaries when compared against PSG in space (Monk et al., 1999). The authors concluded that when PSG was too cumbersome to use that actigraphy was a simple efficient means of evaluating sleep (Monk et al., 1999). Actigraphy has been suggested as a large scale, reliable, easily deployable and low cost estimator of sleep/wake patterns especially where PSG is not an option (Walsh et al., 2009a).

#### **4.4.4.1 Actigraphy Device Placement Issues**

In a comprehensive review, Ancoli-Israel et al. (2003) discussed two studies which found that there were no differences in actigraphy results collected from multiple locations (dominant wrist, non-dominant wrist, ankle or trunk) although the actual actigraphic data might be dissimilar (Ancoli-Israel et al., 2003; Jean-Louis et al., 1997; Sadeh et al., 1994). This review noted that in some studies wrist placement was found to be better at detecting wake than other placements. Although, it was concluded that more studies comparing different actigraph placements against concomitant PSG were needed (Ancoli-Israel et al., 2003; Middelkoop et al., 1995).

#### **4.4.4.2 Circadian Rhythm Analysis**

Wrist actigraphy has been used as an indicator of circadian rhythms as it estimates sleep and wakes states reliably (both for the major sleep episode and for naps) and also provides a measure of daytime activity levels (Littner et al., 2003). A stable circadian rhythm has a consistent 24 hour sleep cycle and this is reflected in periodic cycles of body temperature, hormone production and physical activity over an extended period (Lockley, 2009). Deviations from this pattern may be as a result of circadian rhythm disorders such includes shift work sleep disorder, jet-lag disorder, advanced and delayed sleep phase syndrome, or aging and psychiatric disorders (Lockley, 2009). In some cases, it may be as a result of unusual environmental photo-periods.

#### **4.4.4.3 Limitations of Actigraphic Monitoring**

There are multiple limitations to actigraphic sleep monitoring such as movement artifacts (arising from non-compliance or breathing). Non-compliance can arise due to a forgetful participant or due to an individual who finds the device an irritant (Behan et al., 2008a; Walsh et al., 2008). Motion artifacts can arise due to breathing or from the participant's location (for example, if the subject were on a moving vehicle the actigraphic device would also record the acceleration of the vehicle). Another limitation of wrist actigraphy is a low specificity (a low ability to detect wake) (Paquet et al., 2007) and as such it

has been shown to be problematic when a large proportion of the recording period is spent awake (Sadeh, 2011) (for example, for participants suffering from insomnia (Acebo, 2006)). A particular problem is the detection of the correct sleep start time as subjects often remain still in an effort to fall asleep. The effect of misclassifying wake epochs at the beginning of sleep can be lessened through the introduction of an algorithm which requires an extended continuous period of no movement in order to demarcate the start of sleep (as previously discussed in Section 4.4.2). Multiple scoring algorithms have been developed for certain sleep disorders (Hauri and Wisbey, 1992; Hedner et al., 2004).

#### 4.4.4.4 Advances in Wrist-based Actigraphic Monitoring

Various advances have been made in wrist-based actigraphic monitoring over the years. For example, a wrist-based wireless activity monitoring device which continuously monitors body movements and derives circadian rhythm profiles from sleep/wake patterns has been developed (Sarela et al., 2003; Ltjnen et al., 2003). The data is also automatically transmitted continuously to a base station. The Vivago WristCare<sup>®</sup> (IST International Security Technology Oy, Finland) was found to be comparable to traditional wrist actigraphy devices validated against PSG in a study of 32 subjects (Ltjnen et al., 2003). A lower Pearson correlation coefficient for TST from WristCare<sup>®</sup> and PSG was found when compared to actigraphy and PSG (0.43 and 0.70 respectively) (Ltjnen et al., 2003). A lower TST difference was reported for actigraphy than for WristCare<sup>®</sup> (41 and 59 minutes respectively; validated against PSG) although this difference was not found to be significant (Ltjnen et al., 2003). Both technologies were found to over-estimate sleep. Similar findings were reported from the two devices using a custom nap analysis algorithm (Ltjnen et al., 2003). The system was deployed into a study investigating the circadian activity rhythms in 23 demented and 19 non-demented nursing-home residents for at least 10 days each (Paavilainen et al., 2005a). Daytime alertness and subjective assessments of sleep were also recorded. Weak, but significant, correlations between activity parameters and self-assessments were found. It was concluded that the system was a valid instrument for unobtrusive continuous long-term monitoring of the circadian rhythm and sleep/wake patterns of the elderly (Paavilainen et al., 2005a). This system was deployed in a study investigating its use as a tool for long term monitoring of 16 nursing home residents for several months (Paavilainen et al., 2005b). Findings were presented as case reports suggesting that actual health status was linked to circadian activity rhythm (Paavilainen et al., 2005b). This device has also been shown to provide information regarding the classification of daytime activities (Mattila et al., 2008).

## 4.5 Advances in Sleep Monitoring

### 4.5.1 Automated PSG Sleep Analysis

The specific, well defined rules for sleep scoring were established by Rechtschaffen and Kales (R&K) in 1968 (Rechtschaffen and Kales, 1968) were recently updated by the American Academy of Sleep Medicine (AASM) in 2007 (Iber et al., 2007). This is a time-consuming process in which a trained rater manually reviews various physiological signals recorded during sleep and scores each 30 second epochs as a specific sleep stage. The automation of this process has proven to be extremely challenging. This is made more difficult due to the presence of significant levels of inter-rater variability, as previously noted. The agreement rate between two individuals scoring sleep has been found to be of the order of 82% and 80.6% for the AASM and R&K standards respectively (Danker-Hopfe et al., 2009). Despite these difficulties, current automated sleep scoring systems are able to achieve performances comparable to these inter-rater agreement rates (Danker-Hopfe et al., 2009).

Furthermore, the percentage agreement rate, epoch-by-epoch agreement rate and Cohen's Kappa have been used to validate a classifier's ability to score sleep correctly.

The following section describes recent advances in the field of automated sleep scoring. Approaches are categorised according to the range of signals used; either a partial or a full compliment of PSG signals.

#### 4.5.1.1 Partial PSG Automated Sleep Scoring

The following section describes multiple approaches to PSG scoring using either EOG, EEG, or EOG and EEG data. These are summarised respectively in Tables 4.1, 4.2 and 4.3.

**EOG** Virkkala et al. (2007) investigated using solely EOG data (extracted from the PSG data set) to estimate sleep stages throughout a night in 265 male train drivers and railway traffic controllers. The sample population was split into training and validation data sets, containing 132 and 131 subjects respectively. The PSG data was manually scored according to the R&K standard (Rechtschaffen and Kales, 1968), however stages 3 and 4 were grouped and jointly referred to as SWS. Features were extracted using EOG data from the left eye referenced to the mastoid (EOG L-M1), and for the right eye (EOG R-M1) and for the left eye referenced to the right eye (EOG L-R). A metric corresponding to alpha (0.5-6Hz in this case) power was extracted using EOG L-M1 and

Author	Accuracy	Feature	Classifier	Data
Virkkala et al. (2007)	Agreement rate: 71.8%, kappa = 0.62	EOG	Rule based decision tree and smoothing process	n=265 (train drivers and railway traffic controllers)

TABLE 4.1: A summary of an approaches to automated PSG scoring using EOG data

another metric relating to beta (18-30Hz in this case) power was extracted using EOG L-R. Cross correlation and synchronous activity metrics between both eyes were also used as inputs to the classifier. These resultant features were used to describe slow eye movements, synchronous activity of the eyes and alpha and beta activity. A decision tree used threshold rules to estimate the current sleep stage; later a smoothing process was applied. An agreement rate of 71.8% and a Cohen's Kappa of 0.62 was reported between the smoothed estimated sleep stages and the manually scored results. Subject specific thresholds improved the accuracy further to 72.5% and 0.63 respectively. The article concluded that in comparison to other studies the manually scored data was not consensus scored. It was also noted that this system is more convenient and has the ability to be much more ambulatory than full PSG monitoring though it did note that a feasibility study would be required.

**EEG** Prinz et al. (1994) investigated automated sleep scoring in 115 older adults (70 training, 45 validation) using solely EEG features. The PSG data was scored as per the standard R&K guidelines (Rechtschaffen and Kales, 1968). A Fourier analysis was applied to the EEG data resulting in a spectrum estimation for each signal. The mean and standard deviation of 18 frequency bands were used as features to identify sleep stages. Rules derived from the guidelines were used to separate data and estimate sleep stages. The Pearson correlation (mean proportion of agreement) between the automatic system and the manual scorer was 0.74 with a Cohen's Kappa of 0.57.

Pardey et al. (1996) used EEG data from 9 healthy female adults. A 10th order Autoregressive (AR) model was used to estimate 1 second of EEG data. The resultant 10 AR coefficients were used as feature vectors to train a 10-by-10 Self-organising Map (SOM). For each stage of sleep over the entire data set, SOM are populated. A direct comparison of SOM showed that it was optimal to separate the SOM into wake, REM and stage 4 sleep and subsequently used as methods to infer wake, REM/light sleep and deep sleep respectively. A neural network, with one hidden layer, was trained using the 10 AR coefficients as inputs to produce 3 outputs relating to the probabilities of wake, REM/light sleep and deep sleep. For example, an output of [1,0,0] would signify

the current state is wake. An agreement rate of 80% was reported between consensus (joint agreement between two manual scorers) scoring and this method. An examination of misclassified data showed that this was mainly due to stage 3 and REM EEG data being incorrectly scored as stage 2 data. It was concluded that this analysis should not be used to replace manual sleep scoring but rather to provide detailed diagnostic information regarding sleep fragmentation and sleep disorders for clinicians.

Berthomier et al. (2007) reported on the validation amongst 15 healthy individuals of an automatic sleep analysis system using single-channel EEG. Sleep stages were scored using the R&K guidelines (Rechtschaffen and Kales, 1968) by 2 independent scorers. Epochs jointly classed as belonging to the same stage by both scorers were compared to classes estimated by this system. Features used in this classifier were spectral energies from [0-4 Hz], [4-8 Hz], [8-12 Hz], [12-16 Hz] as well as temporal information and detectors of sleep spindles, K complexes and alpha bursts. An adaptive fuzzy logic system, based on R&K rules, which repeatedly updates the sleep stage classifier was employed in the analysis. Multiple results for increasing numbers of sleep stages showed high agreement levels. An 82.9% agreement rate ( $\kappa = 0.72$ ) was reported for 5 states (Wake/REM/Stage 1/Stage 2/SWS).

Flexer et al. (2005) developed a probabilistic continuous sleep stager based on Hidden Markov Models (HMM) with a single EEG signal. The sleep stages were scored using standard R&K guidelines (Rechtschaffen and Kales, 1968). Two healthy adult populations from separate labs were used to test the system; these consisted of 40 people (20 training, 20 test) and 28 people (14 training, 14 test). Metrics of EEG reflection coefficients and stochastic complexity were used as features in this system. A Gaussian observation HMM (GOHMM) was used to classify the data into wake, deep sleep, and REM/light sleep as per Pardey et al. (1996). High accuracies of 79%, 82% and 68% were reported for wake, deep sleep and REM/light sleep respectively for the first data set; however low classification results were reported for discriminating stage1, stage2 and stage 3 sleep (24%, 36% and 35% respectively). In the second experiment at a separate sleep lab, data from 14 whole nights resulted in accuracies of 25%, 16%, 38%, 27%, 87% and 61% for wake, stage 1, stage 2, stage 3, stage 4 and REM sleep respectively. The lower accuracies reported could not be attributed to a difference in the cohort. A bigger overlap between data in the feature space was found for this data set explaining the poorer performance. This resulted in a conclusion that classifiers need to be customised to individual laboratories through the use of laboratory specific training data. Flexer et al. (2005) concluded that the approach was not intended as an automatic sleep scoring system according to the R&K rules, but rather it aided in the development of an alternative sleep scoring system which focuses on the restorative value experienced during the sleeping period.

Gudmundsson et al. (2005) investigated sleep staging approaches on EEG data from young subjects using nearest neighbour and support vector machine classifiers. The PSG data was scored into traditional R&K sleep stages (Rechtschaffen and Kales, 1968) by an experienced neurologist. Data was used from 4 young subjects (mean age of 5 years); 3 data sets were used for training and the remaining data from 1 subject was used for test data. Hjorth complexity parameters (Hjorth, 1975), spectral energies in various frequency bands and histogram features on waveform measures of EEG data were used separately as features in this system. Both support vector machines (SVM) and nearest neighbour approaches were used to estimate sleep states from inputted features. Five fold cross validation was used to find optimal parameters for the SVM. Posterior probability estimates were used to smooth out transitions between states by measuring the likelihood in changing state from past observations (in the training set). Using the histogram features with the posterior probability method resulted in the highest accuracy (although the exact definition of this accuracy was not provided) at 0.81 or 81%. The posterior probability method improved accuracy for all methods except in one instance where it remained the same.

Ebrahimi et al. (2008) developed a sleep stage classifier based on EEG signals using neural networks and wavelet packet coefficients. EEG data was taken from the Physionet (Goldberger et al., 2000) database from 7 healthy adults and used in this analysis. From a wavelet decomposition of the EEG, the following features were used: mean energy from each band, total energy, ratio of various energy bands to total energy, mean of the absolute values of wavelet coefficients in each sub-band and the standard deviation of these coefficients in each sub-band. A feedforward multilayer perceptron neural network trained using the backpropagation algorithm was used to classify sleep stages. A bootstrap technique was used to even the proportion of samples in each stage. Bootstrapped samples were used to derive feature vectors and were separately used in training and testing. The classifier successfully discriminated between wake/stage 1-REM/stage 2/SWS with a specificity of 94.4%, sensitivity of 84.2% and an accuracy of 93.0%.

McGrogan et al. (2001) further built upon previous work by Pardey et al. (1996) by devising a neural network sleep analysis system which used reflection coefficients derived from the EEG as features to train a classifier. Reflection coefficients provide similar spectrum information as autoregressive (AR) coefficients, derived from an AR model, however they are independent of signal amplitude and have the range  $[-1, 1]$ . The outputs of the classifier inferred the likelihood of a wake, REM/light sleep and deep sleep state and were in the range  $[-1, 1]$ . The feature space was trained using a total of 8,502 30 second epochs of EEG data from 9 individuals. Estimation of sleep stage was achieved using Mahalanobis distance as a tool for selecting the closest cluster in the training

data set to the current set of data. This distance metric allows for different variance in each class in the calculation of distance in contrast to Euclidian distance which does not. A leave-one-out strategy was employed for the generation of results. The data from 8 subjects were used to calculate mean and covariance values in order to find the Mahalanobis distance. This was used to estimate the current sleep stage in the remaining set of data. A comparison against consensus (of three sleep scorers) and individual scoring took place and provided an agreement rate of 72.2% and 63.3% respectively. Caffarel et al. (2006) reported on using this system to assess the sleep/wake and sleep stage epoch-by-epoch comparison against manual scoring in 114 patients with suspected obstructive sleep apnoea syndrome (OSAS). There was poor overall agreement with a Cohen's Kappa ( $\kappa$ ) of 0.305 for differentiating Wake/Light Sleep/Deep Sleep/REM and of 0.445 for the simpler case of Sleep/Wake. It was concluded that using such a system was not sufficiently accurate for sleep study analyses among the OSAS patients.

**EEG and EOG** Hassaan and Morsy (2008) investigated an adaptive hybrid system for automatic sleep staging. PSG data was taken from 10 healthy adults (7 males). A single EEG channel and the left EOG channel were used to estimate sleep stages. Reference sleep stages were estimated using the Alice sleep staging system (Goldberger et al., 2000). The features used for classification were from the EEG: sum of power of spectrum for [0.5-2, 2-7, 8-12, 14-25, 25+ Hz], the mean frequency from each of these bands, the alpha wave index, the theta wave index and the slow wave index. EOG data was used during post-processing to differentiate between stage 1 sleep and REM sleep; specifically measures of summed and max peak-to-peak values (characterising the rapid increase in eye movements) were used. A feed-forward ANN with back propagation training was used to classify the data. Comparison against the Alice automated scoring system reported an accuracy of 69.96%.

**Sleep Spindle and K-Complex Detector** There has been significant research in the development of automated sleep spindle and K-complex detectors (Ahmed et al., 2009; Huupponen et al., 2007; da Rosa and Paiva, 1993; Richard and Lengelle, 1998). These EEG waveforms define stage 2 sleep and are distinguishable by their shape (Iber et al., 2007; Rechtschaffen and Kales, 1968). The distinct high frequency train of sleep spindles results in a distinguishable high frequency component while time and frequency based analyses have been used to identify k-complexes. Often each 30 second epoch is broken into 1 second chunks and each segment is tested for sleep spindles and K-complexes.

Author	Accuracy	Feature	Classifier	Data
Prinz et al. (1994)	Pearson's Correlation Coefficient: 0.74, kappa = 0.57	EEG	Rule based system	n=115 (older adults)
Pardey et al. (1996)	Agreement rate: 80 % (for wake/light sleep and REM/deep sleep discrimination)	EEG	Self organising map and ANN	n=9
Berthomier et al. (2007)	Agreement rate: 82.9 %, kappa = 0.72	EEG	Adaptive fuzzy logic system based on R& K rules	n=15
Flexer et al. (2005)	Mean accuracy: 73.3% (across wake/light sleep and REM/deep sleep discrimination)	EEG	Probabilistic HMM	n=68
Gudmundsson et al. (2005)	Accuracy: 81 % (elusive definition of accuracy)	EEG	SVM and nearest neighbour classifier	n=4 (mean age: 5)
Ebrahimi et al. (2008)	Specificity: 94.4%, sensitivity: 84.2%, accuracy: 93% (for wake/stage 1 and REM/stage 2/SWS discrimination)	EEG	ANN	n=7
McGrogan et al. (2001)	Agreement rate: 72 %	EEG	Automated clustering	n=9

TABLE 4.2: A summary of some approaches to automated PSG scoring using EEG data

Author	Accuracy	Feature	Classifier	Data
Hassaan and Morsy (2008)	Accuracy: 69.96 %	EEG and EOG	ANN	n=10

TABLE 4.3: A summary of some approaches to automated PSG scoring using EEG and EOG data

#### 4.5.1.2 Full Complement Automated Sleep Scoring

The automation of sleep scoring using EEG, EOG and EMG signals has been investigated almost since the development of criteria for the scoring of sleep and sleep apnoea in healthy adults and children (Anderer et al. 2005; Anderer et al. 2010; Goldberg and Beiber 1979; McGrogan et al. 2001; Pardey et al. 1996; Park et al. 2000; Smith and Karacan 1971; Villa et al. 1998). Despite these multiple approaches, computerised analysis has not been accepted by the sleep community. Schulz (2008) emphasised that a shift in the paradigm of sleep analysis is needed as traditional visual scoring techniques have undergone few changes despite major technological advances. It has been argued that NREM sleep should be thought of as a continuous state of differing restorative depth as opposed to a series of discrete stages (Schulz, 2008). A cumulative score based on the quality, composed of both quality and quantity of sleep, was proposed. This could effectively deem that 5 hours of good quality NREM sleep could be the equivalent of 8 hours of poor quality NREM sleep (Schulz, 2008).

The following section briefly describes various automated sleep monitoring approaches using EEG, EOG and EMG data (these are summarised in Table 4.4):

Anderer et al. (2005) developed the Somnolyzer 24/7 automatic sleep classification system using the Siesta (a large EU funded) database of PSG data. Normal and patient (insomnia, Parkinson's disease, PLMS or sleep apnoea) PSG data from 590 subjects was split evenly into training and validation sets (396 normal and 194 patients). Standard sleep scoring was performed according to the R&K guidelines (Rechtschaffen and Kales, 1968) by 2 trained scorers; consensus (a third scorer validated mismatched epochs). A large number of features were extracted from the EEG, EOG and EMG data (101 features in total). This included density, intensity, max intensity, frequency, amplitude, duration, velocity and waveform measures of EEG (in the form of slow wave, muscle artifact, delta, theta, alpha, beta-1, beta-2 omega/total band, possible/probable/certain spindles), EOG (in the form of possible/probable/certain slow eye movements, REM, eye movement) and EMG (minimum, maximum and mean). An LDA decision tree hierarchical classifier was used to break the problem of classifying all the different stages into separate easier to distinguish stages. Thus each classifier uses implicit (learned) and explicit (rule based) knowledge on features to separate stages. For example, the first classifier separates wake/light/REM sleep from SWS; the explicit knowledge taken from choosing which states to separate and the implicit knowledge is taken from training data. The training classification reported an agreement of 79.9% ( $\kappa = 0.74$ ). Validation data without quality control (reviewing scorer) reported an agreement of 78.3% ( $\kappa = 0.71$ ) and an agreement of 79.6% ( $\kappa = 0.72$ ) using a reviewer for quality control. Quality controls included the absence of any REM stage, too many REM stages, artifacts, missing

data, etc. Comparisons between scoring after quality control by two separate scorers had a high accuracy 99.4% ( $\kappa = 0.991$ ) between two partially automated scorings. Further analysis with a smaller sample reported higher Somnolyzer 24/7 vs consensus scores for sleep estimation than inter-rate scores (Anderer et al., 2007, 2010).

Park et al. (2000) used a rule-based scoring technique and a hybrid rule-based and case-based scoring technique to develop a system which reported over 87% and 82% agreement rates between manual scoring (according to the R&K guidelines) and an automated algorithm in a small number of 3 normal adults and 3 OSAS patients respectively. Features used for classification include EEG frequency information on a per second basis for each epoch to derive the temporal distribution of different frequencies (including [0-2 Hz], [3-7 Hz], [7.5-12 Hz], [12-20 Hz], [20-50 Hz] and [11.5 - 15 Hz] for sleep spindles), total EEG power spectrum per epoch, state of the EOG (divided into SEM, drift, delta, quiet and normal) and the tone of the chin EMG. The rule based scoring techniques mirrored the R&K guidelines. Case based reasoning introduces into the system the ability to learn from past examples. If the rule based method cannot sufficiently classify data, the case based reasoning engine is invoked which compares the current data to previous examples and finds a matching solution, and correspondingly an answer, or sleep stage.

Schaltenbrand et al. (1996) developed an automatic sleep scoring system based on test data from 60 subjects (33 males). These were evenly distributed between normal controls, depressed patients and insomniac patients. 17 features were extracted from PSG data sets and applied to a classifier (14 EEG features, 4 EOG features and 3 EMG features). These features included using a mixture of relative spectral power in specific frequency bands, total spectral power, ratio of spectral powers, dispersion of spectral density and mean frequency of spectral density in EEG, EOG and EMG data. The Fast Fourier Transform (FFT) was used to achieve these spectral features. A MLP ANN was trained using these features to predict 6 scored states (wake/stage 1/stage 2/stage 3/stage 4/REM). Manual scoring by consensus agreement between two trained sleep scorers performed according to the R&K guidelines (Rechtschaffen and Kales, 1968) provided the reference stages. Training and test data were kept separate throughout this analysis. 12 participants data were used for training the classifier. These were kept in the same proportion as the testing set (4 normal, 4 depressed patients and 4 insomnia patients). The testing set contained 48 participants (16 normal, 16 depressed patients and 16 insomnia patients). The results were optimised based on the training data and subsequent final testing was performed with the remaining data. Average agreement rates over all stages for each set of subjects was over 80%. The related inter-rater agreement rates were over 85%. Improved performance was achieved by the creation of an uncertainty index. Expert supervision of epochs with a high uncertainty index improved

expert/automatic scoring agreement rates to 89%. The detection and subsequent correction by clustering of results (defined by Euclidian distance to a cluster of data) improved performance to 84%. This affected nearly 20% of all data and improved an average of 30 epochs per night (3% per recording).

Agarwal and Gotman (2001) extracted features from epochs of PSG data from 12 subjects (9 male) and grouped them into their naturally occurring clusters. The number of clusters was reduced to 8 using the nearest neighbour method over the entire data set. The reviewer examined sample data from each cluster and assigned a sleep stage. This sleep stage was applied to all data sets in that cluster. This allowed for the unique PSG data from each subject and the customisation of scoring for each reviewer. PSG records from normal, abnormal, male, female and varying age groups were used to validate this computer-assisted sleep staging approach. The features derived from the physiological data were amplitude, dominant rhythm, and frequency weighted energy for each EEG and EMG channel, presence of spindles in the central EEG channel, alpha-slow-wave index for the occipital EEG, theta-slow-wave index for the central EEG and presence of eye movements in the EOG. An overall concordance of 80.6% with manual scoring of 20 second epochs, according to the R&K guidelines, was reported. This agreement rate was based upon sleep staging data which had been previously scored (with which it had an agreement rate of 76.8%) and was subsequently re-scored (corrected) by the current reviewing scorer. It was concluded that this could introduce a bias as the current reviewer could use the computer assisted scores as a possible indicator of sleep stage when reviewing data. It was also stated that this method approached a high accuracy amongst a variety of individuals, whereas other techniques with higher accuracy had a sample population with a tight population bounds (eg. all male, 18-23) (Kuwahara et al., 1988).

Pittman et al. (2004) developed an automatic PSG scoring system and tested it in a population with suspected sleep-disordered breathing (SDB). A total of 31 subjects (9 women) took part in the study. Standard PSG was recorded in all cases and this was scored independently by 2 trained scorers in the same lab. Automated algorithms extracted features from the EEG, EOG and EMG. Algorithms were developed to locate sleep spindles, K-complexes, movement and electrode artifacts. Additionally high-frequency EEG data as well as movements were deemed to signify wake, low-energy mixed-frequency EEG data referred to stage 1 and Stage REM sleep, high-energy mixed-frequency, spindles and K-complexes highlighted stage 2 sleep and low-frequency EEG patterns and peak-to-peak amplitudes were considered as indicators of delta waves (SWS). Reduced EMG activity, estimated using fuzzy clustering, and REM indicated REM sleep. The EOG signal was analysed and variance and cross-correlations metrics were extracted; segments with high relative energy and low cross-correlation measures

Author	Accuracy	Feature	Classifier	Data
Anderer et al. (2005)	Agreement rate: 78.3 % (79.6 % if uncertain epochs are supervised), Kappa = 0.71	101 EEG, EOG, EMG features	Hierarchical LDA	n=590 (396 normal, 194 patient)
Park et al. (2000)	Agreement rate: 87%	13 EEG, EOG, EMG features	Rule and case based scoring technique	n=6 (3 normal, 3 OSAS)
Schaltenbrand et al. (1996)	Agreement rate: over 80 % (89% if uncertain epochs are supervised)	17 EEG, EOG, EMG features	ANN	n=60 (20 normal, 20 depressed, 20 insomniac)
Agarwal and Gotman (2001)	Agreement rate: 80.6 %	EEG, EOG, EMG	Manual clustering	n=12
Pittman et al. (2004)	Agreement rate: 73.3 %, Kappa = 0.61	EEG, EOG, EMG	Rule based system	n=31 (SBD)

TABLE 4.4: A summary of some approaches to automated PSG scoring using EEG, EOG and EMG data

were used to define REM sleep. EMG signals with high relative energy were used to score arousals as they were noted as EMG bursts. This rule-based system was devised using the R&K guidelines (Rechtschaffen and Kales, 1968). Inter-rater reliability was 82.1% and had a Cohen's kappa ( $\kappa$ ) of 0.73. Consensus scoring was not used in this study. Agreement rates and Cohen's kappa between scorers and against the consensus PSG were reported as 77.7% and 0.67, and 73.3% and 0.61 over all subjects respectively.

## 4.5.2 Contact-Based Sleep Monitoring

Advances in physiological monitoring technology has brought about a move away from traditional electrodes and towards dry electrodes, textile-based sensors and ambulatory physiological monitoring. This section describes some recent advances in this area particularly applied to sleep monitoring.

### 4.5.2.1 Electrode-based Advances

**The Actiocular Monitor of Sleep:** Kaye et al. (1979) reported on a device used to monitor sleep by recording eye movement (via a piezo-electric transducer placed on the eyelid), submental EMG and body movement (via a sensor placed on the back of the hand). Data for 8-9 hours of sleep over 6 subjects were presented showing a high statistical correlation with wake, NREM and REM sleep using standard PSG recordings.

**Nightcap: A Home Based Sleep Monitoring System:** Mamelak and Hobson (1989) devised a method to monitor eye movements using a strain gauge placed on the eyelid. A piezo-electric cartridge was used to monitor body movements. Movement impinged upon a piezo-electric transducer generates a charge and through recording these charges movement was recorded. Data from a total of 4 subjects (2 female), 15 nights in total (4 nights each, except for one subject) were collected for this study. The system used threshold algorithms on eye movement and body movement data to estimate sleep stages. The agreement rate of hand-scored records against PSG was 87.57% for scoring wake, movement, NREM and REM sleep. Mean values for sleep onset and REM latency were within 1.6 and 10.8 minutes, respectively, compared to PSG records.

**Dry Electrodes - myZeo Personal Sleep Coach:** Recently the *Zeo Personal Sleep Coach* (<http://www.myzeo.com/>) was presented as a wireless dry headband technology for automatic sleep monitoring (Blake et al. 2009; Fabregas et al. 2009; Shambroom et al. 2009; Wright et al. 2008a; Wright et al. 2008b). A single channel is acquired using a dry contact electrode sensor integrated into the forehead section of a headband. This data is transmitted wirelessly to a base station which performs a neural network classification to estimate sleep stages. The published results have shown good correlation for TST between this system and two independent trained manual sleep scorers. Good comparisons were found for sleep efficiency, number of awakenings and the estimation of sleep stages. A validation study of this technology against manually scored PSG by two trained scorers showed an agreement rate of 75.8% and 74.7% for all sleep stages, and 92.6% and 91.1% for sleep/wake discrimination (Shambroom et al., 2011). Inter-rater reliability was 83.2% for all stages and 95.8% for sleep/wake discrimination (Shambroom et al., 2011). This device shows promise as an ambulatory, easy to use sleep monitoring system. This system is marketed towards nightly in-home use in order to improve sleep quality. Although this device is in direct contact with the participant, it can easily collect data from one person regardless of whether the bed is shared or not. A review of the comfort of this device is required as anecdotal comments question its suitability.

#### 4.5.2.2 Embedded and Textile Electrodes

**ECG Recording in a Bed:** An indirect method of recording heart rate variability (HRV) was presented by Lim et al. (2007). By placing high-input-impedance active electrodes on a mattress beneath a cotton bed sheet, it was shown that the R-peaks of each heartbeat (a cyclic peak distinguishable from each heart beat) could be recorded. This was compared to a direct contact ECG method and worked regardless of body

position and location on the bed, although in-bed body movement proved problematic. This embedded electrode approach was noted to be adequate for long-term use.

**Monitoring Bed Temperature in Elderly in the Home:** Bed temperature was proposed as a method of characterising the sleep of an elderly population as a circadian variation in temperature occurs (Lim et al., 2007). By placing a thermistor on top of the mattress, below the waist of the participant, temperature was recorded. The resulting signal was also shown to correlate well with bed movement. Two studies (one in a nursing home and one in a participant's domestic house) highlighted how time in bed and movements (calculated using the derivative of bed temperature) can be measured. However, it should be noted the practicality of this method is limited when the bed is shared.

#### 4.5.2.3 Ambulatory Physiological Monitoring

The development of an ambulatory physiological sleep monitoring technology would be very advantageous as it would allow the continuous monitoring of sleep and wake (using physiological signals as well as movement signals). The following section describes various sleep monitoring technologies which can be considered to be ambulatory:

**Ambulatory Cardiopulmonary Monitoring:** Ambulatory monitoring of cardiovascular, respiratory, motor-behaviour and experiential responses has been developed in the form of textile sensors embedded into the *Lifeshirt<sup>TM</sup>* (Vivometrics, Inc., Venture, CA, USA) (Coyle, 2002). Plethymography sensors, a 3-lead, single channel ECG, a 2-axis accelerometer and an user-input device were embedded into a single comfortable garment. The *Lifeshirt<sup>TM</sup>* has been suggested for use in sleep monitoring as well as for other uses, but not for capturing ambulatory PSG (as this requires EEG, EOG and EMG).

#### **WP100 - Arousal Detection using Peripheral Arterial Tonometry, Oximetry and Actigraphy:**

The WP100 (developed by Itamar Medical Ltd. Caesaria, Israel.) is a wrist worn device which records actimetry and also uses finger probes to measure pulse oximetry and Peripheral Arterial Tonometry (PAT) (Herscovici et al., 2007; Pillar et al., 2003). PAT is an indicator of sympathetic tone variations found by measuring pulsatile volume changes at the extremes of the body, in this case in the finger (Herscovici et al., 2007). PAT and pulse rate were shown to be able to monitor arousals from sleep due to an increased sympathetic activation and, hence, peripheral vasoconstriction during

these times (Pillar et al., 2002). Additionally a high correlation between AASM-defined arousals derived from the PSG and arousals estimated from the WP100 across a wide range of values and subjects was reported (68 subjects; 7 healthy volunteers, 61 with suspected OSAS). The WP100 reported a sensitivity of 0.80, specificity of 0.79 and an area under the Receiver Operator Characteristic (ROC) of 0.87 in detecting patients with more than 20 arousals per hour of sleep (Pillar et al., 2003). Herscovici et al. (2007) reported on a technique to identify 30 second epochs of REM sleep using the WP100 and reported a sensitivity of 78%, specificity of 92% and agreement of 89%. 16 features derived from PAT amplitudes and these PAT-derived metrics were aggregated into a prediction function which used an optimised genetic algorithm to determine the likelihood of REM sleep. A similar approach was taken to discriminate between light and deep sleep which reported a sensitivity, specificity and accuracy of 66%, 89%, 82% and 65%, 87%, 80% for training and validation data respectively (Bresler et al., 2008). Using the actigraphic and PAT data, algorithms have been developed to separate wake, light sleep, deep sleep and REM. The performance of the system to discriminate between all sleep stages over a full night was validated with 227 subjects (38 normal, 189 OSAS) and this reported a Cohen's  $\kappa$  coefficient of 0.475 for all stages of sleep and 0.549 for sleep/wake (Hedner et al., 2011).

### 4.5.3 Non-Contact Sleep Monitoring

Significant work is currently being carried out on developing an unobtrusive, non-contact sleep monitoring system. This would be particularly suited to a pervasive, ambient, distributed in-home healthcare monitoring system.

#### 4.5.3.1 Optical Monitoring

Although optical monitoring (including video and non-video sources) is cheap, easily deployable and completely non-contact, subjects often raise privacy concerns finding the technology intrusive. Some technologies currently under investigation are described below.

**Optical Flow:** A video based sleep monitoring technique was devised which extracted apparent velocity of motion, or optical flow, over a sequence of images to observe chest or blanket movement (Nakajima et al., 2001). Upon investigation of the optical flow metric during quiescent periods, a respiration signal was evident despite the participant being covered by a blanket. Large body movement could also be discriminated. Difficulties

with non-uniform ambient light due to sunrise and other changes in ambient conditions over the recording period were noted.

**Difference Images:** Using the differences between subsequent images in a video of a person sleeping, a system was developed to track the movements of a cohort of 5 children (Okada et al., 2008). Movement tracking was performed at a sampling rate of 1Hz and was recorded with concomitant PSG. A decrease in motion was reported as the participant progressed from wake to REM sleep to stage 1/stage 2 sleep to SWS. It was also noted that changes in sleep stage were often accompanied by movement.

**PIR Sensors:** PIR motion detectors were used to monitor body movement in bed in 2 separate studies (Choi et al., 2006; Shin et al., 2003). In one study, a PIR motion detector was placed at each of the four corners of a bed and it recorded movements at 100Hz (Shin et al., 2003). A wavelet decomposition of each of the 4 signals produced multiple coefficients which represented the movement recorded at different frequencies. The 8th detail coefficients (representing a frequency range of 0.19-0.39Hz) were used as features to train an adaptive neuro-fuzzy inference system which discriminated between movement and non-movement epochs (Shin et al., 2003). A high detection rate for instantaneous (entire) and partial body movements was reported over multiple subjects. Another study used a single PIR sensor placed over the bed to discern periods of movement and non-movement (and this was used to infer wake and sleep) (Choi et al., 2006). An agreement rate between the proposed system and actigraphy of 92.2%, 89.7% and 95.4% for three participants over 5, 3 and 4 nights, respectively, were reported. The ability of the PIR to discern movement depends on the specification and manufacture of each individual PIR. The authors also stated the PIR system was more adept at reporting whole-body movement, being less intrusive, providing real-time monitoring and a high time resolution. Disadvantages of the system noted were being less portable and not offering 24 hour participant monitoring when compared to traditional actigraphy. Other disadvantages of PIRs include thermal artifacts due to sunlight and heaters, reduction in movement accuracy due to blankets and inter- and intra-PIR specification variations.

#### 4.5.3.2 Smart Phone Application - Sleep Cycle Alarm Clock/Sleep On It

*LexWare Labs Ltd., Goteborg, Sweden* developed an *iPhone*, (*Apple Inc., CA, USA*) application which monitors the participant's sleep cycle and wakes them at the most opportune time nearest their scheduled wake time (Drejak, 2010). This software, called *Sleep Cycle alarm clock* (<http://www.lexwarelabs.com/sleepcycle/>), uses the in-built accelerometers and microphone within the iPhone to measure body movements

and estimate sleep state. A longitudinal objective comparison of estimated sleep metrics, including the estimated time spent in 4 states (transition to sleep and REM, light sleep, medium/deep sleep and deep sleep) throughout each night is presented to the user. There has been no scientific validation of this application, although it uses similar approaches to existing actigraphic technology. Variations due to individual mattress characteristics which the system does not cater for has been noted by the developers (Drejak, 2010).

Med Help (2012) is another smart phone application which allows users to record their sleep and wake times. It also allows recording of subjective scores of mood, sleep quality and medication. Users are able to review their own sleeping patterns in order to increase their understanding of their own sleeping patterns.

#### **4.5.3.3 Bed Position Monitoring (Bed Sores)**

Some research has taken place to quantify lack of movement in bed and to monitor the development of resulting bed sores (also known as pressure ulcers) (Hsia et al., 2009; Hnatiuc et al., 2009). Hnatiuc et al. (2009) proposed an intelligent mattress to monitor periods of inactivity based on using fuzzy rules and using pressure amplitude data and length of time between movements (Hnatiuc et al., 2009). Hsia et al. (2009) investigated using a force sensing resistor (FSR) based sensor mattress to determine sleep postures for the prevention of bed sores (Hnatiuc et al., 2009). For this, two different sensor layouts were used; one used 16 FSR parallel strips and the other used a FSR matrix with an increased number of sensors around the torso providing greater resolution and FSR strips placed at the other end of the bed. The shape of the pressure distribution was estimated using Kurtosis and Skewness. A second approach used advanced classification methods to discern sleeping posture. Features extracted include using a PCA based decomposition of pressure values, raw data and descriptive statistics such as mean, root mean square, variance and standard deviations of pressure readings over a specific area. An SVM classifier was used to train and test this estimation of sleeping posture. The data was split between a training set (3 people) and a test set (5 people) for the two different sensor layouts across 6 typical sleeping postures. The accuracy of the Kurtosis and Skewness approach was 100% for sleeping postures parallel to the central line of the bed; however this decreased to 81.43% with deviation from this central line. The classification accuracy for the SVM method was 64.66% for the PCA decomposition of the pressure data, 83.5% for the raw data and 77.66% when the descriptive statistics were used. This method catered for sleeping postures other than those parallel to the central line of the bed. The template matching method was shown to perform well. It would be interesting to note how long was spent parallel and off parallel of the central line of the bed.

#### 4.5.4 Non-Contact Vital Signs Monitoring

##### 4.5.4.1 Ballistocardiography

Ballistocardiography (BCG) is a relatively old technique to measure the movement of the heart and blood flow in the body due to the body's incidence on a non-invasive sensor (Rosenblatt, 1957). Techniques for establishing the BCG have used direct and indirect contact methods including lying directly on a suspended frame (Eblen-Zajjur, 2003), pressure sensors under the legs of a table or bed (also known as load cells) (Adami et al., 2009a,b) and pressure sensors placed within, over and under the mattress (Behan et al., 2008a; Seeton and Adler, 2008; Walsh et al., 2008; Watanabe et al., 2005); some of these are detailed further below. A generic standard BCG for a person under 50 years of age, and its ECG counterpart, can be found in (Rosenblatt, 1957). However an abnormal BCG is often found in a significantly large population of individuals over 50 years of age who have no cardiovascular complaints (Rosenblatt, 1957). This limits the widescale deployment and acceptance of BCG devices. Generally, the respiration signal is apparent in the BCG and is often considered an artefact. The reliable extraction of both respiration and heartbeat, ballistocardiography (BCRG), would be advantageous for sleep monitoring, particularly if the sensor was an ambient technology collecting data in a seamless and unobtrusive manner. The following section describes recent advances in movement-based physiological signal estimation including BCG devices. Some work is also described which uses these derived metrics to infer sleep information such as sleep quality, restlessness, number and duration of awakenings, total sleep time and sleep architecture.

**Embedded Electromagnetic Coil:** The placement of a single coil electromagnetic sensor embedded into a mattress as a method of measuring respiration was proposed as a possible non-contact sleep monitoring solution (Seeton and Adler, 2008). By monitoring changes in the conductive substrate within the electromagnetic field of the coil (for example, the lungs) respiration could be accounted for. A single subject was used to investigate the efficacy of the respiratory signal extracted using the system against a pneumotachograph. A theoretical proof of the system was also presented (Seeton and Adler, 2008).

**Force Sensitive Resistors:** A force sensitive resistor (FSR) is a thin film sensor whose resistance decreases as pressure is applied (Van Der Loos et al., 2003). Body movement can be measured by monitoring the change in resistance of a distributed grid

of sensors spread throughout a bed. Two FSR sleep monitoring solutions are presented below.

120 FSRs were incorporated into a single textile sheet, placed on top of the mattress, which records movement and temperature impinged upon a bed (Van Der Loos et al., 2003). Vitals signs monitoring and sleep quality assessment was also achieved. The sensors were distributed in a grid-like fashion with a greater density around the torso of the subject. The sensor was sampled at 100Hz. Wavelet decomposition was used to estimate heart and breathing rates to an accuracy of 0.5 beats per minute averaged over a moving window of length 25 seconds with a 5 second overlap. A restlessness index during sleep was also derived over the 25 second window. This was calculated by integrating the absolute change in the estimated point position of the body. An adjustable bed frame was built to correct for poor sleep posture and to relieve sleep problems by tilting the subject over (Van Der Loos et al., 2003). Respiration and temperature were shown, using diagrams, to be estimated reliably however heart rate was not.

Another technology was developed using 210 FSRs (sampled at 20 Hz) which recorded body movement, articular (or joint) motion, respiration and heart rate and also inferred sleeping posture (Harada et al., 2002). Templates for different postures were used to estimate the subject's posture. Filtering the pressure signal within the respiration and heart rate bands respectively allows for both signals to be extracted. The respiration signal was extracted using time-based methods while a frequency based method was used to extract the heart rate.

For both of these studies, figures were used to show the ability of the system in extracting these physiological signals. However numerical results that would allow quantification of the reliability of this system for extracting body movement and physiological signals over multiple subjects and postures were not provided.

**Air Mattress:** Multiple sensors for reliable estimation of physiological signals during sleep using a thin air-filled pneumatic mattress placed on top of the traditional mattress have been developed (Carlson et al., 1999; Chee et al., 2005; Chow et al., 2000; Shin et al., 2010; Watanabe et al., 2005).

Carlson et al. (1999) used a peak-valley detection method to derive inter-breath interval and breath amplitude from the pulsatile pneumatic signal imparted on an air mattress. In this study data was recorded from 11 subjects (5 male) and compared to a capnograph (a measurement of carbon dioxide at the nose used to record breathing patterns). Overall 11 supine, 11 side-lying and 11 apnoea simulations were compared to visually

observed chest movements. The frequency of the breathing cycles for the proposed system and the visual observation correlated highly ( $r=0.99$ ). All simulated apnoeas had a standard deviation of inter-breath interval of greater than 3 seconds. Body movements, denoted by excessive breath amplitude and frequency, were detected more accurately than using wrist actigraphy (proposed system: 99% of events, Cohen's  $\kappa = 0.90$  vs. Wrist actigraphy: 50% Cohen's  $\kappa = 0.70$ ).

A system monitoring pressure changes impinged on two air cells, located under the abdomen and thorax connected to a balancing tube, within an air-mattress consisting of horizontal air cells was developed to assess sleep (Chee et al., 2005; Shin et al., 2010). The pressure signal was recorded using separate pressure sensing tubes connected between the balancing tubes. A mechanical human body simulator and 1 human subject was first developed to validate the operation of this system (Chee et al., 2005). ECG and respiration (measured by nasal air flow) were compared to the filtered and differentiated air-mattress pressure signals. A visual comparison of the ability of the system to measure cardiac and respiratory signals was shown; however this does not provide a detailed description of the efficacy of this sensor. A future clinical study of this system over a large cohort would provide quantitative statistics needed for the systematic validation of the sensor. It is noted that during high amplitude body movement the reliable extraction of physiological signals becomes extremely difficult; however this movement information itself is deemed an important sleep metric in the form of a movement index.

Shin et al. (2010) furthered this work by reporting on a population of 13 awake participants during the day. Respiratory effort, oronasal airflow, ECG and activity were recorded concomitantly with the air-mattress system. Autocorrelation algorithms were used to estimate respiration and heartbeat. The total mean difference in beat-to-beat heart rate was 0.68 beats/minute for all subjects and the average of the mean error reported was less than one percent. A Bland-Altman analysis showed that the mean difference was almost zero and that 95% of the heart rate differences were between -1.22 and 1.22 beats/minute. The mean difference of the breath-to-breath respiration rate was 0.5 breaths/minute with a relative error of 2.85%. There was an average correlation of 0.92 between the chest belt and the proposed method of respiration estimation. The system also showed good results for the detection of simulated snoring, simulated apnoeas and simulated body movements; reporting a Cohen's Kappa value of  $\kappa$  of 0.93, 0.85 and 0.92 respectively. A home deployment of the system was also performed on one subject for seven days. A restlessness index was derived using the percentage of the epoch for which physiological information could not be detected.

Chow et al. (2000) developed an air-mattress respiratory monitoring system by examining pressure variations on 4 cylindrical air cells. 4 male subjects were recruited for

this study. Over 90% sensitivity and accuracy values were reported for the detection of hypopneas against PSG and respiratory inductance band (RIP) measurements (specificity values were not reported). A low accuracy was reported for the detection of body movement derived from the RIP measurements. However a possible explanation for this would be the detection of non-core body movements (such as feet movements) for which further validation is warranted. The mean error between breathing rates from the RIP bands and the proposed system was noted to be very small (less than 0.2 breaths per minute).

Watanabe et al. (2005) have developed a sensitive and robust air-mattress which has been shown to extract heartbeat, respiration, snoring and body movement signals reliably. The extracted pressure signal was filtered according to the frequency ranges of heart rate, respiration rate and snoring. An analysis of the power spectrum within these bands was used to estimate heart and respiration rate over the epoch. Further research involved the estimation of sleep stages using bio-rhythm metrics derived from the heart rate measurement and also body movements (Watanabe et al., 2005). A sleep stage transition equation tracking the relationship of these metrics to the progression of sleep stages throughout the night was the basis for the sleep stage estimator developed. A mean percentage agreement for estimating sleep stages between the proposed system and an automated sleep scoring method (using training data; 5 subjects over a total of 15 nights) was 42.8%, 82.6%, 70.5% and 38.3% for all stages, NREM sleep, wake and REM sleep respectively. The results for test data (6 subjects over a total of 12 nights) in discriminating sleep stages had a mean percentage agreement of 44%, 83.6%, 44% and 47.5% respectively for all stages, NREM sleep, wake and REM sleep. Other sleep metrics such as sleep latency and total sleep time were calculated using the estimated sleep stage results with low mean difference compared to the R&K sleep scores produced by the automated sleep scoring system. Sensitivity, specificity or other similar performance metrics were not given. Although the accuracy of the system over all stages is low (42.8% on training data, 44% on test data), this system is a non-contact alternative to sleep monitoring and as such provides valuable statistics without impinging on the subject.

**Pneumatic Tubes:** Pneumatic tubes placed across the width of the bed, over the mattress, have been shown to reliably estimate physiological signals and infer various sleep metrics in both clinical (Chen et al. 2005; Mack et al. 2009a; Mack et al. 2009b; Zhu et al. 2005; Zhu et al. 2006) and real-world settings (Chen et al., 2008; Rantz et al., 2009; Zhu et al., 2008). Filtered pressure signals from two pneumatic tubes, sampled at 150Hz, were compared to ECG, oximetry and RIP band data (Mack et al. 2009a; Mack et al. 2009b). Data were collected from 20 training and 20 test subjects in bed (32 males; evenly split between the training and test group). By isolating characteristic features

of the BCG (in a similar manner to the detection of the QRS characteristics of the ECG) and using both a quasi-template matching method and a peak detection method on the differentiated pressure signal from both pneumatic tubes, the heart rate was detected to within 2.72 beats per minute of the direct measurement of ECG. A peak and trough detection method was used to detect the breathing rate to within 2.1 breaths per minute of that obtained using an RIP band. A refractory period and a comparison between the two pneumatic tubes were used to correct for outliers. This system was subsequently deployed into a nursing home for longitudinal analysis of sleep and physiological metrics (Rantz et al., 2009). Several case studies on subjects within this community have shown the reliability, efficacy and the value in this technology by correlating deviations in the produced metrics with serious life events (Rantz et al., 2009). This system was shown to outperform actigraphy in capturing movement information (Cohen's kappa  $\kappa = 0.478$  and  $0.344$  for the proposed system and for wrist actigraphy respectively) when compared to PSG (Mack et al., 2009b).

A similar system has been developed using two pneumatic tubes, sampled at 100Hz, placed under the pillow of the subject to measure physiological data from the near and far neck occiput regions (Chen et al. 2005; Chen et al. 2008; Zhu et al. 2006; Zhu et al. 2005; Zhu et al. 2008). A wavelet decomposition of the pressure signals at specific frequencies of interest has been shown to detect heart and breathing rate reliably. An adapted pulse peak pursuit method refined from the R-wave detection algorithm for ECG devised by Pan and Tompkins (1985) was employed. A threshold crossing method and refractory period were used to detect respiration. A sensitivity of 99.17% and positive predictivity of 98.53% was found for pulse detection and a sensitivity of 95.63% and positive predictivity of 95.42% was found for respiratory rate detection (Zhu et al., 2006). This sensor was deployed into a real-world setting for 6 months on one female subject (Chen et al., 2008; Zhu et al., 2008). The system was modified to transfer summarised data over the internet to a remote location which assessed the suitability of the system for long term remote monitoring.

A significant improvement in physiological data from pillow pressure signal data was achieved using Independent Component Analysis (ICA) and filtering (Uchida et al., 2003). Pulsatile pressure signal data from a pillow was recorded from 6 volunteers for 10 minutes each in addition to PPG data from the finger and respiration data using a thermistor. The ICA method decreased beat detection error from 5.83% to 1.94% and breath detection error from 4.69% to 0.29% (Uchida et al., 2003).

**Load Cells** Sleep monitoring technologies in the form of pressure sensors placed underneath the legs of a bed, known as load cells, have been used to track movement in

the bed (Adami et al. 2005; Adami et al. 2009a; Adami et al. 2009b; Brink et al. 2006; Choi et al. 2006). A similar approach to that used in the analysis of the pressure signals in the extraction of BCG was applied to this data to discern heart rate, respiration and body measurements. This section describes some of this work:

Load cell pressure sensors for movement classification was used in several studies (Adami et al. 2005; Adami et al. 2009a; Adami et al. 2009b). Movements were classified in experimental settings on 15 subjects (7 male) using difference images from a web-cam and a template matching approach which used coloured cloths placed on the limbs of the participants as the gold standard in discriminating between posture shifts (Adami et al., 2009b). Mean squared differences in pressure readings across the 4 load cells and scaling coefficients (gauging motion relative to the centre of the bed) were used as inputs to an optimally tuned likelihood ratio estimation model for discrimination between movement and no movement. The subjects were requested to shift position to a randomly selected pre-defined posture at specified time intervals. The accuracy of this system was evaluated using the Equal Error Rate (EER). The EER is the operating point where the false alarm rate is equal to the misdetection rate. This system reported an EER of 3.22%. It must also be noted that data were recorded when the subjects were awake in bed.

A cantilever type load cell composed of a top and bottom aluminium plate with a reflex light barrier in between has been shown to reliably extract heart, respiration and activity information (Brink et al., 2006). The distance between the two plates decreases as a downward force is applied to the upper plate. The width between these plates is measured using a reflex light barrier at a sampling rate of 100Hz. The BCG interbeat interval is calculated by counting the difference between subsequent minima in windowed pressure signals from each load cell. A similar technique is used to extract respiration rate. Larger amplitude movements are defined as movement. A preliminary simulation of cardiac movements was performed to characterise the sensor. For this a permanent magnet shaker and 33 KG weight was used at a range of different frequencies. This magnet shaker was tuned to a frequency which was found to be similar to a beating heart and the proposed system was reported to agree with these simulated cardiobalistic movements to within 0.2 beats per minute over different mattresses and bed types. Subsequent validation experiments across different subjects and bed, frame and mattress types as well as sleeping postures were performed. In the first validation experiment data from 4 subjects across 72 combinations of mattress, subjects, frames and type of single bed were recorded (Brink et al., 2006). The extracted BCG heartbeat was found to compare well against ECG and it reported a mean difference of  $-0.035$  Beats Per Minute (BPM) with a Standard Deviation (SD) of 0.89 BPM. A second validation experiment across 24 subjects (12 male) over 3 sleeping postures found a mean difference

of 0.33 BPM (SD 0.94 BPM); a mean difference of 0.1 BPM, 0.23 BPM and 0.65 BPM and an SD of 0.75 BPM, 1.17 BPM and 0.8 BPM was found for supine, sideways and prone sleeping postures respectively. The extraction of respiration was validated in a final experiment where 14 people (7 male) in the same 3 sleeping postures as above were compared against abdominal strain gauges. A mean difference of 0.03 breaths per minute with a SD of 0.33 breaths per minute was found. In a comparison of movement detection with wrist actigraphy (Actiwatch, Cambridge Neurotech Ltd., UK), an agreement rate of 84% was found.

A load cell utilizing a strain gauge was compared against PSG and also actigraphically defined sleep/wake (Choi et al., 2006). 10 subjects (6 male) were recruited for this study, PSG, wrist actigraphy (ActiWatch<sup>®</sup>, Mini Mitter, Co. Inc., USA.) and load cell data was collected over an 8 hour period. The load cell data was high pass filtered and movement intensity and duration values were collected and averaged over all sensors. A thresholding algorithm is used to estimate movement using the load cell data. If movement was deemed to be present for longer than 3 seconds within an epoch, the epoch was scored as wake. Epoch-by-epoch sleep/wake agreement rates of 95.2%, 92.9% and 94.3% and Cohen's Kappa  $\kappa$  of 0.61, 0.4 and 0.44 were found for PSG against load cells, PSG against wrist actigraphy and wrist actigraphy against load cells respectively.

Sleep was classified into either SWS or non-SWS (92.5% epoch-by-epoch agreement rate, Cohen's  $\kappa$  0.62) using features extracted from cardiac activity as measured by a load-cell-installed bed (Choi et al., 2009). These features were heartbeat data and heart rate variability data. Distinguishing between the stages within non-SWS (light sleep and REM) remains to be done.

**Static Charge Sensitive Bed (SCSB)** The SCSB is composed of two sheets of material, each with different dielectric constants, which generate a charge when rubbed against each other (Jansen et al., 1991). An antenna sensitive to electric charge is embedded within the SCSB to gauge the movement of these two sheets of material. The SCSB can be placed within, over or under the mattress and measures body movement arising from respiration, heart beat, posture change and bed entry/exit. The device is commonly insulated to avoid external interference. The device does not need to be directly powered as static charges are used to gauge the movement. This device is not a grid and only the amount of movement can be measured; the direction of movement cannot. However a modified version using a matrix of multiple SCSB plates has been developed which enables the measurement of distributed pressure throughout the bed (Kortelainen and Virkkala, 2007). Descriptions of various SCSB and modified SCSB sleep monitoring devices are given below:

One study used a template matching scheme to extract BCG information and a piecewise linear approximation of the respiration signal from the SCSB (Jansen et al., 1991). ECG, abdominal strain gauge and SCSB data, sampled at 200Hz, were recorded from 8 subjects. A segment of 8 minutes data was randomly chosen to calculate the accuracy of the sensor. The first ten BCG complexes were isolated and averaged to form a template. This was then cross correlated with the SCSB data over windows of increasing size with a refractory period of 0.3 seconds and a max window length of 1.5 seconds. BCG peaks were found using a peak detection method. If no acceptable apex of the BCG is found (due to both the oscillatory nature of the BCG and noise) the BCG peak is estimated to occur at a median value of the previous 25 BCG peaks. A *t*-test of the heart rate derived from the SCSB and from an ECG showed that there was a significant difference in only 5 of the 64 cases. Large body movement was associated to 4 of the 5 outlying cases due to a saturated (and thus indistinguishable) BCG, though this was not confirmed during recording, either visually or experimentally (for example, by EMG). A correlation ( $r = 0.76$ ,  $p < 0.05$ ) was found when comparing the heart rate between the SCSB and ECG over 1 minute intervals. 49 of the 64 cases showed a median heart rate difference less than or equal to 1 BPM. This was larger than 5 in 5 cases. Further research investigated the ability of this SCSB technology to predict sleep stages (Jansen and Shankar, 1993). A classification rate of between 52% and 75% was reported for the 5 R&K sleep stages and also for wake. This improved to between 78% and 89% for wake, NREM and REM, and 86% - 98% for sleep/wake classification.

**Electromagnetic Film (EMFi)** The EMFi sensor is an electret foil sensor placed on top of the mattress which can track pressure changes over time and can be considered a form of a SCSB. Charges are generated on the surfaces of the sensor as force is applied to the parallel permanently polarised layers. This allows for the amount of movement to be monitored; however similar to the SCSB the direction of movement cannot be measured. The EMFi sensor has been deployed in experimental (Akhbardeh et al., 2005; Alametsa et al., 2008), clinical (Aubert and Brauers, 2008) and real-world settings (Merilahti et al., 2007).

The sensor has been investigated as a clinical tool to assess heart conditions (Akhbardeh et al., 2005). The BCG was extracted from the EMFi-film sensor and correlated with the visual Starr classification of BCG states (which classifies BCG as either normal, slightly abnormal, markedly abnormal or extremely abnormal) (Akhbardeh et al., 2005). A wavelet decomposition of this EMFi pressure signal and a novel transform method, known as the AliMap transform, were used to distinguish healthy and unhealthy classes from the BCG recordings of 18 subjects. However, the method was not able to recognise heart conditions from two of the subjects.

The FFT was applied to extract the RR interval from the pressure signal in a multi-channel EMFi film sensor (Kortelainen and Virkkala, 2007). The Matsense mattress consisted of 160 EMFi sensors distributed in a grid and was sampled at 50Hz. 15 sets of recordings were taken from 6 male subjects. A time based pulse detection method using edge detection was found to have an average error of 2% in comparison to the RR interval from the ECG. Using an averaging method with a period of more than 10 seconds improved the error in the average heart rate to less than 1%. By applying an FFT based deconvolution method, known as cepstrum (Rosenblatt, 1963), using a time window which contained exactly two consecutive heart beats (determined using the time based pulse detection method) the RR interval error was 0.4%. This method excluded periods when movement occurred as the reliability of the RR interval was reduced dramatically. This corresponded to 12% of the total sleep time for this data set. This system was further improved through the validation of a EMFi-based system which distinguishes between wake/REM/acNREM using a HMM classifier (Kortelainen et al., 2010). The movement features were used to detect wake periods while the cardiac data was used as inputs to the classifier. An accuracy of 79% ( $\pm 10\%$ ; where  $\pm$  relates to 1 standard deviation) and a Cohen's  $\kappa$  of 0.44 ( $\pm 0.19$ ) were reported.

The EMFi has been deployed in a hospital setting and data recorded over 102 nights on 58 participants (Aubert and Brauers, 2008). This included 11 healthy subjects (7 male; over 53 nights), 19 sleep apnea (17 male; over 21 nights), 6 insomnia (3 male; 6 nights) and 22 participants with sleep disorders (13 male; 22 nights). Data from an ECG (sampled at 250 Hz), a thorax strain gauge (250 Hz) and the EMFi (128 Hz) was recorded. A peak and trough detection method was used to extract respiration data from the low pass filtered pressure signal. The estimate of the RR interval was found using the autocorrelation function. The error in the breathing rate was only reported for the healthy and insomnia participant groups (an error of 0.47 and 1.17 breaths per minute respectively). The average error for heart rate was 1.25 BPM. This consisted of 1.17 BPM, 1.57 BPM, 0.67 BPM and 1.27 BPM for healthy, sleep apnoea, insomnia and sleep disorder groups respectively.

A comparison of BCG recorded using an EMFi sensor was compared to both ECG and carotid pulse for sitting and horizontal (including prone, supine and side-lying) postures (Alametsa et al., 2008). The carotid pulse is the pressure signal resulting from the heart beat apparent at the carotid artery at the surface of the neck. Data was recorded from 7 subjects for each of the positions. Various medical metrics involving amplitude and duration scores derived from the ECG, BCG and carotid pulse data were recorded and compared to the EMFi. The successful use of the EMFi in extracting the BCG validates its clinical use. However, an automated method for extracting these metrics is required to provide comparative statistics.

**Fibre Optic Solutions:** Movement impinged on a multi-modal fibre optic perturbs light passing through the fibre optic by different amounts (Spillman et al., 2004). This principle was used to detect respiration, heart rate and body movements when the fibre optic was woven into the bed (Spillman et al., 2004). Data was recorded from one subject over 4 postures (supine, prone, left fetal and right fetal) and a fourier transform of the pressure signal showed the existence of signals in the respiration and heartbeat range. However no comparison was made to a gold standard measure of heart rate or respiration rate.

Another fibre optic technology used as a sleep monitor is the Fibre Bragg Grating (FBG) sensor which is again sensitive to strain caused by varying amounts of pressure applied (Foo Siang Fook et al., 2008). The surface of the fibre is grated in certain places and this affects the transmission of light at that point in the fibre optic as movement occurs. For the system, 12 FBG sensors were distributed evenly throughout the bed. A trace of the respiration was reported using the proposed system in order to demonstrate its sensitivity. A Fourier transform of the band-passed filtered signals displayed a peak relating to the observed respiration. No comparison against gold standards were made.

#### 4.5.4.2 Radio Frequency-Based Solutions:

A non-contact biomotion sensor in which Doppler radar is used for the extraction of movement and the estimation of breathing and heart rates has been developed by *BiancaMed (NovaUCD, Dublin 4, Ireland)* (de Chazal et al., 2008; Fox et al., 2007; Zaffaroni et al., 2009), Lubecke and Boric-Lubecke (2009), and Choi and Kim (2009).

The *BiancaMed* system was developed using Doppler radar to measure movements by a human body. In a study of 20 individuals (9 healthy, 6 severe sleep apnoea, 2 moderate sleep apnoea, 1 chronic obstructive pulmonary disease, 1 childhood obesity, 1 insomnia; 8 of which were males), high sensitivity and specificity values (79% and 75% respectively) of sleep/wake detection were reported when compared to wrist actigraphy (Fox et al., 2007). It was noted that the lower specificity could be due to the proposed sensor capturing entire body motion and the gold standard employed (wrist actigraphy) only reporting wrist movements (Fox et al., 2007). A signal considered to contain a strong respiratory component was presented, however no gold standard statistical comparison was made. Breathing frequencies were attenuated using a 7th order Butterworth filter. In a later study, large movements, bodily movements and, in periods of quiescence, the amplitude and frequency of respiration were used as features for estimating the sleep/wake state of 14 participants (11 male) (de Chazal et al., 2008). The overall accuracy of the system in classifying sleep/wake state was 82%. When broken down

into classifying each sleep stage this ranged from 61% to 98%. A visual comparison of the respiratory effort signals captured using a RIP sensor and the proposed biomotion sensor was also presented. More recent work describes the performance of this sensor in discriminating sleep and wake (using an algorithm built upon an LDA classifier) in 113 subjects being assessed for sleep disordered breathing (De Chazal et al., 2011). An overall per subject accuracy of 78% (Cohen's kappa of 0.38) was reported, however higher AHI ( $\geq 15$  apnoeic events per hour) was found to reduce performance (74.8% vs. 81.3% in subjects with higher vs. lower AHI). Overall, a sensitivity of 87.3% and a specificity of 50.1% was reported. The system also overestimated SE and TST.

The biomotion sensor was also proposed as an innovative device to estimate the apnoea-hypopnoea index (AHI) (Zaffaroni et al., 2009); as discussed later in Section 4.5.6.

Other researchers have also proposed a similar Doppler radar based non-contact sleep monitor (Lubecke and Boric-Lubecke, 2009). For this system, a visual comparison of extracted pulse rate against finger pulse rate, respiration and body movement were provided; however no gold standard statistical comparison was made.

Another RF sensor using an electromagnetic wave for real-time monitoring of heartbeat and respiration rate was realised using peak detection of the power spectral density of the signal (Choi and Kim, 2009). The results were compared against a fingertip PPG sensor. 100% accuracy was reported for this technology however no subject details were given. An evaluation using several subjects over a long period would provide more robust results.

#### 4.5.5 Multi-Modal Solutions:

A wellness monitoring system was developed by researchers at *VTT Technical Research Centre of Finland, Finland* for older adults in out-hospital conditions for the monitoring of physiological and psychosocial variables (Korhonen et al., 2001). RR interval, activity level, blood pressure, weight, temperature, respiration, ballistocardiography (using an SCSB), movements, sleep stages (using the SCSB) and a behavioural diary were recorded. This system was deployed on 2 participants (1 male) for 10 weeks and 14 days respectively. This system was concluded to be suitable for continuous monitoring of several variables over weeks and months.

In a study by Merilahti et al. (2007), subjective and objective sleep data were recorded using sleep logs, the Vivago WristCare<sup>®</sup> (as discussed in Section 4.4.4.4) and an EMFi bed occupancy sensor (Emfit Ltd., Finland, <http://www.emfit.com/>) (as discussed in Section 4.5.4.1) for 17 participants (3 male) over an average of 58 nights each (Merilahti

et al., 2007). A visual sleep analysis of all data was taken as the gold standard for this data set. A high correlation was found between the objective and subjective sleep metrics; however it was noted that long term data collection for sleep monitoring should include bed time and illumination information as previously suggested in the literature (Sadeh et al., 1995).

Movement captured using a webcam (sampled once every 6 seconds) and from a PIR were combined with spectral Heart Rate Variability (HRV) data to form a sleep monitoring system by Peng et al. (2007). This was compared to wrist activity and subjective sleep quality data from a PSQI questionnaire. Data was collected from 1 healthy female participant over 13 days (split into 7 days training, 3 days testing, 3 days validation). An SVM classifier and both pre- and post-processing fusion methods were employed to estimate sleep state. It was not clear how the determination of true sleep state was achieved; although it was assumed sleep state was scored using wrist actigraphy. High detection accuracy was found for the PIR and video sensor, 0.92 and 0.89 respectively. Small differences were found in sleep efficiency between the sleep logs, actigraphy and the proposed system. The machine learning techniques applied in this scenario (SVMs) were shown to achieve better results than a regression model.

#### 4.5.6 Sleep Apnoea Monitoring

**Respiration monitoring through ECG analysis:** Data derived from the ECG has been shown to provide very high accuracy in detecting OSAS (de Chazal et al. 2009; Dorfman Furman et al. 2005; Khandoker et al. 2009b; Khandoker et al. 2009c; Khandoker et al. 2009d; Khandoker et al. 2009a; Mendez et al. 2009; Mendez et al. 2010).

In a study by Dorfman Furman et al. (2005), three methods for ECG derived respiration (EDR) during sleep (from 24 subjects; 15 male) were compared to nasal air flow and abdomen and thorax based modalities for respiration estimation. The three methods estimated respiration using R wave amplitude, R wave duration and area under each QRS complex. The frequency of these signals were shown to be related to true respiration rate; however the amplitude of the respiration was not comparable. The correlation between the extracted respiration signal was above 0.85 for all metrics against each reference measure of respiration. The peak-to-peak values of the estimated respiration signals decreased during both central and obstructive apnoeas, although no automated detection of these apnoeas was attempted.

Another study investigated applying autoregressive models to RR interval and QRS area time series data in order to screen for sleep apnoea (Mendez et al. 2009; Mendez et al. 2010). Features derived included QRS area phase space (modulus and phase

from these signals), power spectral density at 0.003-0.04Hz, 0.04-0.15Hz and 0.15-0.5Hz, coherence, kurtosis and skewness for both time series. ECG data from 50 sets of PSG sleep data were taken from the Physionet database (Goldberger et al., 2000) and scored for apnoeas and hypopnoeas in this study. The best feature subset was found using Sequential Forward Selection (SFS) and by selecting a optimised kNN-model. Leave-one-out cross validation was employed to generate accuracy, sensitivity and specificity values. An ANN model was compared against the kNN model. 25 data sets (50%) were used to train the kNN and ANN classifier. This contained 15 apnoeic, 8 normal and 2 borderline subjects. Results were reported using the validation set which contained 13 apnoeic, 4 borderline and 8 control subjects. Using 10 features the kNN reached an accuracy of 88%, a sensitivity of 85% and a specificity of 90% in detecting individual apnoeas. The ANN reached an accuracy of 88%, a sensitivity of 89% and a specificity of 86%. The classifiers are able to separate completely the normal subjects from the apnoeic subjects. Testing was performed on a separate database of 8 recordings and completely separated the normal subjects from OSAS individuals.

Other advanced nonlinear classification methods, using support vector machines, for the detection of OSAS from ECG data was shown to have a high accuracy using RR interval and EDR (using R wave amplitude) features (Khandoker et al. 2009b; Khandoker et al. 2009c; Khandoker et al. 2009d; Khandoker et al. 2009a). In comparison to the use of autoregressive models elsewhere (Mendez et al. 2010; Mendez et al. 2009), a 14-level wavelet decomposition of RR interval and EDR data were used as features in this system (Khandoker et al., 2009a). SFS was used to select the best candidate features and reported a 100% accuracy, 100% sensitivity, 100% specificity, 100% area under the ROC curve and 100% Cohen's  $\kappa$  on training data (83 subjects) for an SVM classifier in discriminating OSAS positive and negative subjects. The best classifier used 1 RR interval feature and 3 EDR features and a polynomial kernel of degree 3 with a regularisation parameter of 0.8. An accuracy of 92.85% and a Cohen's  $\kappa$  of 0.85 was reported on test data from 42 independent subjects. The posterior probabilities estimates from the SVM was suggested as an indicator of apnoea-hypopnoeas index (AHI).

**Pressure Based Respiration Monitoring:** A pneumatic sleep monitoring system capable of reliably estimating respiration (Mack et al. 2009a; Mack et al. 2009b) has been previously discussed in Section 4.5.4.1. Further work investigated 40 Healthy subjects (32 male) in order to assess the efficacy of this system in detecting sleep apnoeas and arousals (Mack et al., 2006). Apnoeas were detected when the amplitude of the breathing signal was reduced below 75% or when there were gaps in the breathing signal with minimal postural movement. No discrimination was made between apnoeas

and hypopnoeas. A sensitivity of 89.2% and a specificity of 94.6% was reported for the automated detection of sleep apnoeas (Mack et al., 2006).

Load cells were used to discriminate Sleep Behaviour Disorder (SBD) from normal breathing with a sensitivity of 0.77 and a specificity of 0.91 (Beattie et al., 2009). 150 instances of apnoeas, 150 instances of hypopneas and 150 instances of normal breathing (lasting 20 seconds) from 4 subjects was used in this analysis. 8 features were extracted from each instance including: the variance over all samples, normalised average power of the signal in the 0-0.5Hz, 0.6-0.75Hz and 0.75-5Hz frequency bands, spectral entropy, and 3 features from the signal bandpass filtered in the respiration range (0.2-0.33Hz) namely, variance, range and respiratory amplitude (that is, the range of the median amplitude of overlapping windows of data). This system was developed using a Bayesian classifier and validated using 10-fold cross validation. The sensitivity of the system for hypopnoea/obstructive apnoea was 0.65 and 0.90, for central apnoea 0.82 and 0.92 and for normal breathing 0.84 and 0.84. The Sleep Disordered Breathing (SDB) breathing was also discriminated with a sensitivity of 0.77 and specificity of 0.91. A sensitivity of 0.91 and specificity of 0.77 was reported for discerning normal breathing segments.

**Ambulatory SpO<sub>2</sub> Monitoring:** Real-time detection of apnoeas using a pulse oximetry (saturation of peripheral oxygen; SpO<sub>2</sub>) signal was implemented on a Personal Digital Assistant (PDA) and an accuracy of 93% and an ROC area-under-the-curve of 98.5% was reported (Burgos et al., 2009). 70 sleep recordings, annotated for SDB and including a blood oxygen saturation signal from which SpO<sub>2</sub> data was derived, of duration approximately 8 hours each were used in this analysis. The data was split into 66% training data and 33% test data. Features extracted from the SpO<sub>2</sub> signal included : the number of 2%, 3% and 4% dips from a moving baseline and the length of time the saturation decreases below 95%, 90%, 85%, 80% and 70% per segment of SpO<sub>2</sub> data. 10 tests were performed using 10 classification methods (within the Weka machine learning tool (Witten et al., 1999)) on 10 sets of different training data. The area under the ROC curve, the accuracy and the processing time for each method was used to select the optimal classifier. This was chosen by examining the contribution of each feature, the number of decision trees and reducing the cost matrix (for example, minimising the number of false positives). An area under the curve of 99.1%, sensitivity of 96% and specificity of 96% was reported for the training data. An area under the curve of 98.5%, sensitivity of 92.3%, specificity of 93.5% and accuracy of 93% was reported for the test data. A real-time implementation of the detection of apnoeas was developed using a PDA as well as the development of a system which can transfer pulse rate and SpO<sub>2</sub> data over a telecommunications network to a medical professional (Burgos et al., 2009).

**Inspiratory Flow Monitoring:** Morgenstern et al. (2008) described an automatic non-invasive method of breath classification using a nasal canula. Breath contour data from 11 subjects, trained using an SVM and adaboost classifier, against gold standard measures, provided high sensitivity (0.92) and specificity (0.89) results in assessing inspiratory flow limitation during sleep.

**Multi-Modal Solutions** RR interval, EDR (QRS area) and pulse oximetry (saturation of peripheral oxygen; SpO<sub>2</sub>) data were used to jointly discriminate normal breathing from SDB (de Chazal et al., 2009). The SDB consisted of 6 classes: obstructive, central and mixed conditions of both apnoeas and hypopnoeas. A total of 72 ECG features were used including the Power Spectral Density (PSD) of RR interval data (32 features), the PSD of EDR data (32 features), 5 serial correlation coefficients relating to a delay of 1-5 RR intervals, standard deviation of the RR interval data, mean of the RR interval data. 7 additional SpO<sub>2</sub> features were also used: mean SpO<sub>2</sub>, minimum SpO<sub>2</sub>, the number of samples less than 92%, 5-95% spread of SpO<sub>2</sub>, mean of the absolute differences of SpO<sub>2</sub>, the number of samples of  $(SpO_2 - baseline(SpO_2)) < 3\%$  and number of samples of  $(SpO_2 - baseline(SpO_2)) < -3\%$ . Features were generated for 1-minute epochs overlapping by 30 seconds each. A leave-one-out-cross-validation scheme was applied to generate statistical data from the 183 subjects. Linear discriminant classifiers with equal prior probabilities were applied to the ECG and SpO<sub>2</sub> data separately in order to detect the SDB. The results from these two classifiers were combined using a weighted Bayesian addition integration scheme with the oximetry data weighted at 80% and the ECG data weighted at 20%. If data from one device was found to be corrupt, the posterior probabilities were copied from the valid data. If both sets of data were corrupt, the epoch remained unclassified. The class chosen had the highest posterior probability estimate. A hierarchical summing method was used to generate the total number of SDB, apnoea and hypopnoea epochs from the original 6 classes: obstructive apnoeas, central apnoeas, mixed apnoeas, obstructive hypopnoeas, central hypopnoeas and mixed hypopnoeas. The AHI was estimated as the average percentage of SDB epochs per hour. For detecting SDB, a sensitivity of 84%, a specificity of 87%, an accuracy of 86% and a Cohen's  $\kappa$  of 0.66 was reported. Results of a comparison of the ability of the system to discriminate between severe SDB ( $AHI > 15$ ) and normal breathing ( $AHI < 5$ ) using a) oximetry only were sensitivity 94% and specificity 93% b) ECG only were sensitivity 92% and specificity 65% and c) oximetry and ECG were sensitivity of 94% and specificity 94%.

#### 4.5.7 Brain Imaging and Sleep Monitoring

Advanced brain imaging technologies have been applied to sleep monitoring. In a review, Deboer (2007) discussed that functional Magnetic Resonance Imaging (fMRI) and positron emission tomography (PET) offer the possibility of monitoring the neuronal activity in the brain. This has been applied to sleep research using variations in glucose and oxygen consumption and hemodynamic changes to offer insights into brain structures and brain function. However, the practicality of using fMRI is questionable due to the high noise levels within the scanner in spite of continued improvements in the technology (Kaufmann et al., 2006). Although PET does not offer a real-time measurement, it is stable and quieter than fMRI (Braun et al., 1997). This form of sleep monitoring is outside the scope of this thesis, however the interested reader is referred to Braun et al. (1997) and Deboer (2007).

### 4.6 Conclusions

This chapter has provided an overview of the traditional established standards for sleep monitoring, namely PSG in a clinical setting, wrist actigraphy in a non-clinical setting, and subjective scales or sleep diaries (possibly recorded by clinicians) where neither PSG or actigraphy are appropriate or acceptable. This serves as a precursor to the main contribution of the chapter; a review of various recent advances in contact and non-contact sleep monitoring approaches. Contact-based advances in sleep monitoring include electrodes implanted in bed sheets, garments, a wrist-based device or the use of a minimal number of PSG electrodes. Non-contact advances included an optical approach (video or PIR), the use of smartphones (via an unvalidated algorithm using the device's accelerometer and microphone or the manual entry of sleep/wake times), an approach where the sensor was placed under the mattress (for use in sleep/wake monitoring, physiological signal detection, movement detection, or the monitoring of development of bed sores) and a RF system (in which the device is placed on the bed side locker). Technological advancements in sleep measurement modalities is fast-paced and ever-expanding. A direct clinical comparison is difficult across these technologies due to multiple methodological approaches including different performance metrics, different population types, and different recording environments.

The chapter briefly discussed the need for an unencumbering sleep monitoring system suitable for the long-term monitoring of sleep in sensitive populations (particularly older adults and unhealthy individuals). As such, each of the types of technologies described in the chapter were evaluated using three criteria: 1) accuracy, 2) comfort, and 3)

obtrusiveness, as summarised in Table 4.5. The accuracy of the alternative approaches (in sleep/wake detection) varies considerably. Generally, the accuracy decreases as the proposed solution has less physical contact with the individual. The comfort of the device is dependent on how much it hinders directly the sleeping patterns of the user, that is, does the use of the device negatively affect quality of sleep due to its physical design? Examples of this include electrodes placed on the skull (eg. NightCap (Mamelak and Hobson, 1989)) or having to wear a specially designed garment (eg. LifeShirt (Coyle, 2002)). The obtrusiveness of the system is another important factor. Technologies which require continuous vigilance (in terms of wearing, maintenance, or requiring special care) are a hindrance to the user and inhibit long-term adherence by the user. In certain cases, non-physical factors can affect adherence including privacy concerns from cameras (Demiris et al., 2009), or the clinical appearance of the device (as this places a feeling upon the individual that they are unhealthy).

In general the accuracy of systems that are suitable for long term use is traded off against comfort and obtrusiveness. This thesis proposes a technology which is particularly suitable for long-term deployment in sensitive populations (particularly older adults living in their own homes), is easy to install, does not burden the user by being both comfortable and unobtrusive. An introduction to the entire system and its underlying technologies is given in the next chapter. In the subsequent chapters, experimental evaluations, deployments in various settings and sleep stage discrimination using the system is performed.

TABLE 4.5: Suitability of sleep monitoring modalities for long-term use.

	Device	Accuracy	Comfort	Unobtrusiveness	Signals
Established Standards	Full PSG	XXX			EEG, EOG, EMG, ECG
	Partial PSG	XX			Some of above
	Amb. PSG	XX	X		EEG, EOG, EMG
	Actigraphy	X	XX	X	BM
	Subj. Scales		XXX	X	
	Sleep Diary		XXX	X	
Contact-Based Monitoring	AOG Mon.	X	X		EOG
	NightCap	X	X		quasi-EOG, BM
	MyZeo	XX	X		quasi-EOG, BM
	Textile		XX		quasi-EEG, quasi-EOG
	LifeShirt		X		ECG, BM, T
	WatchPat	X	XX		BM, Oxi, PAT
Optical	Video Mon.		XXX	X <sup>1</sup>	Video
	PIR	X	XXX	XX	BM
	Smartphones		XXX	XX	BM (via Acc.), Sound
Under Mattress	Coil Mattress		XXX	XXX	BM, BR
	FSR		XXX	XXX	BM, T, BR, HR, RI
	Air Mattress	X	XXX	XX <sup>2</sup>	BM, BR, HR
	Pneumatic Tube	X	XXX	XX <sup>2</sup>	HR, BR, BM
	Load Cells	X	XXX	XXX	HR, BR, BM
	SCSB	X	XXX	XXX	HR, BR, BM
	EMFi	X	XXX	XXX	HR, BR, BM
	Fibre Optic		XXX	XXX	HR, BR, BM
	Bed Sores Mon.		XXX	XX	BM
	RF System	X	XXX	XX <sup>3</sup>	BR, BM

AOG = Actiocolographic Monitor; BM = Body Movement; HR = Heart Rate; BR = Breathing Rate; RI = Restlessness index; T = Temperature; Oxi = Oximetry; PAT = Peripheral Arterial Tonometry

<sup>1</sup> Use of video in the bedroom can cause privacy concerns in older adults (Demiris et al., 2009).

<sup>2</sup> Some systems are in direct contact with the participant and as such long-term use reduces comfort.

<sup>3</sup> RF systems must be designed not to appear clinical as this reduces suitability for long term use.

Rating system developed by author using data gathered from each relevant study. Increasing numbers of X indicate better results.

## Chapter 5

# The Non-Obtrusive Non-Contact Sleep Sensor

This chapter introduces a novel, non-contact sleep monitoring solution in the form of an under-mattress bed sensor. The sensor was initially designed to monitor bed exits of patients in a nursing home and also to investigate the development of bed sores in bed-ridden cohorts. In this chapter it is shown that the sensor also has the capacity to provide informative data relating to changes in sleeping patterns over extended periods of time. Such data has potential diagnostic value particularly in relation to health issues in older adults. Several studies have demonstrated that older adults are found to have an impaired sleep, and sleep disturbances, especially in older adults, have been found to be indicative of poor health and functional deficits (Miles and Dement, 1980; Manabe et al., 2000).

This chapter provides an insight into the design of the sensor and data collection system. It provides an investigation into the practicality and suitability of the sensor for the long-term collection of sleep data, particularly in older adults. Sample data, collected in test scenarios, recorded by the sensor is shown and this illustrates to the reader the form of the sleep monitoring data. Additionally, results from the deployment of the sensor in experimental conditions and in a community-based study of older adults are presented.

### 5.1 Under-Mattress Bed Sensor

The under-mattress bed sensor (UMBS) is proposed as a non-contact, actigraphy based sleep monitoring solution particularly applicable to long term domestic placement. The

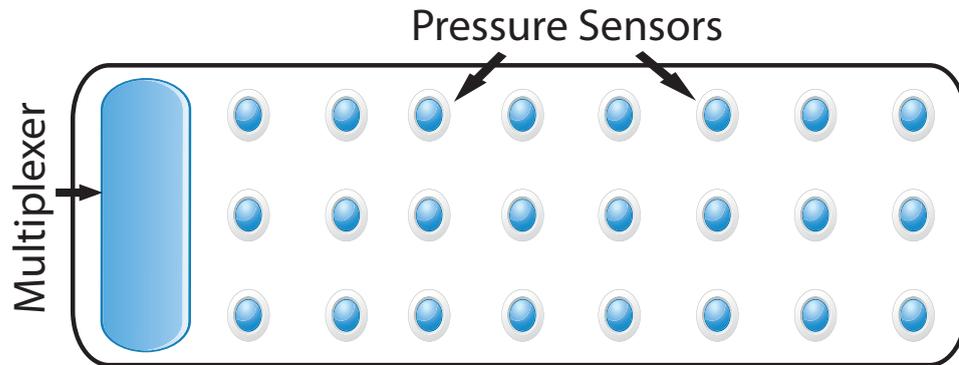


FIGURE 5.1: Layout of the UMBS. Each pressure sensing location is marked by the blue circles. The signal from each tactel is sampled by a multiplexer.

sensor is composed of 24 tactels, or pressure sensing points, evenly distributed throughout the device in a 3 x 8 pattern. The sensor is 90cm x 24cm x 1cm and each tactel is located 10cm away from neighbouring tactels. An outline of the tactel placement within the UMBS is shown in figure 5.1. The typical placement of the UMBS is underneath the upper torso of the subject, underneath the mattress, as given in Figure 5.2.

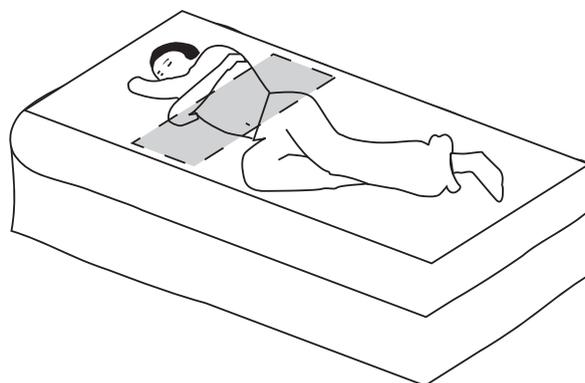
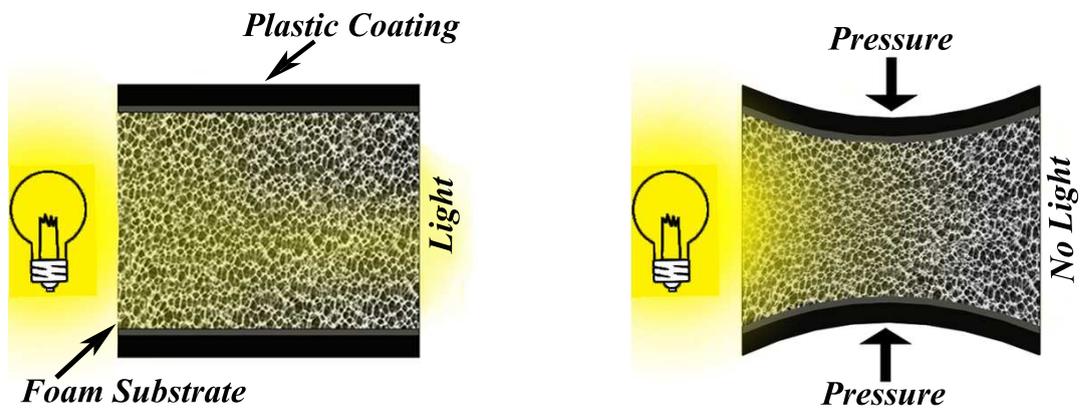


FIGURE 5.2: Typical placement of the UMBS



(a) No pressure applied to the sensor and light illuminates through the substrate from the emitter (on the left) throughout the foam.

(b) The application of pressure reduces the amount of light permeating through the substrate (due to the compression of the substrate).

FIGURE 5.3: UMBS open cell urethane foam substrate mechanism of transducing downward pressure into a signal. The foam is covered with a heavy duty flat plastic material (black). The amount of light permeating throughout the substrate is illustrated by the yellow colouring.

### 5.1.1 Tactels

This sensor uses a light-based pressure sensitive system which is not sensitive to electronic interference. Relative pressure is recorded by measuring the amount of light passing between an emitter and receiver woven into a semi-permeable substrate. Changes in pressure applied to the substrate results in varying amounts of light passing between a fibre optic emitter and receiver (as shown in Figure 5.3).

### 5.1.2 UMBS Sensor Types

The instantaneous pressure placed on each tactel is measured at a rate defined internally, using a piezo-electric crystal, within the UMBS. This rate is either 10Hz or 20 Hz and is set during the manufacture of the sensor. The range of values over which the UMBS measures relative pressure varies between 0 and 255 or 0 and 2047 depending on the model. In the course of this research, data was collected using three types of sensors as defined in Table 5.1.

### 5.1.3 Communications Protocol

The UMBS is designed to communicate with a data aggregator which polls data from the UMBS through a standard RS232 serial port. The communications protocol defines

TABLE 5.1: UMBS Sensor Types

Sensor Name	Sampling Rate	Range of Values
UMBS (v1)	10Hz	0 to 255
UMBS (v2)	10Hz	0 to 2047
UMBS (v3)	20Hz	0 to 2047

requests, as sent by the data aggregator, and replies, as issued by the UMBS. The UMBS will always be the slave device in the system. For this research, two requests were used. The first request ensures that the UMBS is ready for data transfer (by sending a 'ping' with a positive response, 'ack', indicating the device is on). The second request asks for the current pressure values recorded by all tactels. The UMBS will not deal with subsequent requests until it has completed its current request. This effectively resulted in a half-duplex communication protocol.

#### 5.1.4 Data Collection System

A flow chart of the UMBS data collection algorithm is given in Figure 5.4. Initially, a 'ping' message is sent to the UMBS to ensure that it is currently active and that the serial port the device is connected through is working correctly. Subsequently, the current tactel data is requested from the UMBS. The UMBS data packet is sent to the port and buffered. This buffer is read repeatedly with a 10ms delay between readings until a full data packet has been read. After a full packet has been read, the system either stops data collection, if requested by the user, or requests another UMBS data packet. This ensures that there is as minimal a delay as possible between the collection of UMBS data packets. In practice, this system continually collects data for a number of weeks until the study has finished. This system was implemented on two platforms: a PDA (Dell<sup>®</sup> Axim x51v) and a personal computer (model dependent on each study). The data collection system for the PDA was coded using the Visual Basic Programming language. The personal computer data collection system was coded in the C++/CLR programming language. The PDA data collection was implemented using UMBS (v1), while the computer-based data collection system was implemented using UMBS (v2) and UMBS (v3)

#### 5.1.5 Sampling Rate

The sampling rates for both data collection systems can be seen in Figure 5.5. While the values of each tactel was internally refreshed at either 10Hz or 20Hz, the rate at

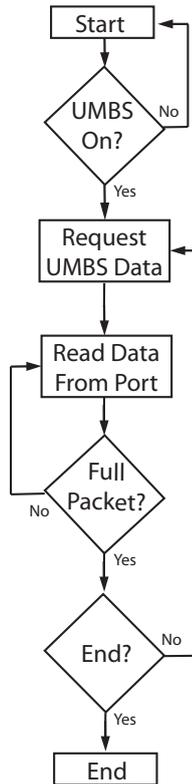
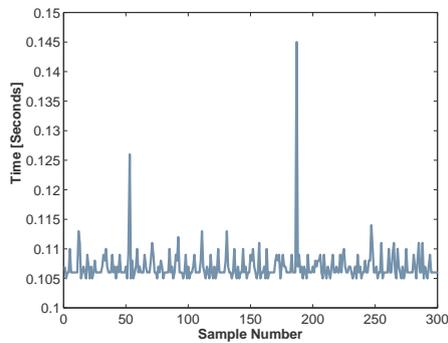
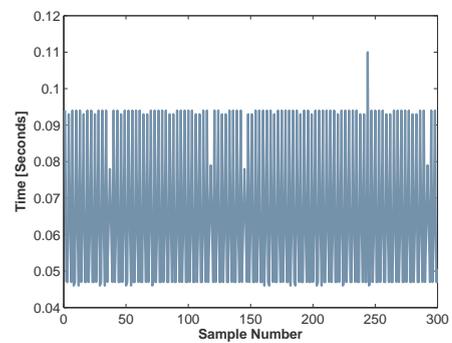


FIGURE 5.4: UMBS data collection algorithm



(a) PDA using 10Hz sensor



(b) Personal Computer using 20Hz sensor

FIGURE 5.5: Sampling rates attained by the data collection equipment.

which this data was accessed and saved was dependent on the interface between the data aggregator and the UMBS. The PDA-based system achieved an average sampling period of  $107.1 \pm 3.5$  ms, with a range from 105 ms to 225 ms. The personal computer-based system achieved an average sampling period of  $65.4 \pm 5.1$  ms, with a range from 46.1 to 109.9 ms.

After data collection, the time-series was interpolated to a constant sampling rate (generally 10Hz) using linear interpolation. This was deemed sufficient as no signal of interest

occurred above 5 Hz. Breathing and voluntary movements (ie. excluding spasms and seizures) occur below this rate (Redmond and Hegge, 1987) .

### 5.1.6 Pressure Measurement

The ability of the tactels to accurately record absolute pressure was investigated by Holtzman et al. (2008b). They found that the tactels did not record absolute pressure (measured in Newtons) and that a non-linear relationship existed between tactel values and pressure. Additionally they investigated the application of pressure at various distances away from the tactel, along the same axis as the UMBS, and also found a non-linear response. The tactel values remained stable over multiple tests, however the elasticity of the internal substrate over extended periods of time was unknown and may degrade when continuous pressure is applied (for example when the UMBS is placed under a mattress). Due to the combination of all of these factors, an algorithm to convert tactel values to absolute pressure values was not developed and this thesis focussed on developing algorithms using the tactel values as directly measured.

### 5.1.7 UMBS Notation

The UMBS contains 24 tactels. The notation used to refer to the UMBS data is as follows. Denoting  $x_{ij}$  as the value of the  $i^{th}$  sample instant of the  $j^{th}$  tactel, and assuming a total of  $N$  samples, the tactel dataset  $\mathbf{X} \in \mathfrak{R}^{N \times 24}$  can be defined:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,j} & \dots & x_{1,24} \\ x_{2,1} & x_{2,2} & \dots & \dots & \dots & x_{2,24} \\ \dots & & & & & \\ x_{i,1} & x_{i,2} & \dots & \dots & \dots & x_{i,24} \\ \dots & & & & & \\ x_{N,1} & x_{N,2} & \dots & \dots & \dots & x_{N,24} \end{bmatrix}, \quad 1 \leq i \leq N, 1 \leq j \leq 24 \quad (5.1)$$

Each row ( $\mathbf{x}_i$ ) of  $\mathbf{X}$  relates to the pressure values recorded over all tactels at one time instant, while each column ( $\mathbf{x}^j$ ) of  $\mathbf{X}$  represents the pressure on one tactel over all times instants.

Sleep data is often examined over non-overlapping time windows, or epochs. In PSG, these windows last 30 seconds. Here, we use the following notation to refer to windows of UMBS data. Given  $m$  sample windows the  $k^{th}$  data window,  $\mathbf{W}^k \in \mathfrak{R}^{m \times 24}$  is defined as:

$$\mathbf{W}^k = \begin{bmatrix} x_{m(k-1)+1,1} & x_{m(k-1)+1,2} & \cdots & x_{m(k-1)+1,j} & \cdots & x_{m(k-1)+1,24} \\ x_{m(k-1)+2,1} & x_{m(k-1)+2,2} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{m(k-1)+i,1} & x_{m(k-1)+i,2} & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{m(k-1)+N,1} & x_{m(k-1)+N,2} & \cdots & \cdots & \cdots & x_{m(k-1)+N,24} \end{bmatrix}, \quad (5.2)$$

$$1 \leq j \leq 24, 1 \leq k \leq (\lfloor \frac{N}{m} \rfloor)$$

Each window consists of  $m$  samples instants and  $N$  is the total number of samples in the entire data set.  $x_{m(k-1)+1,j}$  is the  $m(k-1) + 1^{th}$  sample instant of the  $j^{th}$  tactel. If data from all of these windows are concatenated together, the original data set  $\mathbf{X}$  will be formed (if  $N - m\lfloor \frac{N}{m} \rfloor = 0$ , ie.  $N$  is an exact multiple of  $m$ ).

## 5.2 Initial UMBS Deployment

### 5.2.1 Example Data

Example data collected using the UMBS (v3) is shown in Figure 5.6. When no pressure is placed on the UMBS (unsaturated), the tactel values vary slightly around their minimum values. This value is not constant across all tactels ranging between 200 and 600 units (when the sensor is placed under a mattress), and varies by less than 5 units around this minimum value (as shown in Figure 5.6(a)). When the amount of pressure placed on each tactel exceeds a maximal value (saturated), the tactels saturate (at 2047 for this sensor). When a participant lies passively on the UMBS, clear cyclical patterns of breathing can be seen on some of the tactels, some tactels register no pressure (as the subject is not lying over those tactels) and others are saturated. The last subfigure in Figure 5.6 shows the patterns observed when the subject moves while on the mat. A two dimensional representation of the pressure distribution on the UMBS at a time instant is given in Figure 5.7.

### 5.2.2 Entire Night's Data

Data collected over an entire sleeping period for a normal healthy adult can be seen in Figure 5.8. This data relates to a period when the subject was asleep (as defined by a wrist actigraph). Large shifts in the tactel values define periods when the subject

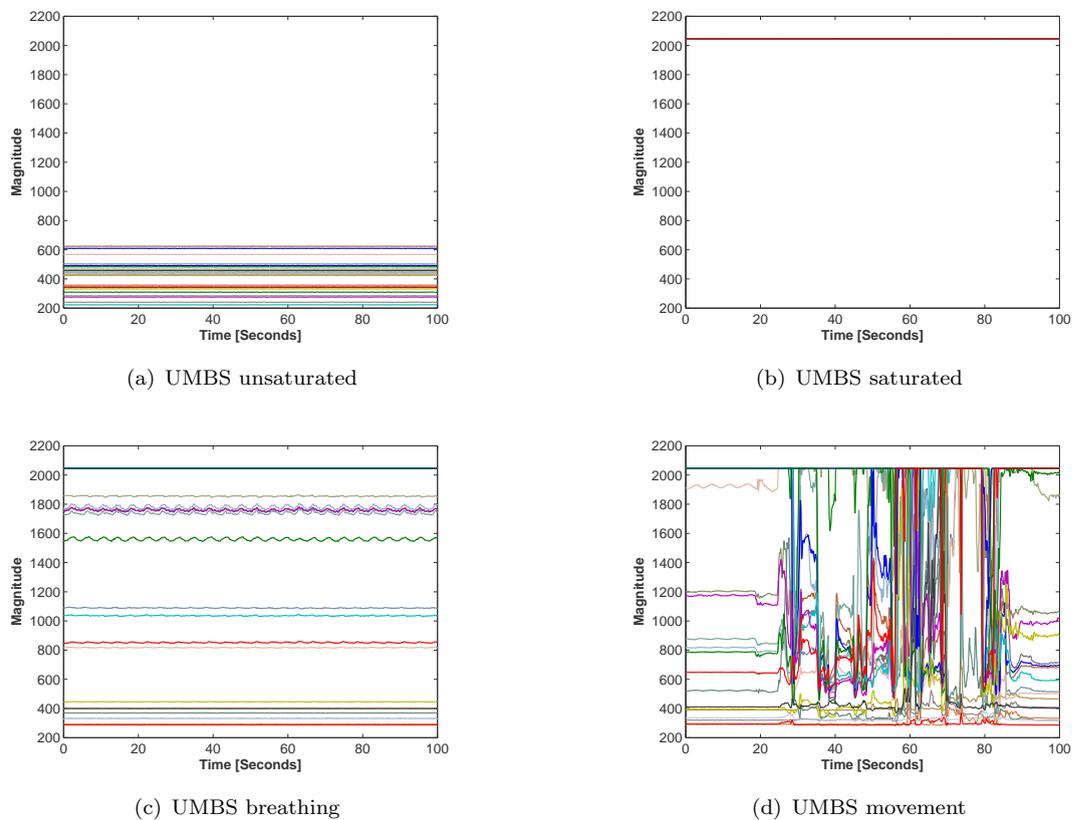


FIGURE 5.6: Sample UMBS data. The signals from each tactel are plotted as separate lines.

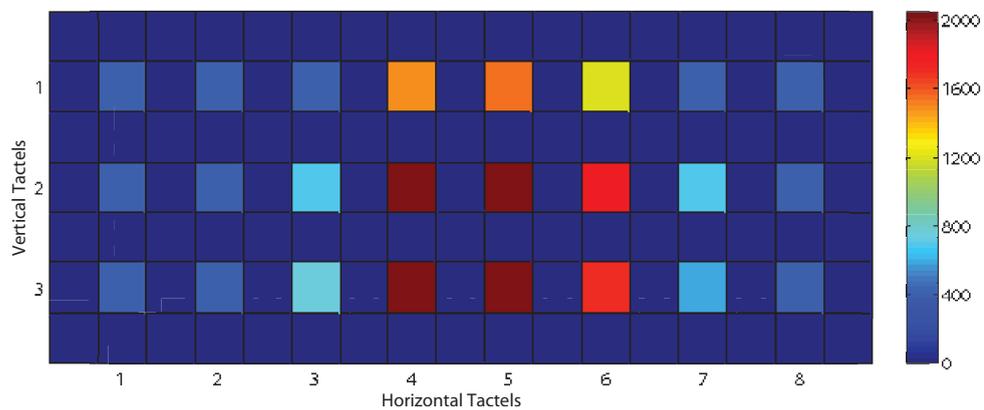


FIGURE 5.7: 2D map of pressure placed on the UMBS at one time instant.

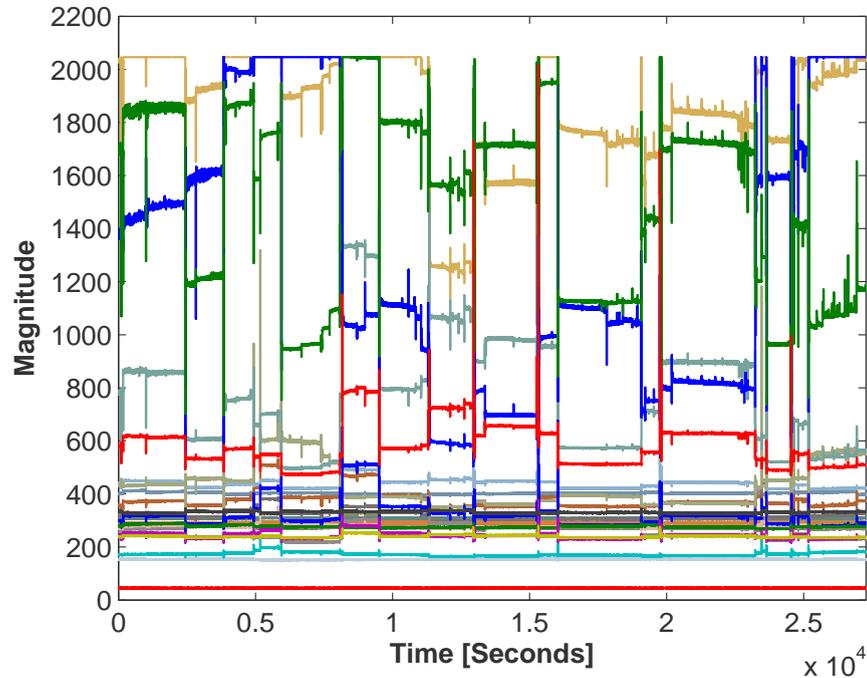


FIGURE 5.8: UMBS data over an entire sleeping period for a healthy adult subject.

shifted their body position. For multiple portions of the night, data from five tactels show pressure exerted on them (each of which are plotted in Figure 5.9). The body weight of the subject is not impinging on some tactels and as such they record values lower than 600.

### 5.2.3 Turns Data Set

Data were collected using the UMBS v3 showing the effect turns in bed had on the UMBS data. The UMBS was placed underneath the mattress. The participant was asked to assume and shift between four typical sleeping postures between lying on their back (B), left side (L), right side (R) and front (F) on the bed. The subject was asked to change postures under three conditions 1) rapid transition without lateral displacement, 2) rapid transition with lateral displacement and 3) slow rolling transitions (inherently including lateral displacement) as shown in Figure 5.10. Between each of these three conditions the subject exited the bed. Each rapid transition lasted approximately five seconds while the slow rolling transitions lasted approximately 15 seconds with approximately 60 seconds between transitions. The data set lasted a total duration of 1,140 seconds and contained 12 changes in posture, 3 bed entries and 3 bed exits. After the subject enters the bed and after each turn, the pressure array reports a new series of values which reflects the position of the subject. A settling time of approximately five seconds

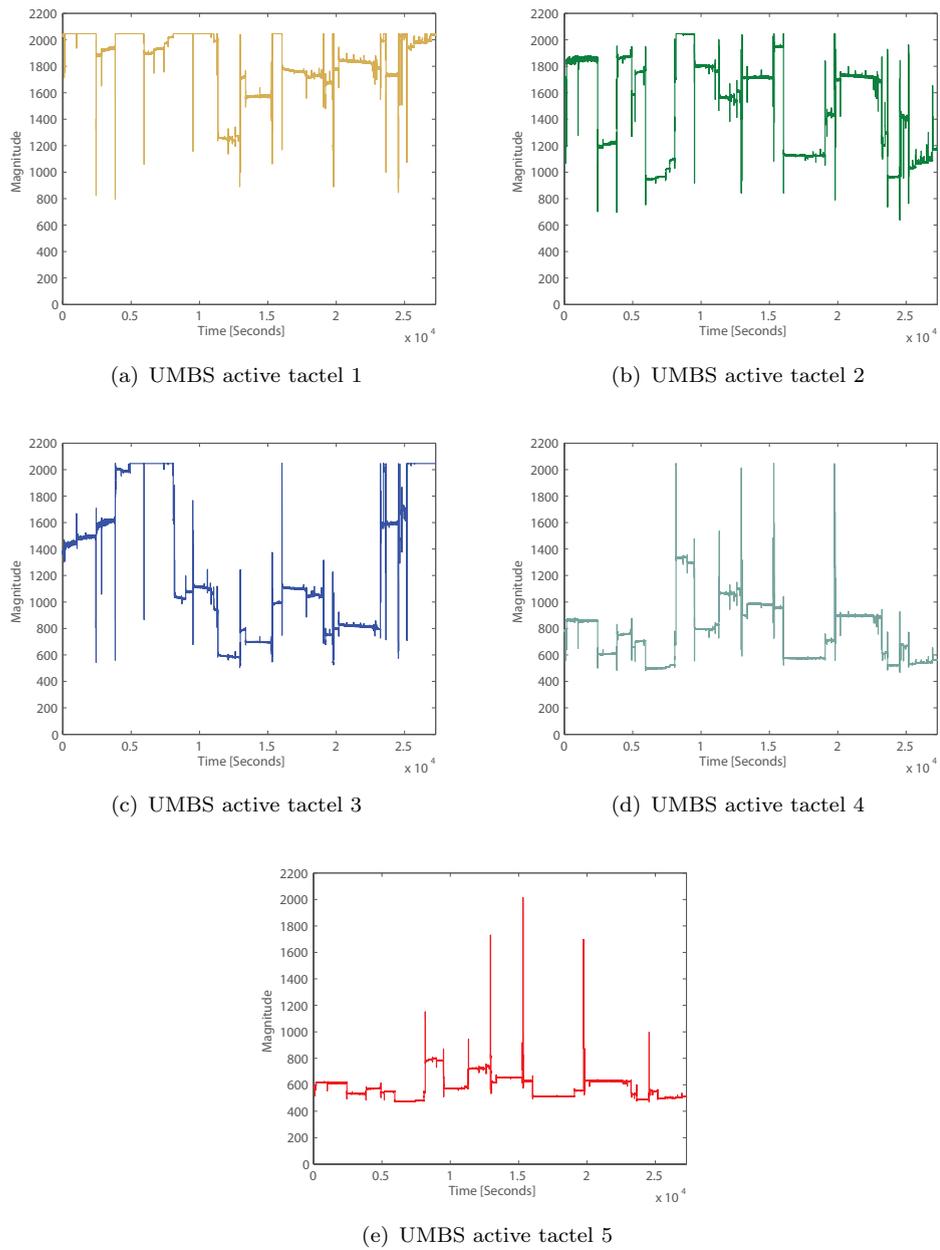


FIGURE 5.9: 5 active tactels over an entire sleeping period for a healthy adult subject.

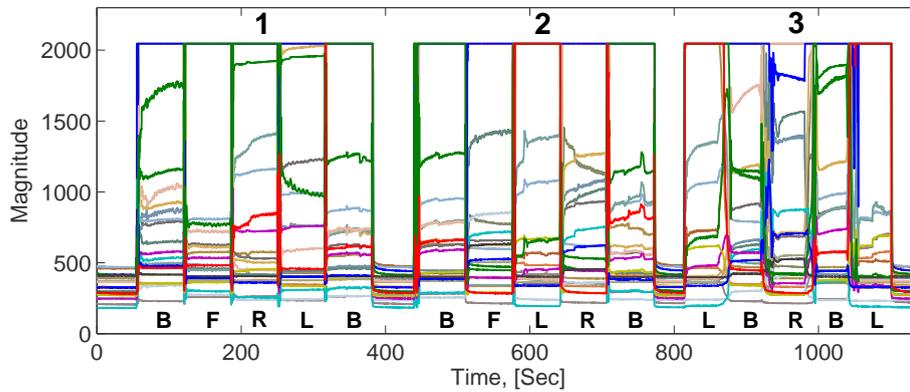


FIGURE 5.10: Sample UMBS Turns Data. The subject is lying on their back (B), left (L), right (R) and front (F) and where 1) represents rapid transition without lateral displacement, 2) represents rapid transition with lateral displacement and 3) represents slow rolling transitions

in the tactels values can be seen. This was extended when the subject performed slow turns.

A long settling time can be seen in Figure 5.10 when the subject exits the bed. This transient effect occurs less when the UMBS is not placed underneath a mattress and as such is thought to be an effect of the particular mattress.

### 5.3 UMBS Community Deployment in Older Adults

A preliminary investigation of sleeping patterns and sociality was carried out in a group of older adults recruited in Summerhill, County Meath, Ireland. This pilot study was carried out in collaboration with the Digital Health Group in Intel Ireland Ltd. All subject interviews were performed by trained ethnographers. This work measured sleeping patterns using both the UMBS and wrist actigraphy. While other metrics were recorded, generally they are outside the bounds of this research, and as such are only briefly discussed. This research evaluated the ability of the UMBS to record sleeping patterns and also its suitability for long-term placement. A preliminary investigation into the comparison of metrics, derived from the UMBS, against wrist actigraphy was also performed. This provided an insight into the application of the UMBS as a sleep monitoring device. Lastly, this study provided a means of investigating whether in-bed restlessness, that is the proportion of time spent moving in bed, is related to daily activity levels in older adults.

TABLE 5.2: Subject details from UMBS community deployment

Sub. Code	Sex	Age	Mean Sleep Length	Health Info.	Restlessness
DH701	F	63	08:35	Sleep Apnoea	16%
DH702	F	69	N/A	Spina Bifida	N/A
DH703	F	62	05:11	Rel. Healthy	3.70%
DH704	F	80	08:55	Rel. Healthy	8.2%
DH705	M	72	07:40	Stroke Sufferer	19.80%
DH706	F	88	N/A	Rel. Healthy	N/A
DH707	F	64	07:25	Rel. Healthy	11.90%
DH708	M	81	06:40	Rel. Healthy	12.10%
DH709	F	65	08:20	Rel. Healthy	9.30%
DH710	F	72	06:50	Rel. Healthy	14.40%

### 5.3.1 Methods

The UMBS (v1) was deployed in a community based study of ten (2 males) older adults (62-88 years old) for a period of two weeks (see Table 5.2). This was part of a study which investigated sleep, social and physical activity. Additional sensors and methods, including wrist actigraphy, a phone use sensor, location sensors (GPS), pedometers (placed on the waist line), as well as audio diaries and ethnographic interviews collected data relating to sleep, social and physical activity. The audio diaries were used to capture data from a modified version of the Pittsburg Sleep Questionnaire (Hislop et al., 2005) and internal state (how the subject felt). The subjects were also allowed to record comments as they wished using the audio diary. The ethnographic interview recorded data in relation to routine behaviours. The UMBS data were continuously recorded using the PDA based platform. The data were linearly interpolated to a constant sampling rate of 9Hz. The data were manually aligned with wrist actigraphy data prior to analysis.

Sleep analysis was not performed on two subjects: 1) DH702 had childhood spina bifida and as such her weight was not evenly distributed throughout the UMBS, and 2) DH706 refused to wear the wrist actigraph. Other problematic issues regarding the wrist actigraph were: 1) three subjects were initially unaware that it should be worn at night, and 2) the device was taken off while one participant was playing cards. It is unknown whether the latter incident was due to comfort or stigma (due to the perception of having to wear a health-type device).

The mean sleep length per subject over all nights of available data is given in Table 5.2. This was calculated using the wrist actigraph.

### 5.3.1.1 Detection of In-Bed Presence and Calculation of Bed Restlessness

The ability of the UMBS to extract daily sleeping patterns including time-to-bed, time-out-of-bed, number and duration of bed exits and bed restlessness was examined during this study. This was achieved using two methods: 1) defining when the subject was in bed, and 2) defining the amount of motion in bed. An empirically defined threshold, applied to all tactels, was used to define whether the subject was in or out of bed using the mean pressure placed on the UMBS (see Equation 5.4) and an *InBed* indicator decision function (see Equation 5.6). For this, the UMBS data was examined in thirty second epochs ( $m=270$  for 9Hz sampling rate) and if the mean value for two or more tactels was greater than a bed presence threshold (that is  $InBed^k \geq 2$ ), the subject was defined to be in bed.

The  $k^{th}$  data window (epoch) corresponds to the  $l = m(k - 1) + i$  to  $mk$  rows of the UMBS data matrix,  $\mathbf{X}$ , that is:

$$\mathbf{W}^k = [w_{i,j}^k] = w_{i,j}^k \triangleq x_{m(k-1)+i,j} \quad (5.3)$$

Defining  $\underline{w}_j^k$  as the  $j^{th}$  column vector of  $\mathbf{W}^k$  (ie.  $\underline{w}_j^k \triangleq w_{:,j}^k$ ), the mean of this vector may be denoted as

$$\bar{w}_j^k = \frac{1}{m} \sum_{i=1}^m x_{(k-1)m+i,j} = \frac{1}{m} \sum_{i=1}^m w_{i,j}^k \quad (5.4)$$

Using this metric, the *inBed* decision function can be defined as

$$InBed_j^k = \begin{cases} 1 & \text{if } \bar{w}_j^k \geq \text{bed presence threshold} \\ 0 & \text{otherwise} \end{cases}, \quad 1 \leq j \leq 24, 1 \leq k \leq \lfloor \frac{N}{m} \rfloor \quad (5.5)$$

For the  $k^{th}$  window, the number of tactels registering bed occupancy is given by

$$InBed^k = \sum_{j=1}^{24} InBed_j^k \quad (5.6)$$

For periods when the subject was in bed, a measure of their in-bed restlessness was calculated. This was calculating using the standard deviation of each thirty second epoch for each tactel  $j$ ,  $\sigma_j^k$  (see Equation 5.7 and Figure 5.11). The movement threshold

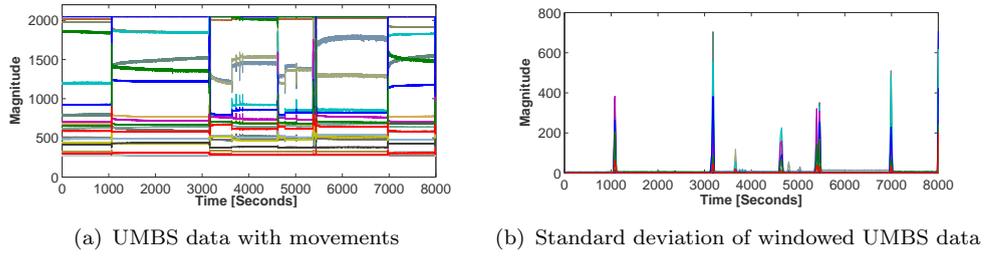


FIGURE 5.11: UMBS data with movements times highlighted using the standard deviation of the windowed UMBS data

which is specific to the mattress and subject, was used to recognise the occurrence of motion on each tactel and was empirically defined for this study. For a given epoch to be classed a 'movement' epoch, two or more tactels had to register motion, that is  $\mathcal{M}^k \geq 2$  (see Equation 5.9).

The standard deviation of the variation in the  $j^{th}$  tactel during the  $k^{th}$  window can be defined as:

$$\sigma_j^k = \sqrt{\frac{1}{m} \sum_{i=1}^m (w_{i,j}^k - \bar{\mathbf{w}}_j^k)^2} \quad (5.7)$$

The associated movement decision function is

$$\mathcal{M}_j^k = \begin{cases} 1 & \text{if } \sigma_j^k \geq \text{movement threshold} \\ 0 & \text{otherwise} \end{cases} \quad (5.8)$$

and the number of tactels registering motion in the  $k^{th}$  window can be defined as

$$\mathcal{M}^k = \sum_{j=1}^{24} \mathcal{M}_j^k \quad (5.9)$$

A restlessness index was generated for each subject over each night of the study from the UMBS data. This was calculated as a ratio of the number of movement epochs over the total number of epochs the subject was in bed. The mean restlessness for each subject over all nights is reported in Table 5.2.

### 5.3.2 Nocturnal Routines and Bed Restlessness

An examination of the suitability of the UMBS for long-term domestic placement and an investigation of the ability of the UMBS to extract bed entry/bed exit times was performed using this data set (Behan et al. 2008a, Behan et al. 2008b).

### 5.3.2.1 Methods

The in-bed indicator, as described in the previous section, was used to extract multiple metrics regarding general sleeping patterns including: 1) time to bed, 2) time out of bed, 3) the number of bed exits during the sleeping period, and 4) the duration of bed exits during the sleeping period in the entire population. These metrics, as well as the restlessness index (also described above), were calculated for all subjects over all nights.

### 5.3.2.2 Results

Sleeping patterns, inclusive of bed restlessness, time in bed and bed exit information, and the changes in these sleeping patterns over time are shown in Figure 5.12. These data were extracted for all participants. The mean time in bed (TIB) for this population was 524.4 minutes ( $\pm 40.8$  minutes) per night per subject as defined using the UMBS.

The extraction of quantitative statistics using this data set was unrealisable due to the low number of participants and also due to the short duration of the study. Generally the routines and sleeping patterns were consistent across the cohort over all nights, however some nights of sleep disturbance occurred and these were found to be resultant from pain or anxiety. Due to the high inter-subject variation, results from one subject are reported on a case-by-case basis. For subject DH701 in the second week of the study (from night 11 on), the bed restlessness was found to decline dramatically (as shown in Figure 5.12). Upon further investigation it was found that this subject had a highly inconsistent bed time and wake time during the first eleven days of the study. In the exit interview, the subject stated that they started an undergraduate degree after the eleventh day. Furthermore, on the night prior to beginning the college course the subject reported "I slept badly" in their audio diary. Upon investigation, an increased restlessness index of 19% was found (mean of study 8.51%) for this night. While variations in sleeping patterns over time were noticeable in the other subjects, these changes were not found to be correlated with any serious life events.

### 5.3.2.3 Discussion

During the study it was found that the UMBS was an easily deployable technology. No complaints were made by the subjects in regard to the use of the UMBS, whereas adherence issues were evident with wrist actigraphy and other modalities which required any subject interaction. As such, the UMBS is suited to a long-term deployment. A number of subjects in this study expressed a discontent in wearing the actigraph (the ambulatory gold standard for sleep monitoring). One subject chose not to wear the

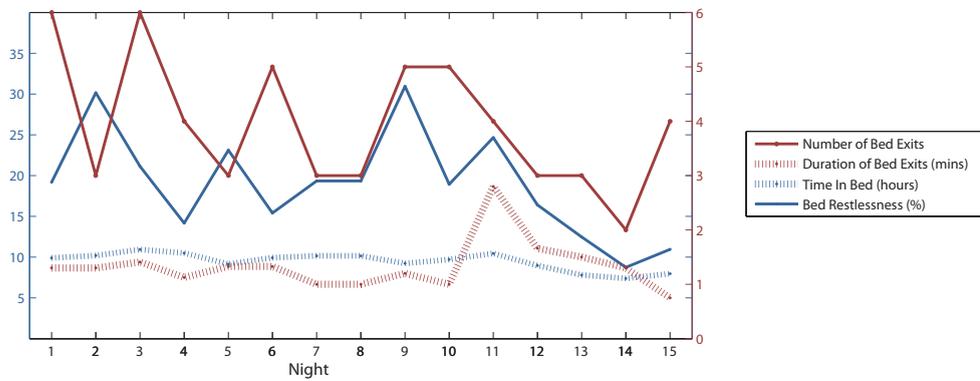


FIGURE 5.12: Sleeping pattern data from subject DH701 over the two week trial. The magnitude of the blue data is given on the left axis, while the magnitude for the brown data is given on the right axis.

watch, while others often forgot to put the device back on after a shower or after a social event where they did not want to wear the device. None of the subjects reported any discomfort resulting from the presence of the UMBS.

The mean time in bed (TIB) for this population (524.4 minutes/night) was longer than that recorded for other studies involving older adults which reported a mean nocturnal time in bed of 475.6 minutes ( $\pm 52$  minutes) (Prinz et al., 1975) and of 468.9 minutes ( $\pm 38.3$  minutes) (Feinberg, 1974). The extended nocturnal sleep duration seen in this study could be because subjects did not partake in any daytime naps or it could be a reflection of their overall healthy status.

This study found that for one subject with a highly irregular routine, the imposition of a regularised schedule served to reduce restlessness. An investigation into the relationship between sleeping patterns and metrics of sociality, routines and daily activity was also undertaken (Behan et al. 2008a, Behan et al. 2008b), however only data relating to the UMBS is reported in this thesis. This research poses the UMBS as a technological system suitable for long term placement as well as a method to quantify changes in daily sleeping patterns. Further studies should examine the this relationship and investigate whether regular sleeping patterns (and schedules) lead to an increase in quality of life metrics, or in health status.

The UMBS was found to report bed times, bed exits and bed-restlessness unobtrusively. Additionally, this system is an ideal modality to investigate nocturia (previously discussed in Section 2.6.2.6). Research by Rantz et al. (2008) using a different pneumatic pressure based sleep sensor placed above the mattress has shown that changes in bed restlessness and bradycardia (specifically, the increased occurrence of large inter-heartbeat-intervals during the night) over long periods of time are indicative of serious

life events. All of the results in this study were reported using a case study analysis. This reflects the investigative nature of this research and the high overhead in the deployment and management of in-home health technologies. Additionally, it is difficult to recruit a cohort of older adults with very similar health levels and co-morbidities. As such, the collection of data from large cohorts is often prohibitive. Large cohorts will help in providing statistically significant results, however the process of answering a strictly defined research question (usually requiring a specific cohort and investigating a specific manifestation of an illness or disease) is often unfeasible.

### 5.3.3 UMBS vs Wrist Actigraphy

A direct comparison between the motion detection capacity of the UMBS and the wrist actigraph was performed using the data from this study (Walsh et al., 2008). This investigation considered wrist actigraphy to be the gold standard for movement detection. While this is a sub-optimal approach in that it assumes that wrist actigraphy is 100% accurate, it was used to provide an indication of the similarity between UMBS derived metrics and wrist actigraphy.

#### 5.3.3.1 Methods

One night of UMBS and wrist actigraphy data was taken from each person (total duration 59 hours and 36 minutes). The UMBS (v1) and PDA data logging platform were used to record the UMBS data. The UMBS data was subsequently interpolated linearly to a constant sampling rate of 9Hz. The UMBS data was manually aligned to the wrist actigraphy data and truncated appropriately. A highpass fifth order butterworth filter with a cutoff frequency of 1.5Hz was applied to each tactel signal, such that only movement related information would remain and the relative pressure impinged on each tactel would be removed. The data was sectioned into multiple non-overlapping windows lasting: 30 seconds, 60 seconds, 90 seconds, 120 seconds, 150 seconds, 180 seconds, 210 seconds and 240 seconds. Two metrics of motion were calculated over each epoch: 1) The standard deviation of the signal over this window for each tactel ( $\sigma_j^k$ ), and 2) the sum of the difference between signal values over the window for each tactel ( $\Delta_j^k$ ) as shown in Figure 5.13. The former is computed according to Equation 5.7, while the latter is defined as

$$\Delta_j^k = \sum_{i=2}^m |w_{i,j}^k - w_{i-1,j}^k| \quad (5.10)$$

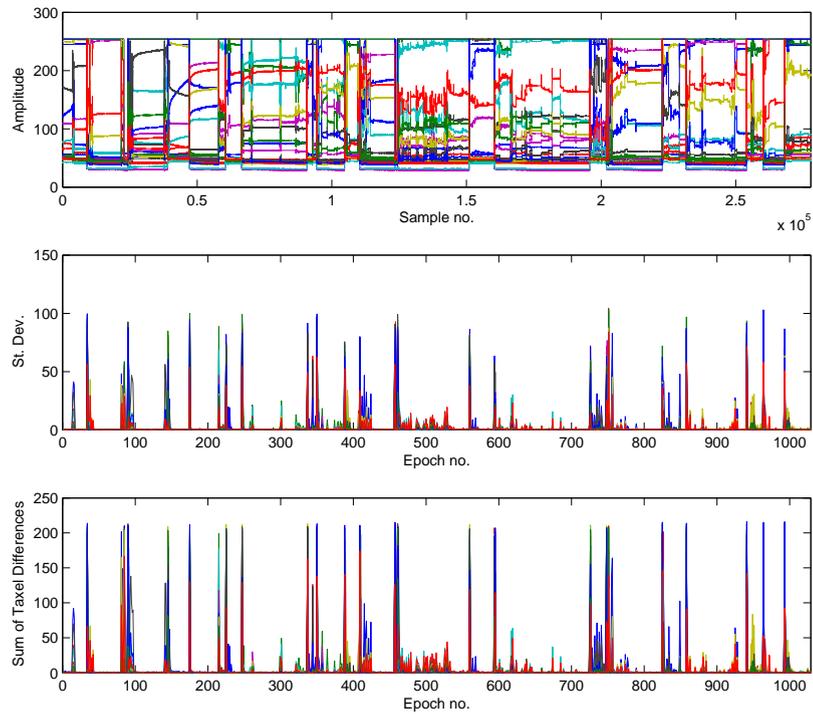


FIGURE 5.13: Comparison of UMBS data, standard deviation metric and sum of difference metric.

The sums of both metrics ( $\sigma_j^k$  and  $\Delta_j^k$ ) over all 24 tactels were also calculated.

The optimal threshold for discriminating movement from each of the UMBS metrics from any wrist actigraph defined motion (activity counts  $> 0$ ) was determined using MCC (see Section 3.2). MCC was used as the data set was significantly biased towards the non-movement state. Furthermore, it was unknown whether a thirty second epoch (which is traditionally used for PSG scoring) is an appropriate epoch length for activity detection. As such, the window length,  $m$ , was also optimised as part of the investigation.

### 5.3.3.2 Results

The best MCC was obtained with a window length of 210 and 240 seconds (see Figure 5.14). However, this was considered too long to give a realistic time resolution for motion. As such, a shorter, more adequate window length of 90 seconds reported an MCC of 0.62 (and using the sum of differences movement decision function). High sensitivity values of 69.6% and very high specificity values of 89.6% were reported for this selection as given in Table 5.3.

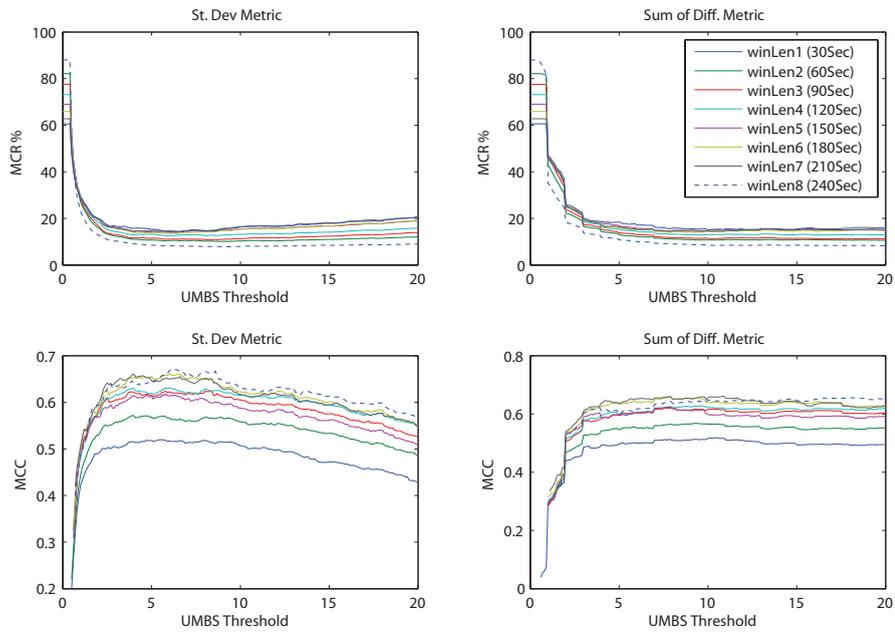


FIGURE 5.14: MCC and MCR for both movement decision functions over 8 window lengths.

TABLE 5.3: Determination of the optimal coefficients

Dec. function	Thres	Win. Len. (seconds)	Sens %	Spec %	MCR %	MCC -1 to +1
Best Classification Results						
St. Dev.	6.3	240	72.6	92.2	14.07	0.67
Sum Diff.	10.8	210	72.6	91.2	14.34	0.66
Optimal Classification Results						
St. Dev.	5.8	90	63.7	94.5	11.07	0.63
Sum Diff.	5.8	90	69.6	89.6	13.3	0.62

This study was repeated with data from one healthy female adult over one night. The tuning process was repeated and specificity and sensitivity values greater than 90% were achieved for a window length of 120 seconds.

### 5.3.3.3 Discussion

This preliminary investigation showed high agreement rates between wrist actigraphy and the derived UMBS motion metrics. This provided similar performance to an alternative non-contact sleep monitoring system (BiancaMed) (Fox et al., 2007) (as detailed in Section 4.5.4.2). When the UMBS system was applied to data collected from a young

healthy adult, sensitivity and specificity rates over 90% (compared to wrist actigraphy) were reported. It is unclear why there is such a large difference between the performance rates found for young and healthy versus older and relatively healthy subjects. In all of this analysis the wrist actigraph was taken to be the gold standard for movement detection, however it is not, and as such, further validation is required. Further studies using an established reference for motion detection (such as a video camera) would elucidate whether metrics derived from the UMBS or from the wrist actigraph are more effective in detecting body movement while in bed.

### 5.3.4 UMBS and Daily Activity

An investigation into UMBS derived restlessness and daily activity was performed (Walsh et al., 2009b). This was performed in order to check if there was a relationship between sleep disturbances and activity levels (as measured using a step counter) the following day.

#### 5.3.4.1 Methods

The daily number of foot steps taken by each subject was recorded using a pedometer attached to the waist line of each subject. This data was directly compared to a restlessness metric, averaged per hour, generated each night by the UMBS. Data was available over a total of 87 nights. No valid pedometer data was available from subject DH705 due to an impaired gait.

#### 5.3.4.2 Results

The mean hourly number of restless epochs and the mean number of steps (the following day), and their standard deviations are given in Table 5.4. The correlation between both of these metrics was computed for each subject over all nights available. While general subject specific patterns are presented, none of these correlations were significant. An inverse relationship was found with an  $r$  value of -0.35 over the entire data set, however this correlation was not significant ( $\rho > 0.1$ ).

#### 5.3.4.3 Discussion

While the associations found in this data set were not significant, it suggests that an inverse relationship between bed restlessness and subsequent daily activity exists. However this was only found in the general case and might not apply to individual subjects

TABLE 5.4: Comparison of UMBS derived restlessness versus activity count (via pedometer).

Subject	No. Days	Avg. Hourly Restless Epochs	No. Steps	Correlation
<b>DH701</b>	12	$19.38 \pm 5.17$	$1697 \pm 971$	-0.115
<b>DH703</b>	10	$6.55 \pm 3.93$	$8551 \pm 4670$	-0.086
<b>DH704</b>	13	$4.26 \pm 0.66$	$5057 \pm 2308$	0.06
<b>DH705</b>	2	$9.32 \pm 3.05$	n/a -	-
<b>DH706</b>	11	$9.11 \pm 2.55$	$8462 \pm 7384$	-0.47
<b>DH707</b>	14	$9.43 \pm 1.69$	$1776 \pm 819$	-0.31
<b>DH708</b>	8	$9.52 \pm 0.59$	$5061 \pm 1998$	0.79
<b>DH709</b>	14	$12.32 \pm 5.1$	$3632 \pm 2027$	-0.01
<b>All</b>	87	$10.26 \pm 5.62$	$4767 \pm 4526$	-0.35

(for example, subject DH708 exhibited a large positive correlation). A larger investigation with a larger cohort over a longer study would uncover any significant correlation between these data, if it exists.

### 5.3.5 Overall Discussion

From the community deployment of the UMBS, it was shown that the UMBS is suitable for long term placement. Adherence issues were reported for the wrist actigraph, however not for the UMBS. Bed presence and bed restlessness indicators were found to provide quantitative statistics of sleeping patterns, particularly bed times, bed exit data and bed restlessness. A comparison of UMBS-derived bed restlessness data and wrist actigraphy showed high agreement rates. However, this investigation defined wrist actigraphy to be an absolute measure of movement which is incorrect. The use of video monitoring and widely accepted motion detection algorithms as the gold standard movement detection modality would provide a more robust comparison of the UMBS-derived metrics and wrist actigraphy. An investigation into the relationship between the UMBS-derived movement metric and daily activity (as measured in step counts) showed a negative correlation, however this was not statistically significant. A larger study may elucidate this claim.

## 5.4 Parallel UMBS Research

A separate research group (based in Carleton University, Ottawa, Ontario, Canada) have investigated the use of the UMBS as a sleep monitoring system, although the focus

of their work is mainly in detecting sleep apnoeas. This has focused on: 1) the detection of breathing signal using a time-based correlation with a respiratory inductance band, 2) the detection of central apnoeas, and 3) an investigation of the attenuation of the respiration signal caused by various mattresses. A brief description of this work is given below.

A cross-correlation based method of respiratory rate estimation, using data from each tactel in the UMBS and compared to respiratory effort (measured using a respiratory inductance band), reported a high accuracy (Jones et al., 2006a). For this study, the UMBS was placed in direct contact with the subject. Furthermore, a reliability metric of the respiratory signal calculated for each tactel was generated using a metric derived from the auto-correlation of the tactel data. This method used the difference between the first peak and trough in the autocorrelation sequence to measure motion. If a suboptimal reliability metric was found for that tactel, it was excluded from the calculation of an overall respiratory rate. The respiratory rate was estimated using data from each tactel which was not saturated or completely under-saturated. The respiratory rate was calculated using the autocorrelation of the signal. The distance between the first two peaks was used to estimate the breathing rate. The respiratory rate estimates from each sensor were fused using a cluster-based voting approach which removed any outliers. Further research was undertaken which used an adaptive linear combiner to correct for, and multiplex, tactel signals which were misaligned in the time domain prior to respiratory rate estimation (Holtzman et al., 2008a, 2011; Townsend et al., 2011b). This was performed for data collected when the UMBS was placed underneath the mattress. This research validated the optimal method for phase correction using the cross correlation of signals from each tactel. A comparison was made to two respiration signals measured using RIP bands. A high correlation against both respiratory effort signals ( $r=0.68$  and  $r = 0.72$ ) was found (this was found to be significant ( $p < 0.05$ )). Furthermore, this research mapped the delay between tactels in a 2-d graph in an effort to provide a quantification of the change in torso movement due to breathing over time.

The effect of different mattresses on the attenuation of the pressure signals was also investigated (Holtzman et al., 2010). This study investigated coil, foam and futon mattresses of various depths. Two UMBS were used for this analysis. One was placed on top of the mattress directly underneath the person, while the second was placed underneath the mattress. A large attenuation in the signal to noise ratio (the power spectral density peak magnitude divided by the total spectral density within the respiration band) was found in the sensor placed underneath the mattress. Additionally, the respiratory time series signal was found to be greatly reduced in amplitude in the sensor placed underneath the mattress. It was found that the identification of physiological signals from the sensor underneath the mattress was much more difficult than when the sensor was in

direct contact with the person. However, the extraction of a respiration signal was still realisable.

A dynamic approach to detecting movement onset times using the UMBS has also been investigated (Jones et al., 2006b). This method calculated the variance and standard deviation in a moving average window from each tactel. Movement was defined to begin when the change in a tactel value goes above three standard deviations (the standard deviation was computed from the previous thirty second window). This method was applied to experimental and simulated results and found to give a high accuracy in determining the starting time of movements. This work was extended to detect periods of low variance in the respiration signal which would indicate the occurrence of central apnoeas (Townsend et al., 2009a, 2010, 2011a). A dynamic relative threshold was used to identify the start and end points of the apnoea. This dynamic approach is ideal for use with different mattress types. This work indicated a positive predictive value of over 0.75 and a Matthew's correlation coefficient (discussed in Section 3.2) of greater than 0.72 in detecting simulated central apnoeas.

An algorithm for detecting rollovers in bed was developed using features derived from the change in subject position over time and an LDA classifier (Townsend et al., 2009b). For this study, a large number of artificially induced rollovers were performed by one healthy volunteer. High sensitivity and specificity values were reported.

## 5.5 Conclusions

This chapter introduces the UMBS as an unobtrusive sleep monitoring system. A description of the pressure sensing mechanism, which make up each individual sensing location in the UMBS, is provided. Additionally, an outline of the communications protocol and the data collection system is given. Descriptions and figures illustrating UMBS data for various scenarios and conditions, including tactel saturation, respiration data, UMBS data recorded over an entire sleeping period, and UMBS data during turns are also provided.

A preliminary investigation of the suitability of the sensor for long term deployment was carried out through a community based deployment amongst older adults. The UMBS was found to be an unobtrusive and valid instrument for the long term monitoring of sleeping patterns, particularly for use in an older adult cohort. Additionally, a comparison of the sensor against wrist actigraphy showed high agreement rates in the detection of nocturnal activity (as defined by wrist actigraphy). Lastly, a bed restlessness metric,

computed from the UMBS data, was shown to correlate to daily activity measurements in older adults (in the form of step counts).

Recently, the ability of the UMBS to identify respiratory patterns, and particularly to identify central apnoeas has been performed by another group (Jones et al., 2006a; Holtzman et al., 2008a, 2011; Townsend et al., 2011b; Holtzman et al., 2010; Jones et al., 2006b; Townsend et al., 2009a, 2010, 2011a). This research identified a high correlation between a dynamically weighted mixture of tactel signals (representing respiratory effort) and a respiration signal measured using an RIP band. However, a comparison of an algorithm which extracts respiration rate would provide a more robust comparison against the current gold standard. In the work presented in this chapter, frequency analyses (using the Fourier transform) were shown to be indicative of manually observed estimates, however the accuracy seemed limited. Further investigation with a larger data set is required to elucidate whether frequency based analyses can derive respiration rate. A high correlation, as reported by Townsend et al. (2011a), does not infer an ability to determine respiration rates.

This chapter discussed a methodology for using the UMBS to extract sleeping patterns over an extended period of time in a sensitive population (older adults). The long-term collection of data could be used to examine any correlation between serious life events (for example adverse health events leading to a reduction in quality of life) and changes in sleeping patterns. Ongoing research has found that long term sleep monitoring, inclusive of the monitoring of bradycardia (an increase in the amount of 'slow' heart beats) and bed restlessness, has been linked to overall health status (Rantz et al., 2008). However, generally, data presented to date has been presented on a case-by-case basis (similar to an approach used in parts of this chapter) and a thorough validation using a large data set would provide more quantitative results. The identification of changes in behaviours and activity needs to be flexible and robust enough to detect any changes in pattern which might relate a negative change in health status (regardless of disease or illness type, or cause). While a case study analysis provides an insightful view of the capability of a technology or approach to health monitoring, it does not provide a statistically valid solution which can be applied to large cohorts. However, the presence of multiple comorbidities in older adults is very common makes the problem increasingly challenging. Additionally, the financial overhead and lengthy duration required for running lengthy studies is often prohibitive. However, the first step in achieving such a solution is to standardise and validate all sensors under analysis. In the context of this research, a rigorous validation of UMBS-derived motion metrics, against a video-based motion metric, is performed in the next chapter. Additionally, automated algorithms for the estimation of respiration rate are also developed.

## Chapter 6

# Experimental Validation of the UMBS

Sleep and wake have traditionally been monitoring using multiple physiological signals (including ECG, respiration, EEG, EMG and EOG). However, the ability to distinguish between active wake (or restlessness) and quiet sleep, based on motion has enabled sleep monitoring to be moved from research settings and from the clinic into the real-world. The use of motion/non-motion as a proxy for wake/sleep has facilitated the acquisition of quantitative data from various cohorts and in many different settings. While Ambulatory Polysomnography (A-PSG) can be more accurate and wrist actigraphy is more accepted than recent advances in sleep monitoring, both impinge upon the subject due to the wearing and active participation required by the subject. Additionally, this is unsuitable for long-term monitoring. Furthermore, these are inappropriate for certain populations and in certain situations (including the elderly, those with mental health issues, large studies and studies in extreme environments). Several competing technologies have been proposed as ambient sleep monitoring solutions suitable for long-term data collection amongst a sensitive cohort (such as residents in an assisted living facility). These solutions have been proposed in the form of an air mattress, a radar based technology placed beside the bed, optical and PIR based solutions (further details on these technologies is given in Chapter 4).

In this chapter the results of a series of experiments are presented which validate the UMBS as an ambient sleep monitoring solution. Firstly, Experiment 1 details the findings of a preliminary study examining the efficacy of the UMBS in the estimation of physiological signals. Subsequently in Experiment 2, a larger investigation was carried out in order to develop and validate algorithms which reliably estimate respiration rate from the UMBS data. Finally in Experiment 3, algorithms were designed to maximise

the ability of the UMBS to discern nocturnal body movement and a comparison of these algorithms against other ambient solutions was also made. Each of these three experiments are discussed separately. The methods, developed algorithms, results and a discussion is given for each experiment. An overall conclusion describing the ability of the proposed sensor to capture relevant information pertinent to inferencing sleep is given at the end of the chapter.

## 6.1 Experiment 1 - Detection of Physiological Signals

Initial findings, reported in Chapter 5, associated large body movements with deviations in the UMBS pressure data. Previous research has shown that small subject movement can be attributed to the physical action of the beating heart and the inhalation and exhalation of the lungs in similar devices (Alametsa et al., 2008; Brink et al., 2006; Mack et al., 2009b; Rosenblatt, 1957). A preliminary investigation was carried out assessing the sensitivity of the UMBS in measuring physiological information.

### 6.1.1 Methods

Three healthy young adult subjects (two male, one female; mean age 25 years old) were recruited for this pilot study. Each subject assumed three sleeping positions resulting: supine (lying down, face up), prone (lying down, face down) and lateral/foetal (lying down on one's side, specifically on the right side during this experiment). Each individual test had a duration of two minutes and was investigated in thirty second epochs. The heart beat (recorded via the radial artery) was located prior to the beginning of each test. The total number of heart beats and breaths (one inhalation and one exhalation) were recorded over each of the four thirty second periods during each test. During this experiment the subjects were asked to lie naturally and to refrain from making any large movements. The UMBS was placed directly underneath and in contact with the subject.

Data were recorded on a Dell Precision laptop with an Intel Dual-Core Processor using a customised C++ software interfacing to the UMBS (v2) as introduced previously in Chapter 5. The UMBS data were subsequently linearly interpolated to a constant sampling rate of  $10Hz$  for further analysis.

An extract of the UMBS data recorded for the female subject is shown in Figure 6.1. The cyclic inhalation and exhalation patterns related to breathing can be clearly observed. This breathing pattern is not evident in each of the 24 pressure sensors within the UMBS. This is because some tactels are saturated (resulting in a maximal value of

255) while other tactels do not report any information as the subject's weight does not impinge on them (resulting in an oscillating value of approximately 45). Tactels containing information were identified using two criteria: firstly, signals must have a standard deviation greater than an empirically defined threshold, and secondly, they must fall within the dynamic range of the sensor, (using notation developed in Chapter 5) that is,

1.  $\sigma_j^k > \tau_{min}$
2.  $\overline{\mathbf{w}}_j^k > \tau_{noload}$
3.  $\text{Max}(\mathbf{w}_j^k) < \tau_{saturated}$

Here  $\tau_{min}$ ,  $\tau_{noload}$  and  $\tau_{saturated}$  are pre-defined thresholds,  $\sigma_j^k$  is the standard deviation along each tactel,  $\mathbf{w}_j^k$  is the data during the  $k^{th}$  window along the  $j^{th}$  tactel, and  $\overline{\mathbf{w}}_j^k$  is the pressure on the  $j^{th}$  tactel during the  $k^{th}$  window instants.

Tactels not satisfying these conditions were not included in any further analysis which resulted in a reduced number of signals. The reduced set of signals corresponding to Figure 6.1 are given in Figure 6.2(a).

The signals from each of the active tactels were split into 30 second segments which contained 300 samples. The mean was removed from each segment and they were normalised to a range of [-1, 1]. The Fourier transform was applied to each set of data as discussed in Chapter 3. Each 30 second segment was zero padded until it contained 512 samples. A spectral analysis of the signals from the active tactels using the Fourier transform can be seen in Figure 6.2(b). A visual examination of the frequency spectrum found peaks in both the respiration and heart rate frequency ranges. The first peak at 0.01 Hertz (Hz) is due to a small linear trend in the data. The second peak occurs at 0.23 Hz (13.8 cycles per minute) and matches the breathing patterns of the subject. The third peak occurring at 0.46 Hz relates to the first harmonic of the peak at 0.23 Hz. The last peak occurs at 1.06 Hz. This matches the manually recorded heart rate of the subject over that test and corresponds to 63.6 beats per minute. A visual analysis of time series and frequency spectrum data for all epochs over all tests was performed and compared to the manual recorded cardiac and respiratory rates.

### 6.1.2 Results

The respiration and heart rates for each of the nine tests (three subjects in three positions), extracted using the time series and frequency spectrum information, can be

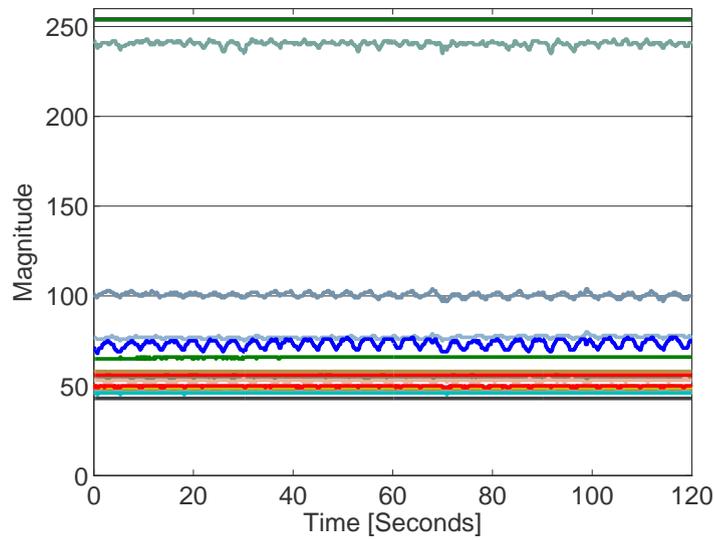


FIGURE 6.1: Example UMBS data recorded from a female subject lying in a prone position

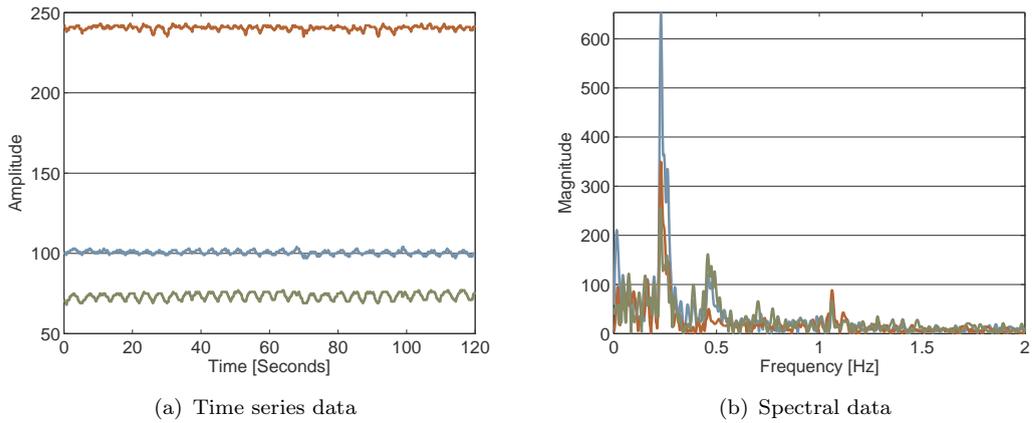


FIGURE 6.2: Active UMBS signals recorded from a female subject lying in a prone position

seen in Table 6.1. The heart rate could not be found during a visual examination of the UMBS time series data and as such is not reported. A clear peak relating to the respiratory rate using the frequency information was clearly evident in most epochs (34 out of 36). Frequency peaks relating to heart rate were found in 12 of the 36 epochs. An investigation of the remaining epochs revealed that noise resultant from body movement contaminated the frequency spectrum significantly.

A comparison of breathing and heart rates between the UMBS time and frequency series data generally showed high agreement rates as shown in Table 6.2. Furthermore, the Pearson's correlation coefficient and mean percentage error for estimated breathing

rates against the manually observed methods reported good results and are given in Table 6.3. A scatter plot of actual vs. estimate breathing rate for both UMBS time series and frequency spectrum data is given in Figures 6.3(a) and 6.3(b) respectively.

Over all tests, a mean heart rate difference of 1.275 ( $\pm 4.02$ ) beats per minute was found. An investigation found that the data collected from Subject 3 in the prone position reported an abnormally high difference between the estimated (using the frequency based visual analysis) and measured cardiac rates. The heart rate could only be estimated when this subject was in the prone position, and not in the supine or side-lying positions. This was presumably due to a movement related artefact due to breathing which was larger relative to the cardiac cycle than for other subjects. Excluding Subject 3, the mean rate difference reduced to  $-0.87 (\pm 0.941)$  beats per minute as given in Table 6.2. A scatter plot of actual vs. estimated heart rate for UMBS frequency spectrum data for all good data over all subjects is given in Figure 6.3(c). The Pearson's correlation and mean percentage error for estimated heart rates against manually observed values for all subjects are given in Table 6.3.

### 6.1.3 Discussion

This experiment was designed to provide a preliminary validation of the efficacy of the UMBS in capturing respiration and heart rates reliably. The results report a high correlation, low mean percentage error and a low mean difference to manually detected actual rates. However it was found that body movement causes significant contamination of the frequency spectrum data. The heart rate could not be estimated in 66% of the epochs; while respiration could not be discerned in 5.05% of the epochs. This is due to respiration having larger perturbations than the heart. For one subject the heart rate information was questionable. This could be due to various potential sources of error including, but not limited to, insufficient contact with the UMBS and faulty manual observation of heart rate. Experiments using longer recording times, longer epochs and a larger cohort will limit the effect of these sources of error. The epoch lengths in this experiment were short, especially with respect to the number of respirations per epoch. Longer epoch lengths would facilitate a greater number of breaths and heart beats per epoch giving increased statistical power and hence more reliable results when reporting their variability. Longer epoch lengths would also lessen the effect of short large body movements on the frequency spectrum. Two scenarios for recording data are proposed: 1) the subject being in direct contact with the sensor, and 2) the sensor is placed beneath a mattress. The latter will investigate the sensitivity of the sensor when placed in a truly ambient position. Additionally, a comparison of these results will quantify the attenuation resultant from using the mattress. The UMBSv1 has a

TABLE 6.1: Heart beat and respiration rates over 3 subjects for 3 positions each. The rates for each epoch are measured manually by an observer, by visually examining the UMBS time series data (UMBS time) and also by visually examining the UMBS frequency spectrum (UMBS freq). A manual determination of the heart rate was not possible by a visual examination of the UMBS time series data (-). For some epochs a visual examination of the frequency spectrum could not discern a clear frequency peak for both respiration and heart rate data (N/A).

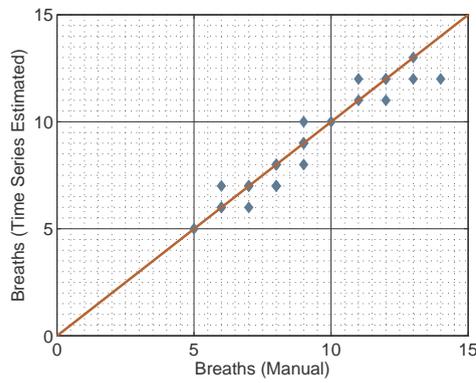
Subject	Position	Measurement Method	Heart Beats				Breaths			
			(per 30 Seconds)				(per 30 Seconds)			
1	Supine	Manual	28	29	31	32	9	8	8	7
		UMBS time	-	-	-	-	9	8	7	7
		UMBS freq	-	-	-	-	8.1	8.1	7.5	7.2
1	Prone	Manual	32	32	32	32	8	8	7	7
		UMBS time	-	-	-	-	7	8	7	6
		UMBS freq	30.9	31.2	-	31.2	7.2	7.5	6.9	6.6
1	Lateral	Manual	32	31	31	31	9	9	9	10
		UMBS time	-	-	-	-	8	9	9	10
		UMBS freq	30.9	-	-	31.2	8.4	9.3	10.5	9.5
2	Supine	Manual	30	31	30	31	9	9	9	9
		UMBS time	-	-	-	-	9	9	10	9
		UMBS freq	-	-	29.7	27.9	9	8.6	9.6	9.6
2	Prone	Manual	32	32	32	33	13	11	14	13
		UMBS time	-	-	-	-	12	12	12	13
		UMBS freq	-	-	31.8	32.4	12.15	12.6	12.9	12.45
2	Lateral	Manual	32	32	32	31	12	11	12	12
		UMBS time	-	-	-	-	12	11	12	11
		UMBS freq	-	-	-	-	11.4	11.25	-	11.85
3	Supine	Manual	38	40	38	37	8	8	7	7
		UMBS time	-	-	-	-	8	7	7	7
		UMBS freq	-	-	-	-	7.5	-	7.5	6.75
3	Prone	Manual	34	34	34	37	6	6	6	5
		UMBS time	-	-	-	-	6	6	6	5
		UMBS freq	39.9	43.2	42	-	6	5.4	6	5.5
3	Lateral	Manual	38	36	35	34	6	6	6	6
		UMBS time	-	-	-	-	7	7	6	6
		UMBS freq	-	-	-	-	6.3	6.3	5.7	6

TABLE 6.2: The difference between respiration rates captured over all subjects, mean  $\pm$  standard deviation breaths per minute. The heart rate difference could not be calculated using UMBS time series data and also excludes Subject 3 in the prone position, mean  $\pm$  standard deviation beats per minute.

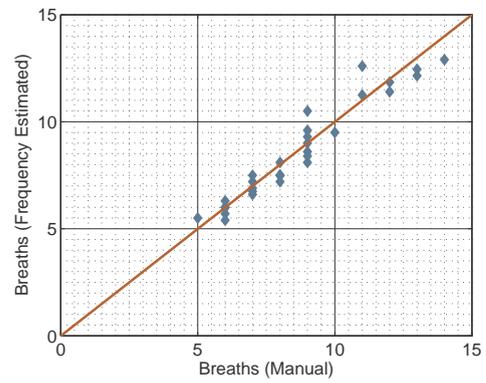
	Difference Against Manual Method	
	Respiration	Heart Rate
UMBS time	-0.14( $\pm$ 0.639)	N/A
UMBS Freq	-0.08( $\pm$ 0.614)	-0.87( $\pm$ 0.941)

TABLE 6.3: Pearson’s correlation coefficient and mean percentage error for the estimated UMBS data against the actual breathing and heart rates over all subjects. Data is also shown for the heart rate estimate excluding Subject 3 Prone position.

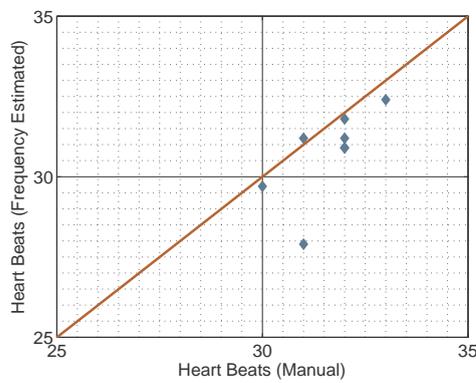
		Respiration All Subjects	Heart Rate All Subjects	Heart Rate w/o Sub 3 Prone
UMBS Time	Correlation [ $\rho$ ]	0.96	-	-
	Error [Mean, %]	-1.10%	-	-
UMBS Freq	Correlation [ $\rho$ ]	0.56	0.88	0.7
	Error [Mean, %]	-6.80%	3.61%	-2.74%



(a) Actual Respiration Against UMBS Time Series Estimate



(b) Actual Respiration Against UMBS Frequency Spectrum Estimate



(c) Actual Heart Rate Against UMBS Frequency Spectrum Estimate

FIGURE 6.3: Scatterplots of UMBS derived physiological measures to actual and respiratory rates. Diagonal line represents point of perfect alignment.

dynamic range of [0-255]; increasing this should positively affect the sensitivity of the UMBS when detecting physiological signals. Additionally, the estimates of breathing and heart rates were calculated manually by an observer in this experiment; however this can raise a source of error and objective measurements should be used.

## 6.2 Experiment 2 - Automated Respiration and Heart Rate Estimation

An accurate and automatic method of estimating respiration and heart rate using the UMBS (v3) was investigated using both time and frequency based methods. Ground truth measurements were used to assess the valid, robust and accurate detection of these physiological signals using the UMBS as proposed previously. A larger cohort was recruited and longer data sets were used. The epoch lengths were also increased to a 5 minute duration as used in similar studies by other research groups (Brink et al., 2006; Carlson et al., 1999; Shin et al., 2010). The dynamic range of the UMBS (v3) is 8 times greater than the UMBS (v1) (varying from [0-2045]), and although the pressure range is similar, the sensitivity of the sensor in detecting movement should be greater.

Figure 6.4 shows concomitant data without movement artifacts from the UMBS, from a strain gauge recording respiratory effort and from a PPG recording cardiac pulse rate. As shown previously, respiratory information can be seen in the UMBS data when compared to the respiratory effort signals directly. In the previous experiment, slight perturbations in the UMBS signals were found to result in a peak in the frequency domain correlated with the estimated heart rate. It is of no surprise that a frequency analysis of UMBS signals in Figure 6.4 results in the similar spike in the cardiac frequency range as well as a peak in the respiratory frequency range. This experiment was designed to compare the respiratory and heart rate measures generated from UMBS data against objective measures of these physiological signals. The creation of an automated algorithm to estimate physiological signals reliably was central to this.

A description of the technologies used to measure the physiological signals, the experimental set-up and the data set is given below.

### 6.2.1 Methods

This experiment recorded data over 8 subjects (4 male; 24-28 years old) lying in direct contact with the UMBS (placed on a solid surface) and subsequently upon a mattress with the UMBS underneath. Subjects were asked to refrain from any large movements

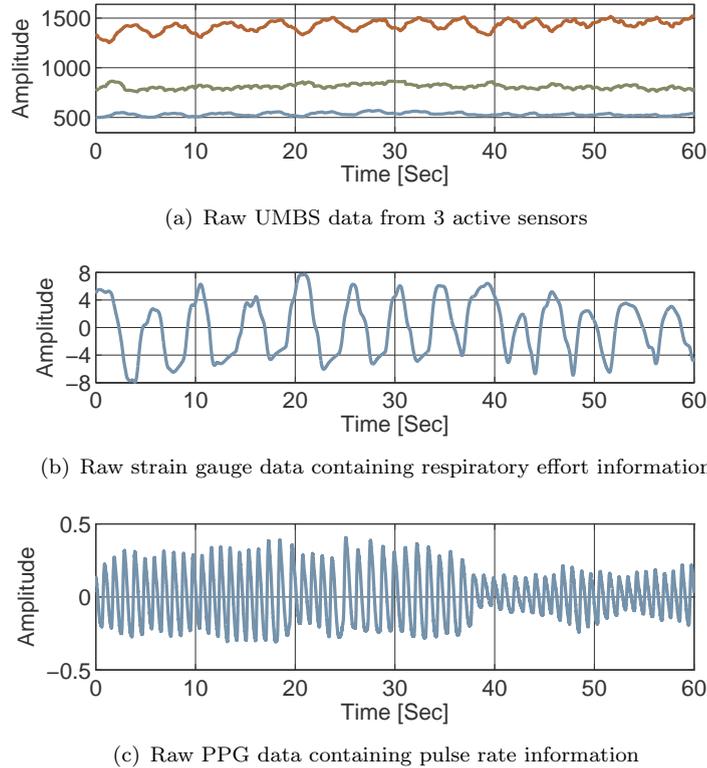


FIGURE 6.4: Concomitant raw UMBS, strain gauge and PPG data

and to stay in the prone position during the experiment. Approximately 30 minutes of data were recorded per subject when in direct contact with the sensor while 60 minutes of data were recorded per subject when the mattress separated the subject and the sensor. Respiratory effort (around the upper torso) was recorded using a strain gauge as shown in Figure 6.5(a). Heart rate was calculated by measuring PPG placed upon the ear lobe of the subject. The placement of the UMBS, PPG and strain gauge for this experiment is shown in Figure 6.5(b). In order to increase the range of the respiration and heart rates recorded, some subjects partook in cardiovascular exercise prior to the start of the experiment (at the subject's discretion). UMBS data was recorded on a computer and time stamped according to the internal Central Processing Unit (CPU) time. All strain gauge and PPG data were recorded on the same computer as the UMBS data using the RSP100C and the PPG100C modules on a Biopac MP150 System (Biopac Systems, Goleta, USA) and customised software. In order to remove noise on the respiratory module (RSP100C), the first low pass was set to allow frequencies below 1.0 Hz, and a second highpass filter was set to allow frequencies greater than 0.05Hz (as per the manufacturer's recommendations). The filter settings applied to the data from the PPG module (PPG100C) were as per the manufacturer's guidelines. The lowpass filter allowed signals less than 3Hz and the highpass filter allowing signals over 0.5Hz. This data was synced to the UMBS data manually after data collection.

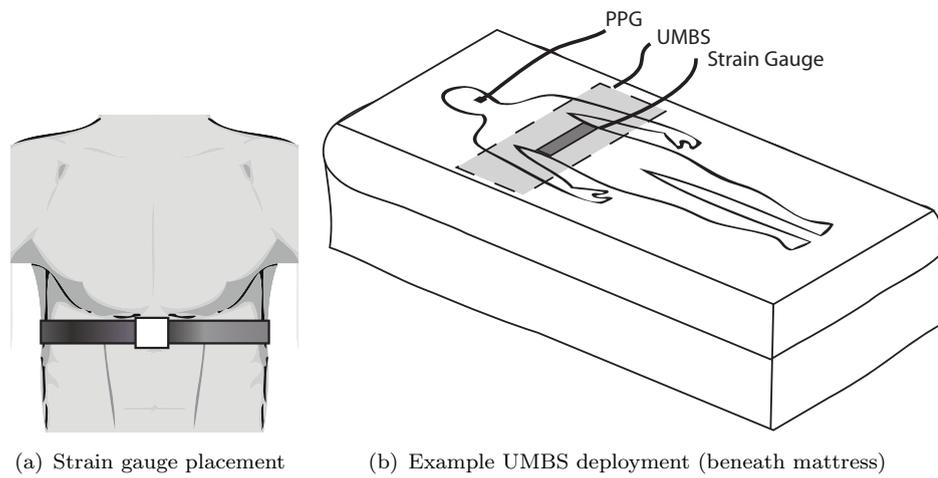


FIGURE 6.5: UMBS and strain gauge placement

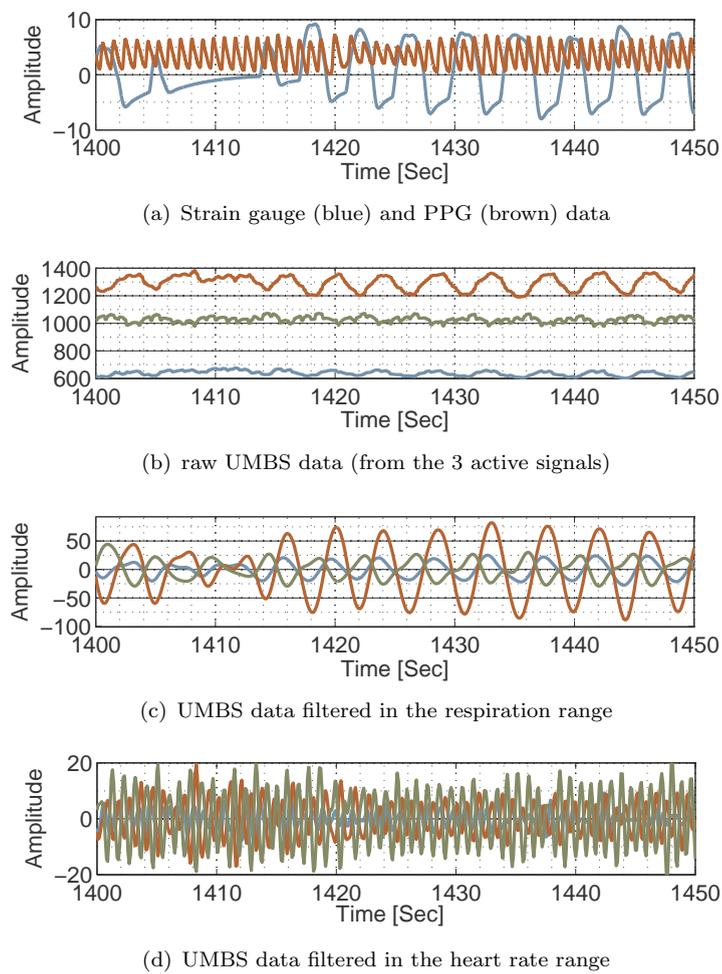


FIGURE 6.6: A sample of the concomitant data recorded from one subject. The respiratory effort is captured using the strain gauge as shown in (a).

Subject	Sex	Sensor Position	Good Epochs	All Epochs
1	F	Direct	5	6
2	M	Direct	3	6
3	M	Direct	2	6
4	M	Direct	6	6
5	F	Direct	3	6
6	M	Direct	5	6
3	M	Mattress	12	12
6	M	Mattress	10	12
4	M	Mattress	5	6
5	F	Mattress	12	11
7	F	Mattress	9	11
8	M	Mattress	10	11
Total	-	-	82	99

TABLE 6.4: Particulars of automated respiration estimation data set.

## 6.2.2 Data Preparation

A number of automated methods for capturing respiration rates were investigated. These included time and frequency based analyses. A total of four types of signals were prepared using the following techniques.

### 6.2.2.1 UMBS Data Preparation

The entire data set (over 495 minutes) was divided into 99 non-overlapping windows of duration five minutes. This epoch length was consistent with similar literature allowing for a large number of complete respiration cycles to occur within each window (Brink et al., 2006; Carlson et al., 1999; Shin et al., 2010). As such, this research estimated the respiration rate based on a 5 minute epoch. Due to large body movement artifacts in the UMBS data (including shifting postures, coughing, etc.) and also due to inadequate clean respiration profiles (inherent from using a strain gauge) 82 windows (82.83%; 410 minutes) were deemed acceptable for further analysis. The UMBS was in direct contact (DC) with the participant for 24 (24.24%; 120 minutes) of the 82 windows and underneath a standard spring mattress (UM) for 58 (58.58%; 290 minutes) of the 82 windows (as given in Table 6.4). A supine posture was maintained by the participants over all tests. Strain gauge data were reviewed manually by an observer and the location of all respiratory peaks were manually stored. Heart rate data (r-r interval) was detected using a standard algorithm (Pan and Tompkins, 1985).

**Signal Extraction** The UMBS tactels which contain no information (saturated tactels and tactels devoid of movement or pressure) were excluded from further analysis leaving *active signals* containing body movement information. Slight temporal phase shifts were found between the UMBS signals measuring respiration and heart rate (as can be seen in Figure 6.7). It was anticipated this was due to the spatial separation of the individual sensors within the device, of particular concern to higher frequency signals such as heart rate.

**Filtering** A median filter, of length 100 samples, was applied to the UMBS data in order to remove any baseline wander (*mFiltData*). Subsequently a high-pass 5th order Butterworth filter with a cut-off frequency of 0.1Hz was applied to remove slow deviations in pressure across the UMBS resultant from gradual postural shifts. This data was subsequently passed through multiple low-pass 5th order Butterworth filters with a varying cut-off frequency of 0.1, 0.2, . . . ,2.0 Hz producing multiple signals (*bpFiltData*). The optimal upper cut-off frequency was unknown and selected during analysis. It was anticipated that the frequency range for heart and respiration rate would be required to estimate each signal. Upper cut-off frequencies were selected for each of the DC and UM conditions.

**Data Fusion** The *mFiltData* and *bpFiltData* contain data from one or more tactels over the recording period. Each set of data were fused together using two techniques; (i) Principle Component Analysis (PCA) (Jolliffe, 2004); and (ii) an Adaptive Linear Combiner (ALC) (Widrow and Stearns, 1985). In PCA, the data is transformed into principal components with each component formed as a weighted mixture of the original signals so that they are successively orientated in the directions of maximum variance in the data. The first principal component is the direction of largest variance, the second component is in the direction of next largest variance (which is orthogonal to the first component) and so on. Since large body movement data segments have been omitted from the data the resulting signal variance can be attributed to respiration and heart rate components. During quiescent periods (ie. epochs devoid of postural shifts), the first principal component was found to contain strong correlations with respiration as can be seen in Figures 6.7 and 6.8(a). A high correlation ( $r = 0.8875$ ) can be seen over 500 seconds of example strain gauge data and the first principal component taken from the relevant UMBS data. In this instance, the first principal component captured 90.71% of the variance (and it was found to be over 80% of the variance for the other sets of data), while the 2nd and 3rd principal component captured 8.36% and 0.93% of the variance respectively. The variance captured by the first principal component is reduced when significant movement occurs. This is due to seemingly random changes

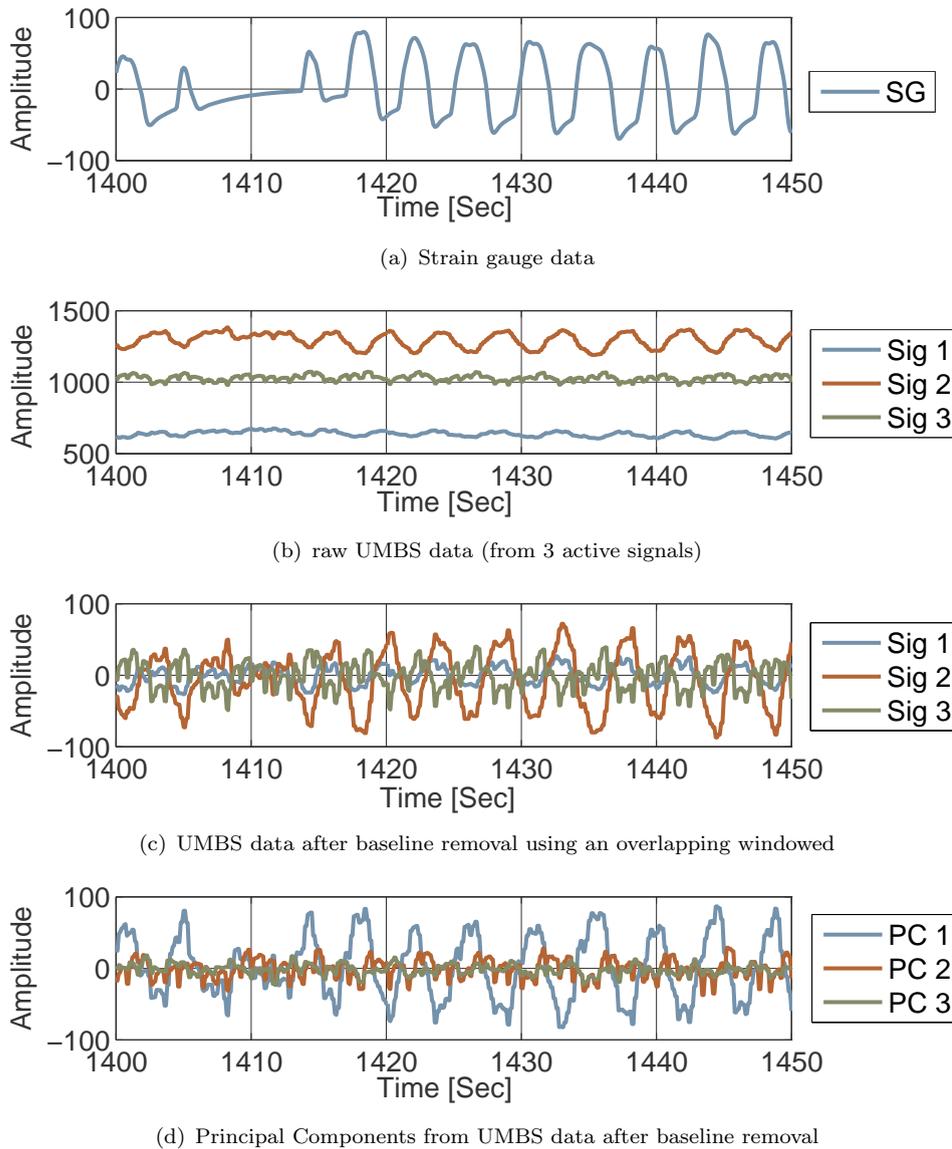
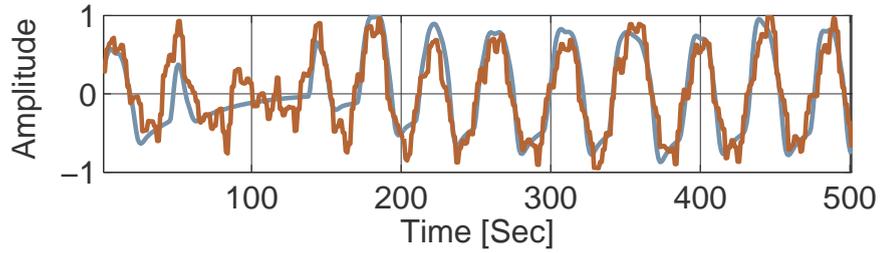


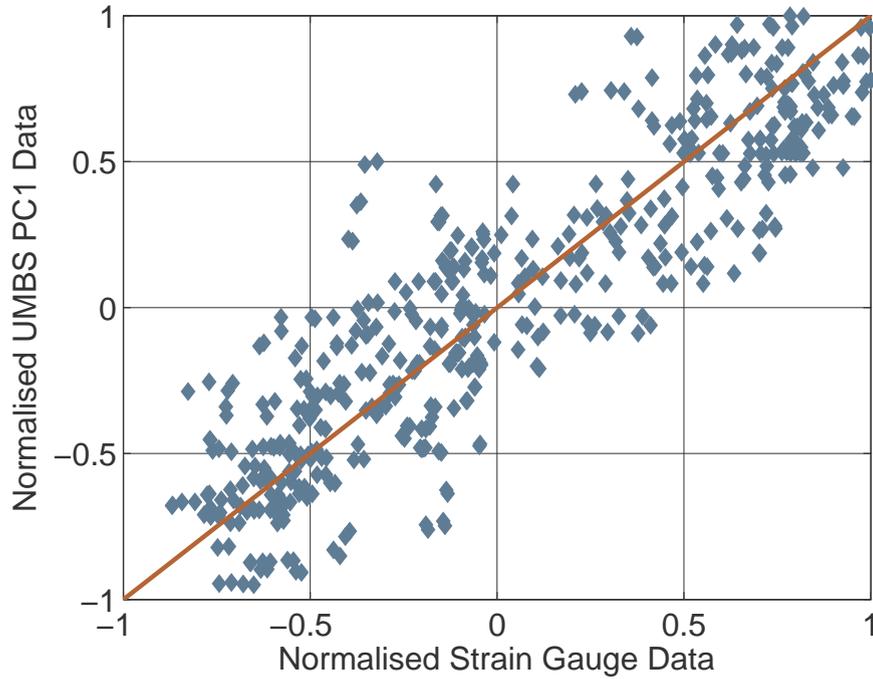
FIGURE 6.7: Concomitant data recorded from one subject. A correlation exists between the first principal component and the strain gauge (SG) data.

in pressure across the sensor. The amount of variance captured by the first principal component may be used as a measure of the amount of movement occurring during that epoch. This is further discussed in Section 6.3.

However, PCA based data fusion does not take into account any phase variation between signal sources arising from the spatial distribution of the tactels within the UMBS. If the phase variation resulted in a 90 degree phase shift between signal sources this would result in a complete loss of signal (as can be partially seen in Figure 6.6(b)). As a result a second method was devised using an ALC to fuse the original signals and 30 phase delayed versions of these signals together catering for any phase difference between the signals. The signal containing the largest variance was taken to be a reference signal. A



(a) Strain gauge and first principal component data



(b) Scatter plot of strain gauge (brown) and first principal component (brown) data

FIGURE 6.8: A correlation can be seen between the first principal component (PC1) and the strain gauge in both figures. The data was normalised to a range of  $[-1, +1]$ .

cancellation signal was then generated from the remaining signals (and 30 phase delayed versions of these signals; one sample apart). These were weighted such that a minimum phase error between the reference signal and the cancellation signal was found. The Least Mean Squares (LMS) algorithm was used to select the appropriate weighting for each input to minimise the Mean Squared Error (MSE) between the sum of the weighted inputs and the desired output during a training phase (after which the weights remained constant). This cancellation signal fused information from all sources (including phase delayed versions of the original signals), and also the reference signal, into one signal. The weights subsequently can be used as a means of mapping the spatial distribution of pressure on the UMBS, however this was not investigated further in this instance. The weights should be re-initialised after each large movement as spread of pressure across

the mat may have changed and, accordingly, new weights may be required.

The ALC technique was tested initially using simulated data. Three signals of various maximum amplitude oscillating at 0.2 Hz were generated and phase shifted (by approximately ten degrees) relative to one other. One signal (randomly selected) was chosen to be a reference (or desired) signal. Thirty phase delayed versions of the remaining signals (1 sample apart) were generated and fed as inputs into the system. The weighted sum of the inputs (iteratively optimised over time) successfully followed the desired signal and the frequency spectrum of the output, as expected, contains a peak at 0.2 Hz (as shown in Figure 6.9). Experimental results with UMBS data using the signal with the highest variance as the desired signal, and subsequently with the strain gauge data as the desired signal can be seen in Figure 6.10. Should a weighted version of the UMBS data be able to completely compensate for the strain gauge data, no error should exist between the output and desired signal. However in practice, this is not often the case when real data is concerned (and may be possibly due to a changing spatial orientation brought about through breathing). Results using a real data set can be seen in Figure 6.10(e). The ALC method serves to minimise the error and the output can be seen to follow the strain gauge data suggesting that the UMBS data (and time shifted versions of this data) can capture respiratory data effectively, as shown in Figure 6.10(d). A low error is generally seen however an increase occurs and diminishes at 100 Seconds. A small movement may be seen in the corresponding UMBS data.

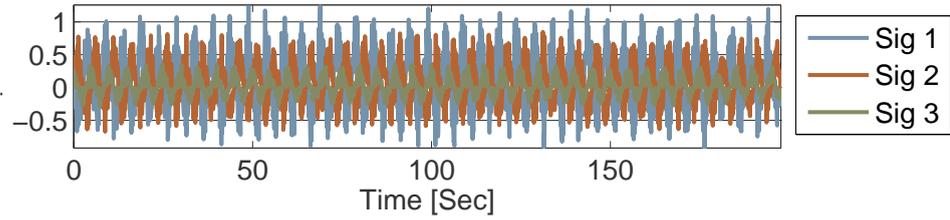
The ALC was applied to both the *mFiltData* and *bpFiltData* data sets generating a total of four signals for later estimation of respiration and heart rate:

- PCA *mFiltData*
- PCA *bpFiltData*
- ALC *mFiltData*
- ALC *bpFiltData*

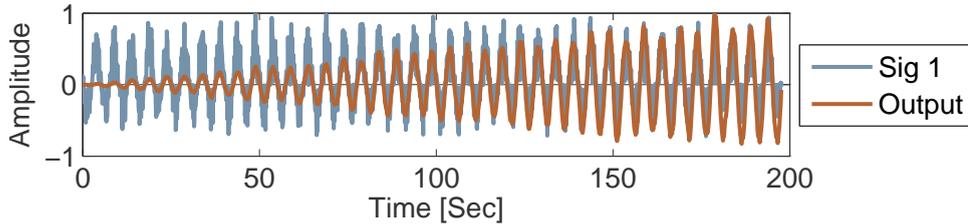
## 6.2.3 Automated Physiological Signal Estimation

### 6.2.3.1 Time-based Techniques

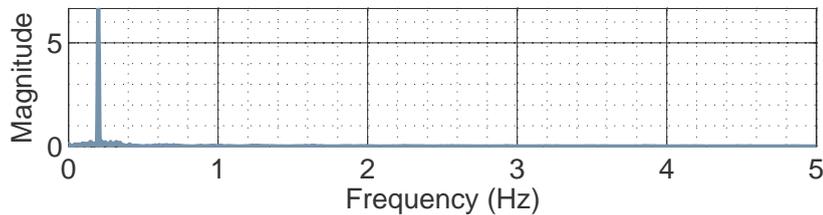
**Peak Detection Algorithm** Two peak detection methods were developed to detect the breathing profiles on both the strain gauge and UMBS-derived data. In the first method, (*peakDet*), all local peaks were found by searching for samples occurring between



(a) Simulated UMBS data; Oscillating at 0.2Hz, out of phase and of different amplitudes



(b) Output from ALC, shown following the 'desired' first UMBS signal

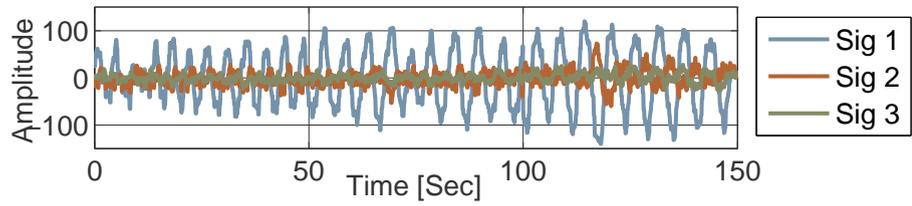


(c) Frequency spectrum of output signal; Peak at 0.2 Hz

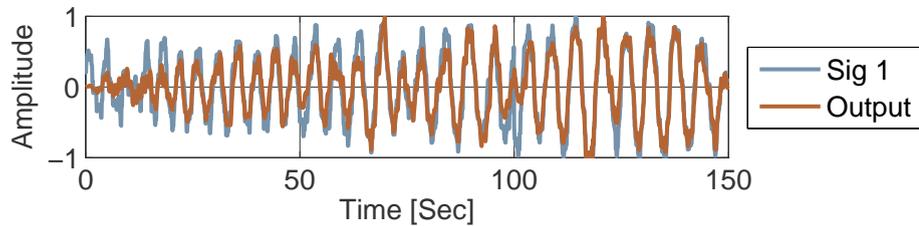
FIGURE 6.9: Simulated UMBS data showing the capacity of the ALC to compensate for the phase variance between UMBS signals. The frequency peak in the output signal is at the expected value of 0.2 Hz.

upward and downward trends which represents the point between the end of inhalation and the start of exhalation. This was achieved through differentiating the signal and finding where the maxima occur. A further requisite involved ensuring that the height of each peak occurred above the mean signal value within that epoch. A fixed (*peakDet fixed*) and dynamic (*peakDet dyn*) minimum refractory period, initially chosen to be two seconds, between each respiration cycle was chosen and this was optimised during testing. The initial window of analysis contained ten peaks. The next peak occurred after the initial refractory period. This refractory period was updated after each new peak was found. The updated minimum refractory length was taken to be half of the median length of time between the previous ten peaks.

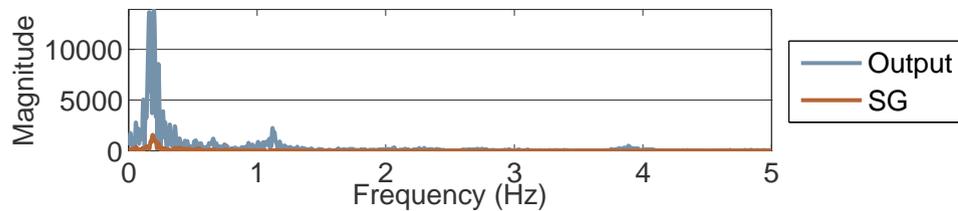
A comparison (using absolute mean difference and standard deviation between the number of manually observed peaks and number of automatically detected peaks) of the dynamic updating method against the fixed refractory length method for a reduced set of 24 five minute windows of data can be seen in Figure 6.11. The reference method was manually scored. The *peakDet* peak detection technique was also applied to the strain



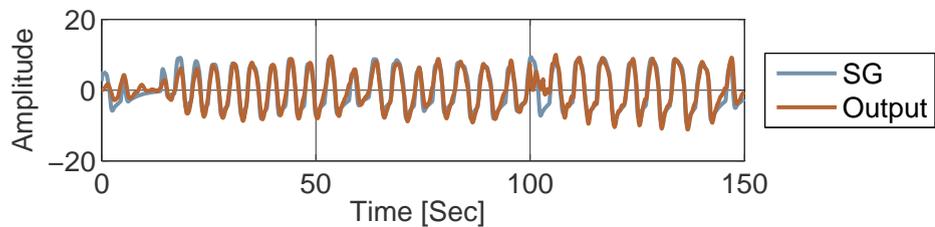
(a) Actual UMBS data; Oscillating at approximately 0.2Hz



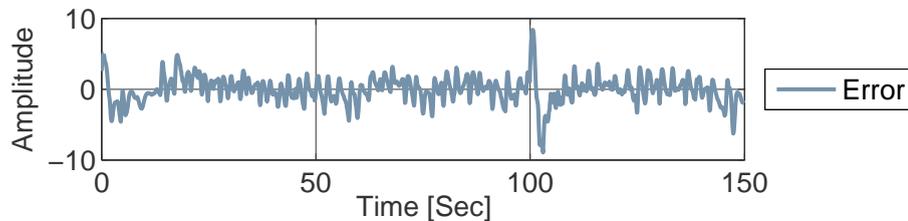
(b) Output from ALC, shown following the 'desired' first UMBS signal



(c) Frequency spectrum of output signal; three peaks are present at 0.16, a slightly larger peak at 0.2 Hz and a peak within the cardiac range at 1.15 Hz.



(d) Strain gauge and ALC output data



(e) Error between the strain gauge and output signals

FIGURE 6.10: Actual UMBS data showing the application of the ALC in compensating for the phase variance between UMBS signals. In Figures (a) and (b), the UMBS signal with the highest variance is used as the 'desired' signal. The output is compared to the actual strain gauge data in Figure (c). In Figures (d) and (e), the actual strain gauge data is set as the desired signal.

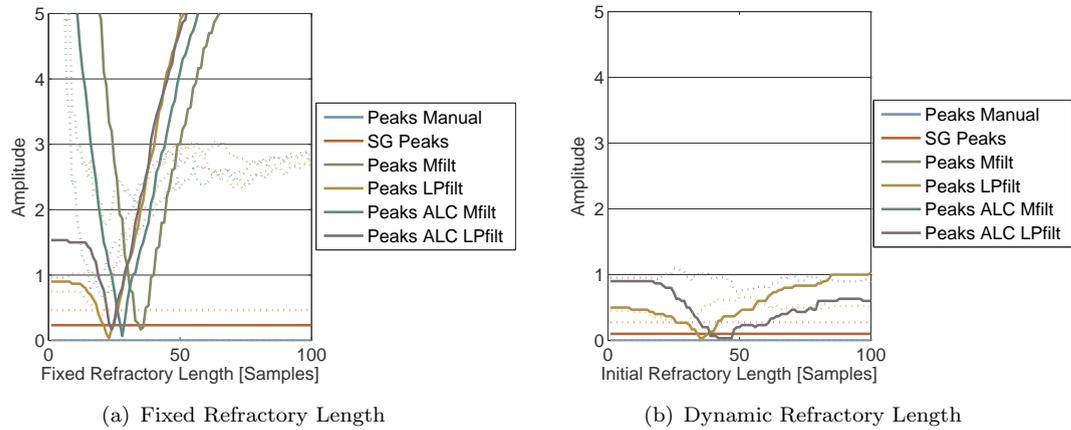


FIGURE 6.11: The absolute mean difference (solid line) and standard deviation (dotted lines) between the respiratory peaks manually observed and automatically detected using the various techniques.

gauge signal and resulted in a low mean difference and low standard deviation when compared to the manually scored results. The dynamic refractory length was shown to correct for an ill-fitting initial window length, however it did not perform well for the ALC data.

A second method was used which sequentially found peaks which were surrounded by lower values on both sides, (*PkDet*) (Billauer, 2012). The minimum height difference between each maxima and the surrounding minima on each side was optimised during analysis.

A time based analysis for estimating heart rate was not applied as the cardiac cycle was not visually discernible as reported in the previous section.

### 6.2.3.2 Frequency Based Analysis

Frequency analysis was applied to all of the UMBS-derived data in order to find frequency peaks relating to respiration and heart rates. Wavelet and windowed Fourier analyses were applied to the active signals. The Lomb-Scargle method (Press and Rybicki, 1989) was applied to the unevenly sampled raw UMBS data. ICA was also applied in an effort to fuse the active signals and extract separate noise and physiological signals. None of these methods outperformed the time based analyses methods when compared to the gold standard measures of respiration during initial analyses.

The frequency-based detection of heart rate was inconsistent and its accuracy was highly variable. As such, heart rate detection was not included in any further analyses.

## 6.2.4 Results

### 6.2.4.1 Respiration Estimation

The mean difference in the number of breaths, for each window of data and for all internal configuration parameters (lpCutOff, refractory length and the reference length), between the derived respiration rate and the gold standard was calculated. The average and standard deviation of this mean difference over all windows, for all internal parameters were subsequently found. The optimal internal parameters were chosen by minimising a performance metric (the sum of the average and standard deviation of the absolute mean difference over all windows of data). These optimal parameters were calculated for each signal (*PCA mFiltData*, *PCA bpFiltData*, *ALC mFiltData*, and *ALC bpFiltData*), for each time-based respiratory peak detection method (*PkDet*, *peakDet fixed*, and *peakDet dynamic*), and for each sensor position (in direct contact with the subject, and under the mattress). A summary of the optimal metrics over all scenarios can be found in Table 6.5. The refractory length remained fixed for the *peakDet fixed* peak detection method, while it was internally modified during algorithm execution for the *peakDet dynamic* peak detection method. The optimal results found are as follows. Over the DC data set, a mean difference of 0.5 (SD of  $\pm 1.85$ ) Breaths per Five Minutes (BrP5M) and a mean percentage error (MPE) of 0.56 % (SD of  $\pm 2.47\%$ ) was found when the participant was in direct contact with the sensor. This occurred using the *PkDet* algorithm applied to the first principal component of the *bpFiltData*, an upper cut-off frequency of 0.4Hz and a refractory length of 308 samples. When the sensor was placed underneath the mattress, the mean difference was -0.12 (SD of  $\pm 2.26$ ) BrP5M and an MPE of -0.16 % (SD of  $\pm 3.12\%$ ). This occurred using the *PeakDet* algorithm applied to the first principal component of the *bpFiltData*, an upper cut-off frequency of 0.8Hz and a refractory length of 19 samples.

### 6.2.4.2 Heart Rate Estimation

Time based analysis was not applied for heart rate estimation from the UMBS data as a visual examination of concomitant UMBS and PPG data could not determine any time-based correlation. As a result, only a frequency based analysis was applied to each window of data. A Fourier transform was applied to the raw UMBS data and the frequency range 0.7-1.8 Hz was investigated for any cardiac activity. A distinct peak was found in the cardiac range for only some of the windows of data. As such, the ratio of the amplitude of this peak to the mean amplitude over the entire cardiac range was used as a quality metric to assess the reliability of using frequency content to report an estimate of heart rate (see Figure 6.12). The was compared to the absolute difference

TABLE 6.5: Accuracy of the respiration rate estimation methods when the UMBS is placed underneath the mattress (UM) and in direct contact with the subject (DC).

Sensor Pos.	Resp. Count Method	Signal	LPCutOff Freq. [Hz]	Ref Len [Samples]	Mean Diff. (BrP5M)	MPE (%)	Perf. Metric
DC	PkDet	PCA <i>mFiltData</i>	-	1000	24.61 ± 22.01	34.78 ± 39.74	46.62
		PCA <i>bpFiltData</i>	0.4	308	0.50 ± 1.855	0.56 ± 2.40	2.33
		ALC <i>mFiltData</i>	-	67	- 0.11 ± 10.51	0.20 ± 12.65	10.51
		ALC <i>bpFiltData</i>	0.4	28	0.38 ± 8.00	0.86 ± 9.82	8.39
	peakDet	PCA <i>mFiltData</i>	-	28	1 ± 9.71	3.14 ± 12.87	10.71
		PCA <i>bpFiltData</i>	0.5	24	-0.38 ± 2.61	-0.48 ± 3.57	3.00
		ALC <i>mFiltData</i>	-	24	0.44 ± 3.60	0.43 ± 4.89	4.045
		ALC <i>bpFiltData</i>	0.6	24	0.16 ± 3.50	0.11 ± 4.72	3.66
	PeakDet	PCA <i>mFiltData</i>	-	313	136.05 ± 147.25	161.42 ± 180.13	283.30
		PCA <i>bpFiltData</i>	0.5	20	-1.83 ± 3.83	-2.04 ± 4.60	5.67
		ALC <i>mFiltData</i>	-	29	0.22 ± 13.11	1.39 ± 17.15	13.33
		ALC <i>bpFiltData</i>	0.5	17	-1.44 ± 3.50	-1.65 ± 4.47	4.94
PkDet	PCA <i>mFiltData</i>	-	783	-0.5 ± 20.07	-1.14 ± 23.53	20.57	
	PCA <i>bpFiltData</i>	1	70	0.13 ± 2.74	0.3136 ± 4.06	2.88	
	ALC <i>mFiltData</i>	-	1	-44.13 ± 31.10	-53.23 ± 36.41	75.24	
	ALC <i>bpFiltData</i>	0.7	1	-43.91 ± 31.11	-52.89 ± 36.61	75.02	
UM	PeakDet	PCA <i>mFiltData</i>	-	31	-0.36 ± 10.04	0.72 ± 13.27	10.40
	Fixed	PCA <i>bpFiltData</i>	0.8	19	-0.12 ± 2.25	-0.16 ± 3.12	2.37
		ALC <i>mFiltData</i>	-	21	-0.50 ± 2.33	-0.72 ± 3.34	2.83
		ALC <i>bpFiltData</i>	1.7	21	-0.22 ± 2.67	-0.28 ± 3.81	2.89
PeakDet	PCA <i>mFiltData</i>	-	170	189.24 ± 126.90	243.54 ± 172.00	316.14	
	Dyn	PCA <i>bpFiltData</i>	0.8	1,2,3,4,5	-0.50 ± 2.42	-0.65 ± 3.28	2.92
		ALC <i>mFiltData</i>	-	4	-1.13 ± 2.69	-1.51 ± 3.70	3.83
		ALC <i>bpFiltData</i>	1.8	1,2,3	-0.72 ± 3.17	-0.88 ± 4.56	3.90

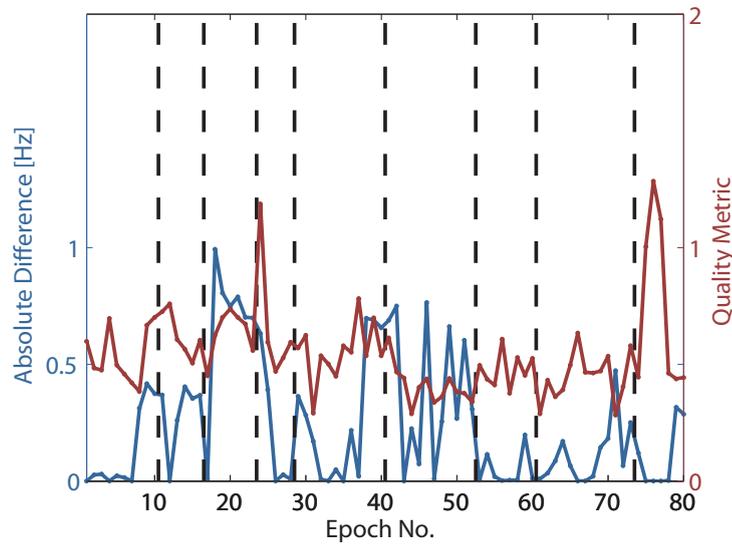


FIGURE 6.12: Absolute difference between the frequency peak of actual heart rate (using PPG) and the estimated heart rate (dotted lines separate data belonging to different subjects). The corresponding derived quality metric (peak frequency height/mean frequency height) is also shown.

(in Hz) against a gold standard measurement of heart rate. This was calculated over all windows of data.

### 6.2.5 Discussion

The UMBS performs well compared to other similar technologies as given in Table 6.6. For the under mattress condition, the mean difference between this system and the gold standard was  $-0.12$  (SD of  $\pm 2.26$ ) BrP5M and an Mean Percentage Error (MPE) of  $-0.16\%$  (SD of  $\pm 3.12\%$ ). A scatter plot of the optimal respiration count estimate for both conditions against the actual count can be seen in Figure 6.13. Shin et al. (2010) investigated direct contact respiration estimation using an air mattress, composed of separate air cells with a balancing tube based pressure sensor, and reported a mean difference of  $0.5 \pm 0.63$  Breaths per Minutes (BrPM) (equivalent to  $2.5 \pm 3.15$  BrP5M) and an MPE of  $2.85\%$ . Estimates of heart beats, body movement, snoring events and apnoeic episodes were also produced with high sensitivity and high positive prediction values. Carlson et al. (1999) developed a non-invasive respiratory monitoring system (NIRMS) for sleeping subjects (in direct contact with the sensor) which monitors pressure changes on an air mattress using a pressure transducer. They reported on eleven subjects over three sleeping postures each (supine, prone and side) for a duration of 5 minutes per condition (33 data sets in total). A mean of absolute differences between the estimated respirations and the actual respirations was  $0.79 \pm 0.6$  BrP5M and a

mean error of 1.38% were reported. Brink et al. (2006) investigated the use of high resolution force sensors (also known as load cells) placed underneath the bed posts for non-contact measurement of respiration, heart rate and body movements during sleep. Data was captured from fourteen subjects (seven male) over a five minute duration in a prone, supine or side position. This produced a mean absolute difference of 0.03 (SD of  $\pm 0.33$ ) BrPM (equivalent to  $0.15 \pm 1.65$  BrP5M) between the estimated and actual breaths and a mean error of 1.2%. Zhu et al. (2006) developed an under-pillow pressure based respiration and heart rate estimation sensor. Data was recorded from thirteen recumbent subjects (8 male) for an average of approximately 115 minutes. A mean absolute difference of 0.04 (SD of  $\pm 0.06$ ) BrPM (equivalent to  $0.24 \pm 0.34$  BrP5M) were reported, resulting in a mean error of 0.38%. However the accuracy of this under-pillow technique might vary over different sleeping positions (supine, prone or side-lying) as the carotid pulse might not be present in all postures. Pressure sensitive pads placed on top of the mattress, in contact with the participant through bed sheets, was put forward as a potential technology for sleep monitoring (Mack et al., 2009a). An algorithm was developed, from data collected from 40 subjects, which detected the ballistocardiogram and accurately reported the heart rate to within 2.72 beats per minute and respiration to within 2.1 breaths per minute. The use of 4 strain gauges, sampling at 128Hz, fitted to a segment of the bed frame was also investigated (Brser et al., 2011). An automated, dynamic technique was developed to extract the heart rate from data collected from 16 individuals. A beat to beat interval error of 1.79% was reported and the algorithm was found to work for over 95% of the data. Load cells, placed underneath the bed posts and sampled at 200Hz, were further investigated and found to extract the ballistocardiogram (Choi and Kim, 2009). Heart rate variability analysis was then applied and the derived features were found to be able to discriminate deep sleep (stage 3 and stage 4 sleep) from the other stages of sleep with a 92.5% accuracy. An under-mattress pressure sensitive capacitive foil electrode grid sampled at 50Hz was found to extract the various heart rate features from the ballistocardiogram (Kortelainen et al., 2010). The combination of a movement detection system and another system, based on the application of HMM on heart beat interval data to discriminate between REM and Non-REM sleep, was found to report an accuracy of 79%.

During the initial analysis, it was discovered that the strain gauge derived respiration data contained some invalid values. This was attributed to issues with the design of the sensor. The strain gauge measures horizontal torsion of an elasticated band at one position along the device, rather than measuring strain throughout the band, as is the case in a Respiratory Inductance Band (RIB) which measures changes in the total circumference of the band. The RIB measures strain (and correspondingly respiration) by measuring the inductance in a wire woven in a sine wave pattern throughout the

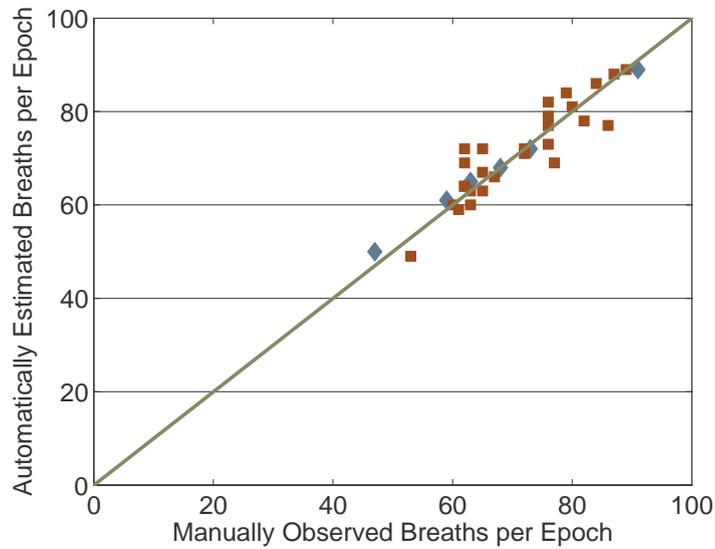


FIGURE 6.13: Comparison of the ability of the UMBS to correctly estimate respiration using direct contact (grey diamonds) and through a mattress (brown squares) against manual respiration counts using a strain gauge over a 5 minute epoch. The diagonal line represents perfect estimation.

TABLE 6.6: Performance of the UMBS compared to similar ambient sensors. Mean Absolute Difference (MAD), Mean Difference (MD), Mean Error (ME), Mean Percentage Error (MPE), Breaths Per Minute (BrPM)

Device	Performance	Cohort	Data Set Details
Load Cells (Brink et al., 2006) <sup>1</sup>	MAD $0.15 \pm 1.65$ BrP5M ME 1.2%	14(7M)	5 min bins
NIRMS (Carlson et al., 1999)	MAD $0.79 \pm 0.6$ BrP5M ME 1.38%	11	5 min bins
Air Cells (Shin et al., 2010) <sup>1</sup>	MD $2.5 \pm 3.15$ BrP5M MPE 2.85%	13	
Pillow (Zhu et al., 2006) <sup>1</sup>	MAD $0.24 \pm 0.34$ BrP5M ME 0.38%	13(8M)	115 min bins
UMBS	MD $-0.12 \pm 2.26$ BrP5M MPE $-0.16 \pm 3.12$ %	8(4M)	5 min bins

<sup>1</sup> Results originally quoted on a per minute basis.

band. Another source of error was found to be as a result of ill-fitted strain gauges; if the circumference of the belt is too large, the point of complete exhalation cannot be measured accurately as seen in Figure 6.6(a). The minima of the strain gauge data does not exhibit a clear point marking the exact location of complete exhalation. A slight upward trend after each minima can also be seen. This is due to the application of a filter used to remove a Direct Current (DC) component. It should also be noted that no clear inhalation peak can be seen between 1405 and 1415 seconds. This absence of an inhalation signal the strain gauge data is also reflected in the concomitant UMBS data. Data sets containing these errors were excluded from further analysis.

A slight temporal phase shift was found when measuring respiration from adjacent sensors on the UMBS as a result of their spatial distribution. The effect of this was considered minimal when investigating respiration as the wavelength concerned was much larger than the temporal phase shift, although it was suggested that this would be more of a concern for higher frequency signals such as heart rate. Methods of signal fusion using cross correlation were initially investigated. However, these was very sensitive to minor body movement (occurring during brief periods of the data) not relating to respiration or heart rate. As such, an ALC was applied in order to investigate whether dynamic phase correction would be more accurate than fixed phase correction offered when using cross correlation. A modified ALC was applied to resolve this problem, however it did not result in any increased accuracy. This may be due to the artificial conditions (not the subject's normal bedroom, etc) or resultant from the subject being asked to remain as still as possible (yet comfortably) during the experiment.

Some subjects partook in cardiovascular exercise (approximately twenty minutes running) prior to the beginning of the experiment, this increased the cardiorespiratory range of the data.

Frequency-based methods (Fourier Analysis) of extracting heart rate and respiration rate were also investigated, however inconsistent results were reported and they were outperformed by the time-based peak detection algorithms. Wavelet analysis showed a large peak occurring in the respiratory range, however further investigation found that while it was indicative of respiration, more accurate estimates were provided by the time-based methods. Time and frequency based methods for detecting heart rate did not provide reliable results.

### 6.3 Experiment 3 - Movement Detection Capacity

This section describes the validation of the UMBS in detecting nocturnal movement. A video based gold standard movement index is used to quantify the true level of motion over each recording. The work investigates the development of an algorithm which accurately and reliably discriminates movement from non-movement epochs. A comparison against other commonly used modalities of measuring movement, namely a PIR and a wrist actigraph, during sleep is also made.

#### 6.3.1 Data Collection and Processing

Data were recorded from four participants (one male) over eight sessions in a standard bedroom. Five sessions were recorded from the male, while a data set was recorded from each of the three females. The eight data sets had a total duration of 1,491 minutes. The UMBS (v3) was placed underneath the mattress of the bed and the data collection system was placed adjacent to the bed. The participants were instructed to lie naturally as they would when going asleep. The PIR and video camera were placed directly over the bed at a height of 1.7 metres. No blankets covered the subject ensuring the detection of any body movement by the video based gold standard used. UMBS data were compared to video gold standard, wrist actigraphy motion metrics, and a motion metric derived from the PIR in 60 second non-overlapping windows.

##### 6.3.1.1 UMBS

**UMBS Motion Metric Estimation Algorithm** A median filter was used to remove outliers in the UMBS data. All data was split into windows of 60 seconds duration. The standard deviation of the signal over this window for each  $j$  tactel ( $\sigma_j^k$  as defined in Equation 5.7), and the sum of the difference between signal values over the window for each  $j$  tactel ( $\Delta_j^k$  as defined in Equation 5.10) were used to generate UMBS motion metrics for each 60 second window of data,  $k$ . Specifically, the sum and standard deviation of these two metrics over all signals were used to generate the following metrics:

$$UMBS\ 1^k = \sum_{j=1}^{24} \sigma_j^k \quad (6.1)$$

$$UMBS\ 2^k = \sum_{j=1}^{24} \Delta_j^k \quad (6.2)$$

$$UMBS\ 3^k = \sqrt{\frac{1}{24} \sum_{j=1}^{24} (\sigma_j^k - \bar{\sigma}^k)^2} \quad (6.3)$$

$$UMBS\ 4^k = \sqrt{\frac{1}{24} \sum_{j=1}^{24} (\Delta_j^k - \bar{\Delta}^k)^2} \quad (6.4)$$

where  $\bar{\Delta}^k$  and  $\bar{\sigma}^k$  are the mean  $\Delta_j^k$  and  $\sigma_j^k$  values over all  $j$  tactels for epoch  $k$ .

### 6.3.1.2 Video

Video based motion detection, using a standard web-cam (*Trust Spacec@m 360*) and motion detection algorithm described below, was used as the reference gold standard motion detection technology (Camurri et al., 2003). Lights were kept on during data collection as required by the motion detection algorithm.

**BioMobius/EyesWeb Video Motion Metric Algorithm** A motion metric was extracted using in-built blocks within the *TRIL BioMobius* platform ([www.biomobius.org](http://www.biomobius.org)) developed as an extension to the Eyesweb (<http://www.infomus.org/EywIndex.html>) data collection and analysis platform. The motion detection algorithm compares the apparent motion between subsequent images within a video. Firstly, some preprocessing steps are performed, the video is converted into black and white and a median filter is used to remove white noise. The motion detection algorithm uses a reference frame (the first frame captured) of the background as a method of detecting the introduction of a foreign object (such as a person) into subsequent frames. This results in the extraction of a silhouette of the new object in the video such as a person. Variations in the shape of this silhouette over a user defined number of previous frames (4 for this experiment; the default value) is found by summing the area difference between the area of the silhouette in the current frame and the silhouette area in each of the previous frames. The resulting 'silhouette motion image' is normalised by the total area of the silhouette in the current image. This results in a 'quantity of motion' metric which avoids scaling issues due to the object being different distances away from the camera. This also compensates for different apparent sizes of objects. It should be noted that a smoothing effect is introduced when multiple previous frames are used in the calculation of the 'quantity of motion' metric. Further details may be found in Camurri et al. (2003). A time-stamped motion metric is produced as each frame is processed. A histogram of the Video Motion Metric data can be seen in Figure 6.14. A manual investigation of the data revealed

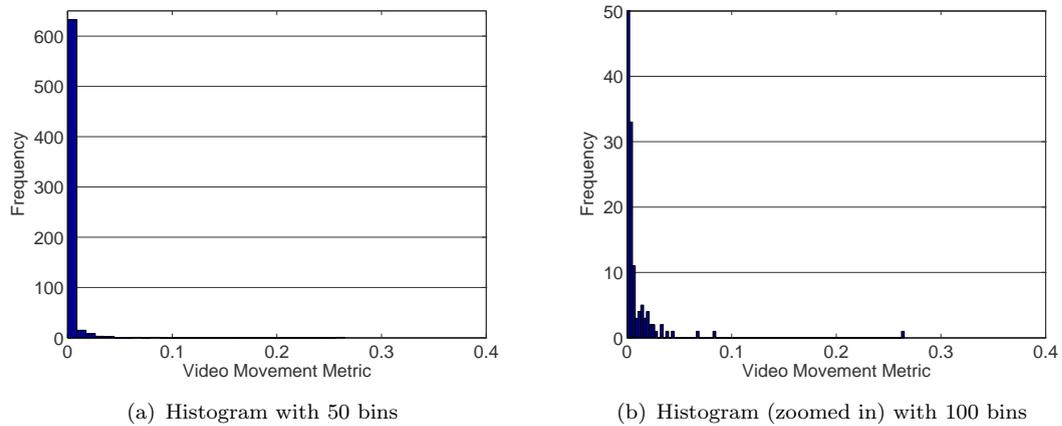


FIGURE 6.14: Video motion metric histogram for the entire data set. The high occurrence of low motion metric values represent illustrates that subjects remained still over extended periods of time.

that using a threshold of 0.01 detected all major and minor body movements without generating false alarms due to random noise.

### 6.3.1.3 PIR SHIMMER

Passive Infra-red (PIR) motion detection is widely used for detecting the presence of an intruder and is commonly seen in house alarms. The lens in the PIR unit focuses environmental infrared radiation onto a pyroelectric sensor. It is the change in the amount of infrared emitted in the environment which triggers an event which is in turn interpreted as motion. Thus, sudden changes across any of the cells in the pyroelectric sensor are interpreted as motion. The circuit is designed using a differential amplifier so that universal changes in environmental infrared levels are not misinterpreted as motion. The PIR device used in this experiment was attached to the Intel Digital Health Group's Sensing Health with Intelligence, Modularity, Mobility and Experimental Reusability (*SHIMMER*) platform (Burns et al., 2010). Data was streamed from this device using a Bluetooth connection and collected using the *TRIL BioMobi* platform in real time. The resulting non-negative sampling error (the delay between the detection of motion and the time-stamping of its occurrence can not be negative) was considered to be negligible when compared to the analysed epoch length of 60 seconds. The PIR module attached to the SHIMMER board was a *Panasonic MP Motion Sensor (AMN 1, 2, 4) Slight Motion Detector*. This had a maximum rated detection distance of 2m, a detection range of 91 degrees both horizontal and vertical, 104 detection zones and a movement speed detection of 0.3 m/s-1.0 m/s.

#### 6.3.1.4 Wrist Actigraphy

In this study a reduced data set (four subjects; 1 male) of concomitant actigraphy and video based motion data (267 minutes) was recorded for direct comparison. Wrist actigraphy motion counts were recorded in one minute epochs and compared to the video motion metric. The movement data between the video motion metric and wrist actigraph were found to be significantly correlated ( $r^2 = 0.3795, p < 0.0001$ ).

#### 6.3.2 Results

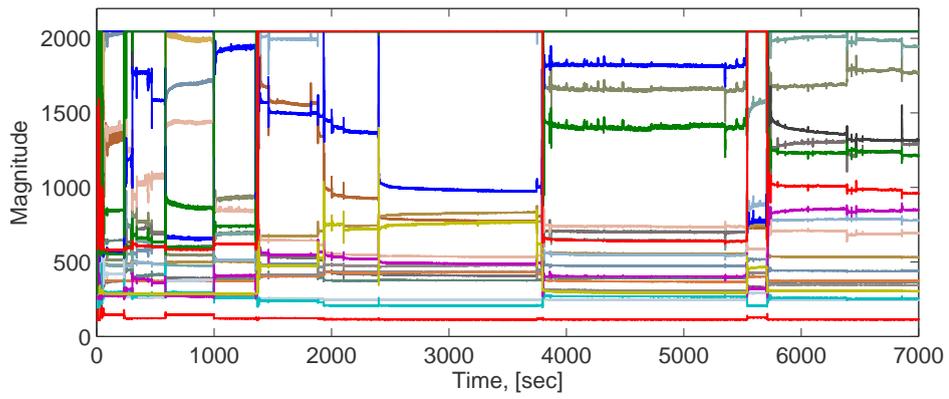
Concomitant data collected from the UMBS (the UMBS 2 motion metric derived from the UMBS data), the wrist actigraph, and the video motion metric for 7,000 seconds are shown in Figure 6.15. A direct comparison between normalised values of two of the UMBS motion metrics, the PIR motion metrics and the video motion metric over 45 sixty second epochs is presented in Figure 6.16.

Specificity and sensitivity values were calculated and a ROC was generated in order to assess the accuracy of the UMBS motion metrics in detecting motion (as defined by the video-based motion metric). The use of specificity and sensitivity caters for any bias in the data set toward either motion or non-motion. The best threshold was found using the greatest distance between the origin and a point on the ROC curve (defined as the ROC distance). This was defined to be the point at which the sensitivity and specificity were closest to their optimal values when given equal significance (Equation 6.5) and was reported as a percentage of the maximal distance ( $\sqrt{2}$ ). An example ROC curve is shown in Figure 6.17 and the optimal point on the curve (furthest from sensitivity and specificity values of 0) is clearly marked.

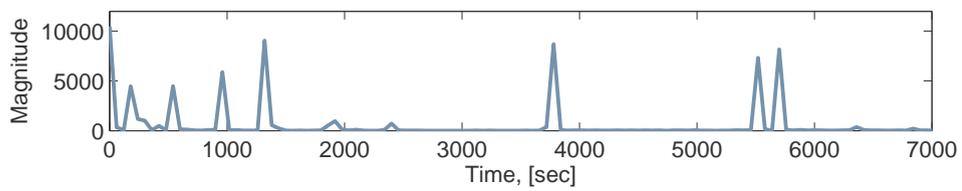
$$ROCDistance = \sqrt{\frac{Sensitivity^2 + Specificity^2}{2}} \quad (6.5)$$

##### 6.3.2.1 Movement Binary Classification Test

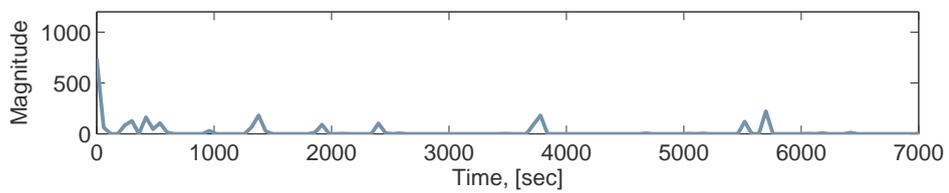
A BCT was performed directly on each of the four UMBS metrics against the true indicator of motion (the video-derived motion metric) in order to select the optimal threshold separately for each UMBS metric. The accuracy, specificity, sensitivity and ROC distance for the optimal ROC point for each of the four UMBS metrics defined above are given in Table 6.7.



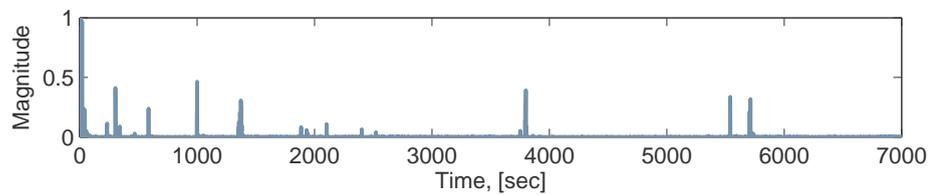
(a) Sample UMBS data showing pressure deviations resultant from large body movement.



(b) UMBS 2 - UMBS Motion Metric



(c) Wrist Actigraphy



(d) Video Motion Metric

FIGURE 6.15: UMBS data and concomitant video and wrist actigraphy data over an extended period

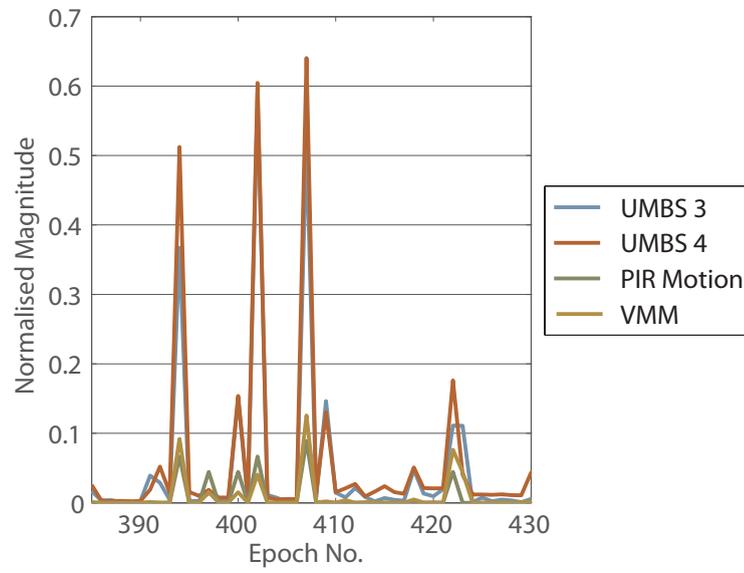


FIGURE 6.16: Direct comparison of 2 derived UMBS motion metrics, the PIR motion metric and the video motion metric over 45 sixty second epochs.

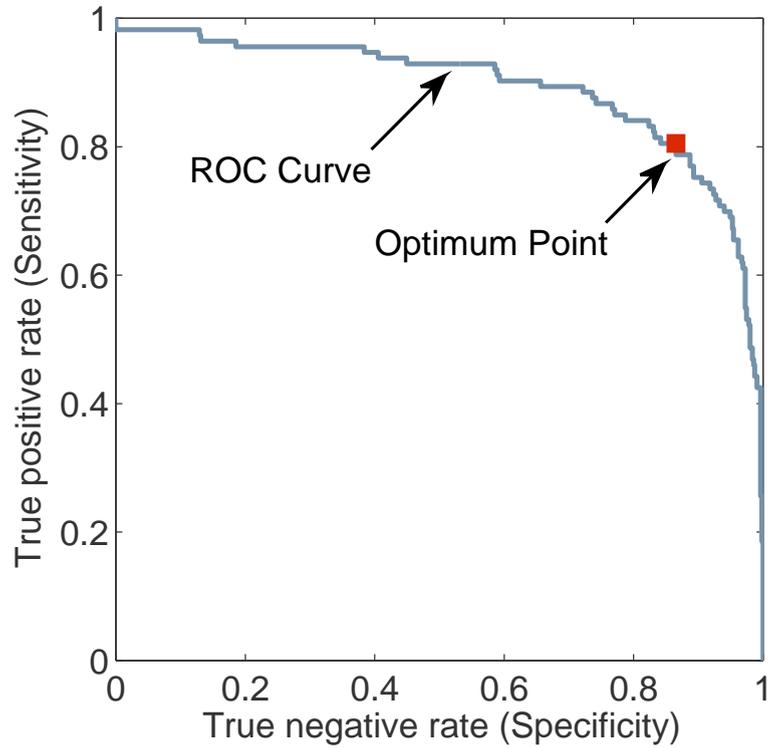


FIGURE 6.17: Example Mirrored ROC Curve (sensitivity vs. specificity)

TABLE 6.7: Results for the binary classification test at the optimal UMBS threshold (as defined by the ROC curve) applied to each UMBS derived motion metric and PIR defined motion against the gold standard video based movement metric.

	Acc	Sens %	Spec %	ROC dist %	UMBS Thresh
UMBS 1	91.48	91.03	91.51	91.28	12.06
UMBS 2	91.48	92.31	91.44	91.88	163.44
UMBS 3	90.07	92.31	89.95	91.15	18.26
UMBS 4	89.74	55.13	91.65	75.63	64.15
PIR	93.36	83.33	93.91	88.79	N/A

### 6.3.2.2 Movement Classification Using Multiple Inputs

Advanced signal processing methods were employed to fuse the data from the four UMBS metrics to maximise motion detection accuracy and reliability. This included linear and nonlinear classifiers using all 4 UMBS metrics as input features. These were LDA, QDA, SVM with linear and non-linear (radial basis functions) kernels, kNN and ANN classifiers.

The data set is biased towards the non-movement class as over 94% of epochs contain no movement. Training classifiers directly with this data would bias the results. For example, if 90 % of data contained no movement and a classifier was tuned so that all epochs reported no movement regardless of the raw data, an accuracy of 90% would still be reported. To avoid this, a reduced data set was created by drawing equal numbers of samples of both classes randomly from the data. This reduced data set was further randomly split into 50% training and 50% test data sets. The original data set contained 1,418 samples and the reduced data set contained 156 samples evenly split between movement and non-movement samples. Classifiers were optimised using the training data and their classification performance evaluated on the test data. This process was repeated one hundred times for each classifier and the mean performance recorded (also known as cross-validation). Samples were replaced between each repetition. The standard deviation of the classification performance over the 100 runs was used to quantify the stability of the accuracies reported by each classifier. These results are given in Table 6.8.

TABLE 6.8: Results (mean  $\pm$  st dev; %) for the accuracy of movement detection when fusing the four UMBS derived motion metrics together and for the UMBS 2 metric.

	Accuracy	Sensitivity	Specificity	ROC Dist.
LDA	80.77 $\pm$ 4.57	96.08 $\pm$ 3.90	65.46 $\pm$ 3.90	82.39 $\pm$ 3.96
QDA	89.48 $\pm$ 3.69	91.54 $\pm$ 5.13	87.42 $\pm$ 6.74	89.67 $\pm$ 3.60
kNN	83.64 $\pm$ 4.60	86.08 $\pm$ 6.69	81.19 $\pm$ 8.16	83.88 $\pm$ 4.52
SVM Linear	87.02 $\pm$ 4.78	95.39 $\pm$ 4.61	78.65 $\pm$ 9.57	87.64 $\pm$ 4.38
SVM Non-linear	90.06 $\pm$ 4.17	93.73 $\pm$ 4.63	86.39 $\pm$ 6.70	90.24 $\pm$ 4.10
NN	89.11 $\pm$ 2.75	83.94 $\pm$ 5.35	94.28 $\pm$ 2.57	89.33 $\pm$ 0.25
UMBS 2	91.48	92.31	91.44	91.88

TABLE 6.9: Wrist actigraphy movement detection performance over various thresholds compared to the video based gold standard. The optimal BCT applied to the derived UMBS data is also given.

Act Thresh	Acc (%)	Sens (%)	Spec (%)	ROC Dist (%)
>0 (all motion)	79.03	93.10	77.31	85.58
>20 (low)	88.39	89.66	88.24	88.96
>40 (medium)	88.76	68.97	91.18	80.84
>80 (high)	89.89	44.83	95.38	74.53
>11 (optimal)	85.77	93.10	84.87	89.09
UMBS 2	91.48	92.31	91.44	91.88

### 6.3.2.3 Alternative Technologies

**PIR Motion Detection** In a direct comparison of PIR motion and video detected motion (gold standard), a high accuracy, specificity and sensitivity was found as given in Table 6.7. PIR motion detection was outperformed by three of the four UMBS metrics.

**Wrist Actigraphy** The wrist actigraph sleep/wake discrimination algorithm uses low (>20 motion counts), medium (>40 motion counts) or high (>80 motion counts) sensitivity thresholds reported on a per minute basis. A comparison against these thresholds and any wrist actigraph defined motion (>0 motion counts) was investigated. The actigraphy threshold with the most accurate motion detection (defined by the maximal ROC distance) occurred at eleven activity counts per minute. The results from a binary classification test applied to these thresholds are given in Table 6.9.

### 6.3.3 Discussion

The UMBS 2 metric accurately detects nocturnal movement in bed with an accuracy, sensitivity and specificity of over 90% using a thresholding method applied to an easily computed UMBS metric (see Table 6.7). The UMBS has been shown to outperform wrist actigraphy, the current real-world gold standard for sleep monitoring, regardless of whether the actigraphy threshold was selected to detect an optimal, low, medium or high levels of motion (see Table 6.9). Additionally, although various advanced classification algorithms were applied to increase accuracy, a binary classification test applied to one set of derived UMBS data still resulted in the greatest accuracy (see Table 6.8). The calculation of the UMBS 2 metric and the application of a thresholding function has a significantly lower computational complexity than the use of advanced classification algorithms and, as such, is easily realisable in a bed-based movement detection system.

The PIR based alternative solution investigated here results in a lower accuracy, specificity and sensitivity than three of the four UMBS motion metrics, yet this remains comparably high. However, no bed sheets were used in this validation and it is conceivable that bed sheets would inhibit the motion detection accuracy of the PIR.

## 6.4 Summary

This chapter has investigated the potential of a pressure based Under Mattress Bed Sensor (UMBS) as a means of non-invasively capturing breathing, heart rate and actigraphic data from a sleeping subject. The novel contributions of this chapter are as follows:

**The development and validation of an algorithm which extracts respiration rate from UMBS data:** While reliable heart rate measurement was shown not to be feasible, validating experiments confirm the sensor’s capabilities for respiration rate estimation. Data was collected from eight subjects over lengthy periods when the UMBS was placed both beneath and above the mattress. The algorithm discussed can reliably estimate breathing rate from UMBS data in the absence of large body movement when the UMBS is placed both above and below the mattress. In particular, a novel algorithm has been proposed involving median filtering to remove the effect of slow postural changes on the UMBS, bandpass filtering to remove unwanted spectral noise, PCA based data fusion to merge respiratory information from all active UMBS signals and an optimised peak detection algorithm which detects peaks related to breathing cycles. When the UMBS is placed underneath the mattress, the mean difference between this system and

TABLE 6.10: Performance of wrist actigraphy, PIR and UMBS metrics in detecting motion compared to a video-based gold standard.

Technology	Acc (%)	Sens (%)	Spec (%)	ROC Dist (%)
Wrist Actigraphy	85.77	93.10	84.87	89.09
PIR	93.36	83.33	93.91	88.79
UMBS 2	91.48	92.31	91.44	91.88

the gold standard was  $-0.12$  (SD of  $\pm 2.26$ ) BrP5M and an MPE of  $-0.16\%$  (SD of  $\pm 3.12\%$ ). This method has been shown to provide a high accuracy compared to a strain gauge based measurement of respiration rate and compares very well against similar ambient approaches to sleep monitoring (as reported in Table 6.6).

**An examination into the extraction of heart rate from UMBS data:** In a preliminary experiment heart rate signals were found to exist in the UMBS data using frequency analysis. However, this information was not available over all subjects. A visual time-based analysis did not reveal heart rate signals. In a further investigation in Experiment 2, it was established that heart rate could not be accurately detected using frequency analysis. The sampling rate of the UMBS (v3) was 20 Hz, however due to the effective simplex polling protocol (discussed in the last chapter) this sampling rate was lower and a mean sampling rate of approximately 15 Hz was found. Other similar pressure-based systems which use a very high sampling rate report a high accuracy in detecting heart rate (Mack et al., 2009a). However in this case data from one sensing position is recorded, whereas in the case of the UMBS data from 24 sensing positions is collected at a lower sampling rate.

**The development and validation of an algorithm which derives multiple motion metrics from UMBS data:** In the investigation of the motion detection capacity of the UMBS, a number of UMBS derived metrics were proposed and shown to detect movement epochs with a high degree of accuracy and reliability (the best, UMBS 2, reporting sensitivity and specificity of over 90%). This was achieved using multiple subjects over extended periods and compared to a video-based gold standard and alternative bed movement sensing solutions. A motion detection algorithm using binary classification on a UMBS derived metric, and an optimally selected threshold, was found to outperform advanced classification techniques, using a combination of all UMBS derived metrics. Additionally, the UMBS motion detection algorithm outperformed wrist actigraphy, the gold standard method of sleep/wake discrimination, and a PIR based ambient technology (as shown in Table 6.10).

By exploiting the movement detection and respiration rate estimation capabilities in parallel, the UMBS system can provide valuable physiological information relating to the sleeping patterns of a subject under study, in an unobtrusive and ambient manner. The technology is suited to data collection over an extended period and will not alienate the participant. This is particularly relevant when monitoring a sensitive cohort, such as the elderly, those suffering from mild cognitive impairment or people with dementia. This, coupled with the low computational complexity of the data processing involved, makes the UMBS an attractive and practical proposition for longer-term sleep monitoring in a non-clinical setting.

## Chapter 7

# UMBS Deployment in Clinical and Domestic Environments

In the previous chapter, metrics derived from the UMBS were shown to reliably and accurately estimate respiration rate and also to detect movement. However, a number of more informative metrics, or features, of motion can be extracted from the raw UMBS data. The UMBS possesses the ability to record temporal, spatial, statistical and spatiotemporal descriptions of in-bed movement throughout the night. In this chapter, algorithms will be proposed which derive these descriptions from the raw UMBS data. These algorithms will subsequently be applied to data collected in both real-world and clinic-based populations. This will provide an insight into the types of information captured by these features. A comparison will be made against wrist actigraphy data in order to investigate whether there is any inherent loss of information due to the discretisation process (that is, in wrist actigraphy, the continuous two axial accelerometer data is summated into a discrete sample over a pre-defined epoch). Additionally, a comparison will be made between the spatial and temporal UMBS-derived features. Furthermore, this chapter describes a method of providing a spatiotemporal description of each movement. A comparison of all UMBS-derived features across data collected in the clinic and the real-world will be investigated.

The research will quantify differences between clinical and real-world measurements of in-bed motion. It will also highlight features which are environment independent (that is, features which will provide similar information regardless of recording condition and bed/mattress type).

## 7.1 Introduction

Traditional approaches of ambulatory sleep monitoring (wrist actigraphy) calculate a movement metric (commonly referred to as *activity counts*) which summates all motion occurring within an epoch (see Chapter 4). This was originally performed as a means of reducing computational complexity in the storing of the data and for the sleep/wake discrimination process. The application of relatively simplistic algorithms have reported high accuracies and reliability of using this derived activity information to estimate sleep and wake states across various cohorts (Kushida et al., 2001; Sadeh et al., 1994). However, the actigraph's placement on the wrist introduces an inherent difference in the type of motion that is being recorded, that is, it is limb movement and not core body movement that is recorded.

While relatively recent advances in technology provide increased computational complexity in contact-based devices which are of appropriate sizes and dimensions, little consideration has been given to the adherence and suitability of wrist actigraphy for long term placement. For example as reported previously in this thesis, one older adult refused to wear an actiwatch, another subject took off the actiwatch for periods of social interaction, and there were problems in the initial days of the study where subjects removed the actiwatch prior to sleep due to confusion about having to wear the watch even whilst asleep (the subjects were asked to wear the actiwatch continuously) (Behan et al. 2008a, Behan et al. 2008b). For sensitive populations, a non-contact modality for sleep monitoring is proposed to provide a similar descriptive index of sleeping patterns and sleep quality as the contact-based approach. Such a solution would be particularly suitable for long-term monitoring. To date, many ambulatory and ambient sleep monitors have focussed solely on temporal descriptions of movement, that is, reporting the magnitude of movement over time (Adami et al., 2009a; Brink et al., 2006; Choi et al., 2006; Fox et al., 2007; Rantz et al., 2008). A limited number of these technologies have the ability to provide a spatial description of sleep, however very few of these take advantage of this capacity (Van Der Loos et al. (2003) use a spatial description of body position to correct for bad posture).

The addition of a spatial description of sleep can be used to quantify the magnitude of physical changes in body posture resultant from nocturnal movements throughout the night. Additionally, spatiotemporal descriptions of each movement, containing both temporal and spatial quantifications, report specific metrics relating to each movement such as duration, magnitude and change in lateral position of that movement. Both of these types of data may quantify in-bed movements better than the more traditional temporal and statistical metrics of in-bed movement.

TABLE 7.1: Derived UMBS Features

Feature	Description
Temporal	The continuous quantification of overall movement over time
Spatial	The continuous quantification of the magnitude of changes in body posture over time
Statistical	A discretised measurement of the magnitude of movement in user-defined epochs
Spatiotemporal	A detailed description (both spatially and temporally) of each in-bed movement

The research questions addressed in this chapter include:

- Can temporal, spatial and statistical descriptions of in-bed movement be derived from the UMBS? And if so, how do they compare across cohorts based in clinical and real world environments?
- Does continuous temporal movement and/or spatial data provide a more informative description of in-bed movement than discretised (binned data, such as *activity counts* recorded by wrist actigraphy) data?
- Can spatiotemporal descriptions of each movement be calculated from UMBS data? And if so, can a spatiotemporal description of postural movements in-bed discriminate between cohorts?
- Are temporal, spatial, statistical and spatiotemporal descriptions of in-bed movement consistent over multiple days?

## 7.2 UMBS Feature Extraction Algorithms

Four types of features relating to movements were derived from the UMBS data: temporal, spatial, statistical and spatiotemporal features (as shown in Figure 7.1 and described in Table 7.1). Statistical data was used to identify periods of continuous movement from which spatiotemporal descriptions of each movement were calculated. Spatial data also informed these descriptions. The algorithms developed to extract these data were validated using a custom collected ‘Turns Data Set’ (previously described in Section 5.2.3).

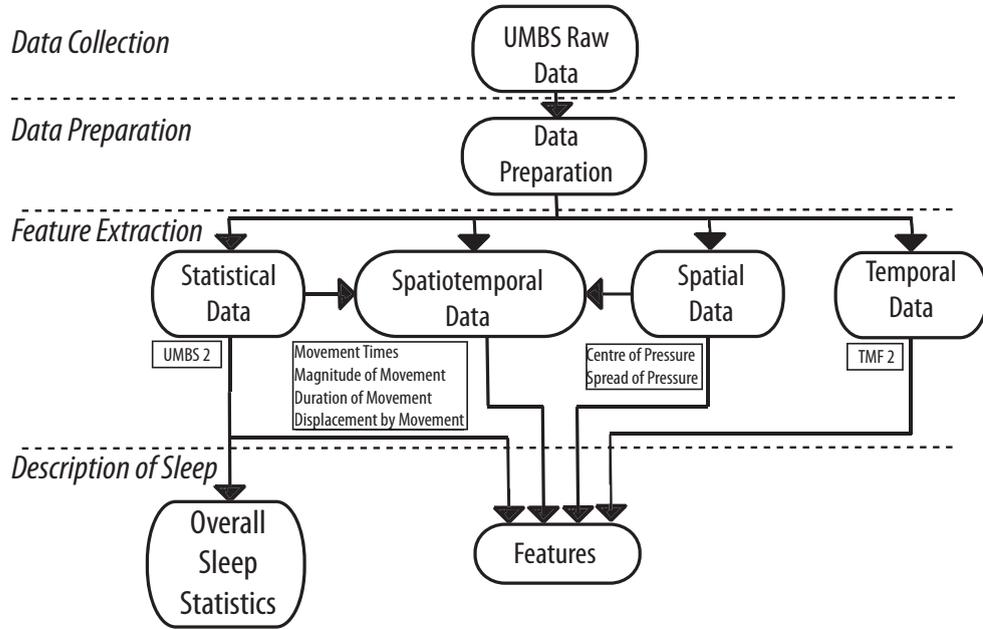


FIGURE 7.1: Extraction of movement features from UMBS data.

### 7.2.1 Temporal Movement Features

Temporal movement features, extracted from the UMBS data, describe the magnitude of motion registered by each tactel in the sensor over time. By monitoring the changes in all tactel values over subsequent sample instants, an overall measure of motion can be calculated. Initially, a continuous temporal activity metric was calculated using the first derivative of the raw UMBS tactel data (see Equation 7.1). This metric represents changes in the pressure placed across each of the 24 tactels in the UMBS over time as shown in Figure 7.2. The standard deviation and mean of this metric was extracted from the data over all tactels as given in Equations 7.4 and 7.5 respectively.

$$\delta_{ij} = [x_{ij} - x_{(i-1)j}] \quad \forall 2 \leq i \leq N, 1 \leq j \leq 24 \quad (7.1)$$

where  $\delta$  is the difference between successive tactel values,  $N$  is the number of samples,  $x_{ij}$  is the value of the  $j^{\text{th}}$  tactel at the  $i^{\text{th}}$  sample instant, and  $\delta_{ij} = 0$  for  $i=1$  and  $1 \leq j \leq 24$ .

$$\delta_i = [\delta_{i1}, \dots, \delta_{i24}] \quad \forall 1 \leq i \leq N \quad (7.2)$$

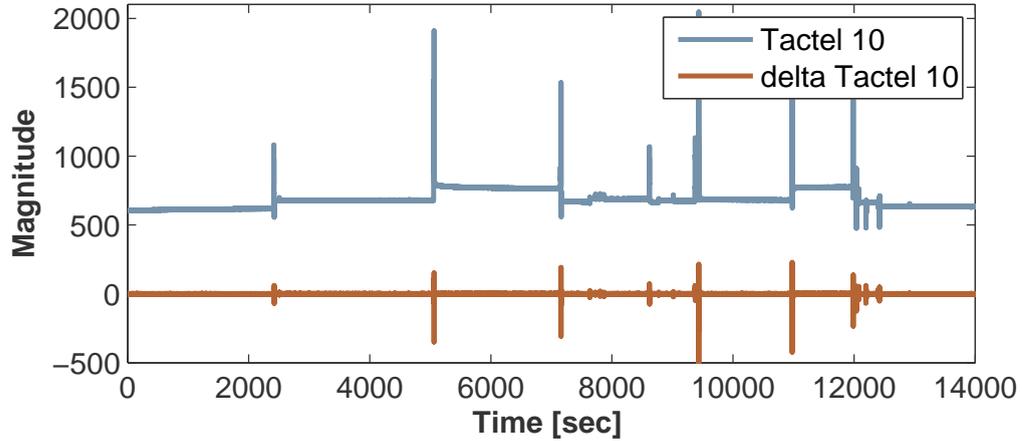


FIGURE 7.2: Data from one UMBS tactel over time and the derivative of that data.

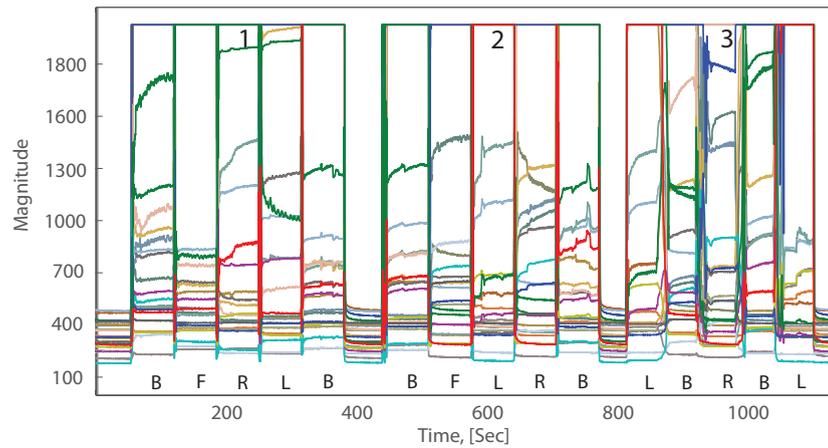
$$\bar{\delta}_i = \text{mean}(\delta_i) = \frac{1}{24} \sum_{j=1}^{24} |\delta_{ij}| \quad (7.3)$$

$$\text{TMF}_1(i) = \text{std}(\delta_i) = \sqrt{\frac{1}{23} \sum_{j=1}^{24} (\delta_{ij} - \bar{\delta}_i)^2} \quad (7.4)$$

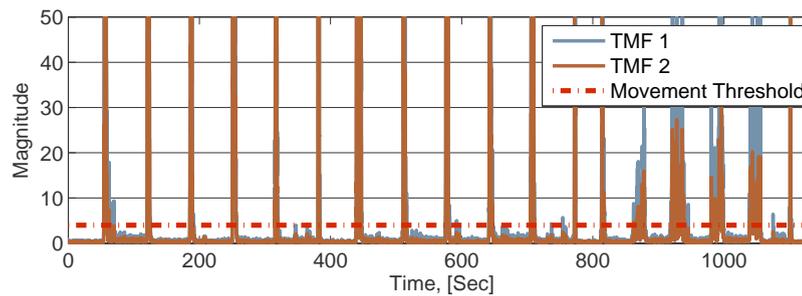
$$\text{TMF}_2(i) = \bar{\delta}_i \quad (7.5)$$

In the turns data set,  $\text{TMF}_1$  and  $\text{TMF}_2$  were found to be highly correlated ( $r=0.96$ ,  $p < 0.001$ ) and as such  $\text{TMF}_2$  was used in further analysis due to its lower computational complexity. The time domain data from the two derived temporal features ( $\text{TMF}_1$  and  $\text{TMF}_2$ ) are shown in Figure 7.3(b) and can be compared directly to the UMBS data as given in Figure 7.3(a). The frequency domain data for  $\text{TMF}_2$  was calculated using the short time Fourier transform. The real part of the log of the frequency component of the signal was used in order to reduce the magnitude of the time-frequency data (as shown in Figure 7.3(c)). No significant difference between the different postural changes was visually discernible from the frequency spectrum, apart from during the slow turns which inherently have a reduced magnitude. However, this reduced magnitude is evident in the time based data. As such, the frequency content of the temporal movement features was not included as a feature in further analysis.

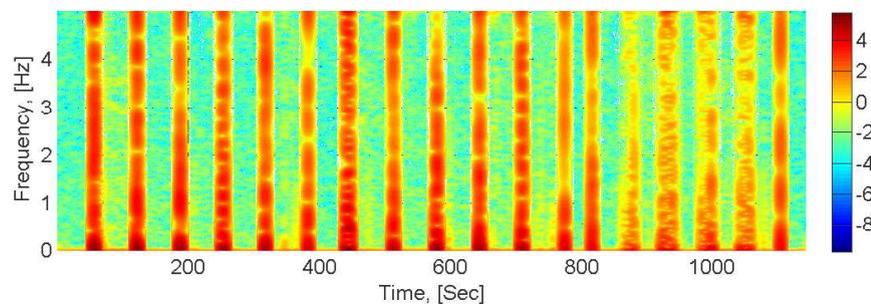
In order to identify movements a threshold of greater than 4 in the time-domain  $\text{TMF}_2$  data was empirically chosen to define when motion occurred (see Figure 7.3(b))



(a) UMBS data collected while the participant was asked to assume and shift between four typical sleeping postures: lying on their back (B); left side (L); right side (R); and front (F). The subject was asked to change postures under three conditions 1) rapid transition without lateral displacement, 2) rapid transition with lateral displacement and 3) slow rolling transitions (inherently including lateral displacement).



(b) Corresponding temporal movement features with movement threshold (time domain).



(c) Temporal movement features (time-frequency domain for  $TMF_2$ ).

FIGURE 7.3: UMBS data and derived temporal features from a participant shifting between postures on the UMBS.

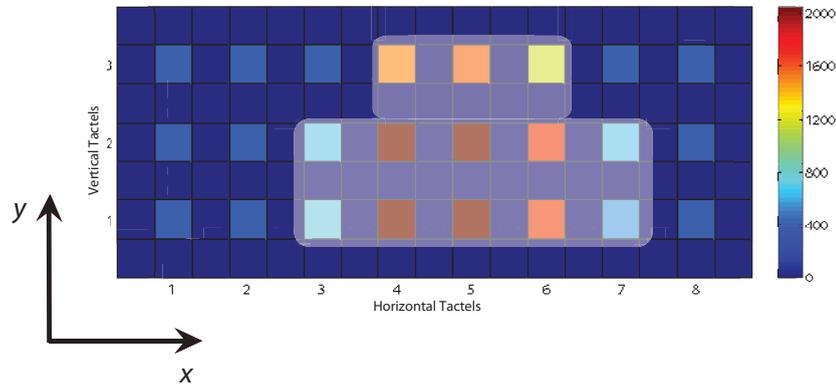


FIGURE 7.4: UMBS pressure data with outline of original tactels registering values above 500.

### 7.2.2 Spatial Movement Features

A continuous description of spatial movement in bed was formed by extracting both the centre and spread of pressure applied to the UMBS. A value of 500 units was used as the threshold to indicate whether pressure was placed on a tactel. This allows us to map where the subject is lying across the UMBS as shown in Figure 7.4. The centre of pressure (COP) is defined as the centre of the outline shown, while the spread is a measure of the width of the outline. In order to increase the accuracy in the centre and spread of pressure measurements, the tactel values were interpolated, that is, for each row of eight original tactels across the UMBS, 71 sample points were linearly interpolated using tactels 1 through 8 in steps of 0.1 (as shown for one row in Figure 7.5). If any of the interpolated sample points registers pressure above 500, it was deemed that pressure was placed upon that point. The absolute centre of pressure was the average point upon which pressure is applied over the three rows of the UMBS in both the lateral (x) and anteroposterior (y) direction (this coordinate system is given in Figure 7.6). These points are referred to as  $COP_{x,y}$  (see Equation 7.8), or  $COP_x$  and  $COP_y$  individually (an example of the  $COP_x$  and  $COP_y$  points for the ‘Turns’ data set is given in Figure 7.7(b)). The value of the centre of pressure point relates to the absolute centre of the subject lying on the UMBS and is calculated for each sample instant as shown in Figure 7.7(b)). The bottom left of the UMBS was taken as the origin of this axis (that is, the corner closest to the right arm when the subject is lying supine).

$$status_{x,y} = \begin{cases} 0 & x_{x,y} < Threshold \\ 1 & Otherwise \end{cases} \quad (7.6)$$

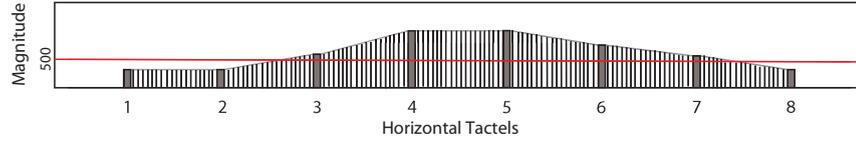


FIGURE 7.5: Interpolated tactel values for one row.

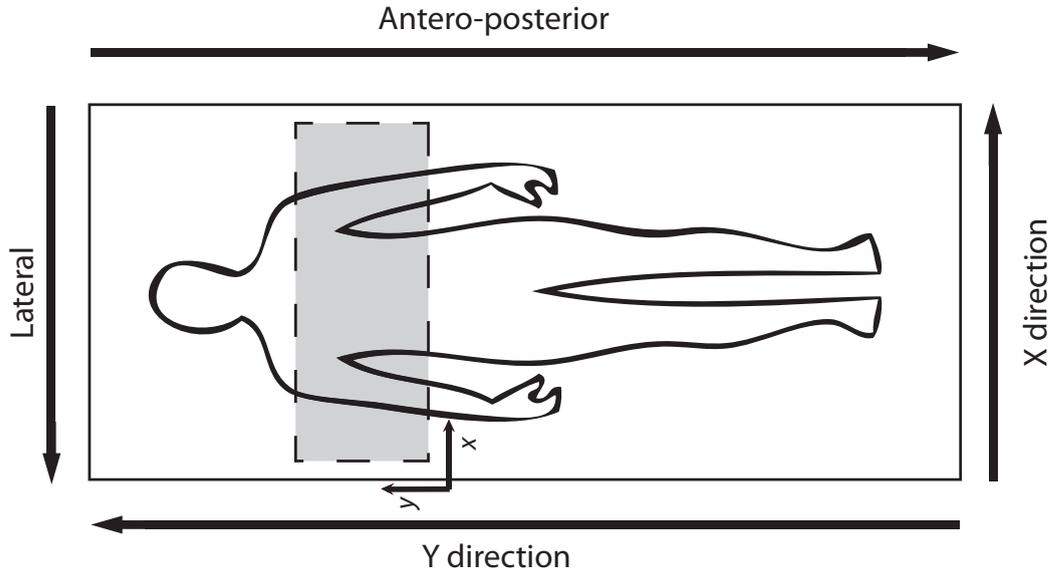


FIGURE 7.6: UMBS terms of location

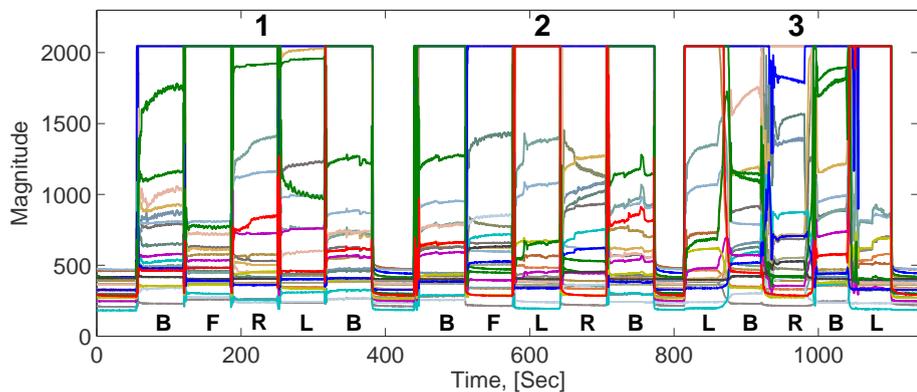
where  $x_{x,y}$  is the pressure placed upon the point  $(x, y)$  and  $status_{x,y}$  is a boolean value which defines whether pressure is placed upon the point  $(x, y)$ . For each point  $(x, y)$ , the pressure value is interpolated from the tactel values in the x-direction only.

$$COP_{x,y} = \frac{\sum_{x=1}^{71} \sum_{y=1}^3 (Pos_{x,y} \times status_{x,y})}{\sum_{x=1}^{71} \sum_{y=1}^3 status_{x,y}} \quad (7.7)$$

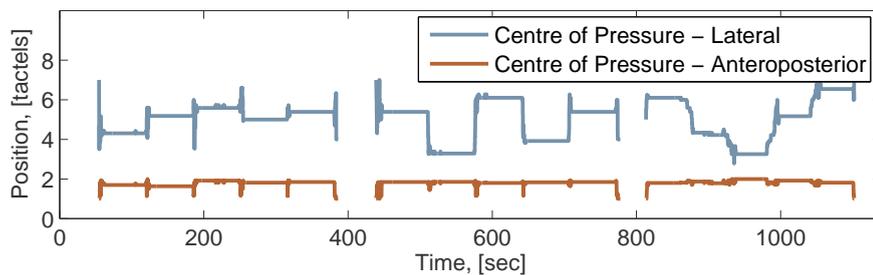
$$COP_{x,y} = [COP_x, COP_y] \quad (7.8)$$

where  $Pos_{x,y}$  is a vector which defines the location of the current point  $(x, y)$ , the numerator is a summation of  $(x, y)$  positions upon which pressure is applied (defined by having a *status* of 1) in both  $x$  and  $y$  directions, and the denominator is the number of positions which register pressure as being applied.

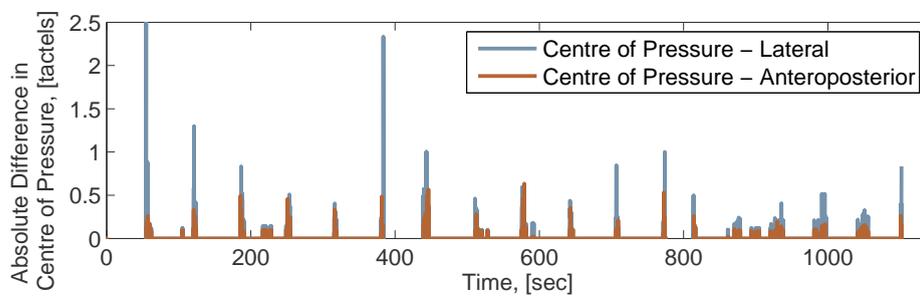
It was unclear whether  $COP_x$  would provide useful information when each sample is examined independently. The absolute value of  $COP_x$  does not possess any relationship



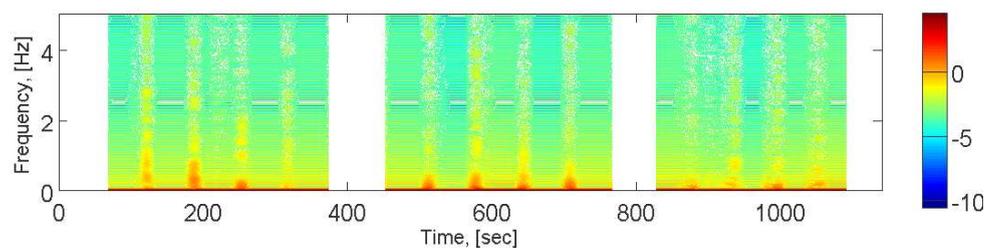
(a) UMBS data collected during conditions 1, 2 and 3.



(b) Spatial movement features - time



(c) Spatial movement features - time difference



(d) Spatial movement features - frequency (no data were plotted when the subject was not present in the bed)

FIGURE 7.7: UMBS data and derived spatial features from a participant shifting between postures on the UMBS.

to sleep quality, solely to position of the subject, however when taken relatively to surrounding values, it provides a spatial measure of in bed movement. The first derivative of the  $COP_x$  data was used to quantify this movement (as per Equation 7.9). A plot of the Spatial Movement Feature (SMF) for the ‘Turns’ data set is shown in Figure 7.7(c).

$$SMF_i = COP_x(i) - COP_x(i - 1) \quad \forall \quad 2 \leq i \leq N \quad (7.9)$$

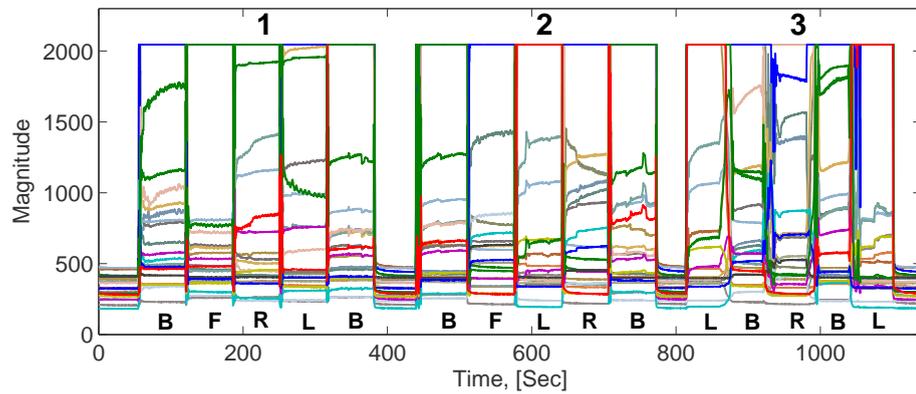
where  $SMF_1$  is 0.

The time-frequency content of the lateral centre of pressure ( $COP_x$ ) is shown in Figure 7.7(d). Similarly to the temporal movement features, the frequency spectrum of the spatial movement features did not offer any increase in information content above the time based features.

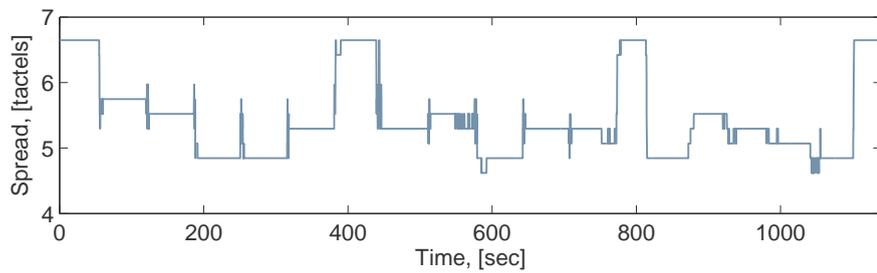
Changes in centre of pressure in the anteroposterior direction were not investigated as it was deemed that changes in posture in this direction would not provide useful information.

$TMF_2$  (Equation 7.5) and SMF (Equation 7.9) were found to be significantly correlated ( $r=0.43$ ,  $p<0.001$ ) for the turns data set. However, while it was significant, it was not found to be a strong correlation. This provides a justification for the use of this metric in addition to  $TMF_2$ .

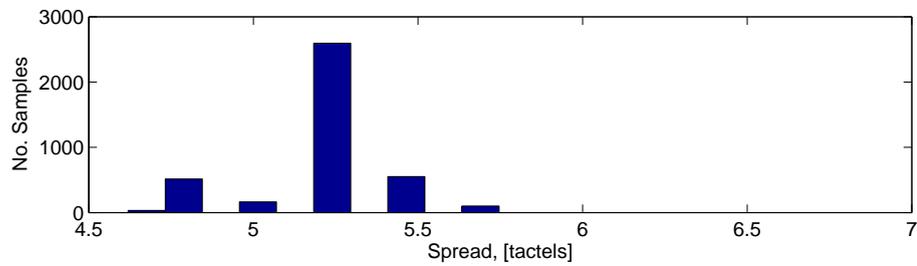
The spread, or width, of the pressure impinged on the UMBS by the subject was also investigated as a means of quantifying in-bed movements. It was defined as the number of tactels taken to cover 80% of the overall pressure applied to the UMBS centered around the  $COP_x$  point. This algorithm begins at the lateral center of pressure point ( $COP_x$ ) and iteratively works outward in both directions until 80% of the overall pressure applied to the UMBS is measured. The width of this window is defined as the spread of pressure and is calculated for each sample instant (see Figure 7.8(b)). A partial function of the spread of pressure was calculated in order to differentiate between side lying and front/back lying postures. However, this metric also measures movement as the spread will change with postural shifts in position. A histogram of the data from these postures was plotted in order to investigate the differences between the postures (see Figures 7.8(c) and 7.8(d)). No discernible variation between the two histograms relating to the different types of postures was present in the ‘Turns’ Data Set (see Figure 7.8(c)). A two-dimensional comparison of tactel values (top-down view) for both prone/supine and side-lying showed little difference between both postures. This may be due to an increase in pressure being directed radially outward for the side-lying case (as the same amount of pressure is focussed in a smaller area).



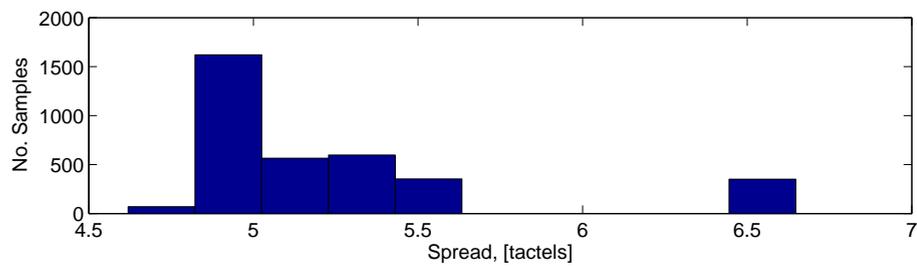
(a) UMBS data collected during conditions 1, 2 and 3.



(b) Spatial movement features - spread of pressure across the UMBS



(c) Histogram of spread of pressure - lying on the front or back



(d) Histogram of spread of pressure - lying on the side

FIGURE 7.8: UMBS data and spread of pressure from a participant shifting between postures on the UMBS.

### 7.2.3 Statistical Movement Features

Discretised UMBS 2 data measuring the magnitude of motion within a predefined epoch length of 60 seconds was calculated using Equation 6.2 (as discussed in Section 6.3 and shown in Figure 6.15). This metric is similar in form to the activity motion count calculated by wrist actigraphy although the specificity, sensitivity and accuracy of the UMBS 2 metric was found to be higher. Time frequency analysis was not applied to this data set as there was an inadequate number of samples.

### 7.2.4 Spatiotemporal Movement Features

Spatiotemporal UMBS movement features were calculated from data in which a complete movement occurred (that is, a single motion event or multiple motion events occurring in quick succession). In order to locate the beginning and end points of the motion event, each epoch deemed to contain motion (as defined by UMBS 2 in Section 7.2.3) was analysed over all time samples within that epoch. Upon conditions where a movement occurred over multiple epochs, all of these epochs were concatenated and examined together. A fifth order butterworth low-pass filter, with a cut off frequency of 0.5 Hz, was applied to the tactel data in order to remove any higher frequency related artifacts, which would not relate to large body movement (Redmond and Hegge, 1987). Within this, possibly concatenated, filtered data, movement was defined to have occurred if the absolute difference between a tactel value and its predecessor (as per Equation 7.10) exceeded an empirically defined threshold (see Equation 7.11 and Figure 7.9(b)).

$$\delta'_{ij} = [x_{ij} - x_{(i-1)j}] \text{ for } 2 \leq i \leq N, j = 1, 2, \dots, 24 \quad (7.10)$$

where  $\delta'_{ij}$  is the difference between successive tactel values for the spatiotemporal window under investigation,  $N$  is the number of samples,  $x_{ij}$  is the value of the  $j^{\text{th}}$  tactel at the  $i^{\text{th}}$  sample instant (within the window under investigation), and  $\delta'_{ij} = 0$  for  $i=1$  and  $j = 1, 2, \dots, 24$ .

$$\text{movement}' = \begin{cases} 1 & \text{if } \delta' \geq \text{threshold} \\ 0 & \text{otherwise} \end{cases} \quad (7.11)$$

Under circumstances where movement briefly ceases, a refractory period of ten seconds was used to concatenate these separate multiple movements together in order to analyse them as belonging to the same movement event. The majority of movements lasted less than 20 seconds. As such, frequency analysis was not applied to the spatiotemporal data.

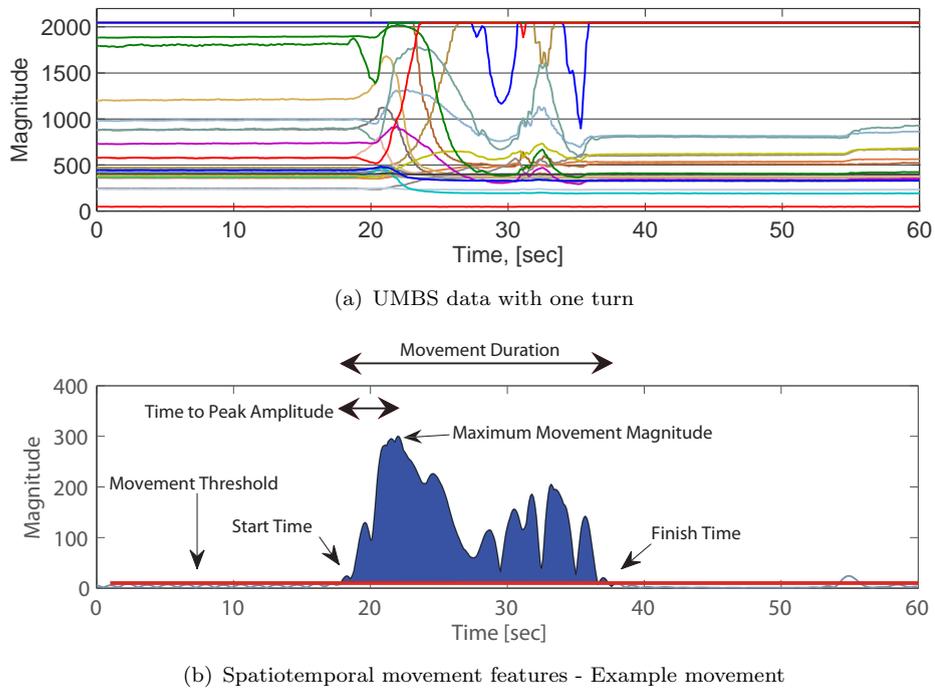


FIGURE 7.9: UMBS and spatiotemporal data from a participant shifting between postures on the UMBS.

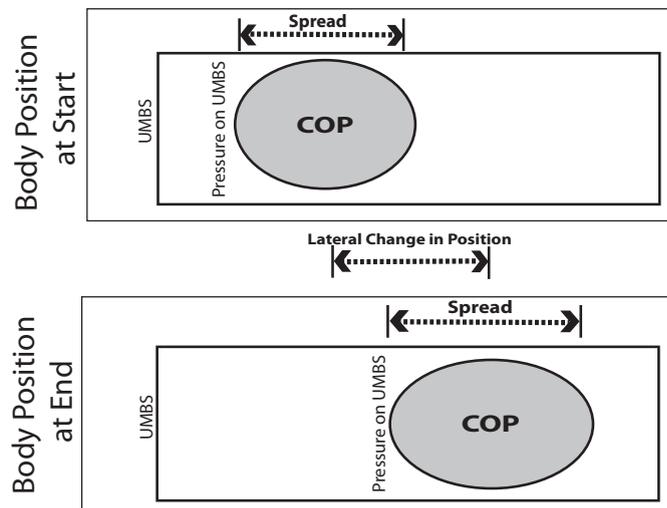


FIGURE 7.10: Pressure impinged on the UMBS at the beginning and end of a movement

A number of metrics were extracted from each of these movement events as listed in Table 7.2.

TABLE 7.2: Spatiotemporal Features

Feature	Description
Start Time	The absolute start time of the movement (see Figure 7.9(b)).
Finish Time	The absolute end time of the movement (see Figure 7.9(b)).
Movement Duration	The duration of the movement (see Figure 7.9(b)).
Lateral Change in Position	The displacement in posture across the UMBS between the start and end of the movement (see Figure 7.10).
Change in Spread	The change in the 80% spread of pressure across the UMBS between the start and end of the movement (the change in the spread before and after the movement as shown in Figure 7.10).
Movement Area	The absolute sum of the movement metric throughout the movement period (shown in blue in Figure 7.9(b)).
Maximum Movement Magnitude	The maximum (or peak) movement value during the movement (see Figure 7.9(b)).
Time to Peak Movement	The time taken to reach the maximum movement magnitude point from the start of the epoch (see Figure 7.9(b)).
Percentage to Peak Movement	Time to peak movement taken as a percentage of the total duration of the movement.
Magnitude of Movement	Movement area normalised by the duration of that movement.
Spread Movement Index	The sum of the absolute difference of the 80% spread of pressure across the UMBS.

## 7.3 Methods

The UMBS was deployed in four settings: 1) the homes of healthy adults, 2) the homes of relatively healthy older adults, 3) a sleep clinic where participants were assessed for a sleep disorder (most commonly sleep apnoea) and 4) a general clinical research centre where healthy adults are undergoing sleep research studies.

### 7.3.1 Summerhill Data Set

Data were collected from the homes of 10 older adults for a period of two weeks as previously discuss in Section 5.3 (see Table 7.3 for participant information). The 10Hz 256 resolution UMBS (v1) and PDA data collection system was deployed in the homes

TABLE 7.3: Data from the home of relatively healthy older adults.

Subject	Sex	Age	Nights
SH101	F	63	12
SH102 <sup>1</sup>	F	79	-
SH103	F	62	11
SH104	F	80	13
SH105	M	72	2
SH106 <sup>2</sup>	F	88	-
SH107	F	64	12
SH108	F	65	13
SH109	M	81	8
SH110	F	72	14

<sup>1</sup> this participant had spina bifida since childhood.

<sup>2</sup> This participant refused to wear the actigraphy watch.

TABLE 7.4: Data from the homes of 3 healthy young adults.

Subject	Sex	Age	Nights
MH101	M	27	1
MH102	M	27	2
MH103	F	27	5

of older adults. Mid-study data integrity checks were carried out. Data were collected successfully over 75% of nights. The primary reason for sensor failure was power outage. This was due to either loose connections or the participant unplugging the sensor. Participants were asked to wear a wrist actigraph (Actiwatch, Cambridge Neurotechnology, UK) for the duration of the study. One participant was excluded from analysis as she had spina bifida since childhood and as a result her weight was not evenly spread across the bed. One participant, aged 88 years old, refused to wear the watch for the duration of the study. Another regularly removed the watch for any periods of social interaction, each of these periods were excluded from analysis.

### 7.3.2 Maynooth Data Set

Data were collected from the homes of 3 healthy adults over a number of days (see Table 7.4). The 20Hz 2048 resolution UMBS (v3) was deployed in their bedrooms. A customised C++ program collected data continuously on a laptop (Dell Precision M6300).

TABLE 7.5: Data from a Sleep Clinic from individuals with a suspected sleep disorder.

Subject	Sex	Age	Nights
PT101	F	63	1
PT102	M	59	1
PT103	M	44	1
PT105	M	52	1
PT106	M	30	1
PT108	F	43	1
PT109	M	39	1
PT110	M	74	1
PT112	F	n/a	1
PT114	M	n/a	1

Age data not available for two subjects.

### 7.3.3 Peamount Data Set

Data were collected from the Sleep Clinic at Peamount Hospital (Newcastle, Co. Kildare, Ireland) from 10 adults referred with a suspected sleep disorder for one night (see Table 7.5). The 20Hz 2048 resolution UMBS (v3) was deployed in the clinic continuously until an entire data set was recorded. A customised C++ program collected data continuously on a laptop (Dell Precision M6300).

### 7.3.4 Boston Data Set

Data were collected from the General Clinical Research Center in the Division of Sleep Medicine at Brigham and Womens' Hospital, Harvard Medical School (Boston, MA, USA) over multiple nights from 12 subjects (see Table 7.6). These subjects were undergoing a protocol under which the phase and length of day varied (this is discussed in further detail in Section 8.2.1). Data were taken from the initial baseline period as sleep occurs during the normal sleep time of the subject. The 20Hz 2048 resolution UMBS was deployed in the clinic continuously. A customised C++ program collected data continuously on multiple computers (Dell Optiplex 745).

### 7.3.5 Pre-Processing

The UMBS data were interpolated after data collection to a constant sampling rate of 10Hz. All data were manually truncated to remove wake periods at the beginning and end of the sleeping episode. Sleep diary reports and either wrist actigraphy or UMBS

TABLE 7.6: Data from a research clinic from healthy younger and older adults.

Subject	Sex	Age	Nights
BN101	M	21	2
BN102 <sup>1</sup>	M	19	3
BN103	F	56	3
BN104 <sup>1</sup>	M	70	6
BN105 <sup>1</sup>	M	23	3
BN106 <sup>1</sup>	M	23	3
BN107 <sup>1</sup>	F	19	3
BN108	F	21	1
BN109 <sup>1</sup>	M	24	3
BN110 <sup>1</sup>	M	27	3
BN111 <sup>1</sup>	F	19	3
BN112 <sup>1</sup>	M	24	3

<sup>1</sup> data available for the initial sleep period.

derived movement periods were used to define the beginning and end of analysis. As this chapter examines the movement periods in bed these edge times were deemed suitable.

## 7.4 Results

### 7.4.1 Distribution Plots

Distribution plots (Dorn, 2008) were used to provide a visual description of the features recorded across subjects and across cohorts. These are vertical forms of histograms akin to the boxplot, however they additionally show the distribution of the data along one axis (as shown in Figure 7.11). The width of the distribution refers to the frequency of that sample in that set of data. This width is normalised to the largest occurrence of a particular sample across the entire data set. In this chapter, the data was often divided into distinct sections, and then plotted separately. This is because the data is biased towards lower values.

### 7.4.2 Comparison Across All Subjects and Cohorts

#### 7.4.2.1 Temporal, Spatial and Statistical Features

The distribution of the temporal (TMF<sub>2</sub>), spatial (SMF) and statistical (UMBS 2) movement features, over the first night of each recording, are shown in Figures 7.12(a),

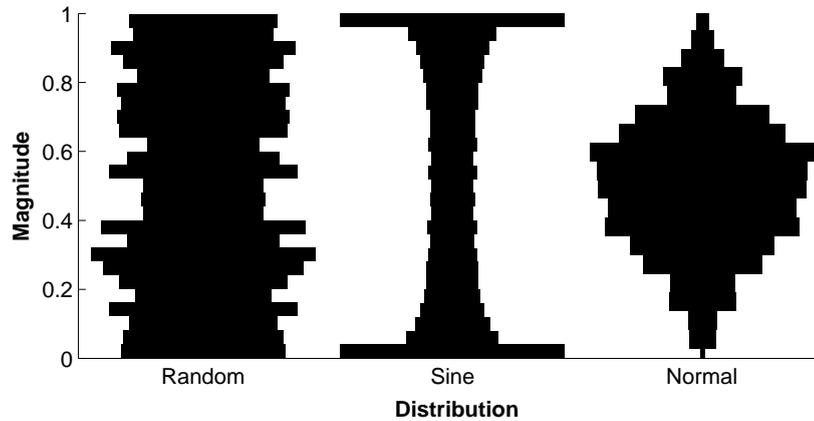


FIGURE 7.11: Distribution plots for uniform random (left), sinusoidal (centre), and normally distributed random time series data.

7.13(a) and 7.14(a) respectively. Only  $\text{TMF}_2$  and SMF values greater than zero were included in these results.  $\text{TMF}_2$  and SMF values equal to zero related to no movement change between two sample points and due to their abundance were excluded from further analysis. The labels of each level of movement in these Figures were termed relative to the magnitude of movement of the adjacent levels. Comparisons across the four cohorts (Maynooth, Summerhill, Boston and Peamount) are also given in Figures 7.12(b), 7.13(b) and 7.14(b) for the temporal ( $\text{TMF}_2$ ), spatial (SMF) and statistical (UMBS 2) movement features. These features, computed as percentages, were compared across each cohort, as shown in Table 7.7.

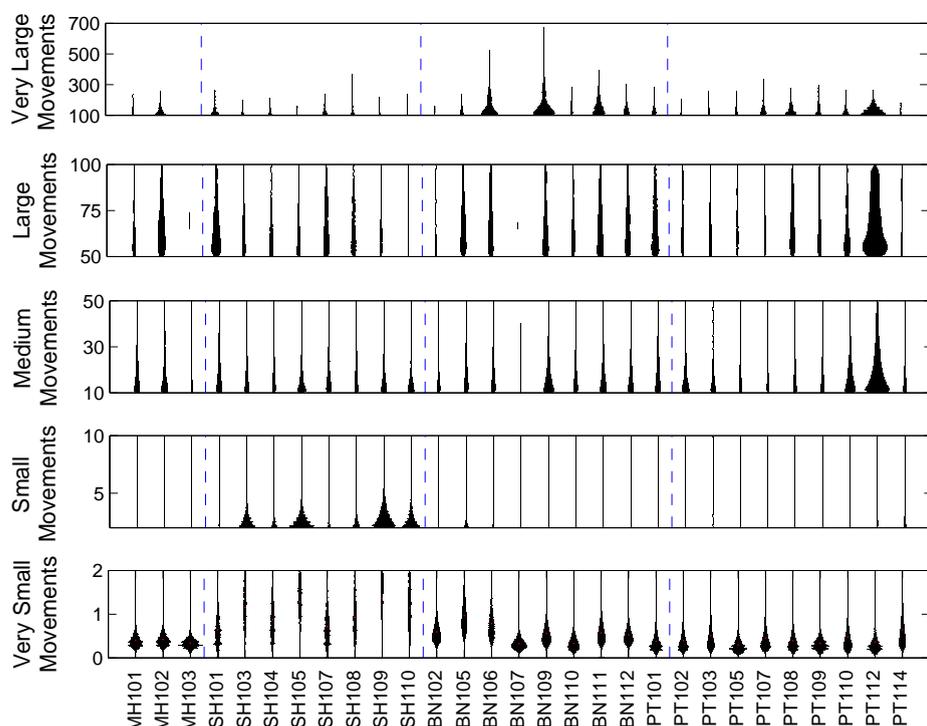
#### 7.4.2.2 Spatiotemporal Features

The spatiotemporal descriptions of movements, as defined in Table 7.2, were also extracted across the entire cohort, and are included in Appendix A for completeness.

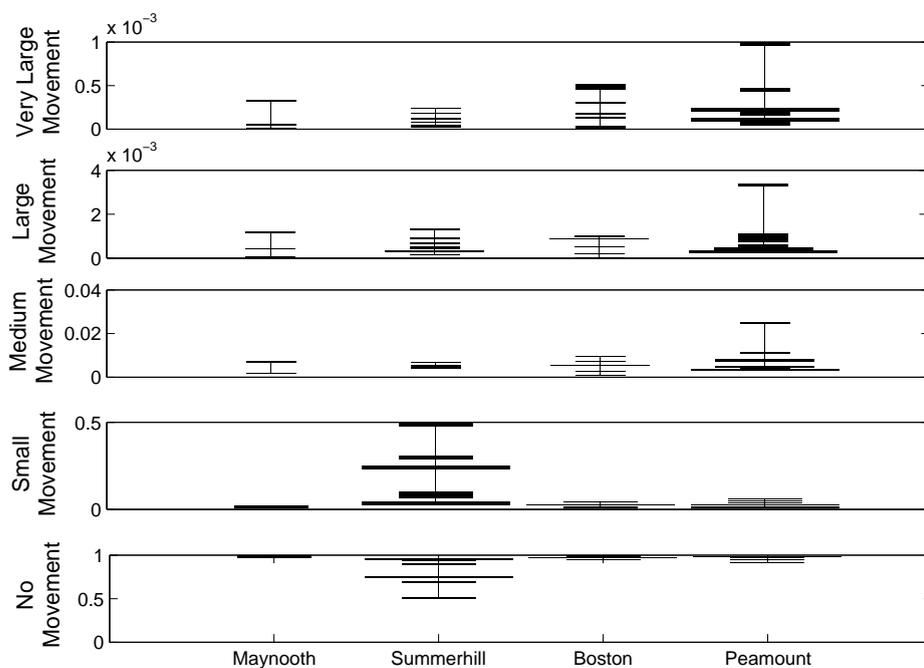
The absolute magnitude of all metrics were calculated. The lateral change in position values contained a large number of zeros over all cohorts. As such, only values greater than zero were included in further analysis. The spatiotemporal features, computed as percentages, were compared across each cohort, as shown in Table 7.7. The variation in the number of spatiotemporal movements is shown in Figure 7.15

#### 7.4.3 Variance Over Multiple Days

The variance in each metric over five days for eight community dwelling older adults and one healthy community dwelling adult was examined. Distribution plots for the  $\text{TMF}_2$ , SMF and UMBS 2 are provided in Figures 7.16, 7.17, and 7.18, while those for

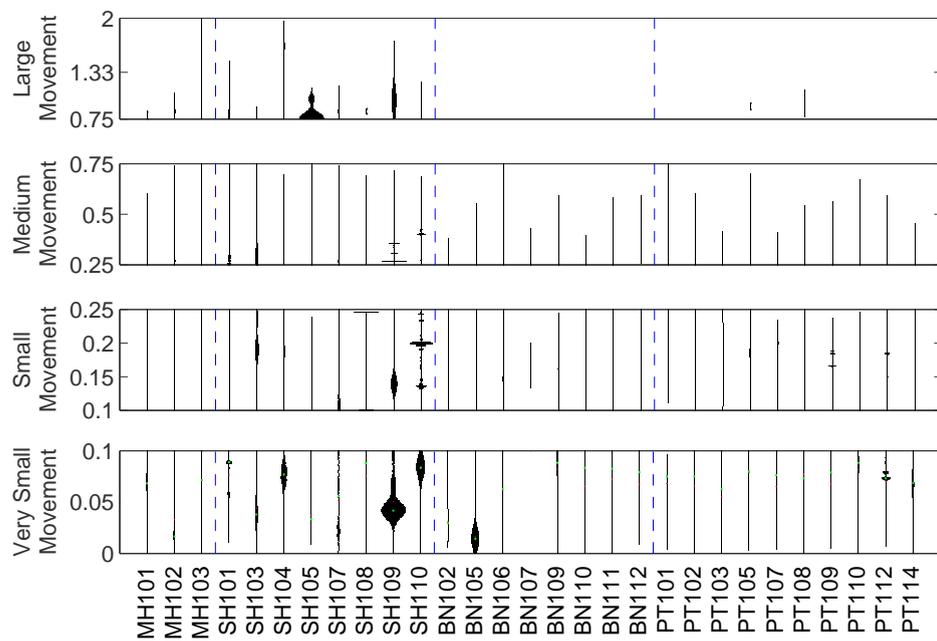


(a) Distribution of  $TMF_2$  temporal movement feature over all subjects.

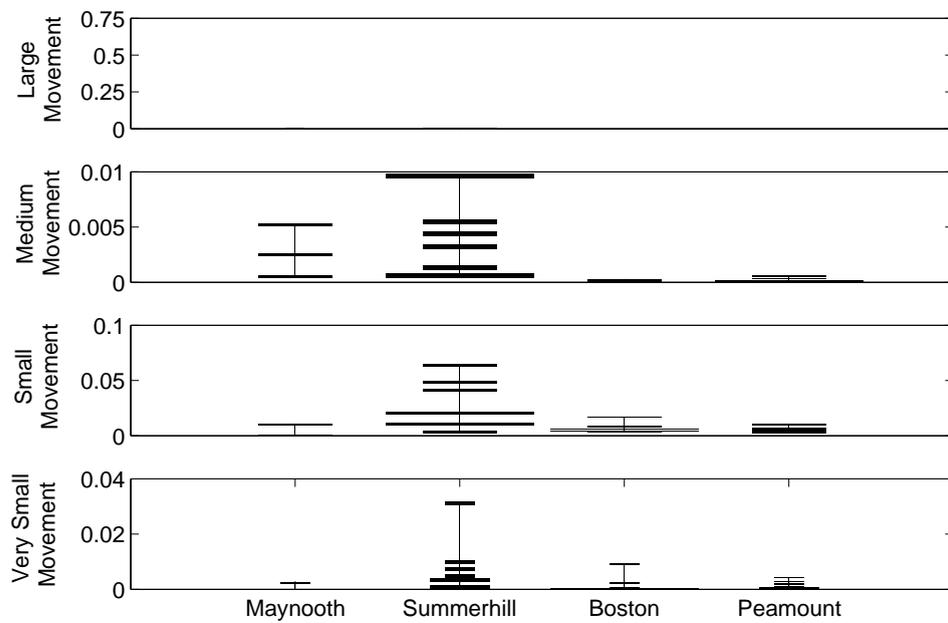


(b) Distribution of  $TMF_2$  temporal movement feature over the 4 cohorts (fraction of the sleep period). For example, for over 95% (0.95) of the sleeping episode the Maynooth participants experienced no movement. Each horizontal line represents data from a subject, wider lines indicate data from more than one subject (the width is scaled accordingly). Thicker lines represent closely aligned data collected from multiple participants.

FIGURE 7.12: Distribution of  $TMF_2$

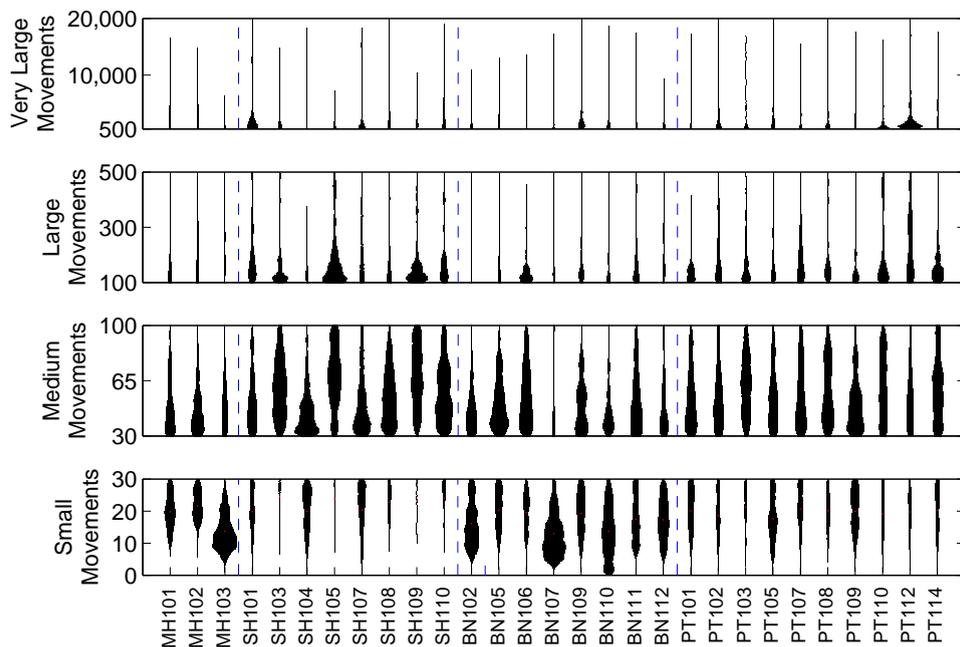


(a) Distribution of SMF spatial movement feature over all subjects.

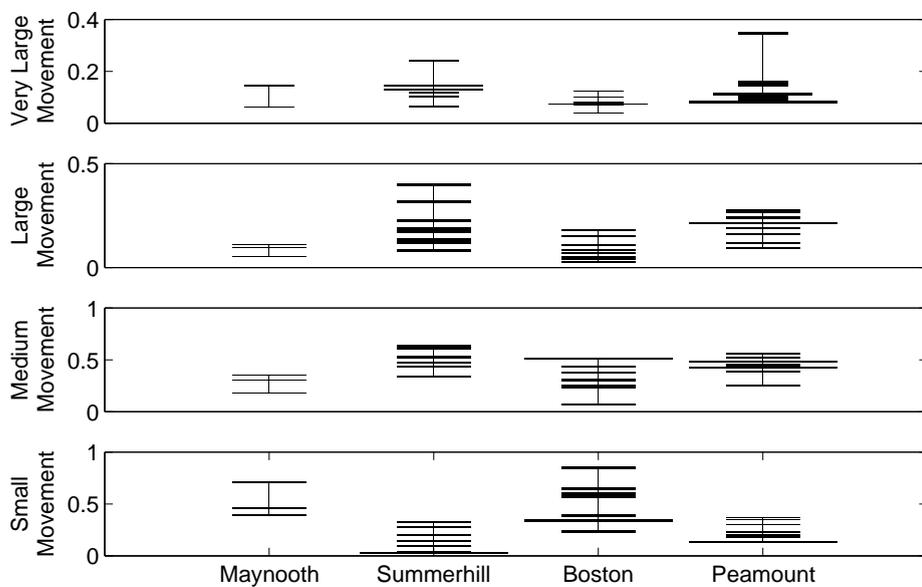


(b) Distribution of SMF spatial movement feature over the 4 cohorts (fraction of the sleep period). A significant number of SMF values equal to zero were excluded from these results.

FIGURE 7.13: Distribution of SMF



(a) Distribution of UMBS 2 statistical movement feature over all subjects.



(b) Distribution of UMBS 2 statistical movement feature over the 4 cohorts (fraction of the sleep period).

FIGURE 7.14: Distribution of UMBS 2

TABLE 7.7: Comparison of temporal, spatial, statistical and spatiotemporal movement features for each metric averaged across the four cohorts.

Movement Feature	Description	Maynooth	Summerhill	Boston	Peamount
Temporal (TMF <sub>2</sub> )	Very Large Movement	0.013	0.01	0.026	0.026
	Large Movement	0.054	0.057	0.067	0.086
	Medium Movement	0.51	0.51	0.55	<b>0.75</b>
	Small Movement	1.3	<b>18.70</b>	1.98	2.62
	No Movement	98.13	<b>80.72</b>	97.38	96.52
Spatial (SMF)	Large Movement	0	0	0	0
	Medium Movement	<b>0.27</b>	<b>0.43</b>	0.01	0.02
	Small Movement	0.67	<b>2.73</b>	0.70	0.57
	Very Small Movement	0.15	<b>0.79</b>	0.17	0.14
Statistical (UMBS 2)	Very Large Movement	11.74	13.49	7.91	13.26
	Large Movement <sup>1 2</sup>	8.50	<b>20.26</b>	8.90	<b>20.23</b>
	Medium Movement <sup>1 2</sup>	27.90	<b>51.93</b>	33.56	<b>44.13</b>
	Small Movement	<b>52.86</b>	14.32	<b>49.64</b>	22.37
Spatio-temporal	Number of Movements <sup>1 2</sup>	12.64	<b>37.62</b>	11.18	<b>38.53</b>
	Movement Area <sup>1 2</sup>	15.92	<b>32.18</b>	16.62	<b>35.28</b>
	Change in Spread	14.22	<b>63.01</b>	6.67	16.10
	Duration <sup>1 2</sup>	10.59	<b>33.21</b>	12.43	<b>43.77</b>
	Max. Movement Mag.	14.75	25.48	22.59	<b>37.18</b>
	Lateral Change in Pos.	21.05	<b>64.75</b>	3.27	10.93
	Magnitude of Movement <sup>1 2</sup>	16.32	<b>29.98</b>	18.05	<b>35.66</b>
	% to Peak Movement <sup>1 2</sup>	11.66	<b>39.43</b>	11.57	<b>37.34</b>
	Spread Movement Index <sup>2</sup>	11.71	<b>61.11</b>	8.80	18.37
Time to Peak Movement <sup>1 2</sup>	11.48	<b>34.63</b>	13.08	<b>40.81</b>	
Total		260.36	515.35	231.00	329.95

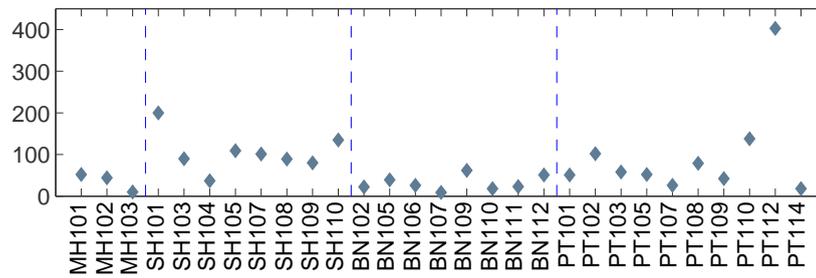
<sup>1</sup> denotes metrics for which the Summerhill (community dwelling older adults) and Peamount (clinic based sleep clinic patients) data sets are jointly high in comparison to the Maynooth (young healthy adults) and Boston (clinic based healthy young adults) data sets.

<sup>2</sup> A Student's *t*-test found a significant difference between the Maynooth and Boston cohorts against the Summerhill and Peamount cohorts ( $\rho < 0.05$ ). Data was grouped from the two separate cohorts and a Kruskal-Wallis test was used to ensure validity in grouping the data.

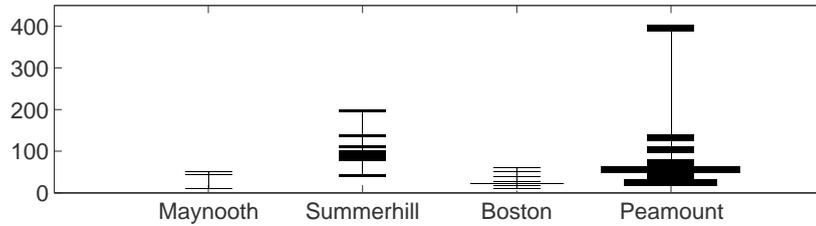
Bold denotes values which are noticeably different to others across that metric.

The temporal, spatial and statistical features sum to 100 over each cohort (vertically), however the *No Movement* case for the spatial feature ( $SMF = 0$ ) was excluded due to its abundance.

The spatiotemporal features sum to 100 across all cohorts over each feature respectively.



(a) Number of spatiotemporal movements over all subjects.



(b) Number of spatiotemporal movements over the 4 cohorts.

FIGURE 7.15: Number of spatiotemporal movements

the spatiotemporal features are included in Appendix B. SMF values equal to zero were excluded from these data as they were overly abundant in comparison to other values.

#### 7.4.4 Cohort Classification of Spatiotemporal Movement Features

Linear and quadratic discriminant classifiers were applied to the spatiotemporal movement features in order to discriminate between cohorts. 103 movements were captured from the three Maynooth participants, 836 movements from the eight Summerhill participants, 250 movements from the eight Boston participants and 969 movements from the nine Peamount participants. 101 movements were randomly chosen from each cohort and split into 51 training and 50 test data sets. Each data set contained nine spatiotemporal features: movement area, change in spread, duration, max. movement magnitude, lateral change in position, magnitude of movement, percentage to peak movement, spread movement index and time to peak movement. This process was repeated one hundred times in order to ensure consistent results and values were averaged over the 100 repetitions. Confusion matrices were generated from the test data for both the linear and quadratic methods (see Tables 7.8 and 7.9). With ideal classification, corresponding classes would report 25% as a quarter of the samples belonged to each class. The non-diagonal elements of these confusion matrices illustrate where incorrect classifications

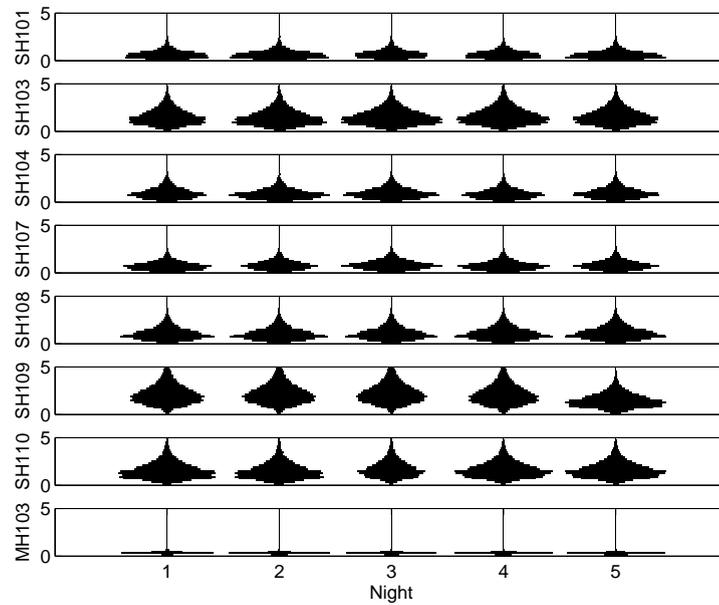


FIGURE 7.16:  $TMF_2$  CompOverMultNights

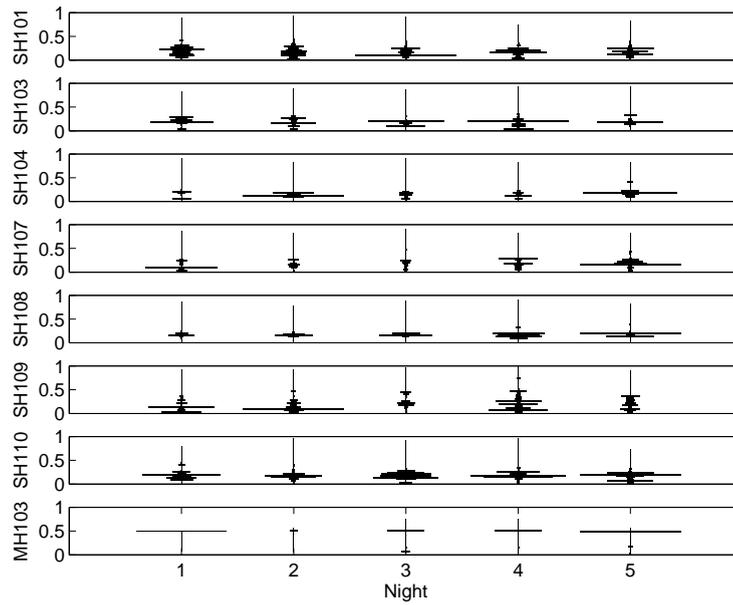


FIGURE 7.17: SMF CompOverMultNights

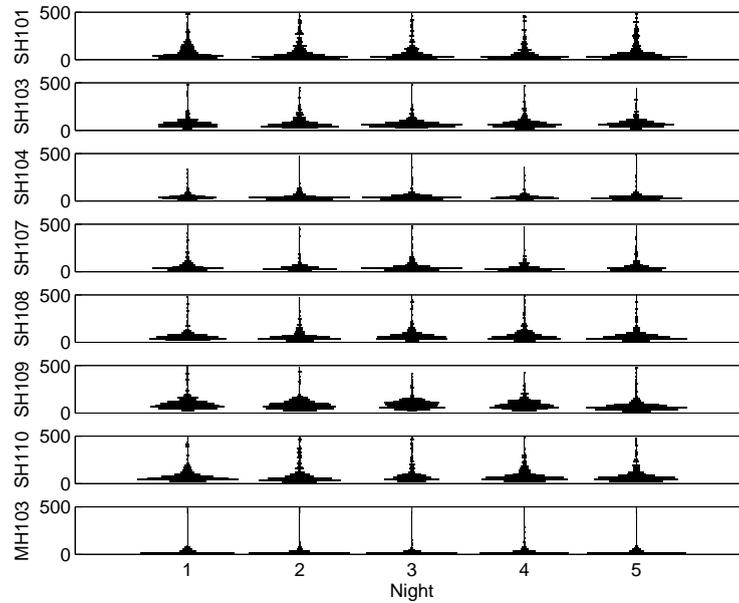


FIGURE 7.18: UMBS 2 CompOverMultNights

TABLE 7.8: Confusion Matrix discriminating cohorts using spatiotemporal features using linear discriminant analysis (percent)

		Actual Data Set			
		Maynooth	Summerhill	Boston	Peamount
Predicted Data Set	Maynooth	5.53	3.52	3.52	2.51
	Summerhill	4.52	10.55	8.54	6.03
	Boston	10.05	5.03	8.04	13.57
	Peamount	4.52	4.02	5.03	5.03

are made. For both the linear and quadratic discriminant analysis, low accuracies were reported and considerable amounts of misclassifications were made. A Normalised Misclassification Rate (NMCR) of 71% and 68% were found for the linear and quadratic classifiers respectively.

## 7.5 Discussion

### 7.5.1 Movement Features Across All Subjects and Cohorts

A comparison of features across the various cohorts was made using data from the initial night of data collection. This was performed due to the ‘*first night effect*’, whereby participants initially feel alienated by being in an artificial situation (eg. being monitored)

TABLE 7.9: Confusion Matrix discriminating cohorts using spatiotemporal features using quadratic discriminant analysis (percent)

		Actual Data Set			
		Maynooth	Summerhill	Boston	Peamount
Predicted Data Set	Maynooth	3.03	2.53	1.01	2.53
	Summerhill	6.57	6.06	3.03	2.53
	Boston	0	0.51	1.52	2.02
	Peamount	13.64	11.11	22.73	21.21

or environment (such as a non-domestic bedroom). The first night was selected as data from only one night was available in the Peamount data set (where ‘first night effects’ were deemed likely to occur).

#### 7.5.1.1 Temporal Movement Feature - $\text{TMF}_2$

The distribution of  $\text{TMF}_2$  over the four cohorts showed some distinguishing features both across subjects and between cohorts (see Figure 7.12(a)). Three of the four cohorts showed a high level of ‘no movement’ mostly ranging close to one (where on this scale one represents one hundred percent). This reflects the almost continuous occurrence of no shifts (or very small shifts, eg. breathing) in posture throughout the night. It must be noted that some range of movement ( $\text{TMF}_2$  greater than zero, eg. small, medium, large or very large movement) was reported for most subjects for some quantity of the sleeping episode. The Summerhill data set reported lower amounts of ‘no movement’ due to a much larger occurrence of small movements (also shown in Table 7.7). This equates to increased restlessness levels in the Summerhill cohort. It was expected that a similar pattern would be seen in the Peamount data set considering there is a high likelihood for sleep disorders in the Peamount data set. However, there is no such difference against the Summerhill data set. While this may suggest that the older adult Summerhill participants experience a more disturbed sleep than the Peamount population, it may be related to a possible high sleep efficiency in the Peamount population (due to sleep deprivation induced by the possible sleep disorder). Further research is necessary in order to provide a robust explanation.

It was anticipated that the clinic-based populations would have a more disturbed sleep than normal due to the artificial environment imposed upon them. The Boston (healthy clinic-based) populations reported higher levels of small, medium, large and very large movements than the Maynooth (healthy home-based) participants. However this difference was small and larger sample sizes would be required to test for significance. Unfortunately, no baseline (home-based) data set exists for the Peamount (sleep-clinic-based)

participants which would provide a direct comparison of domestic and clinic-based sleep. However, it should be noted that the medium movement levels in the Peamount data set were distinctly higher than the Maynooth, Summerhill and Boston data sets.

#### **7.5.1.2 Spatial Movement Feature - SMF**

The SMF distributions were similar across the two clinical populations (Boston and Peamount), which was unexpected as the Boston population contained a healthy cohort. A possible explanation for the small difference is the occurrence of the 'first night effect'. An examination of follow-up nights would justify this claim.

The magnitude of spatial deviations (postural changes) are much larger for the Summerhill data set than for other populations (see figure 7.13 and Table 7.7). The medium movement SMF in the Maynooth participants was much larger than either clinical population, although similar to the Summerhill data set. These patterns of larger spatial movement may be a result of feeling unencumbered while sleeping in their normal, natural, environment. Whereas subjects may be less likely to have large movements in a clinical environment while wearing polysomnographic equipment. The number of large spatial movements captured was negligible over all cohorts.

#### **7.5.1.3 Statistical Movement Feature - UMBS 2**

The statistical movement features (UMBS 2) are skewed towards lower values in both the Maynooth and Boston data sets. The Summerhill and Peamount participants report higher movement values in the medium and large ranges, although these differences are not as large compared to the relative difference seen in the spatial and temporal small movements reported (see figure 7.14 and Table 7.7). This relative similarity suggests a generalisability whereby algorithms can be designed and implemented across cohorts without requiring separate algorithms for specific populations. This suggestion is further supported by the current and successful application of sleep/wake detection algorithms in commercially available wrist actigraphs amongst various population types (apart from insomniacs where the phenomenon of sustained quiescent wake occurs at an increased rate).

#### **7.5.1.4 Spatiotemporal Movement Features**

Spatiotemporal movement features were extracted for each movement over all participants. The Summerhill and Peamount populations had much larger spatiotemporal

movement features than the Maynooth and Boston populations, notably in some specific metrics (see Table 7.7). These included:

- Number of movements
- Movement area
- Duration
- Magnitude of movement
- Percentage to peak movement
- Time to peak movement

Figure 7.15 shows the number of movements reported over all subjects and provides a description of the distribution of spatiotemporal movement features over the four cohorts. The Maynooth and Boston data sets reported lower numbers of movements compared to the Summerhill and Peamount data sets (see Table 7.7). This would indicate a more restful sleep due to the lower numbers of arousals.

An extremely large number of movements (approximately 400) was reported for one Peamount participant. Upon further inspection it was found that this individual had a large number of movements (31 per hour). A similar pattern of a high number of arousals would be expected in a subject exhibiting a large apnea/hypopnea index (as noted in Section 2.6.2.1).

The lateral change in position during movements was larger in the Summerhill population than for the other cohorts which tended to have similar ranges (see Table 7.7 and Figure A.5). Larger lateral changes in position would suggest a decrease in sleep depth (due to the larger postural shift). It would be interesting to examine the sleep/wake state during all spatiotemporal movements and investigate whether a threshold applied to the lateral change in position values discriminates between sleep stages. It may be the case that only large lateral changes in position occur during light sleep (stage 1 and stage 2).

It was also noticed that the distribution in the change in spread values is significantly larger for the Summerhill cohort than for the others (see Table 7.7 and Figure A.2). It was shown in Section 7.2.2 that the change in spread metric does not measure body posture. The significantly increased values of change in spread and lateral change in position are likely related to overall movement.

The maximum movement magnitude increased from Maynooth to Boston to Summerhill to Peamount, in that order (see Table 7.7 and Figure A.4). It is interesting to note that

the maximum movement magnitude is higher in the Peamount participants compared to the Summerhill population. This seems in conflict with the previous suggestion where a larger lateral change in position is reported in the Summerhill data set. However, a large change in spread is reported in the Summerhill data set providing a possible explanation. Additionally, it should be noted that a direct relationship need not exist between all these metrics.

Overall, distinct patterns in spatiotemporal features were found between cohorts, as well as between subjects. Using this data we see that there are cohort specific patterns, although larger studies are required for a full evaluation. Variations may also occur on an environmental basis (that is, a situational context for example older adults in a domestic setting), as well as an inter-daily basis. Environmental studies, carried out over the long-term, are difficult to implement as they must control for many variables (such as functional and cognitive capacity). However, they could be used to uncover whether the transfer from domestic to clinical (including long-term care homes) plays a positive or negative influence on sleep and sleep quality. Inter-daily variations may correspond to short term changes in the quality of sleep. A preliminary investigation of this was carried out and is discussed in the next section.

### 7.5.2 Consistency of Measurements Over Multiple Days

The variance in all movement features over multiple days was investigated for consistency. The Summerhill data set, which contained an appropriate number of multiple days data, was matched by data from the Maynooth participants. This gave a data set consisting of five nights. While long data sets were collected in the Boston cohort, only the initial three days of the study were appropriate for investigation. After this baseline period, the participants were subjected to a modified day length. These artificially imposed sleeping periods would not be appropriate as the circadian rhythm of these individuals is not in alignment with their environment. The Peamount data was also excluded as only one night for each participant was collected. The data set covered a period of 5 days.

The variance in the movement features across the five nights were found, for some participants, to be inconsistent. These included:

- SMF (see Figure 7.17)
- movement area (see Figure B.2)
- percentage to peak movement (see Figure B.8)

- time to peak movement (see Figure B.9)
- and spread movement index (see Figure B.10)

However, many metrics were found to be consistent. These included:

- $TMF_2$  (see Figure 7.16)
- UMBS 2 (see Figure 7.18)
- Number of movements (see Figure B.1)
- Change in spread (see Figure B.3)
- Duration (see Figure B.4)
- Maximum movement magnitude (see Figure B.5)
- The lateral change in position (see Figure B.6)
- Magnitude of movement (see Figure B.7)

Minimal inter-daily variations in sleeping patterns and sleep quality were reported by the participants for this period (using an informal audio questionnaire). However, natural inter-daily variations in sleep would occur. Thus, the inconsistencies in some movement metrics over the five days were attributed to this natural variation. The set of data collected from each night reflects that specific sleeping episode.

Interestingly,  $TMF_2$  and UMBS 2, temporal measurements of in-bed movement, were found to be consistent over multiple days. This suggests that these variables are stable over multiple nights and may be indicative of general sleep performance. However,  $SMF_2$ , a spatial measurement of in-bed movement, was found to be variable over the multiple sleeping episodes. This may suggest that this metric is more reflective of day-to-day changes in sleep quality. As such, two forms of sleep metrics may be derived from the UMBS: 1) a short-term highly changeable metric, derived from the inconsistent metrics, which provides a description of sleep that is highly variable between days, and 2) A stable measurement of sleep, derived from the consistent metrics, which allows a characterisation of the general aspects of an individual's sleep. A longer term study using larger and more varied cohorts is required to validate such metrics.

### 7.5.3 Spatiotemporal Movement Feature Derived Cohort Classification

The performance of LDA and QDA classifiers built to discriminate between the four cohorts using the spatiotemporal features is summarised in the confusion matrices given in Tables 7.8 and 7.9 respectively. These confusion matrices highlight where misclassifications occur, that is, they show the distribution of incorrect classes. The results are clearly very poor with a NMCR of 71% and 68% respectively for LDA and QDA demonstrating that the cohort cannot be classified effectively.

These results would suggest that each spatiotemporal description of in-bed movement does not have a cohort specific identifier or signature which discriminates between cohorts. This suggests that there is no distinct pattern in how movements occur and it is fact of their occurrence that is important.

It should be noted that the number of spatiotemporal movements was not included as a feature in this analysis. This was chosen so that the spatiotemporal features extracted from each movement, across all cohorts, could be directly compared, and not the number of movements which occurred. In this data set, three times as many spatiotemporal movements occurred in the Summerhill and Peamount data sets than in the (healthy) Maynooth and Boston data sets. Further investigation would elucidate whether the number of nocturnal movements (regardless of any feature of those movements) could be used to distinguish between healthy and unhealthy cohorts.

### 7.5.4 Non-Environmental Specific Features

The majority of features described in this chapter are environment specific, that is, they are dependent of the mattress/bed type and the weight/height/dimensions of the subject under investigation. For example, UMBS 2 is dependent on the thickness of the mattress. While the empirical selection of thresholds has catered for much of this variation, only the standardisation of the environment could ensure the absolute comparison of data from multiple subjects. It has been ensured, where possible, that similar environments were selected, especially within a cohort. Catering for specific environments is not a realistic, nor a wanted, proposition for the development of a device which will ideally be applied to large populations (and, inherently, an almost countless number of environmental variables).

However some features are non-environment specific, such as the spatial movement feature, SMF2, (as long as the threshold is selected so that subject presence is detected).

Additionally, some of the spatiotemporal movement features, including movement duration, lateral change in position, change in spread, time to peak movement, percentage to peak movement and spread movement index, were deemed to be non-environment specific.

## 7.6 Conclusions

This chapter focusses on the derivation of temporal, spatial, statistical and spatiotemporal movement features from the UMBS. These algorithms were then applied to experimental, domestic and clinic-based data sets and a comparison of the derived features across all cohorts performed. The variability observed in these features between subjects, between cohorts and across multiple days was then discussed.

In this context this chapter makes a number of novel contributions as follows.

### **The formulation and validation of algorithms deriving novel UMBS features and the application of these features on domestic and clinic-based data sets:**

Custom algorithms were developed to extract temporal, spatial, statistical and spatiotemporal ways of describing in-bed movement (Walsh et al., 2011b). These algorithms were validated using an example data set . Subsequently, these algorithms were applied to data collected in domestic and clinic-based environments from old and young cohorts ranging from healthy and relatively healthy to those at a high risk of having a sleep disorder. Multiple differences in these derived movement features were shown between the four cohorts. The combination of these features provide a rich description of the movements of individuals whilst in-bed.

The home-based older adults (Summerhill) were found to have increased movement levels (noted through multiple features) and this increase was also evident in the sleep-clinic-based population (Peamount) who have a high likelihood of suffering from a sleep disorder (most likely apnoea or hypopnoea) and thus, a more disturbed sleep. Older adults are widely documented to have a lower sleep quality than younger subjects and often report frequent awakenings, poor sleep quality and a lower sleep efficiency (Miles and Dement, 1980). In an Irish context, the Technolgical Research for Independent Living (TRIL) Centre cohort of over 600 Irish older adults, 67% were found to have poor sleep (PSQI > 5) (McHugh et al., 2011). However, similar profiles seen in these subjects indicate that community dwelling older adults suffer from sleeping disturbances to a similar degree to individuals with a potential sleep disorder (the Peamount cohort). The Peamount population were referred to the sleep clinic either through their general

physician or through a respiratory consultant. As such, their likelihood of suffering from a sleep disorder is high. A larger comparison of the sleep disturbances of community dwelling older adults compared to healthy and unhealthy (subjects with a sleep disorder) participants would elucidate this finding. Additionally, the longitudinal monitoring of these disturbances, with a possible intervention strategy through a feedback mechanism showing these disturbances back to the subject, may serve to increase the quality of life of older adults. This would be particularly applicable to older adults who tend to have a large number of complaints regarding their sleep (Miles and Dement, 1980).

Significant differences ( $\rho < 0.05$ ) were found for a number of spatiotemporal movement features (number of movements, movement area, duration, magnitude of movement, percent to peak movement, and time to peak movement) between the Summerhill and Peamount subjects when compared against the Maynooth and Boston subjects. However, the sample sizes available for this analysis is small, and the collection of larger data sets (both in terms of the number of sleeping periods and number of subjects) would allow for a more valid comparison. The results presented herein suggest an ability to discriminate between the healthy (Maynooth and Boston) and less healthy (Summerhill and Peamount) cohorts.

The statistical movement feature of the UMBS 2 metric is akin to the *activity count* metric reported by wrist actigraphy devices (however it should be noted that the UMBS 2 metric is generated from core body movement as opposed to limb movement). These metrics were previously shown to be highly comparable to each other (see Section 6.3.2). The movement detection capacity of UMBS 2 has also been shown to outperform wrist actigraphy earlier in this thesis in Section 6.10. Interestingly, relatively consistent patterns of activity were found across the four cohorts. This affirms and validates the application of wrist actigraphy across separate populations as wide documented in the literature. Categorical differences were found in the temporal movement feature,  $TMF_2$ , between the different cohorts. The difference in these movement features would suggest different information is being measured by the two metrics. While  $TMF_2$  provides a higher time resolution of movement data, this does not necessarily convert to higher resolution information. A more smoothed low frequency data (such as UMBS 2) might align better with subjective or PSG-defined objective assessment of sleep quality better. Alternatively, the opposite may be true.

The spatial movement feature (SMF) provides a spatial description of movements in-bed. Distinct differences can be seen between the home-based (Maynooth and Summerhill) and clinic-based (Boston and Peamount) populations. The movements of the clinic-based populations do not contain large spatial deviations. This may suggest that the clinic-based cohort are more constrained by sleeping within a smaller area and are less

active than the home-based populations in the home based populations. This could be a conscious factor mediated through either limited freedom of movement (through the application of multiple electrodes and sensors) or artificially self-imposed inactivity due to the desire to sleep during the assigned time. Further research should be performed in order to investigate the association between wrist actigraphy and UMBS derived movement features and subjective or PSG-defined objective assessments of sleep quality in order to ascertain the benefit of the highly sampled movement and spatial data.

Frequency analyses were applied to the TMF<sub>2</sub>, SMF and UMBS 2 data in order to investigate whether patterns existed in the occurrences of movements throughout the sleeping period. This did not show any increased performance in detecting discernible patterns across the cohorts.

**An investigation of the variation of features over multiple days:** An investigation of the consistency of movement features over multiple days was also performed. This showed both stable and variable patterns in the metrics. For example, stable patterns were found in the UMBS 2 movement feature. This was as expected as consistent patterns in wrist actigraphy data, which has been shown to be highly similar to UMBS 2, across various cohorts have been reported in the literature. However, the variability in some movement feature data suggests that data should be collected from multiple nights in order to get a valid baseline of participant's spatial and spatiotemporal movements. Further research is required in order assess whether capturing these variations adds further insight into the individual's sleep. The low level of variation between multiple days in the Summerhill data indicates that sleeping patterns are relatively consistent over time. This consistency between multiple days was expected in these data as no serious life events occurred during data collection. As this device is suited to long term placement, it would be interesting to capture data surrounding serious life events. Long term analyses and outlier detection methods applied to all of the temporal, spatial, statistical and spatiotemporal features could provide proactive and preventative methods of highlighting an individual's deteriorating health status.

**An analysis of spatiotemporal differences in movements between multiple cohorts:** Spatiotemporal descriptions of movements during sleep provides an insight into the formulation of each nocturnal movement. There were noticeable differences in the average magnitudes of these features on a per cohort basis (see Table 7.7). This prompted an investigation into whether there were differences between individual movements. Linear and quadratic classifiers were applied to the spatiotemporal features in order to discriminate between the cohorts. Interestingly, the application of classifiers

did not prove successful (see Tables 7.8 and 7.9) in discriminating between cohorts using the spatiotemporal features. It should be noted that although the sample size of cohorts were small, the number of movements in each class were relatively large. This analysis provides a basis for the suggestion that there is no difference between how these populations move in bed, and it is more the occurrence and frequency of such movements that is an important factor.

## Chapter 8

# UMBS Sleep Classification

### 8.1 Introduction

In this chapter, concomitant UMBS and PSG data collected from healthy subjects is used to develop discriminating functions which separate sleep/wake state and identify sleep stages. In order to realise this, a large data set consisting of multiple sleeping episodes from younger and older adults was required. This data set facilitates the training and testing of these functions. This chapter details the collection of this large data set, gives an overview of the features (described in the previous chapter) extracted from the UMBS data, details the performance metrics used, and reports the accuracy of dimensionality reduction mechanisms used to reduce the number of features. Multiple sleep/wake discriminating classifiers are tested on both cohort and subject specific data, and the optimal sleep/wake classifying function is validated using independent unseen data. This robust approach is employed to ensure that the developed system is applicable to individuals whom the system has not been exposed to, which in turn justifies its larger scale deployment. An investigation into identifying specific sleep stages is also carried out using multi-class classifiers and hierarchical binary classifiers.

### 8.2 Methods

#### 8.2.1 Study Protocol

Data were collected during 2 research protocols (described below) in the Intensive Physiological Monitoring Unit of the General Clinical Research Center at the Division of Sleep Medicine, Brigham and Womens' Hospital, Boston, MA, USA. Full ethical approval was granted by the Institutional Review Board at the Brigham and Womens' Hospital. These

protocols investigated the effect artificially imposed day lengths and ambient light levels have on the circadian and biological systems of healthy human adults. The time at which the scheduled sleep episode begins is referred to as *lights off* and the time at which the sleep episodes ends is referred to as *lights on*. Prior to each scheduled sleep episode, electrodes were placed on the subject in the standard montage: EEG recorded from four derivations (C3, C4, O1, O2) and EOG (LOC and ROC) both referenced to the contralateral mastoid and the submental EMG was also recorded. These PSG data were captured using a Vitaport Digital Sleep Recorder (*Temec Instruments, Kerkrade, Netherlands*). The electrodes were applied approximately two hours prior to the scheduled *lights out* and recording began approximately one hour prior to the scheduled *lights out*. The electrodes were removed after the end of the scheduled sleep episode, after *lights on*. The PSG records were scored in 30 second epochs by a trained technician using standard methods (Rechtschaffen and Kales, 1968). The PSG data were scored into one of eight states: wake, stage 1 sleep, stage 2 sleep, stage 3 sleep, stage 4 sleep, REM sleep, movement or undefined. The first six of these related to the sleep/wake stage of the subject. Movement related to periods in which the subject was clearly moving. Movement was deemed to be awake for the purposes of this research. Undefined related to periods where not enough information was available to score the data (this only occurred during application and removal of the electrodes). Thus, all epochs within *lights out* and *lights on* were either scored as sleep or wake.

The 20Hz UMBS was placed under the mattress for the duration of data collection. Data were recorded continuously over the entire study using a customised C++ program on a desktop computer (*Dell Optiplex, Dell Inc., Tx, USA*). The timestamps for each data packet was referenced to a consistent lab time held between all devices and all computers. This ensured that PSG data was concomitant.

Each participant in this protocol was subjected to prescribed periods of *lights on* and *lights off* over multiple days. The subject was instructed to sleep, or attempt to sleep, during the lights off periods. This process of scheduling the sleep and wake times was external to this analysis. An example of these periods is shown in Figure 8.1 using a double raster plot. On this type of graph, data for successive days are plotted beneath each other. The data is then offset by one day and plotted beside the first plot. This eliminates the discontinuity in the data at midnight (24:00). The subjects were housed in locations and circumstances which are specially modified to ensures that the subject will become disassociated with the outside world and with the absolute time of day. The subject is continually monitored through a team of specialist healthcare clinicians and technicians. The details of the research protocols the participants are voluntarily subjected to is given below:

**Protocol A - PPG-CSR** In this protocol, the natural nocturnal sleep length was extended to twelve hours during the initial three days of the study (see Figure 8.1). Additionally, a midday nap was scheduled however sleep was not included in any analysis. These extended sleep times and naps were inserted into the protocol in order to ensure the subject had no short term sleep debt. Over the next three days, the sleep length was reduced to approximately nine hours. Over the remaining thirteen days the day length was extended to twenty-eight hours, while the sleep length was reduced further to six and a half hours. This protocol meant that the subject became sleep restricted after the initial six days. As such the wake time during scheduled sleep episodes diminished as the protocol continued.

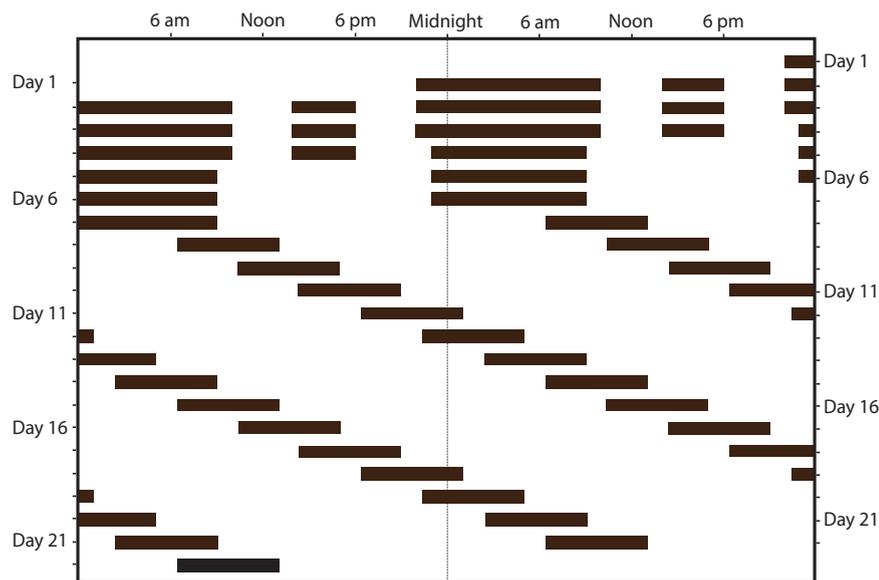


FIGURE 8.1: Double raster plot for protocol A (black represents scheduled periods of sleep).

**Protocol B - Circ-Gen** For this protocol, data were collected for over 21 days, however only data from the initial 21 days were used for analysis. Baseline data were collected during the initial 3 nights of the study (see Figure 8.2). Subsequently the subject was kept awake for a period of forty hours in order to disassociate their conscious mind from the absolute time of day. They are then allowed to sleep for approximately 10 hours. The day length was then modified to a twenty-eight hour schedule with approximately seven hours of sleep per sleeping episode. The subject was re-entrained to a sleep time aligned with environmental norms for the final two days of the research study.

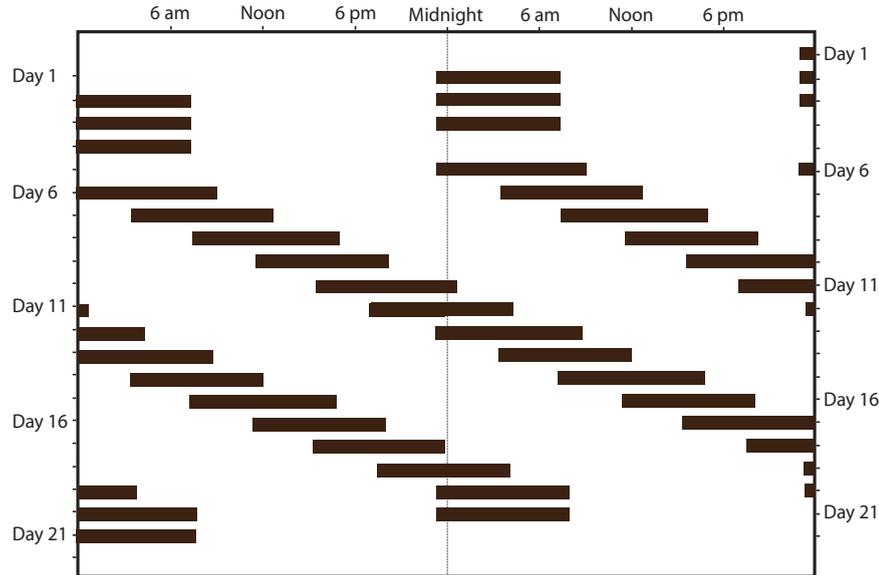


FIGURE 8.2: Double raster plot for protocol B (black represents scheduled periods of sleep).

### 8.2.2 Subject Details

Data were collected from 5 cohorts for this analysis as detailed in Tables 8.1, 8.2, 8.3, 8.4 and 8.5. The cohorts in Tables 8.1 and 8.2 were used to train and validate the classifiers. The total distribution of sleep stages per subject over the sleep episodes for these subjects is given in 8.3. The cohort in Tables 8.3 and 8.5 related to data from a group of young adults over one night and an older adult over 3 nights. These data were used to test the performance of the optimal classifier on unseen subjects and were referred to as data sets C and E respectively. The cohort in Table 8.4 did not have any corresponding PSG data. As such, its sole purpose was used to provide quantitative statistics to remove the approximate mean from the A, B, C and D data sets. It was also used for dimensionality reduction.

### 8.2.3 Data Preparation

Only data recorded between the *lights on* and *lights off* periods were included in the analysis. This eliminated all epochs of type undefined. PSG and UMBS data were analysed in non-overlapping 60 second epochs. Two consecutive 30 second epochs of PSG data were concatenated together and included in analysis only if the sleep/wake state did not change between both epochs. Otherwise they were discarded from any analysis. UMBS-derived respiration rate was estimated using the method described in Chapter 6. A sliding window is used for this analysis. As such, the initial and final two

TABLE 8.1: Data Set A - young healthy adults used for classifier training and validation.

UMBS Subcode	Study	Age yrs	Sex m/f	records scored
BN101	Protocol B	19	M	13
BN102	Protocol B	23	M	13
BN103	Protocol B	21	F	13
BN104	Protocol B	24	M	13
BN105	Protocol B	19	F	13
BN106	Protocol B	21	M	13
BN107	Protocol A	20	F	13
BN108	Protocol A	24	F	13

TABLE 8.2: Data Set B - older healthy adults used for classifier training and validation.

UMBS Subcode	Study	Age yrs	Sex m/f	records scored
BN109	Protocol A	56	F	14
BN110	Protocol A	70	M	14
BN111	Protocol A	56	M	14
BN112	Protocol A	58	F	14
BN114	Protocol A	64	M	14
BN115	Protocol A	55	M	14
BN116	Protocol A	60	F	14

TABLE 8.3: Data Set C - young healthy adults used for classifier testing. \* indicates subjects with less than twenty 60 second epochs of wake.

UMBS Subcode	Study	Age yrs	Sex m/f	records scored
BN117	Protocol A	19	F	1
BN118	Protocol A	18	F	1
BN119	Protocol A	27	M	1
BN120	Protocol A	24	F	1
BN121*	Protocol B	19	F	1
BN122*	Protocol B	24	M	1
BN123*	Protocol B	23	M	1
BN124	Protocol B	27	M	1

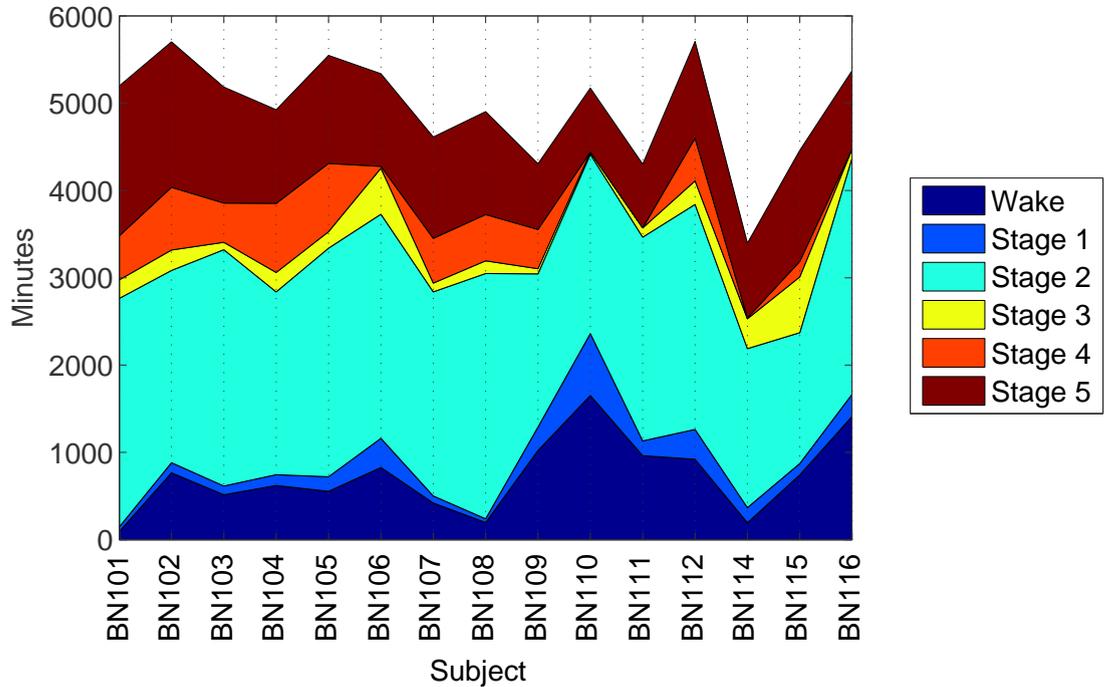


FIGURE 8.3: Distribution of sleep stages per subject over all sleep episodes. BN101-BN108 belong to the younger adult cohort, while BN109-116 are part of the older adult cohort.

TABLE 8.4: Data Set D - young healthy adults used for extracting coefficients used in pre-processing steps. No corresponding sleep stages associated.

UMBS Subcode	Study	Age yrs	Sex m/f	records scored
BN125	Protocol B	23	F	0
BN126	Protocol B	22	F	0
BN127	Protocol B	21	F	0
BN128	Protocol B	19	M	0
BN129	Protocol B	21	F	0
BN130	Protocol B	24	M	0

TABLE 8.5: Data Set E - older healthy adult used for classifier testing.

UMBS Subcode	Study	Age yrs	Sex m/f	records scored
BN113	Protocol A	55	F	3

TABLE 8.6: UMBS-Derived Features Used to Discriminate Sleep from Wake.

No.	Feature Type	Feature Description
1	Respiration	Number of Respiratory Peaks (see Section 6.2)
2	Spatial Movement Feature SMF (Eqn. 7.9)	Standard Deviation
3		Maximum
4		Mean
5		Time Greater Than 0
6		Num. Distinct Movements
7		Temporal Movement Feature
8	TMF <sub>2</sub> (Eqn. 7.5)	Maximum
9		Mean
10		Median
11		Time Greater Than 4
12		Num. Distinct Movements
13		Statistical

60 second epochs of UMBS data were excluded from classification. If any PSG or UMBS data were missing or excluded, that entire sample was removed from any analysis.

Data set D (see Table 8.4) were excluded from analysis as it did not have any corresponding PSG data. The data were used to guide statistics to give each feature an approximate zero mean. Additionally, it was used to guide the dimensionality reduction of data in the other data sets.

#### 8.2.4 Features

13 features were derived including the temporal ( $TMF_2$ ), spatial ( $SMF$ ) and statistical ( $UMBS\ 2$ ) descriptions of motion as well as an estimate of respiration rate, all derived solely from the UMBS (as given in Table 8.6). It was also found that a large number of features were correlated with each other (see Table 8.7) using data set D. As a result, it was conceivable that a number of features could be removed from the data set. However, it was unknown which features were the primary source of information. Dimensionality reduction procedures discussed later will investigate which features are redundant and the impact of removing these features on classification performance will also be evaluated

TABLE 8.7: Correlation ( $r$ ) Between the UMBS Features Averaged over the 6 Subjects in Data Set D.

		Resp. Peaks		SMF			TMF			Statistical UMBS 2		
	Peaks	St. Dev.	Max	Mean	TMT <sup>2</sup>	NCM <sup>3</sup>	St. Dev.	Max	Mean	Median	TMT <sup>2</sup>	NCM <sup>3</sup>
Resp.	1.000											
SMF		1.000										
St. Dev.	-0.145	<b>0.810</b>										
Max	-0.159	1.000										
Mean	-0.107	<b>0.923</b>	<b>0.569</b>	1.000								
TMT <sup>2</sup>	-0.094	<b>0.854</b>	<b>0.517</b>	<b>0.942</b>	1.000							
NCM <sup>3</sup>	-0.093	<b>0.847</b>	<b>0.509</b>	<b>0.937</b>	<b>0.997</b>	1.000						
TMF							1.000					
St. Dev.	-0.137	<b>0.376</b>	<b>0.606</b>	0.174	0.140	0.138	<b>0.979</b>	1.000				
Max	-0.144	<b>0.353</b>	<b>0.577</b>	0.161	0.131	0.130	<b>0.958</b>	<b>0.909</b>	1.000			
Mean	-0.120	<b>0.372</b>	<b>0.584</b>	0.182	0.137	0.134	<b>0.529</b>	<b>0.484</b>	<b>0.700</b>	1.000		
Median	0.020	0.156	<b>0.259</b>	0.070	0.027	0.025	<b>0.897</b>	<b>0.842</b>	<b>0.936</b>	<b>0.593</b>	1.000	
TMT <sup>2</sup>	-0.157	<b>0.412</b>	<b>0.618</b>	0.215	0.170	0.171	<b>0.661</b>	<b>0.622</b>	<b>0.708</b>	<b>0.494</b>	<b>0.851</b>	1.000
NCM <sup>3</sup>	-0.141	0.390	<b>0.539</b>	0.231	0.178	0.179	<b>0.638</b>	<b>0.627</b>	<b>0.623</b>	<b>0.420</b>	<b>0.663</b>	<b>0.647</b>
Stat. UMBS 2	-0.124	0.274	<b>0.463</b>	0.112	0.076	0.078						1.000

<sup>1</sup> Standard Deviation (St. Dev.)

<sup>2</sup> Total Movement Time (TMT)

<sup>3</sup> Number of Continuous Movements (NCM)

 Bold relates to correlations which are statically significant ( $p < 0.05$ )

TABLE 8.8: Confusion Matrix Showing Determination of the Number of True Positives, False Positives, True Negatives and False Negatives.

		Actual Class	
		True	False
Predicted Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

### 8.2.5 Performance Measures

The performance of the classifier can be viewed in a similar way to that of a binary classification test where four metrics capture the performance of the test. These metrics are: 1) the classifier positively identifies a true result (True Positive), 2) positively identifies a false result (False Positive), 3) negatively identifies a true result (False Negative) and 4) negatively identifies a false result (True Negative). Over all samples, the optimal classifier is the one which has a minimal number of false positives (FP) and false negatives (FN) and a maximal number of true positives (TP) and true negatives (TN). A blind approach to maximising these (see Equation 8.1) can lead to a sub-optimal selection should the data be largely skewed toward either class. In order to take account of this bias, the specificity (Equation 8.2) and sensitivity (also referred to as recall) (Equation 8.3) are often reported. These metrics quantify the correct classification of false and true samples respectively. A common approach for choosing a classifier with maximal results involves maximising the F-Score (Equation 8.5) which is the harmonic mean of the recall (Equation 8.3) and the precision (Equation 8.4) of the test. The precision of the test is the number of correct results classified as a ratio of all results the classifiers reports as being correct. The use of the harmonic mean ensures that the result is heavily skewed toward the lower value and this avoids the situation where there is a large difference between the precision and recall values.

For cases where there are more than two classes, the overall accuracy of the test is often used. This is the ratio of correct classifications the algorithm makes over the total number of samples in that test.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8.1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (8.2)$$

$$Recall = Sensitivity = \frac{TP}{TP + FN} \quad (8.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (8.4)$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8.5)$$

## 8.2.6 Feature Reduction Mechanisms

A high correlation was found between a number of the thirteen original features. As such, three dimensionality reduction methods were investigated in order to remove any redundant features and to simplify the classification process.

### 8.2.6.1 Principal Component Analysis

Principle Component Analysis (PCA) orthogonally transforms data from a number of possibly correlated variables into a series of uncorrelated variables known as principal components (Jolliffe, 2004). Each principal component captures the largest variance in the data under the condition that it is orthogonal to the preceding principal components. Thus, a series of principal components are calculated which capture the variance (or internal structure) of the original data set. The variance of each principal component decreases with each successive principal component calculated. PCA was applied to this data and it was also found that over 99.5% of the variance in data set D could be explained using the first 4 principal components. While the top three principal components explained over 96.5% of the variance in this data set. Data set D (see Table 8.4) was used to calculate the coefficients which would be applied to the remaining data sets in order to reduce their dimensionality. This dimensionality reduction procedure is not ideal as the coefficients used do not pertain to the data set under investigation. However, this procedure was used in order to avoid any overfitting on the training and test data.

### 8.2.6.2 Forward Selection Component Analysis

Forward Selection Component Analysis (FSCA) arranges signals in successive descending order of how each signal describes the overall variance of the data set (Prakash et al., 2012), as per Algorithm 8.1. Initially, the signal which explains the largest variance on the data set is chosen (step 2) and the effect of its variance is removed from the data set (step 3). This process is repeated until either the last signal is reached or a stopping criterion (such as explaining 99.9% of the variance in the data set). Thus, a prioritised list of variables that are most representative of the data set, of decreasing contribution, is produced.

**Algorithm 8.1** Forward Selection Component Analysis Algorithm

1. Given a data set  $\mathbf{X}$ , set  $\tilde{\mathbf{X}} = \mathbf{X}$ .
2. Select  $\tilde{\mathbf{x}}_i$  (the Forward Selection Component)  $\in \tilde{\mathbf{X}}$  such that  $\tilde{\mathbf{x}}_i^* = \arg \min_{\tilde{\mathbf{x}}_i \in \tilde{\mathbf{X}}} \left( \left\| \tilde{\mathbf{X}} - \hat{\mathbf{X}}(\tilde{\mathbf{x}}_i) \right\|_F^2 \right)$  where  $\hat{\mathbf{X}}(\tilde{\mathbf{x}}_i) = \tilde{\mathbf{x}}_i \frac{\tilde{\mathbf{x}}_i^T \tilde{\mathbf{X}}}{\tilde{\mathbf{x}}_i^T \tilde{\mathbf{x}}_i}$ ,  $i = \text{index}(\tilde{\mathbf{X}}, \tilde{\mathbf{x}}_i^*)$
3. Remove contributions of  $\tilde{\mathbf{x}}_i^*$  from  $\mathbf{X}$   
 $\tilde{\mathbf{X}} = \left( I - \frac{\tilde{\mathbf{x}}_i^* (\tilde{\mathbf{x}}_i^*)^T}{(\tilde{\mathbf{x}}_i^*)^T \tilde{\mathbf{x}}_i^*} \right) \tilde{\mathbf{X}}$
4.  $S = [S, \tilde{\mathbf{x}}_i^*]$ ,  $Q = [Q, i]$   
 where S is the set of orthogonal Forward Selection Component(s) (FSC) components, prioritised in decreasing contribution, and Q is the set of corresponding indices.
5. Repeat from 1 until a stopping criterion is reached, or until there are no variables left.

TABLE 8.9: Top four FSCs.

FSC No.	Feature No.	Feature
1	1	Estimated respiration rate
2	6	The SMF defined number of distinct movements
3	8	The max of the TMF
4	12	The TMF defined number of distinct movements

In this instance, FSCA was employed as a means of selecting a minimal number of original features which explain a large proportion of the variance in the data set. If there is high redundancy in the data, that is a large number of highly correlated signals, then FSCA can achieve similar performance to PCA. PCA will always outperform or equal FSCA in the explanation of variation on a per feature basis. However, the principal components generated often do not relate to any tangible metric, but rather are a weighted mixture of several of the input features. The use of the raw input features is sometimes preferential when trying to understand the system and how the actual inputs relate to the output state or when trying to reduce the computational load as only a low number of original features may need to be measured or computed.

Training and test data (data sets A and B) were analysed using FSCA. It was found that the four top FSC per subject explained over 97% of the variance in each subject's case. In 95% of these cases, it was found to be four features as defined in Table 8.9.

### 8.2.6.3 Feature Subset Selection

Feature subset selection is a method for determining the best subset of features which results in the best discrimination of classes. It consists of both a search strategy and an objective function (Cantu-Paz et al., 2004). The search strategy selects candidate features iteratively while the objective function evaluates each selection. The search

strategy can be exhaustive where all possible subsets of candidate features are chosen. However, there are  $2^N$  possible subsets of candidate features, where  $N$  is the total number of features (if the optimal number of selected features is unknown), making it computationally unfeasible as  $N$  becomes large. Various strategies have been developed to minimise the number of searches while maintaining maximal performance of the classifier. The objective function is a measure of the "goodness" of the chosen subset of features. The objective function can be split into two groups: 1) filters, and 2) wrappers. Filters evaluate the chosen subset of features based on their information content (often using distance metrics, correlation metrics or, alternatively, non-linear approaches such as the determination of mutual information between features). Filters are quick to execute and their results are very generalisable as the intrinsic properties of the chosen subset of features is directly related to the output or class. However, there is a tendency for filters to select large subsets of features (Kohavi and John, 1997). Wrappers use a classifier each time to evaluate the performance of the selected subset of features (using accuracy, F-score or other similar measures). Wrappers tend to have higher accuracies than filters as the classifier is included in the process of choosing the optimal subset of features (Kohavi and John, 1997). Cross-validation is generally used to avoid any cases where over-fitting occurs ensuring high generality. Wrappers tend to have longer execution times than filters. This is because the classifier must be trained upon each iteration. The inclusion of the classifier means that the subset of features chosen are only optimal for that classifier and thus the chosen features might be different should another classifier be used. This may lead to a reduction in the overall generality of the solution.

In this analysis, the Sequential Forward Selection (SFS) wrapper method (Kohavi and John, 1997) (see Algorithm 8.2) was investigated as a means of reducing the number of features. This method begins with an empty subset of features and iteratively adds one feature to the subset. At each iteration, the feature which maximises a pre-defined cost function is added to the subset of features. This process is repeated until the addition of any new feature does not result in an increase in accuracy. The cost function is a performance metric which tests the ability of the classifier to correctly label the inputs using the chosen subset of features. For this analysis, the cost function used was  $(1 - \text{F-score})$  as this caters for biased data in addition to placing emphasis on reducing the number of false positives and false negatives. The main drawback of this algorithm is that it is unable to remove features which have become obsolete after the later addition of features. Additionally, this algorithm might not necessarily converge on the optimal selection of features which an exhaustive search would uncover. Other wrapper methods have also been developed, such as Sequential Backward Selection, Bidirectional Search and Sequential Floating Selection, however only SFS was investigated in this analysis.

---

**Algorithm 8.2** Sequential Feature Selection Algorithm

---

1. Start with an empty subset of features,  $Y_0 = \{\emptyset\}$  and  $k=1$ .
2. Select the feature which maximises the performance of the classifier,

$$x^a = \arg \max_{x \in Y_k} J(Y_k + x)$$

3. Add  $Y$  to include  $x^a$  and increment  $k$ ,  $Y_k = Y_{k-1} + x^a$  and  $k = k + 1$
  4. Repeat from step 2 until the performance of the classifier does not improve.
- 

### 8.3 Sleep/Wake Classification

Data from two types of participants were collected in this study. They were used to develop classifiers which would discriminate between sleep and wake on younger adults (data set A), older adults (data set B) and a combined data set of both younger and older adults. For this analysis, sleep was defined as any of the five sleep stages: stage 1, stage 2 stage 3, stage 4 or stage 5 sleep. The ability of a classifier to successfully discriminate between sleep and wake was tested on both a per cohort (cohort specific classification) and a per subject (subject specific classification) basis. The optimal classifier type was chosen from five types of classifiers: 1) LDA, 2) QDA, 3) kNN, 4) ANN, and 5) SVM with various internal configurations as shown in Table 8.10. A brief explanation of these is given in Chapter 3.3.

LDA and QDA classifiers were trained using a proportion of each data set as training data to divide the  $n$ -dimensional space into linearly or quadratically defined subspaces which optimally discriminate between sleep and wake. The performance of each classifier was then found by dividing the remaining data (the validation data set) into these subspaces and then calculating the number of correct classifications.

Three internal states of the kNN classifier were chosen: 1)  $k=1$ , 2)  $k=5$ , and 3)  $k=9$  where  $k$  represents the number of nearest neighbours required to classify a sample as belonging to one state or another. The euclidian distance between each test sample and each training sample over all dimensions was used as the distance metric to find the nearest neighbours. The distance from the test sample to all samples in the training data set was computed. The test sample was labelled with the majority class of the nearest  $k$  training set samples. As such, kNN is a non-linear classifier which creates a non-linear boundary which discriminates between states. The effect of noise on the accuracy of the discriminating hyperplane is greater for lower values of  $k$ . Values of  $k$  which are too high can create a boundary which is too smooth (and not specific enough to the data set). A preliminary investigation determined that the optimal  $k$  value was less than 10.

Two internal states for the ANN classifier were chosen: 1) a ANN with 20 neurons in the hidden layer, and 2) a ANN with 40 neurons in the hidden layer. Both of these networks had one neuron in the output layer, with high and low values reporting each state (wake and sleep). There are several options for optimising the topology of a neural network, for example varying the number of hidden layer neurons or even the number of hidden layers. Due to time constraints, only the two options listed above were included in this analysis.

A lengthy optimisation procedure was followed for the SVM whereby the internal states for sigma ( $\sigma$ ) and the box constraint (C) were optimised. The box constraint is a parameter which controls the level to which misclassifications can occur. If this is kept too strict and no misclassifications are allowed to occur this can result in a loss of generality.  $\sigma$  is a parameter which controls the potential minimum width between classes (or the margin). The non-linear kernel used for this analysis was the RBF. Initially a large once-off sweep was performed using a large range of values for sigma and the box constraint. It was found during a preliminary analysis that the optimal performance (as noted using *F-score*) would be found for sigma ( $\sigma$ ) values in the range 0.01 to 4 and box constraint (C) values in the range 0.01 to 20. This SVM optimisation procedure is an exhaustive search and as such is computationally inefficient. Accordingly, the sigma and box constraint parameters were increased by 0.5 units per iteration in order to reduce the running time of this optimisation process. The optimal classifiers were chosen based on their performance (as measured using F-score, see Equation 8.5) when trained and tested on a per cohort (young, old, all) basis and a per subject basis.

The classifiers were trained and tested using the original features and the Principle Component(s) (PC) derived from the features. PCA was used to reduce the dimensionality of the data, and FSCA (as described in Section 8.2.6) was used to select the signals which give the highest contribution to the variation in the data set. SFS was applied in the cohort specific classification as a large number of samples existed for each cohort under analysis. It was felt that SFS would not be appropriate in the subject specific classification case due to the low number of wake samples available.

### 8.3.1 Cohort Specific Classification

Three data sets were assembled and split into training and validation data for cohort specific classification (see Table 8.10). The first data set contained only the younger adults (data set A), the second data set contained only the older adults (data set B) while the third data set contained both cohorts (data set A and data set B). This data was then used to train a number of classifiers (listed above) and provide performance

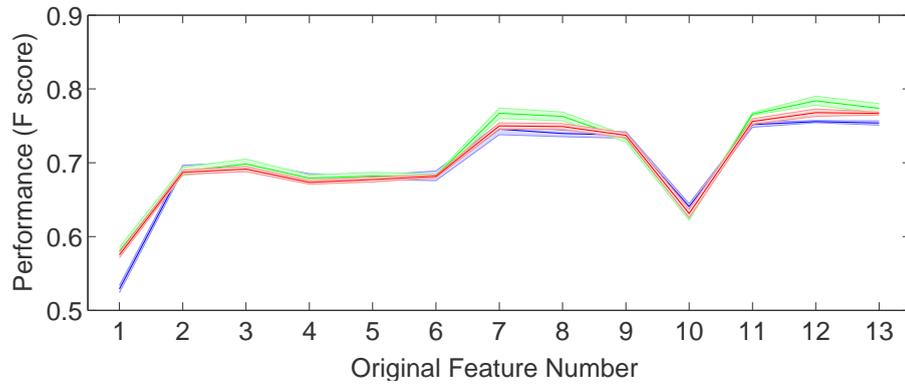
results based on the validation data. The optimal classifier was later tested on unseen data (data sets C and E). In most cases a large proportion of samples were used to train the classifier, while the remainder validated the performance of that classifier. However, a significantly lower number of samples were used to train the SVM classifier as its processing time increases exponentially with the number of training samples (Joachims, 1999) and also due to the tuning process for selecting the optimal internal parameters. Thus, a trade-off between the search for optimal internal configuration values and the duration of the optimisation procedure was made.

### **8.3.1.1 Classifier Training and Validation Preprocessing Procedure**

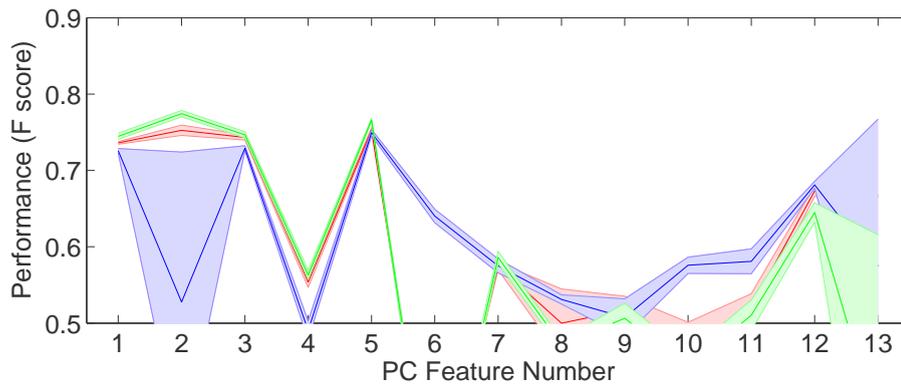
Initially, the set of 13 input features along with their corresponding sleep or wake stage were split into subsets of sleep and wake epochs on a per subject basis. 400 epochs of sleep and 400 epochs of wake randomly chosen from each person were then selected for classification. For two of the younger subjects (BN101 and BN108) and one older adult (BN113), random sampling with replacement was used to select the wake epochs as there were less than 400 wake epochs available (95, 198 and 196 wake epochs available respectively as shown in Figure 8.3). For all other subjects and for the selection of sleep epochs, sampling without replacement was used as the number of available samples allowed it. These data were then grouped into three sets: data set A) younger adults, data set B) older adults or data set AB) all subjects. Each data set consisted of an equal proportion of wake and sleep epochs per subject. The PCs of this data was then approximated using the coefficients derived from data set D. A separate data set of younger adults were used to approximate these coefficients in order to avoid overfitting. This process resulted in 13 original and 13 principal component features available for classification.

### **8.3.1.2 Single Feature Classification**

The ability of each feature taken individually to discriminate between sleep and wake using LDA was also investigated (see Figure 8.4). The number of samples used to train and validate the classifier was 4266 and 2134 for the younger adult cohort, 3732 and 1868 for the older adult cohort and 8000 and 4000 for the younger and older cohort respectively. As above, 400 epochs of wake and 400 epochs of sleep were randomly chosen from each subject. The results were generated for the younger adult, older adult and all subject cohorts. This process was repeated five times in order to examine the consistency of the results. Minimal standard deviations in the classification performance



(a) Performance of LDA classification using the original features.



(b) Performance of LDA classification using the PC features.

FIGURE 8.4: Performance of each single feature in discriminating between sleep and wake over the three cohorts: younger adults (green), older adults (blue) and all adults (red). The standard deviation in the performance over 5 repetitions is shaded in.

occurred when using the original features as inputs, whereas this standard deviation was dramatically larger for some of the PCs in some of the cohorts.

### 8.3.1.3 Multi-Feature Classification

Various classifiers were trained and validated on all the PC and/or using all of the original features as combining information for multiple features would be likely to increase performance. Conversely, some features might not provide any additional capacity over other features in discriminating between sleep and wake. As such, dimensionality reduction techniques were applied. If successful, this procedure would also serve to reduce the computational load of the final algorithm and to reduce the number of features which need to be calculated. Classifiers were trained and validated using all the PC or using all the original features, or with data containing the top three and four PCs as well as the top three and four FSC. The classification procedure was repeated multiple times using

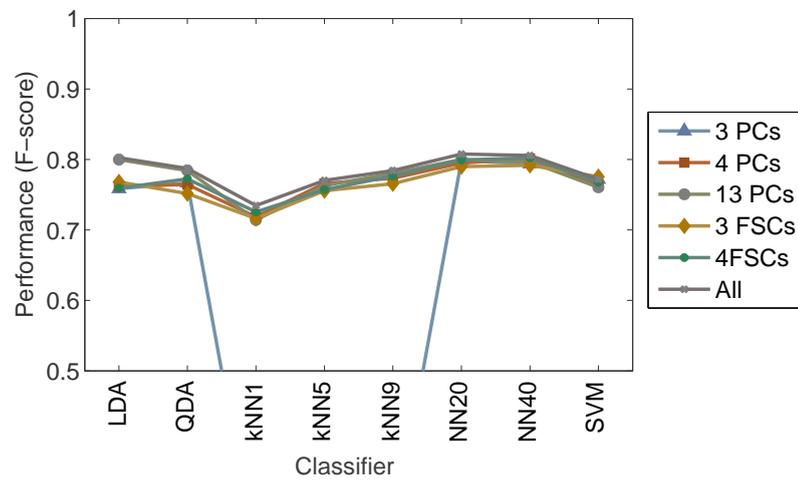
different randomly selected samples in order to ensure consistent results (also known as cross validation). The number of repetitions as well as the number of training and test samples for each classifier and each cohort analysed is given in Table 8.10.

TABLE 8.10: Classifiers and samples sizes used for sleep/wake discrimination.

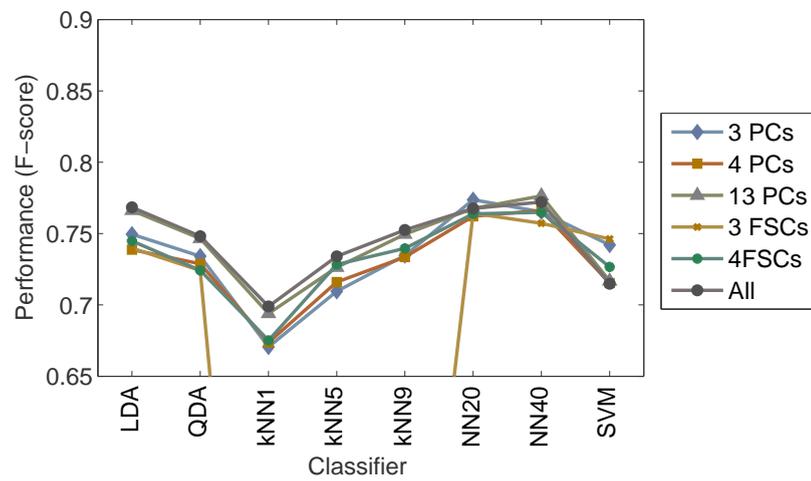
Classifier	No. Samples - GSC							No. Samples - SSC		
	NR	A		B		AB		NR	All Subjects	
		Tr	Va	Tr	Va	Tr	Va		Tr	Va
LDA	5	4266	2134	3732	1868	8000	4000	5	532	268
QDA	5	4266	2134	3732	1868	8000	4000	5	532	268
kNN1	5	4266	2134	3732	1868	8000	4000	5	532	268
kNN5	5	4266	2134	3732	1868	8000	4000	5	532	268
kNN9	5	4266	2134	3732	1868	8000	4000	5	532	268
NN20	5	4266	2134	3732	1868	8000	4000	5	532	268
NN40	5	4266	2134	3732	1868	8000	4000	5	532	268
SVM	5	320	6080	280	5320	300	11700	5	400	400

Group Specific Classification (GSC), Subject Specific Classification (SSC), Number of Repititions (NR), Number of Training Samples(Tr), Number of Validation Samples (Va),  $k$ -Nearest Neighbour where  $k=x$  ( $k$ NN $x$ ), Neural Network with  $y$  neurons in the hidden layer (NN $y$ ).

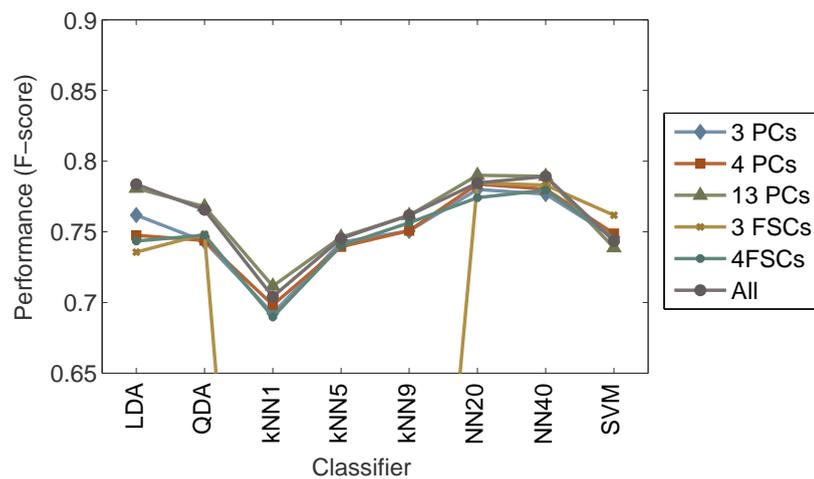
The mean performance of the classifiers (F-score, see Equation 8.5) over all repetitions is reported in Figure 8.5 and also in Tables 8.11, 8.12 and 8.13 for the younger adults, older adults and all subjects respectively. The internal configurations (sigma,  $\sigma$ , and the box constraint, C) for the SVM were tuned and the results from the optimal configuration reported (see Figure 8.6). The performance score values had a standard deviation of less than 0.025 over all repetitions. The three optimal classifiers (C-younger, C-Older and C-all) for each cohort (younger, older and all subjects respectively) reported a high performance (F-score) along with a low standard deviation.



(a) Younger adult cohort



(b) Older adult cohort



(c) Entire cohort

FIGURE 8.5: Performance of the cohort specific classification over the three cohorts.

TABLE 8.11: Performance (F-score) of the classifiers in discriminating sleep and wake over the younger adult cohort

	3 PCs	4 PCs	13 PCs	3 FSCs	4 FSCs	13 FSCs
<b>LDA</b>	0.768	0.762	0.802	0.758	0.759	0.799
<b>QDA</b>	0.751	0.764	0.787	0.766	0.772	0.784
<b>KNN 1</b>	0.716	0.718	0.734	0.232	0.725	0.714
<b>KNN 5</b>	0.755	0.765	0.770	0.245	0.756	0.762
<b>KNN 9</b>	0.765	0.773	0.784	0.266	0.776	0.780
<b>NN 20</b>	0.789	0.794	0.807	0.798	0.799	0.800
<b>NN 40</b>	0.791	0.801	<b>0.805</b>	0.793	0.802	0.796
<b>SVM</b>	0.775	0.762	0.755	0.771	0.767	0.760

All performance scores had a standard deviation of less than 0.025 over all repetitions.

TABLE 8.12: Performance (F-score) of the classifiers in discriminating sleep and wake over the older adult cohort

	3 PCs	4 PCs	13 PCs	3 FSCs	4 FSCs	13 FSCs
<b>LDA</b>	0.749	0.738	0.766	0.739	0.744	0.768
<b>QDA</b>	0.734	0.728	0.746	0.724	0.724	0.748
<b>KNN 1</b>	0.670	0.673	0.693	0.214	0.675	0.698
<b>KNN 5</b>	0.709	0.715	0.72	0.312	0.728	0.734
<b>KNN 9</b>	0.734	0.733	0.749	0.353	0.739	0.752
<b>NN 20</b>	0.773	0.762	0.767	0.764	0.763	0.767
<b>NN 40</b>	0.765	0.766	<b>0.776</b>	0.757	0.764	0.772
<b>SVM</b>	0.742	0.715	0.716	0.746	0.726	0.714

All performance scores had a standard deviation of less than 0.025 over all repetitions.

TABLE 8.13: Performance (F-score) of the classifiers in discriminating sleep and wake over the entire cohort

	3 PCs	4 PCs	13 PCs	3 FSCs	4 FSCs	13 FSCs
<b>LDA</b>	0.761	0.747	0.780	0.735	0.743	0.783
<b>QDA</b>	0.743	0.743	0.767	0.748	0.747	0.765
<b>KNN 1</b>	0.692	0.698	0.711	0.167	0.689	0.704
<b>KNN 5</b>	0.742	0.739	0.746	0.229	0.740	0.745
<b>KNN 9</b>	0.750	0.750	0.761	0.256	0.756	0.761
<b>NN 20</b>	0.780	0.783	<b>0.790</b>	0.784	0.774	0.784
<b>NN 40</b>	0.776	0.780	0.789	0.783	0.779	0.789
<b>SVM</b>	0.746	0.748	0.738	0.761	0.745	0.743

All performance scores had a standard deviation of less than 0.025 over all repetitions.

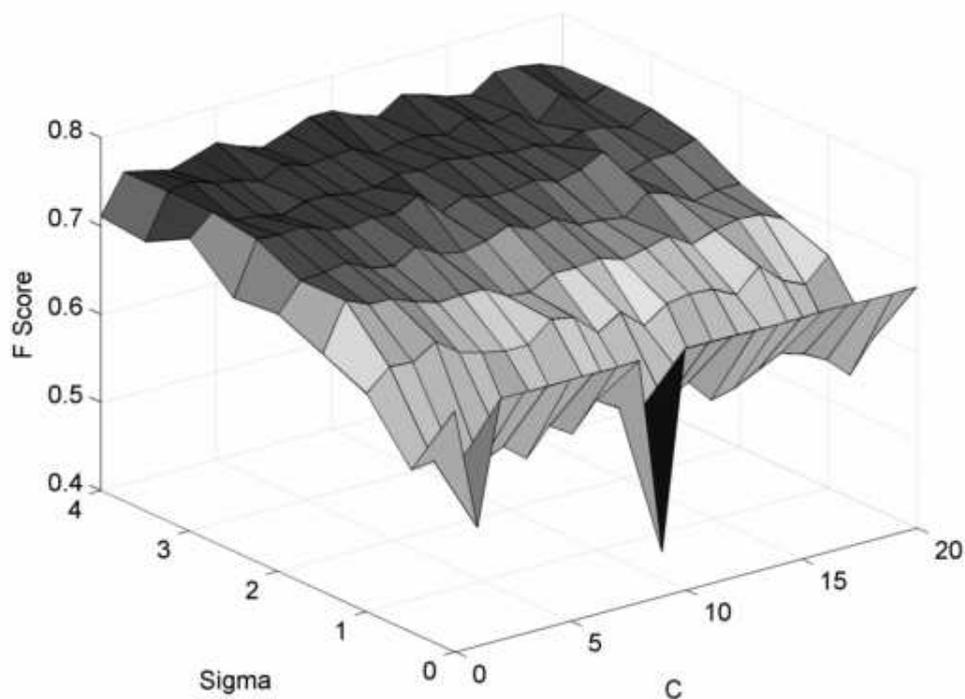


FIGURE 8.6: SVM Performance over all box constraint ( $C$ ) and sigma ( $\sigma$ ) values applied to the younger adult cohort (cohort specific classification).

#### 8.3.1.4 Sequential Feature Selection

SFS in the forward direction was applied to the younger, older and entire data sets using either the original features or the 13 approximated PCs. LDA was used as the classifier in this analysis as it's more computationally efficient (more suitable for implementation on low complexity systems) and its performance only slightly inferior to the optimal classifier for each cohort (NN20 for the younger cohort, NN40 for the older cohort, and NN40 for the entire cohort all using the 13 original features). A preliminary investigation using QDA was carried out and classification using LDA reported a slightly higher performance. In all cases and for each number of selected features, the original features outperformed the PCs consistently (see Figure 8.7). A list of the chosen features per cohort is given in Table 8.14, the numbering of these features is taken from Table 8.6.

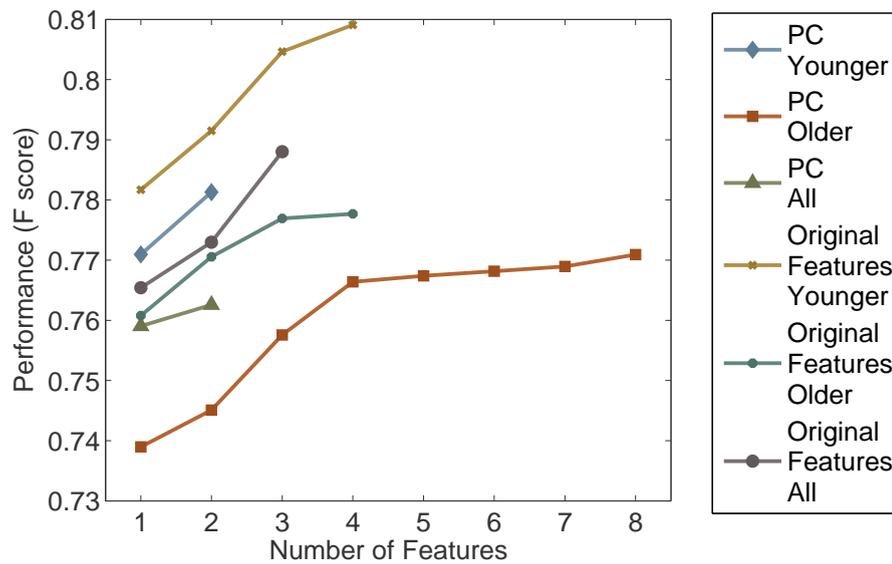


FIGURE 8.7: Sequential feature selection performance per number of input features.

TABLE 8.14: Sequential feature selection chosen features per cohort.

Features	Cohort	Chosen Features	Performance
PCs	Younger Adult	2, 11	0.781
	Older Adult	5, 2, 1, 7, 9, 8, 3, 11	0.770
	All Subjects	2, 11	0.761
Original	Younger Adult	12, 13, 10, 1	0.809
	Older Adult	12, 13, 10, 2	0.777
	All Subjects	12, 13, 10	0.788

Feature numbers relate to the features as per Table 8.6.

Additionally, the 13 original and 13 PCs were combined and SFS was applied. No greater performance was found when using these 26 features. The majority of features came from the subset of the original 13 features and additionally the effect of the PC features in increasing classifier performance was minimal. This was consistent with the results found above where all 13 original and 13 PC features were taken together and a classifier trained and validated. Additionally it should be noted that no difference in performance should be expected as the PC should be transformed versions of the original features (however this may not be the case the PCA transformation matrix is derived from data set D). No increase in performance in taking the original or PC features separately was found.

TABLE 8.15: Best group specific classifier.

Cohort	F-score	Sens.	Spec.	Classifier	Details
Younger	0.807	0.724	0.880	LDA	SFS, Features: 12, 13, 10, 1
Older	0.772	0.690	0.840	LDA	SFS, Features: 12, 13, 10, 2
All	0.790	0.876	0.657	NN20	All 13 PCs

TABLE 8.16: Optimal group specific classifier.

Cohort	F-score	Sens.	Spec.	Classifier	Details
Younger	0.799	0.902	0.645	LDA	All Original Features
Older	0.768	0.856	0.627	LDA	All Original Features
All	0.783	0.880	0.633	LDA	All Original Features

### 8.3.1.5 Optimal Classifier Selection

The optimal classifier over each cohort is given in Table 8.15. However for these classifiers, the sensitivity values are lower than most commercial sleep/wake detection products, conversely the specificity is much higher for the SFS classifiers. In the non-SFS classifiers, this is reversed which is more preferable as the subject is in the true positive (sleep) state more often throughout the night (and correspondingly this will result in a higher overall accuracy). As such, the non-SFS classifiers were chosen. The optimal non-SFS classifier was a neural network as per Tables 8.11, 8.12 and 8.13. However the performance of an LDA classifier (see Table 8.16) was slightly lower (by less than 0.008) and within one standard deviation, but had very high sensitivity values and adequate specificity values (making it more suitable due to the bias typical of the bias in the data from each night's sleep). Additionally due to the simpler implementation in low-level hardware of an LDA classifier, it was chosen as the optimal classifier for sleep/wake classification. Low standard deviations (less than 0.02) were found for the F-score, sensitivity and specificity.

### 8.3.1.6 Cohort Independent Classifier Testing

The younger adult classifier, the older adult classifier and the all subject classifier with the optimal performance (as described in Table 8.16) were applied to data from an unseen cohort of 5 younger adults over one night (data set C), to data from an unseen cohort of one older adult over three nights (data set E) and to a subset of data from both of these cohorts. This process provides an independent testing of the optimal classifier. The size of these independent testing data sets (data sets C and E) are relatively small,

however an independent report of the performance of each classifier using this data is still valid. Equal proportions of sleep and wake were randomly selected from each of these cohorts ensuring a valid comparison of results from training, validation and testing data.

**Younger Adult Testing Cohort** For the younger adult cohort, fifty samples of sleep and fifty samples of wake were randomly selected from each of the five subjects in data set C and compiled into the younger adult testing data set. Random sampling with replacement was used to select the fifty wake epochs due to the low number of wake epochs available in this data set. A much lower number of samples was used here compared to the training and validation stage as the number of wake epochs was below 50 for subjects BN117 and BN120 (23 and 25 epochs respectively). The subjects were only awake for 13% of the recordings. Random sampling without replacement was used to select fifty of the sleep epochs in this data set. The subsets of chosen sleep and wake epochs from all subjects were combined to creating a set of 500 chosen epochs.

**Older Adult Testing Cohort** For the older adult cohort, 200 samples of sleep and 200 samples of wake were randomly selected for each subject over the three nights and compiled into the older adult testing data set. Random sampling without replacement was used to select the sleep and wake epochs in this data set as the data set contained over 41% wake epochs. The subsets of chosen sleep and wake epochs from all sleeping episodes were combined to create a set of 400 chosen epochs.

**All Subject Testing Classifier** The data from one sleeping episode from one younger subject (BN119) and one older subject (BN113) were compiled into a testing set for older and younger subjects. The data from the younger subject contained 113 wake epochs and 472 sleep epochs. The data from the older subject contained 264 wake epochs and 233 sleep epochs. 200 samples were randomly selected, without replacement where possible, to create a data set of 400 sleep epochs and 400 wake epochs. This data set contained an equal contribution of data from younger and older adults.

The mean performance (F-score) of each of the three classifiers applied to the three data sets above over 10 trials is reported in Table 8.17. The standard deviation of the performance of each classifier was less than 0.025. The F-score, sensitivity, specificity and accuracy of each classifier on their respective cohort is given in Table 8.18. A comparison of the performance of the classifiers during training, testing and validation on their respective data sets is given in Table 8.19.

TABLE 8.17: Testing performance of the optimal cohort specific classifiers on all three cohorts.

		Cohort		
		Younger Adult	Older Adult	All Subjects
Classifier	Younger Adult	0.758	0.867	0.801
	Older Adult	0.750	0.813	0.776
	All Subjects	0.738	0.823	0.788

TABLE 8.18: Performance of the optimal cohort specific classifiers on their cohorts' independent unseen data (data sets C and E).

Cohort/Classifier	F-score	Sens.	Spec.	Acc.
Younger Adult	0.758	0.879	0.560	0.719
Older Adult	0.813	0.980	0.570	0.775
All Subjects	0.788	0.955	0.532	0.743

TABLE 8.19: Performance of the optimal cohort specific classifiers on their cohorts' data during training, testing and validation.

		Data Set		
		Training	Validation	Testing
Classifier	Younger Adult	0.805	0.796	0.758
	Older Adult	0.775	0.780	0.813
	All Subjects	0.779	0.787	0.788

## 8.3.2 Subject Specific Classification

### 8.3.2.1 Pre-Processing

A similar sampling procedure to the Cohort Specific Classification approach above was used to select 400 sleep and 400 wake samples from each subject. However, the data from each subject was kept separate and randomly assigned to either train or validate the classifier. Equal numbers of sleep and wake epochs were used to train and validate each classifier. Each classifier was trained and validated on each of the six types of data previously described (the 3, 4 and 13 PCs, the 3 and 4 FSCs and the 13 original features). Again, cross validation was used to examine the consistency of the results.

### 8.3.2.2 Multi-Feature Classification

Performance results for each subject and the six types of input features over the 8 classifiers can be found in Figures 8.8-8.15. The mean performance of the SVM classifier

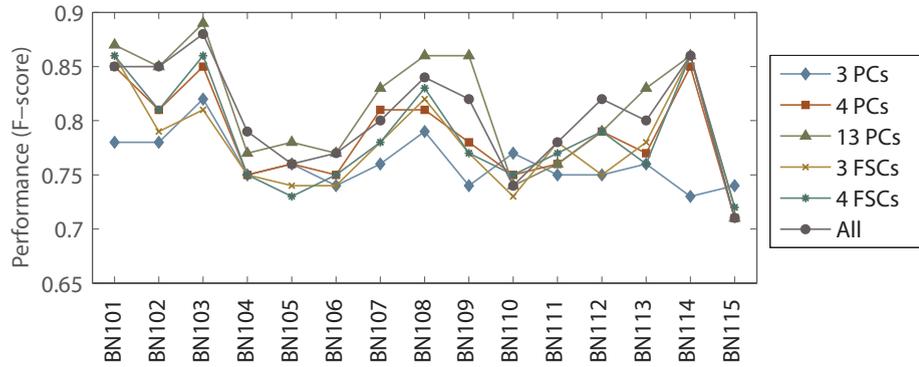


FIGURE 8.8: LDA classifier performance over six sets of input features.

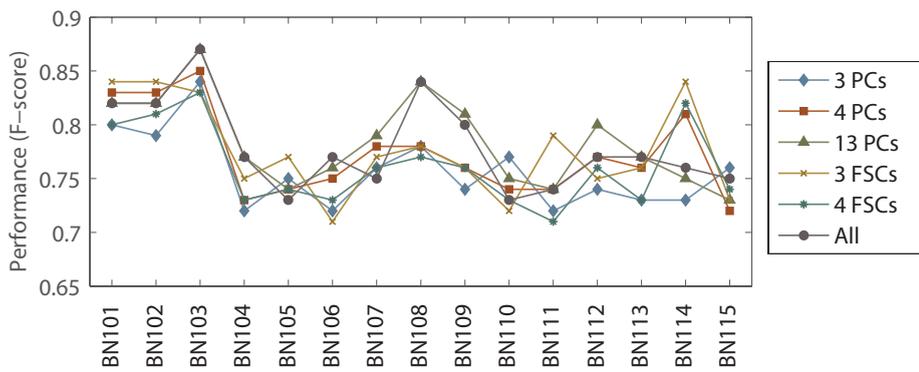


FIGURE 8.9: QDA classifier performance over six sets of input features.

reported is the optimal result over all sigma ( $\sigma$ ) and box constraint ( $C$ ) values per subject averaged over all cross-validations. The optimal internal SVM parameters varied on a per subject basis, and as such the performance at the best combination of internal parameters is reported.

The average subject performance reported on a per cohort (younger adults, older adults and all subjects) basis was then evaluated using the mean across the relevant cohort (see Figure 8.16). Averaging the best subject performance per cohort was not suitable for the SVM case as the internal parameters for each subject is different. As such, the best performance per cohort over all combinations of sigma and box constraint values were reported. As a result, the mean SVM performance across each cohort will be equal or lower than the values previously reported in Figure 8.15. In this case, the mean performance across each cohort is lower.

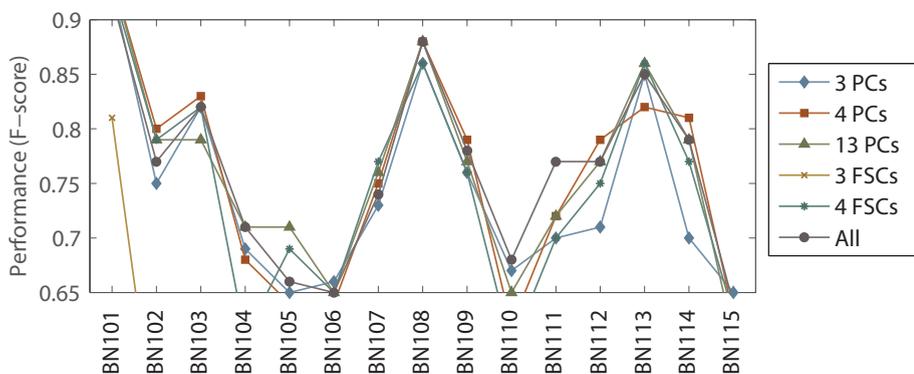


FIGURE 8.10: kNN1 classifier performance over six sets of input features.

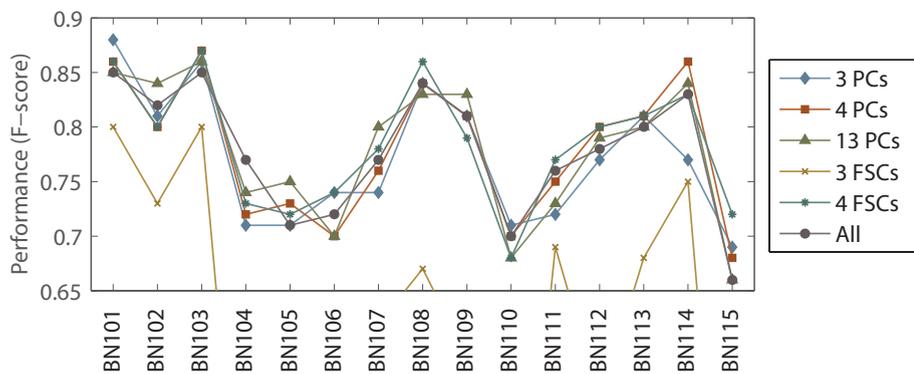


FIGURE 8.11: kNN5 classifier performance over six sets of input features.

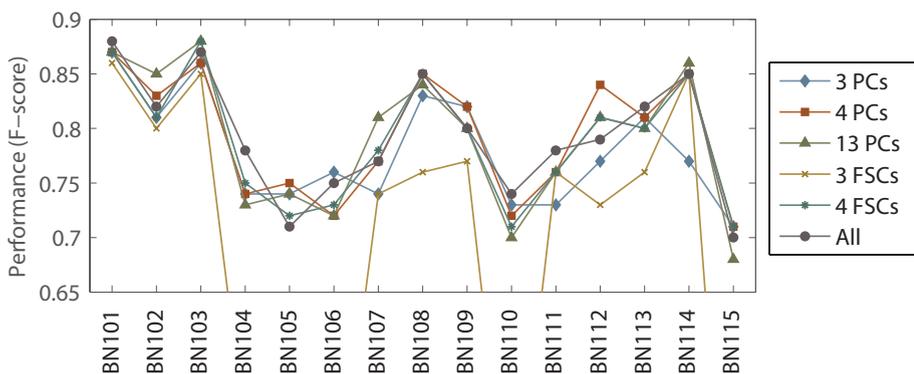


FIGURE 8.12: kNN9 classifier performance over six sets of input features.

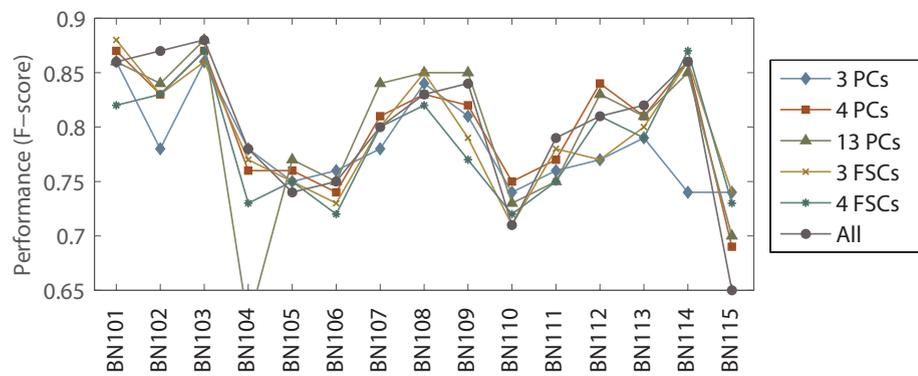


FIGURE 8.13: NN20 classifier performance over six sets of input features.

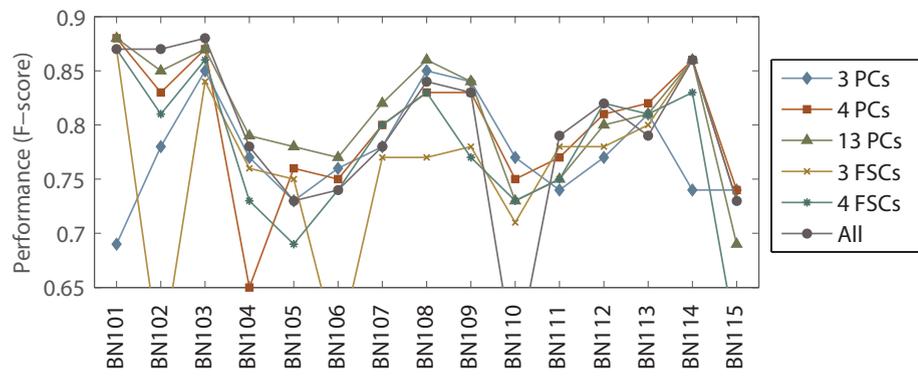


FIGURE 8.14: NN40 classifier performance over six sets of input features.

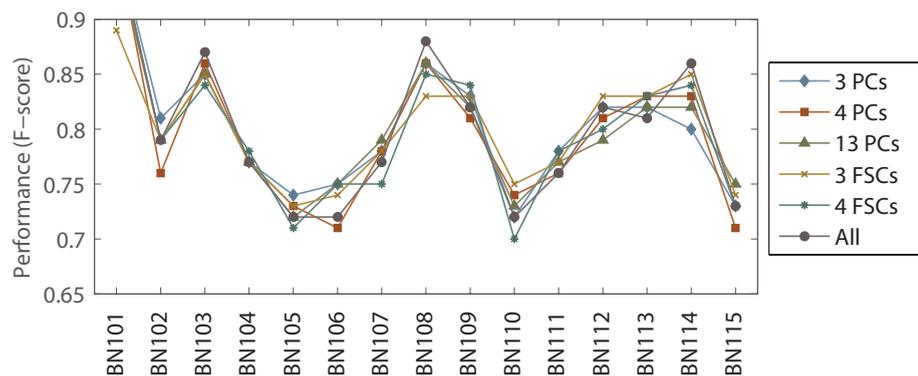
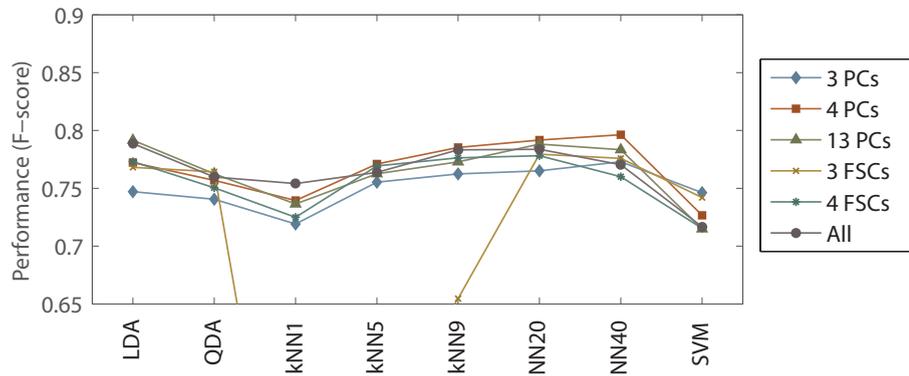
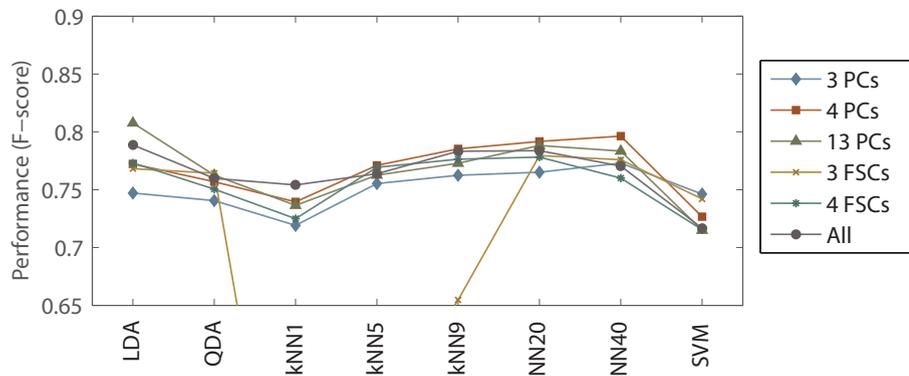


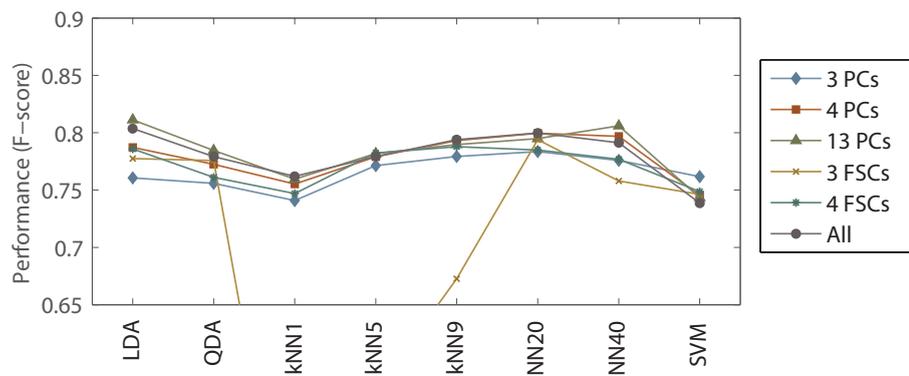
FIGURE 8.15: SVM classifier performance over six sets of input features.



(a) Mean subject specific performance over six sets of input features for the younger adult cohort.



(b) Mean subject specific performance over six sets of input features for the older adult cohort.



(c) Mean subject specific performance over six sets of input features for the entire cohort.

FIGURE 8.16: Mean subject specific classifier performance.

TABLE 8.20: Optimal subject specific classifier (results reported per cohort).

Cohort	F-score	Sens.	Spec.	Classifier	Details
Younger	0.828418	0.879665	0.729294	LDA	13 PCs
Older	0.807631	0.877827	0.696587	LDA	13 PCs
All	0.811247	0.8788	0.714031	LDA	13 PCs

TABLE 8.21: Validation performance of the optimal subject specific classifiers on both cohorts.

		Cohort	
		Younger Adult	Older Adult
Classifier	Younger Adult	0.7462	0.8358
	Older Adult	0.7299	0.8217

### 8.3.2.3 Optimal Classifier Selection

The optimal classifier over each cohort is given in Table 8.20. This reports a slightly better performance compared to the optimal cohort specific classifier. When examined on a per subject basis, very high performance rate are reported in some cases. This leads to a much larger standard deviation when results are averaged over each cohort analysed. The standard deviation in the F-score is 0.049, 0.054 and 0.050 for the younger adults, older adults and all subjects respectively.

### 8.3.2.4 Subject Specific Classifier Validation

The younger adult classifier and older adult classifier with the best performances (BN103 and BN109 respectively) were both applied to the unseen data from a younger and older population. As these classifiers were subject specific, an older/younger adult classifier could not be formed. The formation of the older adult and younger adult data sets used for this analysis was discussed previously in Section 8.3.1.6.

The mean performance (F-score) of each classifier applied to both data sets above over 10 trials is reported in Table 8.21. The standard deviation of the performance of each classifier was less than 0.025. The F-score, sensitivity, specificity and accuracy of each classifier on their respective cohort is given in Table 8.23. A comparison of the performance of the classifiers during training, validation and testing on their respective data sets is given in Table 8.22.

TABLE 8.22: Performance of the optimal subject specific classifiers on their cohorts' data during training, testing and validation.

		Data Set		
		Training	Validation	Testing
Classifier	Younger Adult	0.895	0.880	0.746
	Older Adult	0.822	0.847	0.821

TABLE 8.23: Performance of the optimal subject specific classifiers on their cohorts' testing data.

Cohort/Classifier	F-score	Sens.	Spec.	Acc.
Younger Adult	0.746	0.788	0.676	0.732
Older Adult	0.821	0.945	0.645	0.795

### 8.3.3 Discussion

This investigation found positive results for the discrimination of sleep and wake in both healthy younger and healthy older adults using LDA classifiers applied to features derived from UMBS data. A performance similar to commercially available actigraphic sleep/wake monitors was found (Fox et al., 2007). The UMBS system consists of a classifier which was trained, validated and tested using a large data set collected over an extended period of time using a relatively large cohort over multiple sleeping episodes. High performance has been reported when the optimal discriminating method was applied to a completely independent cohort in a testing phase. This validates the performance of the UMBS sleep/wake detection algorithm system when applied to new subjects.

The subjects under analysis were defined as healthy using a strict inclusion criteria and were also screened for potential sleep disorders such as sleep apnoea, insomnia, and circadian rhythm sleep disorder. As such, the data investigated in this analysis relates to healthy subjects, and not to those whose sleep would warrant monitoring due to the presence of co-morbidities and/or a sub-optimal health status. This research focussed on the ability of the UMBS to discriminate between wake and sleep in healthy cohorts as an initial step. Later steps may include unhealthy cohorts, where ethically appropriate. Furthermore, it is unrealistic to expect all of those with a reduced health status to have an identical pattern of degradation in their sleep. More illness specific changes in sleeping patterns would be expected (Happe, 2003). As such, it may be beneficial to analyse each cohort separately.

The sleep and wake patterns of each subject studied is not representative of natural variations in sleep and wake. All subjects complied to a strictly defined protocol of sleep and wake over the recording period. Large periods of the studies did not have a twenty four hour day. While this resulted in the collection of sleep and wake data throughout various circadian phases of the subject, this does not affect the PSG-definition of the sleep/wake state of the subject (Rechtschaffen and Kales, 1968). As such, the relationship between the UMBS collected features and the PSG-defined sleep/wake state should not be impacted. In this research, it was taken that the UMBS activity patterns should relate solely to sleep or wake state and not be modified due to circadian phase or the time of day. There was not sufficient sleep and wake data available for all circadian phases to examine whether circadian phase affected the ability of a classifier to discriminate between sleep and wake. The propensity of the subject to sleep at different circadian phases varies (Czeisler et al., 1992). This results in a greater amount of wake time during certain circadian phases, however, this should not affect the relationship between the UMBS-data and the PSG-defined sleep/wake state either. All subjects underwent one of two protocols enforcing sleep and wake times. While these protocols were different, most notably during the initial baseline period, this also should not affect the relationship between UMBS data and sleep/wake state. While the extended day length and sleep restriction are artificially enforced, this should have no bearing on the relationship between UMBS-derived data and sleep/wake state. As such, this data is adequate for the investigation of the classification of wake and sleep stages from UMBS data. No notable difference was found in the performance of any classifier when comparing protocols using the younger adult cohort (BN101-BN106 belonged to protocol B; BN107-BN108 belonged to protocol A) as seen in Figures 8.8-8.15. The older adult data set related to only one protocol and as such a comparison between protocols could not take place.

Naturally the subject might not be asleep for the entire period that they are prescribed to sleep. The quantity of sleep they get during this period only affects this research when a significant lack of either sleep or wake occurs. This occurred in some subjects where a low number of wake epochs were recorded. Consequently, random sampling with replacement was used to cater for this condition. Subjects for whom an inadequate number of wake epochs occurred were excluded from analysis. This only happened during the classifier testing phase for three subjects. While this reduced the amount of testing data available, there was still an adequate amount of data to produce valid testing results.

Normally, the subject is likely to be asleep for the majority of that period as they are not allowed to sleep during other times. However, periods exist where the subject is awake trying to get to sleep (exhibiting periods of *quiescent wake*). Due to the presence of PSG recording equipment, subjects are less likely to move as they naturally would in

their domestic environment which may exaggerate the amount of quiescent wake. This is a problematic condition for all activity based sleep monitors. Subsequently, the data collected might not be representative of data where no PSG is being collected (ie. during typical natural domestic sleep where quiescent sleep may occur less). However this is a constraint of all research where the subject is aware of being monitored (regardless of the nature of that monitoring). In any case, any system developed using data with high amounts of quiescent wake is likely to be more specialised and more sensitive to classifying epochs of quiescent wake correctly. While this may result in the development of a system more sensitive to the condition of quiescent wake, only a home-based investigation using a very minimally intrusive gold standard sleep/wake monitor would fully be able to answer this question. Additionally, it must be noted that all the data in this investigation was collected in an artificial environment and as such may not be fully representative of data collected in the home. Again, there exists an inherent constraint of this nature for any type of clinical research.

Methods of dimensionality reduction were investigated in order to assess if any redundancy in the features could be removed from the proposed UMBS sleep/wake detection system. Although the number of features is low, it would be advantageous if a final system had as low a complexity as possible (in order for the system to be implemented on low cost hardware). Additionally the calculation of some features (which can be a lengthy process) could be avoided. FSCA and PCA were initially investigated. FSCA and PCA modify and/or arrange the feature set in terms of the decreasing variance in the data. PCA transforms this data to a new basis in a manner which ensures that each new signal, or principal component, is orthogonal to all previously generated principal components. FSCA arranges the signals in terms of decreasing variance explained. These methods of dimensionality reduction do not necessarily compare well with the optimal selection of features which would result in the best discrimination of the output classes. For example, feature 12 (the number of distinct movements as defined using Temporal Movement Feature (TMF)) was found to provide the highest performance in discriminating between sleep and wake (see Figure 8.4), however it was not always identified as the first FSC or as the major contributing factor in the determination of the first principal component. In fact, the feature 12 was found to be the fourth FSC, and its inclusion as a feature significantly increased the accuracy of kNN classifiers (see Tables 8.11, 8.12, and 8.13). The use of all of the original features outperformed the use of the best three or four FSC (see Figures 8.8-8.15). Additionally, the use of all of the PC outperformed the use of the top three or four PC. This was in spite of the very low contribution of the remaining PC to the variance of the data set. The use of some of the features, found using a feature subset selection method, in the discrimination of sleep and wake provided good results. SFS was used to achieve this. For this the class each data sample corresponded to was

involved in the determination of the optimal selection of features. The determination of these features is specific to the classifier used (which in this case was LDA). The SFS classifiers were found to provide the highest mean performance of all the older adult and younger adult classifiers, despite using only four of the original features (as given in Table 8.14). The best mean performance for the mixed cohort was a non-SFS classifier, however the SFS classifier performed only marginally worse. The best three and four FSC and the top three and four PC were outperformed by the SFS selected features consistently. The FSC did not correspond exactly to the features SFS determined as the best features (although feature 12 was in common). FSCA and PCA are often used as unsupervised machine learning techniques where the output state is either unknown or doesn't exist. In the context of this research, the supervised approach of using SFS provided better results than the aforementioned unsupervised methods. Overall, the use of all of the original features was chosen as the optimal method to use (see Table 8.16). The results from the application of these classifiers to training, validation and test data report high performance (as given in Table 8.19).

It should also be noted that a tradeoff was made during this research where the PC were approximated using an external data set (data set D) in order to reduce the occurrence of overfitting on the training and validation data sets and also to increase generality. It was felt that the data set used to calculate the PCA transformation matrix was large enough to provide results robust enough to be considered generally applicable. It is of importance to note that this data pertains to younger adults and as such might not be ideal for data sets containing older adults. This could be a potential explanation for the consistently lower results found in the sleep/wake discrimination of older adults (see Figure 8.5). However, lower accuracies in the discrimination of sleep and wake of older adults than in younger subjects have been previously reported in the literature (see Section 4.5). There are many potential sources of explanation for this including the increasing number of subclinical co-morbidities in older adults or the higher proportion of wake in the sleeping episode (which increases the occurrence of quiescent wake, a problem for activity based sleep/wake monitors). A comparison of the performance of an LDA classifier using the 13 original features, 13 PC features, and all 26 jointly as inputs found no improvement in performance. This was expected as the 13 PC should just be a transformed version of the 13 original features (assuming that the PCA transformation matrix used was generally applicable).

In this analysis, both linear and non-linear classifiers were investigated. The most basic linear classifier examined (LDA) performed very well in comparison to more advanced, and optimised, non-linear techniques. This was the case for both the cohort specific scenario and the subject specific scenario. While LDA was marginally outperformed by some of the non-linear techniques, LDA was chosen as the optimal cohort specific

classifier under the condition that all thirteen original features were used as inputs. This was due to the ease of implementation of the classifier in a finalised system which is of specific importance when using low-level hardware and software. The non-linear classifiers under examination in this analysis did not significantly outperform the more basic linear approaches despite comparatively lengthy optimisation processes. While an exhaustive search for a better solution using different kernels in the SVM or alternative ANN topologies might prove fruitful (for example using a low number of neurons in the hidden layer, such as 5), the results from this analysis suggest that the linear approaches provide a near optimal sleep/wake discrimination and that a more comprehensive search for the ideal classifier and topology would not significantly improve performance. Additionally it must be noted that SVM classifiers were trained using significantly less samples than that applied to the other classifiers. The number of samples used to train the SVM classifier was reduced due to the high computational overhead and also due to the lengthy process of optimally tuning the internal parameters. A tradeoff was made between the number of samples used to train the classifier and the resolution of the search for optimal internal parameters.

Figure 8.6 shows the performance of the SVM classifier throughout a range of sigma and box constraint values. It suggests that a larger range of values for sigma should be investigated. However in a preliminary investigation it was found that the performance of the classifier declined above a sigma value of 4. As such, the performance metrics reported herein are considered accurate for this data set.

In addition to classifying cohort specific data (all younger subject, all older subjects, and all subjects), subject specific classification was also performed and results were averaged over the younger adult and older adult cohorts (as given in Tables 8.22 and 8.23). The performance of the subject specific classifiers outperformed the cohort specific classification over the training and validation data. This is unsurprising as each subject specific classifiers were trained and tested on each subject. The group specific classifier marginally outperformed the subject specific classifier for the younger cohort using validation data. In this scenario, the best performing subject specific classifiers (optimised on a subject's data) were applied to independent testing data. As such, it is unsurprising that much lower performance was reported.

Applying the subject specific classifiers to certain individuals, very high validation performances were reported (F-Score over 0.85), however other individuals reported lower performance rates (F-Score of 0.75). The high results illustrate that some subjects present specific patterns (exhibited in the UMBS features), pertaining only to that individual, which can discriminate sleep from wake very well. The lower results illustrate

that such specific patterns do not exist for all subjects (as quantified using the UMBS features).

A low variation in the average subject specific performance per cohort was found over all classifiers (as seen in Figure 8.16). Higher variations are evident when comparing the input features to the classifiers. Again, the LDA classifier was found to be the optimal choice for subject specific classification. The use of all 13 PC, followed by all 13 original features, performed best. This suggests that effort should be directed towards deriving useful features, rather than an exhaustive search of the optimal configuration of the optimal classifier.

## 8.4 Sleep Stage Classification

### 8.4.1 Multi-Class Classification

A subset of the younger adult data set was randomly selected to contain 1000 samples of each sleep stage. This created a data set of 6000 samples using the 5 sleep stages as well as the wake stage. There was an even contribution of data per sleep stage from each subject included in this analysis. The data was split into 66.6% training and 33.4% test data. This investigation used LDA classification to distinguish between the sleep stages as it was easy to implement in a finalised system and has a low computational overhead. This test was repeated 100 times using a different combination of samples from the cohort in order to assess the consistency of results. Results were reported using the mean over all repetitions and as a ratio of the number of samples per stage (see Table 8.24). The standard deviation over all repetitions was less than 0.04. The accuracy of this method in correctly classifying the wake or sleep stage is reported along the diagonal. The accuracy is less than or equal to 50% in all cases. This system performed relatively poorly, although still significantly better than random chance (given that there are 6 stages under examination). The process of correctly classifying sleep stages is not an easy one, even using standard PSG data and human scorers. In such cases, inter-rater reliability is relatively low as 80.6% for the Rechtschaffen and Kales (5 sleep stages) standard (Danker-Hopfe et al., 2009).

In order to reduce the complexity of the system, the number of classes were reduced. Stage 1 and stage 2 sleep were combined in a category called '*light sleep*', while stage 3 and stage 4 were combined in a category called '*deep sleep*'. The wake and stage 5 (REM) stages remained as before. In a similar approach as that above, 1000 samples per stage were selected randomly with an equal distribution from each subject. Again, cross validation was applied and the standard deviation in results was less than 0.075%.

TABLE 8.24: Confusion matrix of LDA classification of wake/sleep stages using the 13 UMBS features.

		Predicted					
		Wake	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Actual	Wake	0.50	0.19	0.12	0.02	0.09	0.06
	Stage 1	0.17	0.25	0.29	0.06	0.10	0.10
	Stage 2	0.02	0.14	0.45	0.07	0.18	0.10
	Stage 3	0.01	0.10	0.27	0.12	0.35	0.12
	Stage 4	0.00	0.09	0.30	0.10	0.38	0.09
	Stage 5	0.04	0.17	0.31	0.08	0.24	0.13

TABLE 8.25: Confusion matrix of LDA classification of a reduced number of wake/sleep stages using the 13 UMBS features.

		Predicted			
		Wake	Light Sleep	Deep Sleep	REM
Actual	Wake	0.56	0.18	0.12	0.11
	Light Sleep	0.13	0.32	0.28	0.25
	Deep Sleep	0.02	0.19	0.53	0.25
	REM	0.06	0.27	0.36	0.28

TABLE 8.26: Confusion matrix of LDA classification of wake/NREM/REM stages using the 13 UMBS features.

		Predicted		
		Wake	NREM	REM
Actual	Wake	0.61	0.18	0.21
	NREM	0.06	0.59	0.34
	REM	0.09	0.50	0.40

Results were reported using the mean over all repetitions and as a ratio of the number of samples per stage (see Table 8.25). This classification process was repeated using just three classes: Wake, NREM and REM (see Table 8.26). In these results the standard deviation was than 0.062%.

The overall accuracy of these systems in distinguishing between different states of sleep is relatively poor (see Table 8.27). QDA was also compared and it was found to consistently be outperformed by LDA. This could be due to the more complex classification task for QDA where a greater number of additional parameters have to be estimated over LDA. LDA makes stronger assumptions during derivation and as such has smaller variance in these estimates. However, LDA uses linearly defined decision boundaries which may not always be ideal. Non-linear classifiers, such as ANNs and SVMs, could potentially provide better accuracies, however these were not investigated further due to the very low accuracies reported by LDA and QDA.

TABLE 8.27: Overall accuracy (total percentage of correct classifications) of LDA and QDA classification over the three sets of data.

	LDA	QDA
W/1/2/3/4/REM	0.31	0.27
W/LS/DS/REM	0.28	0.26
W/NREM/REM	0.51	0.49

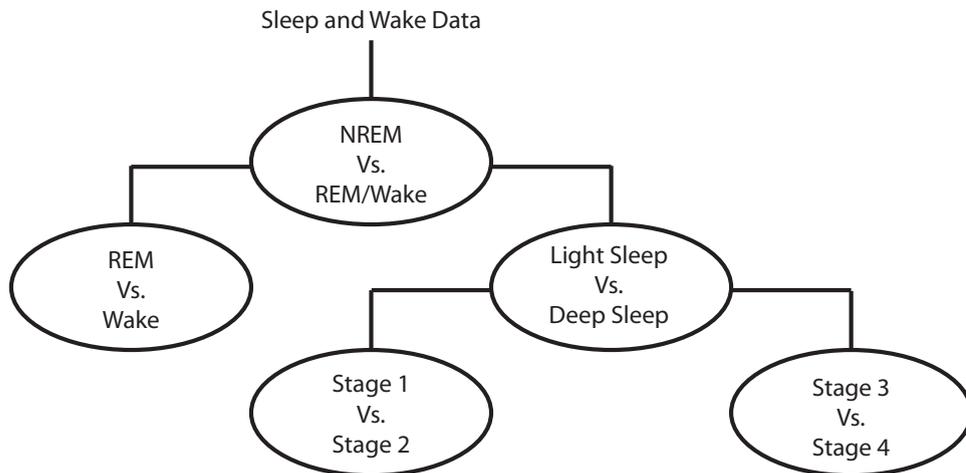


FIGURE 8.17: Hierarchical sleep classifier.

#### 8.4.2 Hierarchical Binary Classification

Hierarchical classification approaches have been suggested in the literature as a means of successively discriminating between sleep stages (Anderer et al., 2005). Initially a classifier is designed to distinguish between groups of classes which display similar data patterns. Subsequently, this process is repeated until each stage has been identified (see Figure 8.17). For example in PSG data, Wake and REM sleep often have similar physiological profiles which can be differentiated from NREM sleep. Thus, a classifier can be applied to separate these stages. At a second level, a classifier will be applied to separate REM sleep from wake. Another classifier will be applied to the NREM sleep samples in order to distinguish between light sleep (stage 1 and stage 2 sleep) and deep sleep (stage 3 and stage 4) sleep. At a third level, classifiers will be used to identify each stage of NREM sleep.

However while this technique was found to have merit on PSG data, it will only work where there is sufficient differences in the data between groups of sleep stages. In order to investigate whether any such groupings exist, a clustering approach was applied. Initially, the younger adult data set was divided into clusters using a common unsupervised technique. The  $k$ -Nearest-Neighbour clustering algorithm using the Euclidian distance metric was used to split the data into between two and ten clusters. The optimal number

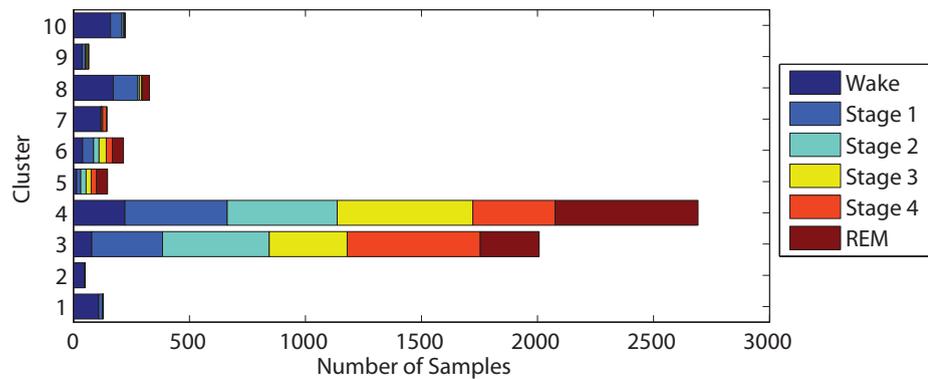


FIGURE 8.18: Distribution of sleep stages per cluster.

of clusters was unknown. A subset of the younger adult data set was used to form these clusters with an equal proportion of data (1000 samples) belonging to the wake, stage 1, stage 2, stage 3, stage 4 and stage 5 epochs. An equal proportion of data was taken from each subject. While some small clusters identified wake epochs, the larger clusters contained approximately equal distributions of data from all of the sleep and wake stages (for example, see Figure 8.18). Similar results were found when the numbers of cluster varied between 2 and 10. Due to the consistent overlapping of most sleep stages in clusters, with wake (sometimes accompanied by stage 1) often being identified in separate clusters, it was decided that a hierarchical approach to sleep stage classification could not achieve greater sleep stage classification accuracy for the features in this data.

### 8.4.3 Discussion

An investigation into the potential of the UMBS to discriminate between the stages of sleep was also performed (see Tables 8.24, 8.25 and 8.26). While an accuracy much larger than random chance was found, it did not report an accuracy of an adequate level to be realisable in a final solution. Subsequent analysis using an unsupervised clustering technique showed that large overlaps of sleep stages occurred in most clusters. This clustering approach was employed in order to inform whether a hierarchical approach to discriminating between sleep stages would be beneficial. The optimal hierarchical approach to distinguish between sleep stages (similar to that shown in Figure 8.17) was not evident. This may explain the low performance in discriminating between sleep stages. This result is not altogether surprising as the physiological basis for discriminating between sleep stages is not largely defined in terms of body movement, but mostly a combination of EEG, EMG, and EOG (Rechtschaffen and Kales, 1968). While neurological changes between REM and NREM (the control of the sympathetic/parasympathetic nervous system) do occur, this does not generally result in changes in movement patterns

but rather changes in the rhythmicity of physiological functions such as heart rate and respiration rate (Rostig et al., 2005). An investigation into using advanced non-linear techniques, such as Gaussian mixture models, to gauge this rhythmicity may provide better accuracy.

## 8.5 Overall Discussion

In various other actigraphic sleep/wake detectors the sleep state has been defined using data from a current epoch and its surrounding epochs (Mullaney et al., 1980; Sadeh et al., 1994). However, the sleep or wake state of the subject during an epoch does not have a causal relationship to the state of the subject during surrounding epochs, despite having a strong correlation with these surrounding epochs. For example, if a subject is awake during one epoch that does not determine their sleep/wake state during the next epoch despite having increased the probability of their being asleep or awake during that subsequent epoch. Some technologies have used an approach where the state a current epoch is assigned is altered based on surrounding epochs (Cole et al., 1992). However, it is not physiologically accurate to determine the sleep/wake state of the subject based on previous or future epochs (Rechtschaffen and Kales, 1968). Altering the assigned sleep stage, post analysis, can increase the over all accuracy of results by correcting misclassification. However such an approach loses the sensitivity of detecting discontinuities in one state, particularly the subject waking briefly. Recent research by Lim et al. (2012) has found that such discontinuities can be used to identify subjects who are cognitively impaired. The focus of this research was to solely investigate the relationship between the current set of features and the current state of the subject. This was performed so that the classifier would report the causal relationship between the UMBS features and the PSG-defined sleep/wake state.

Often the accuracy of actigraphy is measured using specificity (Equation 8.2) and sensitivity (Equation 8.3). However, these might be sub-optimal techniques as sleep data sets are generally heavily biased toward the sleep state. For example, let's say that the number of misclassifications (false negatives and false positives) that the optimally trained classifier reports are equal. Due to the bias in the data set where there are significantly more sleep epochs (true positives), the sensitivity of the classifier will be very high. However, the specificity the classifier reports will be much lower, despite the same number of misclassifications, due to fewer wake epochs (true negatives) in the data. Furthermore, this bias is not fixed for all sleeping episodes under analysis and can change substantially both between subjects and between sleeping episodes. One approach which addresses this is to select a subset of samples, randomly chosen, with an

equal number of actual sleep and wake epochs and to calculate the performance of the discriminating algorithm using that data. Cross validation should be used to quantify the accuracy of the performance metrics reported. An alternative approach could be to weight the sensitivity or specificity values in order to correct for this bias. A consistent approach across all future literature should be taken in order to compare the accuracy of multiple technologies. The reporting of overall sleep metrics should be avoided when validating a sleep/wake monitoring device.

## 8.6 Conclusions

This chapter analysed the ability of UMBS-derived features to predict the wake or sleep state of an individual. Additionally, the degree to which sleep stages could be determined using UMBS-derived data was also investigated. The performance and reliability of this technology is comparable to the current ambulatory sleep/wake monitoring gold standard, wrist actigraphy, and an alternative non-contact sleep/wake monitor (BiancaMed SleepMinder (De Chazal et al., 2011; Fox et al., 2007)). However, this system is suitable for long term placement in domestic homes and is ideal for the non-intrusive collection of health data, particularly amongst sensitive populations (such as those with mild cognitive impairment or dementia).

The collection of a large database or concomitant sensor and PSG data is a significant logistical barrier to validating any proposed novel sleep monitoring system. While the deployment of the UMBS device is relatively straightforward, the collection of PSG data has a much larger overhead. The EEG montage must be applied and removed at the beginning and end of each sleeping episode by a trained technician. Additionally, the PSG data must be subsequently scored manually by a trained scorer which is an expensive and time consuming process. In order to collect a large enough data set to allow an accurate investigation of the discrimination of sleep and wake, data must be collected from many subjects. The relatively low percentage of wake in each sleeping episode means that data must be collected over many sleeping episodes in order to get an adequate number of wake samples. Through the collection of this data set two main novel contributions were possible, they are as follows:

**A method to predict the sleep/wake state of a person:** This research identified an optimal sleep/wake discriminating algorithm and tested its performance on a set of individuals external to the training/validation stage. Very high sensitivity and relatively high specificity values were reported for the optimal classifier. These results compete favourably against wrist actigraphy which is the current gold standard for ambulatory

TABLE 8.28: Comparison of wrist actigraphy, BiancaMed SleepMinder and UMBS for sleep/wake discrimination

	Sensitivity	Specificity
Wrist Actigraphy ((Paquet et al., 2007) in (Van De Water et al., 2011))	91%	65%
BiancaMed SleepMinder(De Chazal et al., 2011)	87.3%	50.1%
UMBS - Younger Adult	87.9%	56%
UMBS - Older Adult	98%	57%
UMBS - All Adults	95.5%	53.2%

sleep monitoring and an alternative non-contact sleep/wake monitoring device (BiancaMed SleepMinder). However, it might be unfair to provide a direct comparison of results as the UMBS-derived results are based on an unbiased data set (equal numbers of actual sleep and wake epochs), while the other studies use a biased data set calculated using the proportion of sleep and wake in each sleep episode under analysis. The problem this highly variable sleep/wake bias introduces has been discussed in the previous section.

It should be noted that while the optimal solution used all thirteen UMBS derived metrics, a dimensionality reduction technique identified four signals which would achieve a slightly lower accuracy. Three metrics provided the majority of that accuracy. These were derived from the  $TMF_2$  and UMBS 2 data. The fourth metric varied between the UMBS-derived respiration rate and the SMF metric for the younger and older adults respectively. It may be advantageous to remove this fourth metric as the computational overhead of such a system will be considerably lower. This may be particularly applicable to a low-level hardware/software solution.

**An investigation into the ability of UMBS-derived data to discriminate between sleep stages:** This research has also investigated the ability of UMBS-derived metrics to discriminate between wake and a number of sleep stages. The accuracy of the results was relatively low. A hierarchical method of classification was also investigated using a unsupervised clustering technique. This did not prove fruitful, but did highlight that each cluster does contain data from different sleep stages. This suggests that it is not feasible to discriminate between sleep stages using UMBS-derived data.

## Chapter 9

# Conclusions and Future Work

This thesis presents the novel development of an ambient sleep monitoring system suited to long-term domestic placement, particularly in the homes of older adults. This work is centred around the development of assistive technologies which allows older adults to live independently and *age in place*. Additionally such advances deliver a higher quality of life while reducing economic and societal costs. A review of the recent advancements in sleep monitoring modalities (as well as an overview of more traditional methods) provides a justification of the need for an ambient sleep monitoring system. The proposed system provides indices relating to nocturnal movements, bed restlessness levels, sleep metrics, and respiration rate. Experimental, domestic, and clinical deployments, using both younger and older adults, validated algorithms devised to generate these statistics and quantified their accuracy against clinical and widely accepted standards.

This section details the overall conclusions and contributions of this research as well as providing possible avenues for future investigation.

### 9.1 Overall Conclusions

Sleep is a fundamental physiological process where significant decrements in its quality or quantity may represent a degradation in overall health status. While the relationship between the afflicting condition/illness and poor sleep may be causal or correlational, the ability to track changes in sleeping patterns (inclusive of bed restlessness, nocturnal movements, and respiratory rate) over time may provide important clinical information. Such information may be used to better direct clinical interventions benefitting the overall health of the individual (for example, improving quality of life, physical functioning, and mental health). Poor sleep may be expressed multifactorially perhaps

through deviations in bed restlessness levels, longitudinal inconsistency in sleep/wake routines, overall sleep quantity as well as through other general sleep metrics (such as SL, WASO, number and duration of bed exits, etc.).

In Chapter 2 sleep stages and typical sleep profiles for healthy individuals are introduced. Typical changes in sleeping patterns due to ageing and illness are presented which provide an argument for using measurements captured during sleep as a proxy for overall health status. A review of the more traditional methods of sleep monitoring (namely PSG, sleep diaries, sleep tests and subjective measures, and wrist actigraphy) in Chapter 4 provides a justification of the need for the development of modalities suitable for long-term deployment. Only recent advances in sleep monitoring technologies are appropriate for long-term placement in the homes of older adults.

The remainder of this thesis focusses on: 1) the development of an ambient sleep monitoring system from describing the sensor through to the validation of the technology and the implementation of a data collection platform, 2) the development of feature extraction algorithms (inclusive of bed restlessness, nocturnal movement and respiration rate), 3) exploratory work comparing clinical and domestic data amongst older and younger adults, and 4) the development of an ambient sleep/wake discriminating system that competes favourably with wrist actigraphy and alternative non-contact systems (such as BiancaMed's SleepMinder (De Chazal et al., 2011; Fox et al., 2007)). An investigation into predicting sleep stages was made, however a low performance was found. Some competing systems report better accuracy, however such systems use the heart rate, or the rhythmicity of the heart cycle, a feature which was not possible with the UMBS.

The core contributions of this thesis are summarised as followed:

### 9.1.1 Literature Review

A literature review, presented in Chapter 4, provides a comprehensive review of multiple types of sleep measurement modalities. It begins with the traditional clinical method of PSG and describes the types of signals which need to be recorded so that a trained manual scorer can determine the sleep stage the individual is currently experiencing (as detailed in Section 2.2). Typically, PSG is performed while under observation of technicians, however unattended and even ambulatory systems have been instantiated and tested. While this may allow a more unencumbering system (by allowing the person to sleep in their own homes or continue to perform activities of daily living after donning electrodes), it increases the likelihood of recording more movement artifacts and lowering signal quality (as a result of inefficient electrode attachment through application by non-specialists). Additionally during clinical PSG, when electrodes become unattached

a technician can be alerted and the fault corrected (most likely through re-applying the electrode). A description of typical sleep diaries, commonly administered sleep tests and subjective scales has also been provided.

The ambulatory gold standard for sleep/wake monitoring, wrist actigraphy, was also introduced, described and its clinical performance (compared to PSG) detailed. The low specificity rates reported have been attributed to the high prevalence of *quiescent wake* in individuals attempting to fall asleep and lying still, yet remaining awake. An overview of a common algorithm catering for this was detailed, however while specificity increases through its application, the overall specificity rate is still relatively low. Due to the bias in the data set (as it generally contains a majority of sleep epochs), the overall performance of the system remains high.

A detailed and wide ranging description of the recent advances in sleep monitoring was also presented in this chapter. The advances in automated methods of predicting sleep stage using either full or partial compliment PSG was provided. A significant number of these methods utilise machine learning algorithms and often focus on neural networks, advanced classification methods or some derivation from rule-based methods. Optimal feature extraction and selection methods have also often been employed. Other electrode advances include the novel placement of electrodes, the use of dry electrodes (which are more suitable for unsupervised placement), embedding electrodes in bed linen and bed clothes, and the detection of less obtrusive physiological signals which can be related to sleep/wake state (such as peripheral arterial tonometry).

Non-contact sleep monitoring remains a challenging problem although the numerous advances include the use of video cameras in the bedroom during sleep, the use of smartphones on/near the bed, and bed posture detection. Significant research has focussed on the detection of physiological signals (mainly respiration and heart rate) through a system which is completely unobtrusive. These have mainly focussed around the placement of pressure sensitive sensors in/on/under a mattress or under the bed posts, and radar-based systems. Many systems have recently produced reasonably accurate solutions to measure respiration and heart rate; these have subsequently been used to predict sleep/wake state. Machine learning techniques are often used to generate algorithms to accurately and reliably predict sleep state. The large overhead (in terms of time, expense and discomfort for the subject) in collecting PSG data is a considerable barrier in collecting concomitant data and generally limits large-scale validation studies. As such, patients suffering from various sleep disorders are often recruited as they are attending sleep clinics; this is often at the expense of collecting data from healthy normal adults.

The initial results from the deployment of some of these systems as part of smart homes/ambient assisted living research projects are also discussed. Additionally, an overview

of the development of advanced systems for detecting sleep apnoea, suited to non-clinical settings, is also given. A brief discussion regarding brain imaging systems during sleep is also provided.

Initially evident in this review was a lack of suitable technologies for long-term sleep monitoring. As such, the focus of this thesis was in the development of such a system. However, concomitant research was undertaken by many other groups in similar areas (and their technologies are included in this review) justifying the need for advances in ambient sleep monitoring modalities.

### **9.1.2 Experimental Validation of Movement and Respiration Extraction Algorithms**

Chapter 5 provides a detailed description of the unobtrusive under mattress bed sensor (UMBS), that is the main focus of this thesis, and its inner workings, the data produced by the device, the communications protocol and sampling rate, and introduces some notation. The data captured by the sensor is further illustrated through initial small scale experimental deployments which provide an insight into the sensitivity of the device for physiological signal and nocturnal movement detection. The deployment of the device in the domestic homes of ten older adults provides insights into the ability of the sensor to capture rest/activity cycles, bed restlessness, and to estimate a number of standard sleep metrics (such as TIB). Comparisons against wrist actigraphy and daily activity levels provide a description of the utility of the sensor in capturing nocturnal activity levels directly, or daily activity levels indirectly.

A comprehensive validation of the detection of physiological signals, the development of algorithms to automatically estimate respiration rate, and the validation of UMBS features quantifying motion levels is detailed in Chapter 6. Initial experiments show the presence of respiration and heart rate signals in the time and/or frequency domain UMBS data; however heart rate is not present over all instances. This provides the basis for an algorithm which automatically derives respiration rate. The optimal algorithm was selected and tuned using data collected over two conditions (sensor placed both over and under the mattress) in a sample cohort of eight healthy young adults. Peak detection (time-based) techniques were found to outperform frequency based techniques over the entire data set and the system was found to be comparable to other recently developed systems.

Four UMBS motion metrics were proposed and subsequently validated against a gold standard video based system. For comparative purposes, the UMBS metrics were also compared to wrist actigraphy and PIR based systems. Three out of four optimally tuned

metrics were found to reliably and accurately detect movement, and to outperform the wrist actigraphy and PIR based systems.

### **9.1.3 Deployment of the UMBS in Clinical and Domestic Environments**

The deployment of the UMBS in multiple settings allows us to contrast its usefulness across cohorts and environments as discussed in Chapter 7). This is illustrated through studies in clinical and domestic environments amongst healthy young, healthy old, and individuals at high risk of having a sleep disorder (most likely sleep apnoea). This produced four UMBS data sets which were directly compared:

1. Summerhill: 10 relatively healthy community dwelling older adults over approximately two weeks.
2. Maynooth: 3 healthy community dwelling younger adults over less than five nights.
3. Peamount: 10 individuals at high risk of having a sleep disorder (mixed age) assessed in a sleep clinic for a sleep disorder over one night.
4. Boston: 12 healthy younger and older adults participating in a research study in an intensive physiological monitoring unit over approximately three nights.

Comparisons between these cohorts were facilitated through the extraction of movement features from the UMBS data. These included discretely sampled temporal, spatial, and statistical descriptions of in-bed movement occurring throughout the night. Additionally, spatiotemporal descriptions of major movements in bed were extracted from the sensor. Comparisons of each of these features both within and between cohorts were suggestive of distinguishing characteristics. As a result, features such as these may facilitate the detection of decrements in sleep quality resultant from a degrading overall health status. An inter-daily investigation of the features allowed an investigation of the consistency in these metrics across multiples days (although this was only available for a reduced set of individuals). Some features were found to be consistent within a cohort and/or over multiple days while others were not. It was concluded that some features are environment specific (for example some features being dependent on body weight and mattress thickness), while others are not (particularly spatial movement features). The non-environment specific movement features may be particularly suitable for use in pressure based systems when directly comparing populations while all features may be used to identify longitudinal decrements in sleep quality or quantity. Longer-term

studies may be used to elucidate which metrics, if any, relate to short and long-term deviations in sleep quality.

An investigation into the ability to distinguish the cohort based on spatiotemporal movement features proved not to be fruitful. As such, it was deemed that no discriminating characteristic of each individual movement exists; however the number of occurrences of these movements were very different between cohorts. This may warrant further investigation as it suggests that the occurrence of such movements, and not how the movement is performed, disturbs sleep and may discriminate cohorts. A large and more varied cohort is required to investigate this fully.

#### 9.1.4 UMBS Sleep Classification

A comprehensive data set was collected in order to assess the capacity of the UMBS in discriminating both sleep from wake and the sleep stages in a cohort of older and younger adults (as discussed in Chapter 8). In order to pursue this rigorously, the sizeable data set over more than 12 sleeping episodes from a large cohort was split into training, testing and validation data sets. This data was collected concomitantly with PSG and as such a direct comparison was performed. These data were collected during both naturally and artificially imposed day times and day lengths, however this only affect the subject's sleep state and, as such, should not affect the relationship between sleep stage and UMBS data.

The features described in previous chapters were calculated for each epoch of PSG data and used as inputs into multiple classifying functions which were tuned for optimal performance. Results from training and testing using each feature and a combination of all features were generated on cohort-specific and subject-specific data and the optimal sleep/wake discriminating function found. Subsequently, feature reduction methods (including sequential forward feature selection and dimensionality reduction methods) were imposed in order to provide an insight into the redundancy in the data and also to evaluate the classifiers with respect to computational efficiency. Results were generated individually for each cohort (younger adults only, older adults only, all subjects) and also tested using independent cohorts in order to provide an insight into the performance of each classifier for unseen data sets. The subject-specific classifiers performed better than cohort-specific classifiers as each set of data was trained and subsequently tested on data from each person singly. The cohort-specific classifier performed comparably against wrist actigraphy and a radar based system and reported favourable rates of sensitivity. However, a distinction must be made when reporting this comparison as all data in this investigation was taken from a healthy population whereas the RF

system used a population undergoing assessment for a sleep disorder (de Chazal et al., 2008). Additionally it is worth noting that the classifiers in this thesis were generated, and tested using an unbiased data set containing equal proportions of sleep and wake.

An investigation into sleep stage classification using UMBS data was also performed, however it did not provide sufficient accuracy to be comparable to PSG. The multiclass classification approaches used were implemented using a *One-Versus-All* scheme and a hierarchical classification approach. Due to intermingled data belonging to each sleep stage this process was considered overly complex, and advanced classification methods (for example, using Kernel methods) were deemed likely to overfit and thus not considered appropriate for further investigation. Some systems (mainly contact based technologies) report higher accuracies in predicting sleep stages, however these systems often use heart rate as a feature (which was not possible in this case), and also investigate the rhythmicity of respiration and heart rate (which was not undertaken as the sampling rate of the UMBS was too low).

## 9.2 Further Work

The research in this thesis provides several avenues for further research. Some of these topics include the following:

### 9.2.1 User-Acceptance Testing of Domestic Sleep Monitoring Technology

Multiple modalities for sleep monitoring have been developed in recent years (as discussed in Chapter 4). These have ranged from modifications of PSG (in the forms of advances in electrodes or automated scoring techniques) to non-contact vital signs and movement monitoring (including PIR, video, pressure, and RF based systems). While the accuracy and reliability of these technologies is sufficient for sleep monitoring, the user acceptance of such modalities remains to be investigated. Potential pitfalls for such technologies may range from the physical discomfort when using a potential device to increased anxiety and decreased *humanisation* of the environment (resultant from the user feeling they are being monitored, becoming excessively mindful of a potentially decreased health status, or from the stigma associated with availing of such a system by visitors to the environment). For example, technologists should be mindful when choosing a video monitoring application which may not be adopted by all users due to privacy concerns. Townsend et al. (2011c) investigated the trade-offs older adults are willing to make between privacy and autonomy (living independently), in particular they noted

“The concept of video monitoring seemed to participants to be a violation of privacy, but when alternatives and usefulness were considered, some concerns were diminished. In all cases, video compared favorably to a nursing home environment”. This suggests video monitoring technologies are only appropriate when all other options for staying in the environment are removed. Qualitative data from focus groups and pilot deployments should be used to assess which instantiation of sleep monitoring technologies are optimal for user (particularly older adults) adoption.

### 9.2.2 Addition of Markov Models in Predicting Sleep State Transitions for UMBS data

In this thesis, an approach is taken which relates the current state of the subject (wake or sleep) to the UMBS derived features measured from that person at that instant. However, many wrist actigraphy scoring approaches use data from previous and future epochs to define the current wake/sleep state in an effort to reduce the errors from scoring periods of quiescent wake as sleep (as discussed in Section 4.4). This implies that an individual is definitely currently awake if they move excessively in the next one or two minutes. While this is untrue, it is not unjustified to use data from surrounding epochs to inform current sleep/wake state. Ljtinen et al. (2003) proposed a series of rules, for example re-labelling short arousals as sleep, and short sleep fragments as wake. However, *Markov models* may also be used to implement such a system.

Markov models define the state of a random variable which changes over multiple iterations using previous states of that variable. For example, if a random variable represents when the subject is awake or asleep, the associated Markov model could be used to predict the likelihood of the person remaining in that state or transitioning to the other state. Future work could investigate specific configurations of Markov models which use UMBS features and previous sleep/wake states to predict the current sleep/wake state.

### 9.2.3 Assessment of Ambient Methods for Long-Term State Transition Analysis

Recent work has investigated the transitions between rest and activity using wrist actigraphy in a sample of community dwelling older adults (Lim et al., 2012). Increased levels of fragmentation in rest and activity were associated with cognitive impairment in these individuals, independent of total amount of rest or activity. Ambient sleep monitoring may provide an unobtrusive means to assess and longitudinally track the fragmentation of rest periods and subsequently be used to infer a decline in cognition. Future work

may involve the large scale deployment of inexpensive, unobtrusive sleep monitoring technologies which may be able to validate the longitudinal association between the fragmentation of rest periods and cognitive decline.

#### **9.2.4 An Investigation into Behavioural Change Technologies for Adopting Healthy Sleep Patterns**

While much work has focussed on the development of the technologies to capture quantitative statistics of sleep, little emphasis has been placed on how this data will be subsequently used. Providing clinicians with such information may result in an overburdening of data, especially as extra sensors (inclusive of non-sleep sensors) collect data. Feeding the sleep data back to the user may become ineffective over time, even when providing a context of 'healthy' boundaries for each metric are included, due to waning interest levels. Recent work investigating the development of persuasive pervasive technologies which provide peripheral feedback (perhaps through the background or wallpaper of a user's smartphone) relating to a user's attainment of their goals has shown positive results in increasing levels of physical activity (Consolvo et al., 2009). Subsequent more recent work includes educating users of healthy sleep behaviours (in terms of caffeine use, alcohol use and timing of exercise) and relating these to their own patterns highlighting causes of possible sleep disruptions (Bauer et al., 2012). However such a system requires active user journalling of their activities (and sleep) which may not be sustainable over extended periods of time (due to waning user engagement). Ambient methods of sleep monitoring may be used to provide the objective sleep data which may, in turn, inform the user of better sleep practices (such as maintaining a consistent bed time). Long-term studies should be used to elucidate whether such interventions result in increased quality of life, wellness and health status.

#### **9.2.5 An Investigation into the Electronic Delivery of Cognitive Behavioural Therapy for Insomnia**

Insomnia is associated with an inability to either initiate or remain asleep. It is the most common sleeping disorder and has a particularly high prevalence among older adults. Insomnia is associated with co-morbid conditions, medication use, circadian rhythm disturbances, primary sleep disorders and daytime dysfunction. Chronic insomnia has been found to be an independent risk factor for cognitive decline. Currently, the assessment of insomnia in a sleep laboratory is inaccessible due to high operational costs, and its clinic-based nature is inappropriate for vulnerable populations. The cautious use of sleep

medications for older adults is recommended since they are associated with poor health outcomes and mortality in older adults.

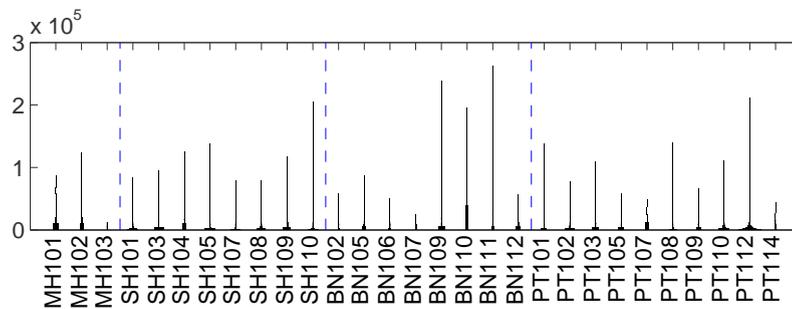
Ambient sleep monitoring modalities (such as the system described in this thesis) and subjective reports may facilitate the development of a home-based, inexpensive insomnia assessment system for older adults. Such a system may provide an insomnia risk factor to clinicians and will inform potential medication / non-medication treatment options. A subsequent further avenue of research may investigate the novel electronic delivery of cognitive behavioural therapy for insomnia. Additionally, there is evidence to suggest that when shown their actual sleep, insomniacs anxiety and sleep-related pre-occupations are reduced, and sleep quality improves. Such a project would investigate the diagnosis and treatment of insomnia in older adults delivered through a non-invasive, inexpensive, home-based system.

### 9.3 Overall Summary

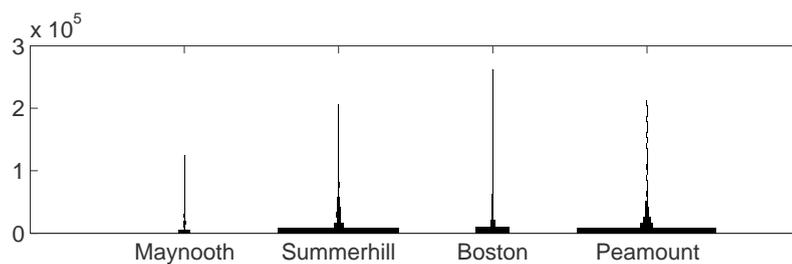
The fundamental focus of this thesis is the development of an appropriate technology for the long-term monitoring of sleep, particularly suited for placement in the homes of older adults. In its original clinical instantiation (mainly in nursing homes), the Under Mattress Bed Sensor (UMBS) provides measurements of presence in bed and quantifies the timing between bed movements for inferring the development of bedsores. The contributions of this work are in the development of algorithms which extend the capacity of this sensor, and are shown to accurately provide metrics of in-bed movement levels, respiration rates and sleep state. This was performed using data collected in experimental, domestic, clinical, and clinical research settings. Through the long-term deployment, it is envisaged that this system may be used to provide unobtrusive insights into overall health status, the progression of the symptoms of chronic conditions, and allow the objective measurement of daily (sleep/wake) patterns and routines.

## Appendix A

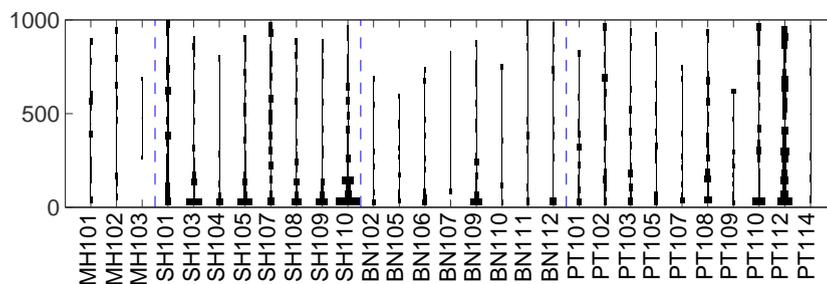
# UMBS Spatiotemporal Features - Variance Between Multiple Cohorts



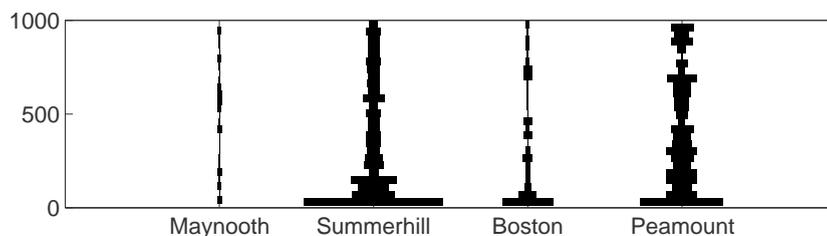
(a) Movement area over all subjects.



(b) Movement area over the 4 cohorts (% of the sleep period).

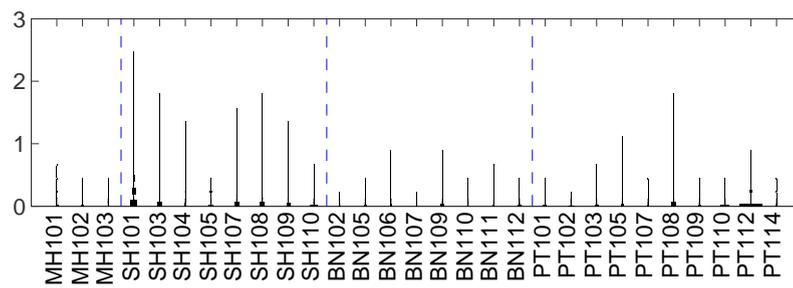


(c) Movement area over all subjects (reduced range).

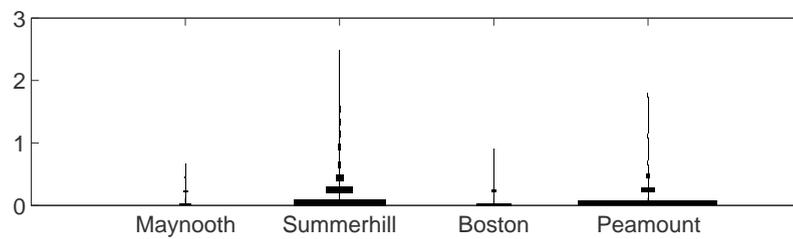


(d) Movement area over the 4 cohorts (reduced range).

FIGURE A.1: Movement Area

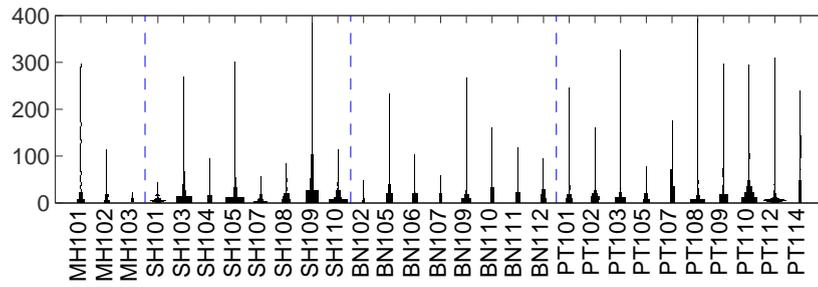


(a) Change in spread over all subjects.

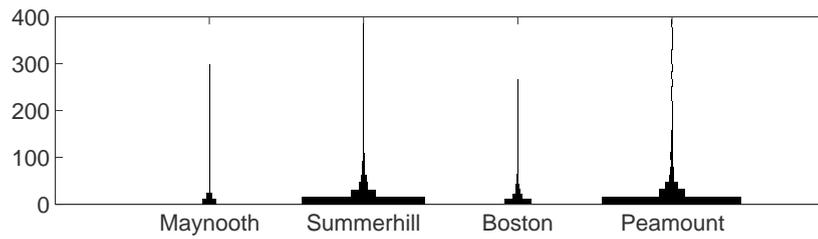


(b) Change in spread over the 4 cohorts.

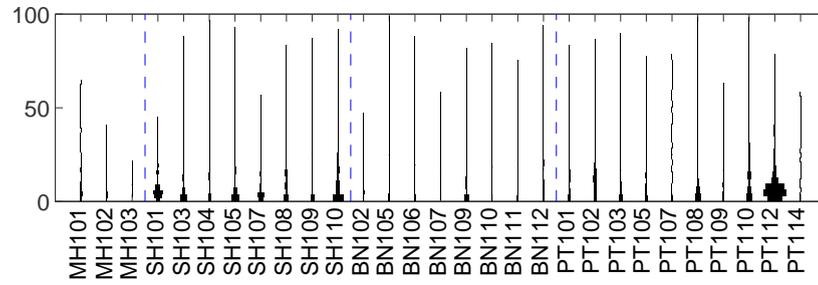
FIGURE A.2: Change in spread



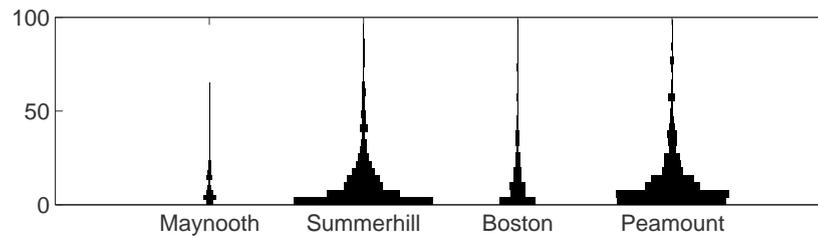
(a) Movement duration over all subjects.



(b) Movement duration over the 4 cohorts.

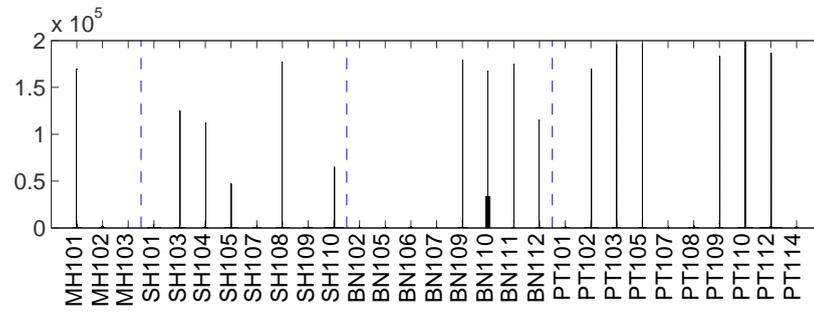


(c) Movement duration over all subjects (reduced range).

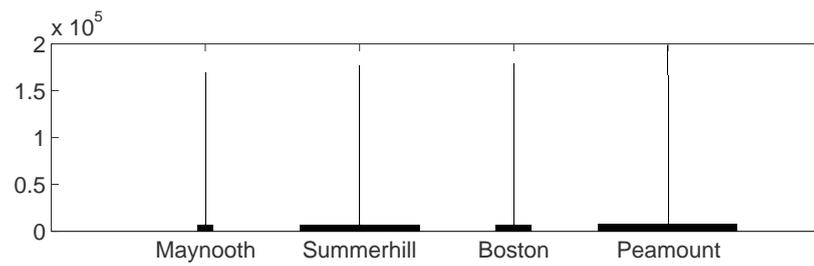


(d) Movement duration over the 4 cohorts (reduced range).

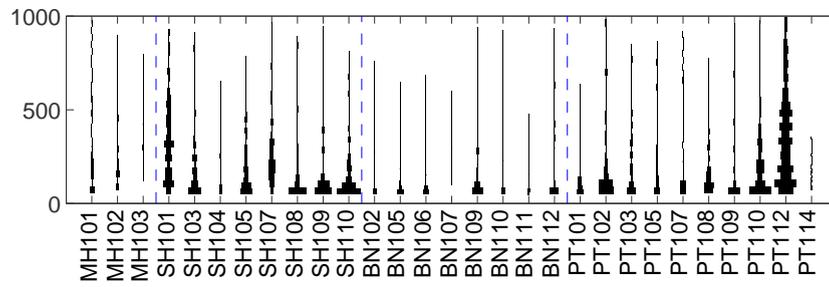
FIGURE A.3: Movement duration



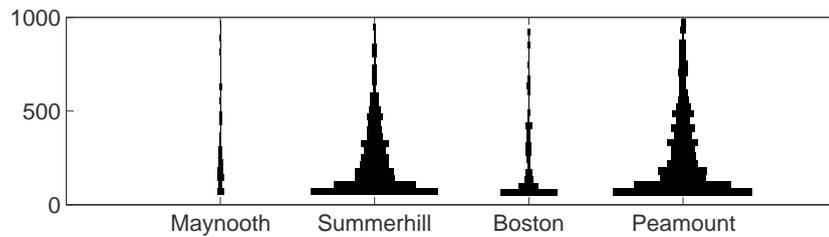
(a) Maximum movement magnitude over all subjects.



(b) Maximum movement magnitude over the 4 cohorts.

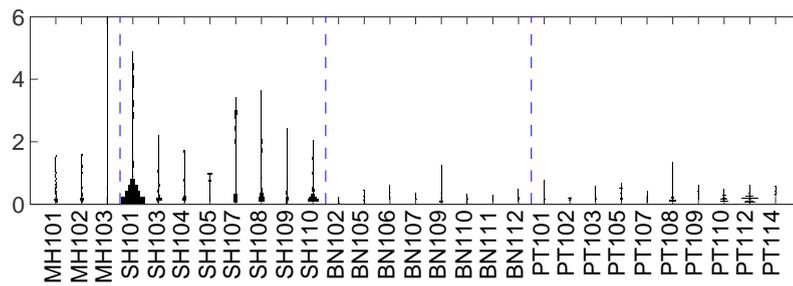


(c) Maximum movement magnitude over all subjects (reduced range).

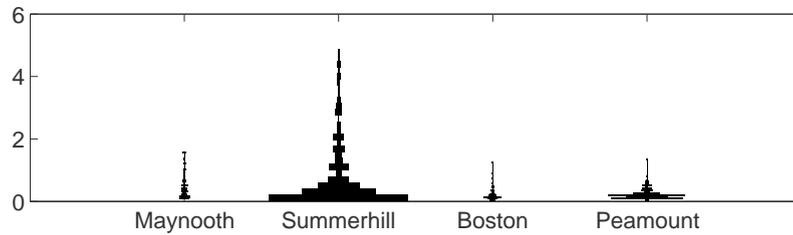


(d) Maximum movement magnitude over the 4 cohorts (reduced range).

FIGURE A.4: Maximum movement magnitude

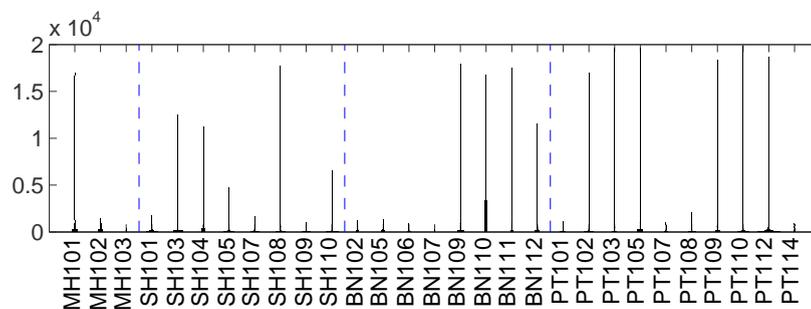


(a) Lateral change in position over all subjects.

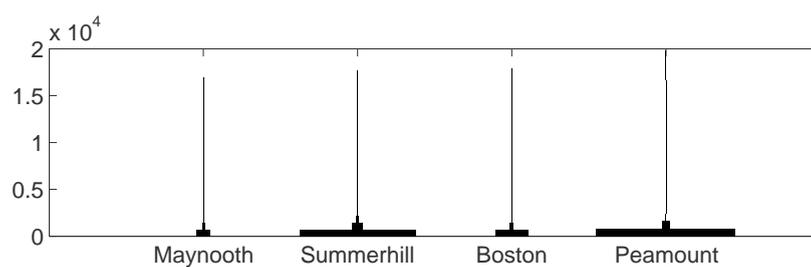


(b) Lateral change in position over the 4 cohorts.

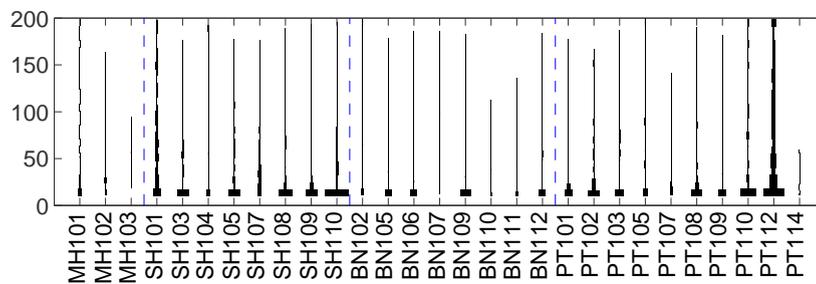
FIGURE A.5: Lateral change in position



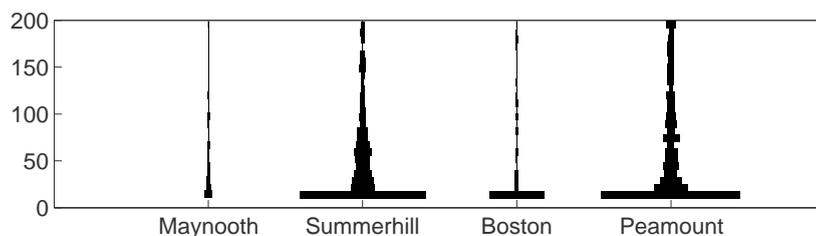
(a) Magnitude of movement over all subjects.



(b) Magnitude of movement over the 4 cohorts.

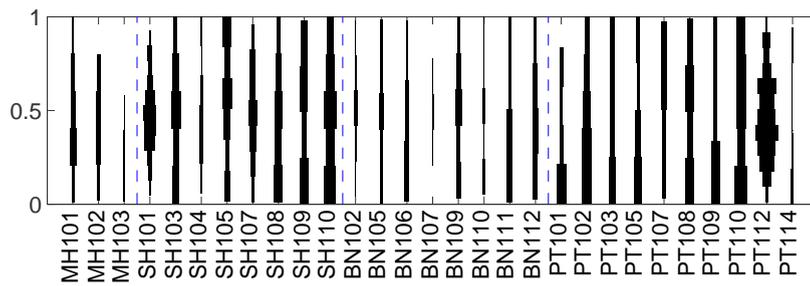


(c) Magnitude of movement over all subjects (reduced range).

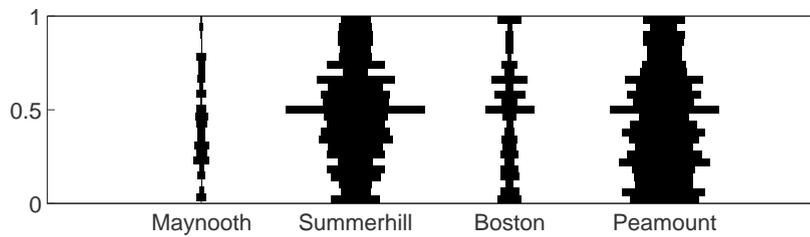


(d) Magnitude of movement over the 4 cohorts (reduced range).

FIGURE A.6: Magnitude of movement

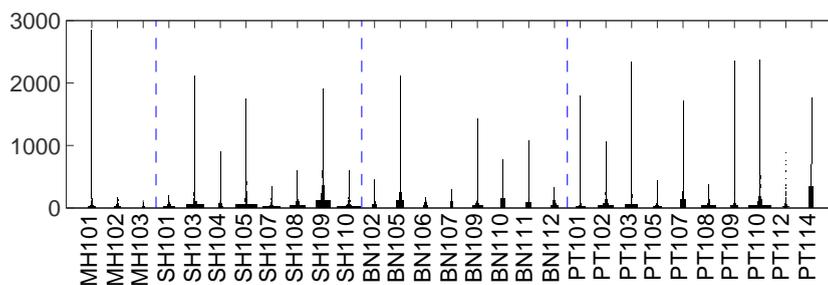


(a) Percentage to peak movement over all subjects.

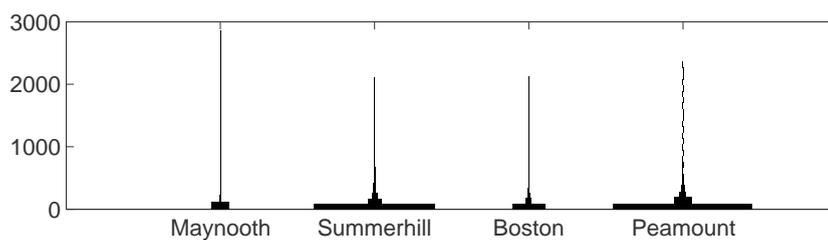


(b) Percentage to peak movement over the 4 cohorts.

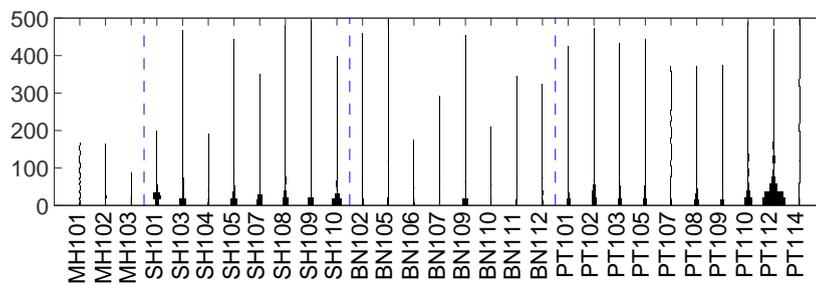
FIGURE A.7: Percentage to peak movement



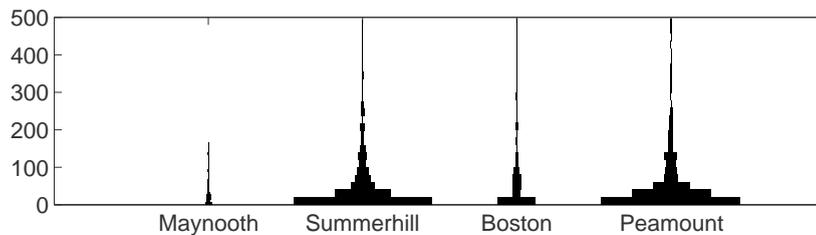
(a) Time to peak movement over all subjects.



(b) Time to peak movement over the 4 cohorts.

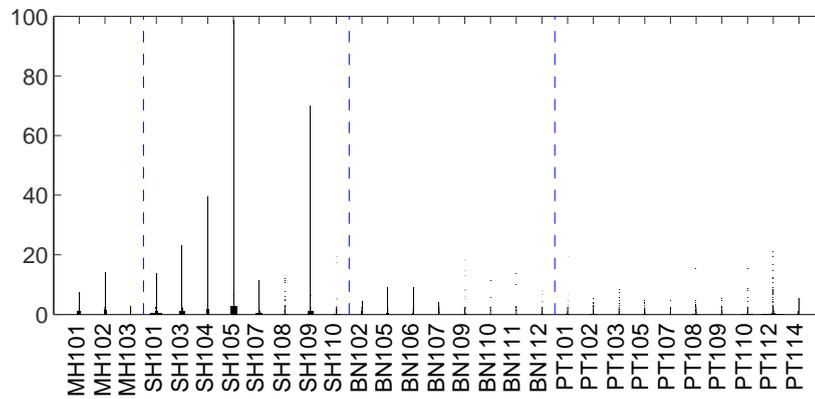


(c) Time to peak movement over all subjects (reduced range).

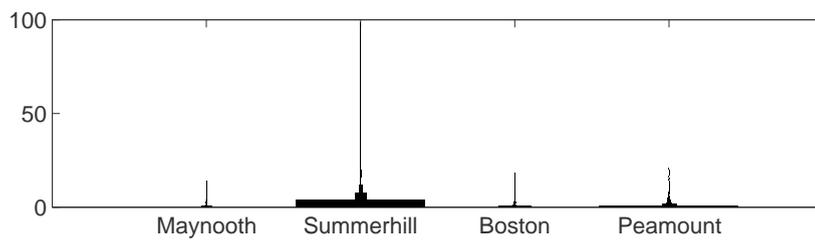


(d) Time to peak movement over the 4 cohorts (reduced range).

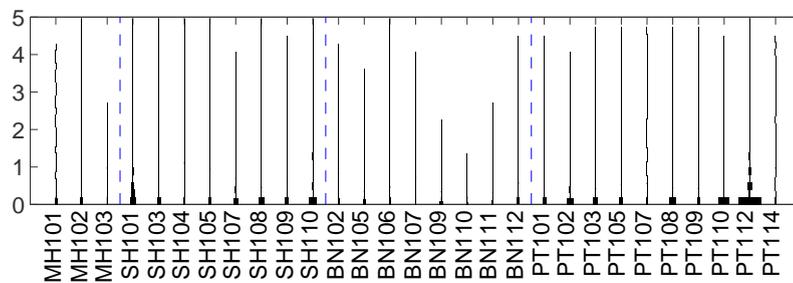
FIGURE A.8: Time to peak movement



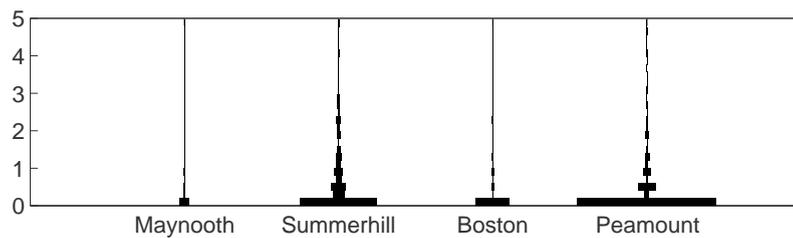
(a) Spread movement index over all subjects.



(b) Spread movement index over the 4 cohorts.



(c) Spread movement index over all subjects (reduced range).



(d) Spread movement index over the 4 cohorts (reduced range).

FIGURE A.9: Spread movement index

## Appendix B

# UMBS Spatiotemporal Features - Variance over Multiple days

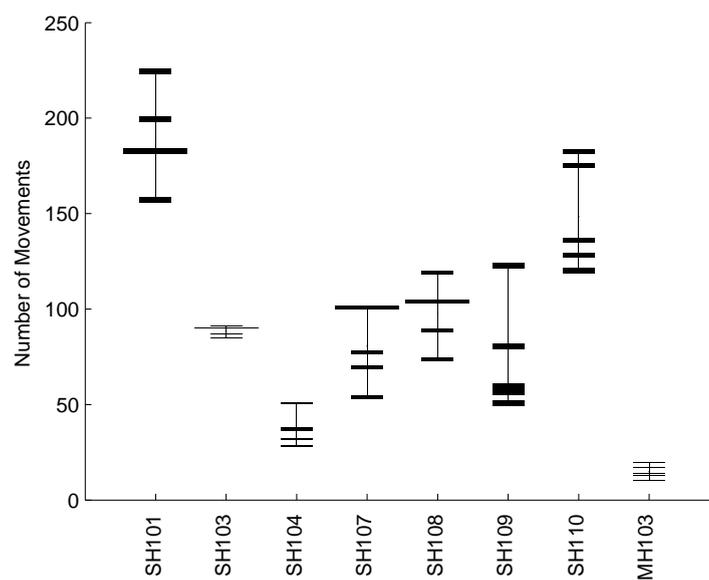


FIGURE B.1: numMovements CompOverMultNights

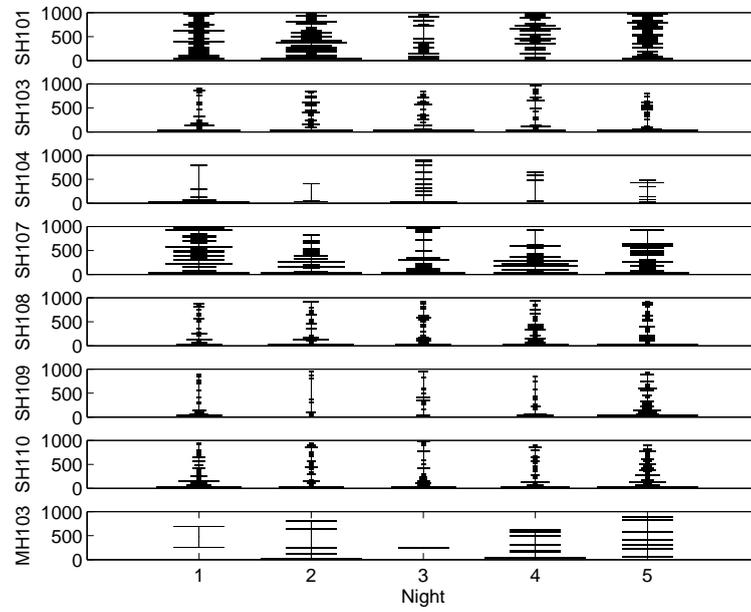


FIGURE B.2: movementArea CompOverMultNights

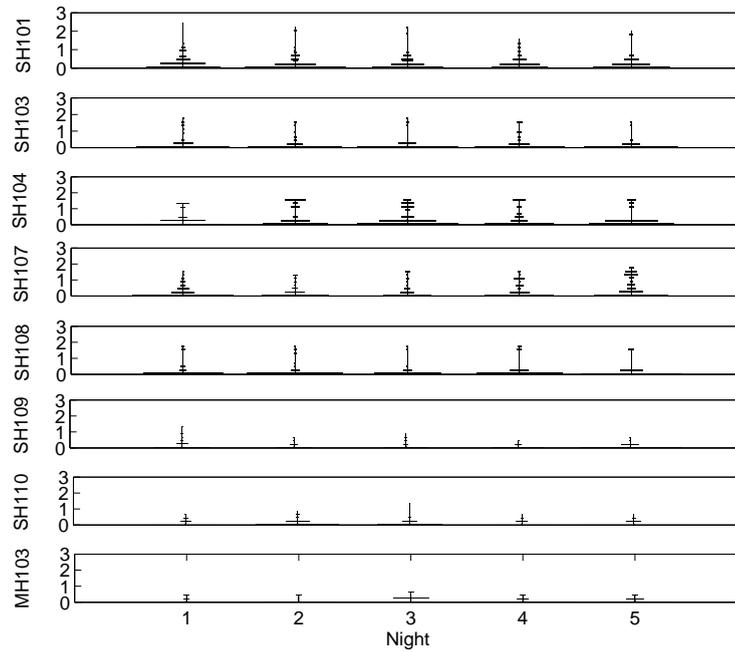


FIGURE B.3: deltaSpread CompOverMultNights

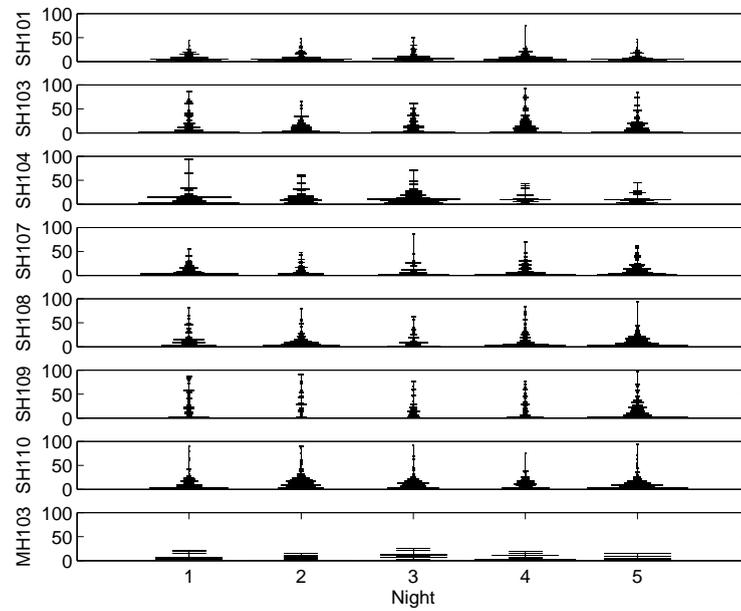


FIGURE B.4: duration CompOverMultNights

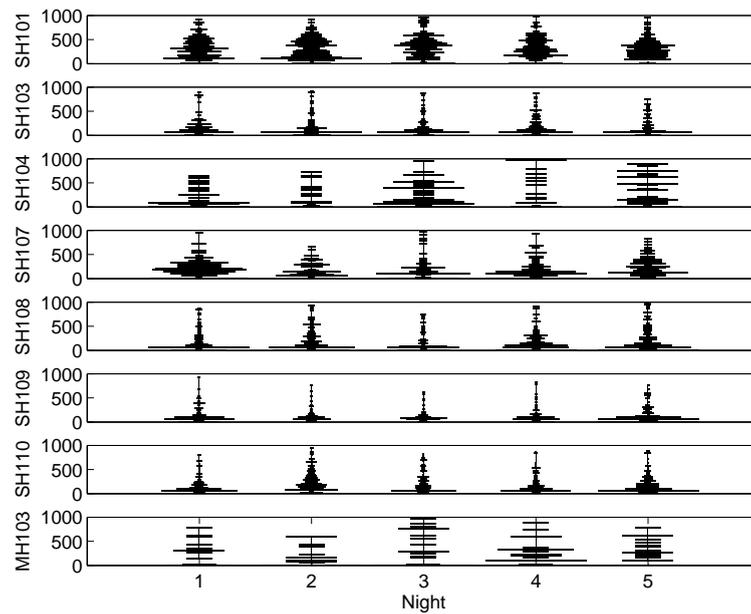


FIGURE B.5: magMovement CompOverMultNights

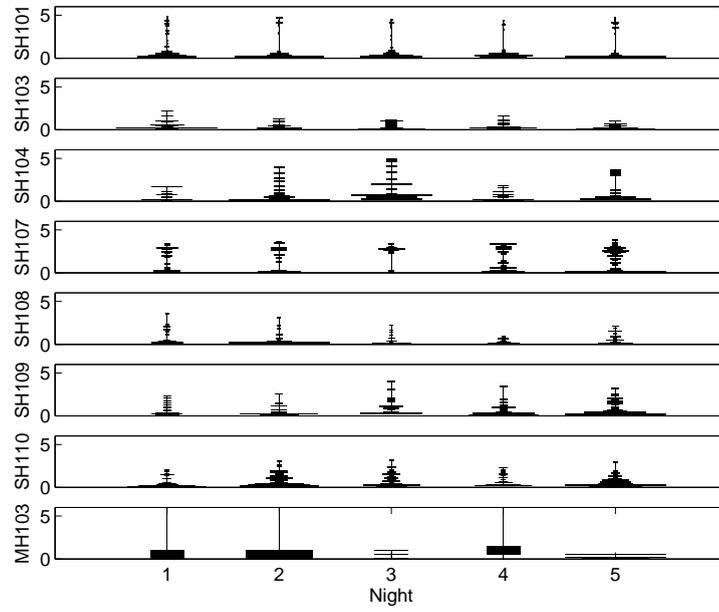


FIGURE B.6: mlChange CompOverMultNights

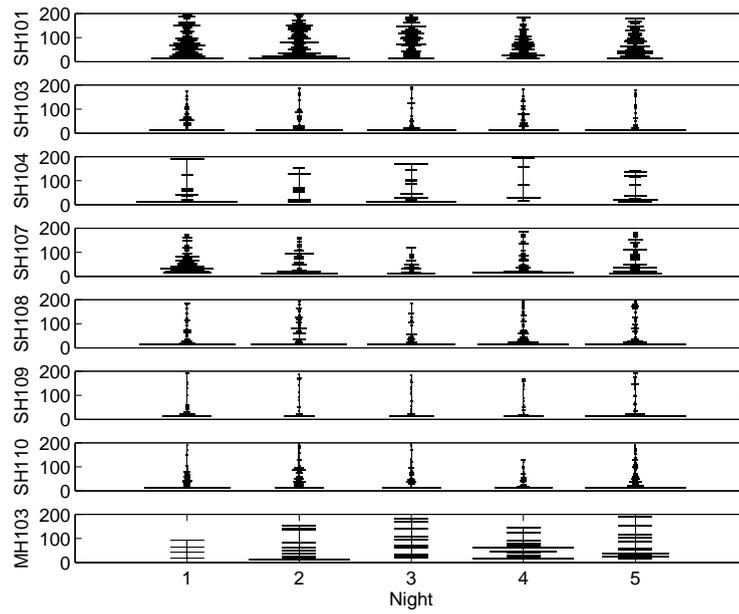


FIGURE B.7: mx CompOverMultNights

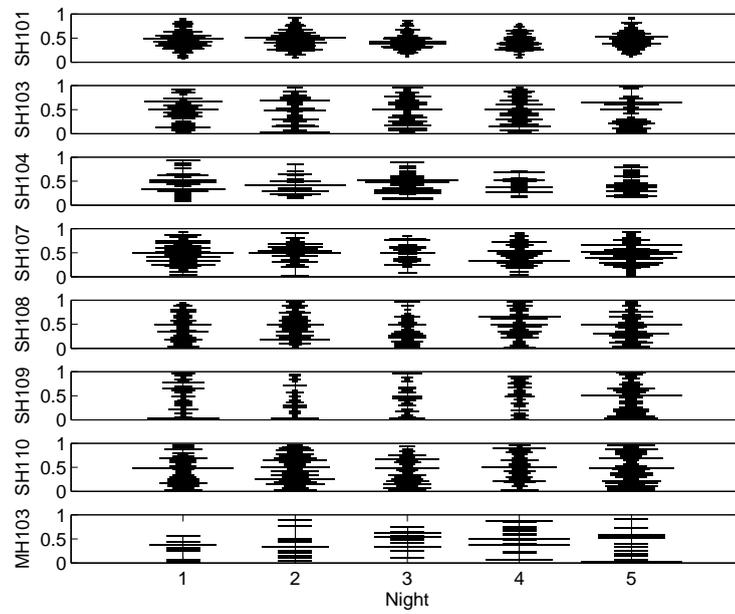


FIGURE B.8: percentToePeak CompOverMultNights

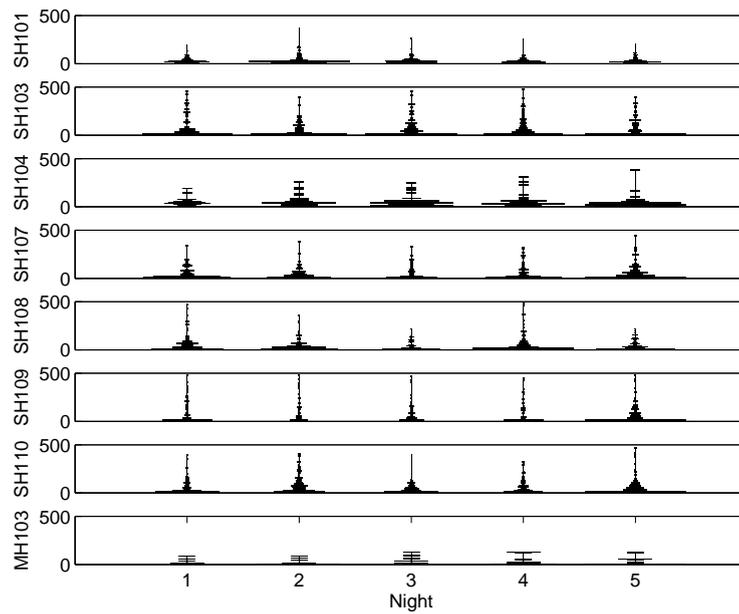


FIGURE B.9: timePeak CompOverMultNights

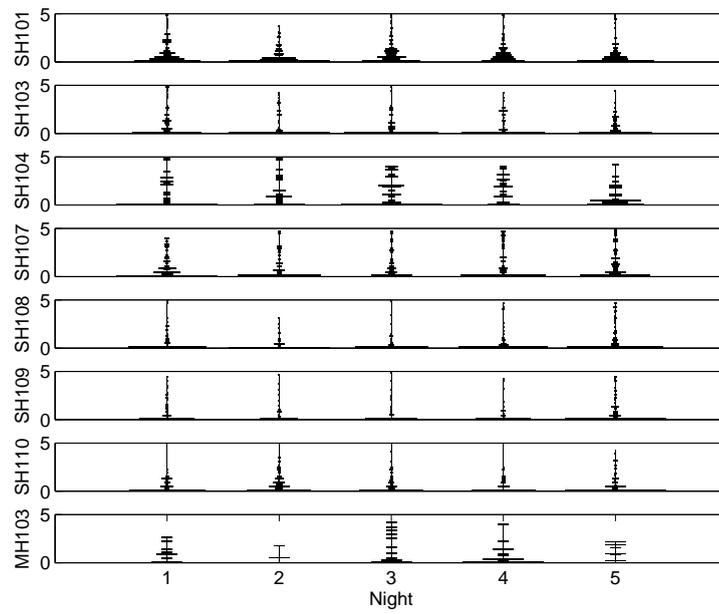


FIGURE B.10: sumDiffSpread CompOverMultNights

# Bibliography

- C Acebo and MK LeBourgeois. Actigraphy. *Respir Care Clin*, 12:23–30, 2006.
- Christine Acebo. Actigraphy. In T Lee-Chiong, editor, *Sleep: A Comprehensive Handbook*, chapter Actigraphy, pages 19–23. John Wiley & Sons, Inc, 2006.
- A. Adami, T. Hayes, M. Pavel, and C. Singer. Detection and classification of movements in bed using load cells. *Conf Proc IEEE Eng Med Biol Soc*, 1:589–592, 2005. doi: 10.1109/IEMBS.2005.1616481. URL <http://dx.doi.org/10.1109/IEMBS.2005.1616481>.
- A. Adami, M. Pavel, T. Hayes, and C. Singer. Detection of movement in bed with unobtrusive load cell sensors. *IEEE Trans Inf Technol Biomed*, Jan 2009a. doi: 10.1109/TITB.2008.2010701. URL <http://dx.doi.org/10.1109/TITB.2008.2010701>.
- Adriana M Adami, Tamara L Hayes, Misha Pavel, and Andre G Adami. Comparison of load cells and wrist-actigraphy for unobtrusive monitoring of sleep movements. *Conf Proc IEEE Eng Med Biol Soc*, 1:1314–1317, 2009b. doi: 10.1109/IEMBS.2009.5332577. URL <http://dx.doi.org/10.1109/IEMBS.2009.5332577>.
- R Agarwal and J Gotman. Computer-assisted sleep staging. *Biomedical Engineering, IEEE Transactions on*, 48(12):1412–1423, 2001.
- N. Agoulmine, M.J. Deen, Jeong-Soo Lee, and M. Meyyappan. U-health smart home. *Nanotechnology Magazine, IEEE*, 5(3):6–11, sept. 2011. ISSN 1932-4510. doi: 10.1109/MNANO.2011.941951.
- Beena Ahmed, Amira Redissi, and Reza Tafreshi. An automatic sleep spindle detector based on wavelets and the teager energy operator. *Conf Proc IEEE Eng Med Biol Soc*, 2009:2596–2599, 2009. doi: 10.1109/IEMBS.2009.5335331. URL <http://dx.doi.org/10.1109/IEMBS.2009.5335331>.
- A. Akhbardeh, S. Junnila, T. Koivistoinen, T. Koobi, and A. Varri. Ballistocardiogram classification using a novel transform so-called alimap and biorthogonal wavelets. *Intelligent Signal Processing, 2005 IEEE International Workshop on*, pages 64 – 69, sept. 2005. doi: 10.1109/WISP.2005.1531634.

- J. Alametsa, J. Viik, J. Alakare, A. Vrri, and A. Palomki. Ballistocardiography in sitting and horizontal positions. *Physiol Meas*, 29(9):1071–1087, Sep 2008. doi: 10.1088/0967-3334/29/9/006. URL <http://dx.doi.org/10.1088/0967-3334/29/9/006>.
- RP Allen, D Picchiatti, and WA Hening. Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. a report from the restless legs syndrome diagnosis and epidemiology workshop at the national institute of health. *Sleep Medicine*, 4(2):101–119, 2003.
- American Sleep Disorders Association. *International classification of sleep disorders, revised: diagnostic and coding manual*. American Sleep Disorders Association, Rochester (NY), 1997.
- S. Ancoli-Israel. Sleep problems in older adults: Putting myths to bed. *Geriatrics*, 52(1):20–29, 1997.
- S. Ancoli-Israel. Sleep and aging: prevalence of disturbed sleep and treatment considerations in older adults. *The Journal of Clinical Psychiatry*, 66(Suppl 9):24–30, 2005.
- S Ancoli-Israel, R Cole, C Alessi, M Chambers, W Moorcroft, and C.P. Pollak. The role of actigraphy in the study of sleep and circadian rhythms. *Seep*, 26(3):342–92, 2003.
- S. Ancoli-Israel, GS Richardson, RM Mangano, L Jenkins, P Hall, and WS Jones. Long-term use of sedative hypnotics in older patients with insomnia. *Sleep Medicine*, 6:107–113, 2005.
- P Anderer, G Gruber, S Parapatics, M Woertz, T Miazhyńska, G Klosch, B Saletu, J Zeitlhofer, MJ Barbanj, H Danker-Hopfe, SL Himanen, B Kemp, T Penzel, M Grozinger, D Kunz, P Rappelsberger, A Schlogl, and G Dorffner. An e-health solution for automatic sleep classification according to rechtschaffen and kales: validation study of the somnolyzer 24 x 7 utilizing the siesta database. *Neuropsychobiology*, 51(3):115, 2005.
- P Anderer, A Moreau, M Woertz, M Ross, G Gruber, S Parapatics, E Loretz, E Heller, A Schmidt, M Boeck, D Moser, G Kloesch, B Saletu, GM Saletu-Zyhlarz, H Danker-Hopfe, J Zeitlhofer, and G Dorffner. 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, 2010.
- Peter Anderer, Georg Gruber, Silvia Parapatics, and Georg Dorffner. Automatic sleep classification according to rechtschaffen and kales. *Conf Proc IEEE Eng Med Biol Soc*, 2007:3994–3997, 2007. doi: 10.1109/IEMBS.2007.4353209. URL <http://dx.doi.org/10.1109/IEMBS.2007.4353209>.

- E Aserinsky and N Kleitman. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 118(3062):273–274, September 1953.
- Xavier L Aubert and Andreas Brauers. Estimation of vital signs in bed from a single unobtrusive mechanical sensor: algorithms and real-life evaluation. *Conf Proc IEEE Eng Med Biol Soc*, 2008:4744–4747, 2008. doi: 10.1109/IEMBS.2008.4650273. URL <http://dx.doi.org/10.1109/IEMBS.2008.4650273>.
- L Ayalon, L Liu, and S Ancoli-Israel. Diagnosing and treating sleep disorders in the older adult. *Med Clin North Am*, 88:737–750, 2004.
- Liat Ayalon and Sonia Ancoli-Israel. Normal sleep in aging. In T. Lee-Chiong, editor, *Sleep: a Comprehensive Handbook*, chapter Normal Sleep in Aging, pages 3–9. John Wiley & Sons, Inc, 2006.
- S. I. Barbar, P. L. Enright, P. Boyle, D. Foley, D. S. Sharp, H. Petrovitch, and S. F. Quan. Sleep disturbances and their correlates in elderly japanese american men residing in hawaii. *J Gerontol A Biol Sci Med Sci*, 55(7):M406–M411, Jul 2000.
- Laura K Barger, Brian E Cade, Najib T Ayas, John W Cronin, Bernard Rosner, Frank E Speizer, and Charles A Czeisler. Extended work shifts and the risk of motor vehicle crashes among interns. *N Engl J Med*, 352(2):125–134, Jan 2005. doi: 10.1056/NEJMoa041401. URL <http://dx.doi.org/10.1056/NEJMoa041401>.
- R. Barttiti and F. Masulli. Bfgs optimization for faster and automated supervised learning. In *International Neural Network Conference*, volume 2, pages 757–760, 1990.
- M Basta, S Schiza, M Mauridis, C Lydakis, A Plaitakis, ME John, and MN Siafakas. Sleep breathing disorders in patients with idiopathic parkinson’s disease. *Respiratory Medicine*, 97(10):1151–1157, October 2003.
- C H. Bastien, A. Vallires, and C M. Morin. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med*, 2(4):297–307, Jul 2001.
- Jared Bauer, Sunny Consolvo, Benjamin Greenstein, Jonathan Schooler, Eric Wu, Nathaniel F. Watson, and Julie Kientz. Shuteye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, CHI ’12*, pages 1401–1410, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2208516.2208600. URL <http://doi.acm.org/10.1145/2208516.2208600>.
- W Baust, E Holzbach, and O Zechilin. Phasic changes in heart rate and respiration correlated with pgo-spike activity during rem sleep. *Pflugers Arch*, 331:113–123, 1972.

- Zachary T Beattie, Chad C Hagen, Misha Pavel, and Tamara L Hayes. Classification of breathing events using load cells under the bed. *Conf Proc IEEE Eng Med Biol Soc*, 1:3921–3924, 2009. doi: 10.1109/IEMBS.2009.5333548. URL <http://dx.doi.org/10.1109/IEMBS.2009.5333548>.
- J. Behan, D. Prendergast, B. Quigley, and L. Walsh. Determining the relationship between sleep and social activity in community dwelling older individuals. In *ISG 2008*, pages 024–029, 2008a.
- Julie Behan, David Prendergast, Lorcan Walsh, and Brenda Quigley. Social rhythms and nocturnal routines in community dwelling older adults. In *ICOST*, pages 73–80, 2008b.
- Selim R. Benbadis. Introduction to sleep electroencephalography. In T. Lee-Chiong, editor, *Sleep: A Comprehensive Handbook*, chapter Introduction to Sleep Electroencephalography, pages 989–1024. John Wiley & Sons, Inc., 2006.
- C Berthomier, X Drouot, M Herman-Stocia, P Berthomier, J Prado, D Bokar-Thire, O Benoit, J Mattout, and M d’Ortho. Automatic analysis of single-channel sleep eeg validation in healthy individuals. *Sleep*, 30(11):1587–1595, 2007.
- E. Billauer. Peakdet: Peak detection using matlab (non-derivative local extremum, maximum, minimum). Internet, July 2012. URL <http://www.billauer.co.il/peakdet.html>.
- SK Blake, SD Pittman, MM MacDonald, K Sun, B Lanzi, D clark, L Hueser, SE Fabregas, JR Shambroom, and DP White. Assessment of a wireless dry sensor to detect sleep in healthy volunteers and subjects with sleep disorders. *Sleep*, 32 (Suppl.):A370. Abstract 1132, 2009.
- JM Bland and DG Altman. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8:135–160, 1999.
- DL Bliwise. Dementia. In MH Kryger, T Roth, and WC Dement, editors, *Principles and Practice of Sleep Medicine*, chapter Normal Aging, pages 26–42. Saunders, 2000.
- DL Bliwise, AC King, RB Harris, and WL Haskell. Prevalence of self-reported poor sleep in a healthy population aged 5065. *Social Science and Medicine*, 34:49–55, 1992.
- Paolo Bonato. Wearable sensors and systems. from enabling technology to clinical applications. *IEEE Eng Med Biol Mag*, 29(3):25–36, 2010. doi: 10.1109/MEMB.2010.936554. URL <http://dx.doi.org/10.1109/MEMB.2010.936554>.

- G. S. Brassington, A. C. King, and D. L. Bliwise. Sleep problems as a risk factor for falls in a sample of community-dwelling adults aged 64-99 years. *J Am Geriatr Soc*, 48(10):1234–1240, Oct 2000.
- AR Braun, TJ Balkin, NJ Wesensten, RE Carson, M Varga, P Bardwin, S Selbie, G Belenky, and P Herscovitch. Regional cerebral blood flow throughout the sleep-wake cycle. an h2 150 pet study. *Brain*, 120:1173–1197, 1997.
- Maayan Bresler, Koby Sheffy, Giora Pillar, Meir Preiszler, and Sarah Herscovici. Differentiating between light and deep sleep stages using an ambulatory device based on peripheral arterial tonometry. *Physiological Measurement*, 29:571–584, 2008. doi: doi:10.1088/0967-3334/29/5/004.
- Mark Brink, Christopher H Mller, and Christoph Schierz. Contact-free measurement of heart rate, respiration rate, and body movements during sleep. *Behav Res Methods*, 38(3):511–521, Aug 2006.
- Christoph Brser, Kurt Stadthanner, Stijn de Waele, and Steffen Leonhardt. Adaptive beat-to-beat heart rate estimation in ballistocardiograms. *IEEE Trans Inf Technol Biomed*, 15(5):778–786, Sep 2011. doi: 10.1109/TITB.2011.2128337. URL <http://dx.doi.org/10.1109/TITB.2011.2128337>.
- CJC Burges. A tutorial on support vector machines for parttern recognition. *Data Mining and Knowledge Discovery*, 2(2), 1998.
- A. Burgos, A. Goni, A. Illarramendi, and J. Bemudez. Real-time detection of apneas on a pda. *IEEE Trans Inf Technol Biomed*, Nov 2009. doi: 10.1109/TITB.2009.2034975. URL <http://dx.doi.org/10.1109/TITB.2009.2034975>.
- L.E. Burney, W.A. Steiger, and T. Georges. Implications for comprehensive health care. *The Milbank Memorial Fund Quarterly*, 42(4):45–55, October 1964.
- Adrian Burns, Emer P. Doheny, Barry R. Greene, Timothy Foran, Daniel Leahy, Karol O’Donovan, and Michael J. McGrath. Shimmer: an extensible platform for physiological signal capture. *Conf Proc IEEE Eng Med Biol Soc*, 2010:3759–3762, 2010. doi: 10.1109/IEMBS.2010.5627535. URL <http://dx.doi.org/10.1109/IEMBS.2010.5627535>.
- Daniel J. Buysse, Charles F. Reynolds III, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, 28:193–213, 1989.
- Daniel J. Buysse, Martica L. Hall, Patrick J. Strollo, Thomas W. Kamarck, Jane Owens, Laisze Lee, and Steven E. Reisand Karen A. Matthews. Relationships between the pittsburgh sleep quality index (psqi), epworth sleepiness scale (ess), and

- clinical/polysomnographic measures in a community sample. *Journal of Clinical Sleep Medicine*, 4(6):563–571, 2008.
- J Caffarel, G. John Gibson, Harrison J. Phil, and Michael J. Griffiths, Clive J. and Drinnan. Comparison of manual sleep staging with automated neural network-based analysis in clinical practice. *Med Biol Eng Comput*, 44(1-2):105–110, 2006. doi: 10.1007/s11517-005-0002-4.
- A. Camurri, I. Lagerlof, and G. Volpe. Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *Int. J. Human-Computer Studies*, 59:213–225, 2003.
- E. Cantu-Paz, S. Newsam, and C. Kamath. Feature selection in scientific applications. In *Proceedings, ACM International Conference on Knowledge Discovery and Data Mining*, pages pp 788–793, Seattle, WA, August 2004.
- Francesco P. Cappuccio, Lanfranco D’Elia, Pasquale Strazzullo, and Michelle A. Miller. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*, 33(5):585–592, May 2010.
- B. W. Carlson, V. J. Neelon, and H. Hsiao. Evaluation of a non-invasive respiratory monitoring system for sleeping subjects. *Physiol Meas*, 20(1):53–63, Feb 1999.
- Neil R. Carlson, William Buskist, and G. Neil Martin. *Psychology: the Science of Behaviour (European Adaption)*. Allyn & Bacon, 2000.
- MA Carskadon, WC Dement, MM Mitler, T Roth, PR Westbrook, and S Keenan. Guidelines for the multiple sleep latency test (mslt): a standard measure of sleepiness. *Sleep*, 9(4):519–524, December 1986.
- Mary A Carskadon and William C Dement. Cumulative effects of sleep restriction on daytime sleepiness. *Psychophysiology*, 18(2):107–113, 1981.
- P.A. Cerny. Data mining and neural networks from a commercial perspective. In *ORSNZ Conference Twenty Naught One*, 2001.
- K R Chaudhuri, S Pal, A DiMarco, C Whately-Smith, K Bridgman, R Mathew, F R Pezzela, A Forbes, B Hgl, and C Trenkwalder. The parkinsons disease sleep scale: a new instrument for assessing sleep and nocturnal disability in parkinsons disease. *J Neurol Neurosurg Psychiatry*, 73:629–635, 2002.
- K Ray Chaudhuri, Suvankar Pal, K Bridgman, and C Trenkwalder. Acheiving 24-hour control of parkinson’s disease symptoms: Use of objective measures to improve nocturnal disability. *Eur neurol*, 46((suppl 1)):3–10, 2001. doi: DOI:10.1159/000058047.

- Yongjoon Chee, Jooman Han, Jaewoong Youn, and Kwangsuk Park. Air mattress sensor system with balancing tube for unconstrained measurement of respiration and heart beat movements. *Physiol Meas*, 26(4):413–422, Aug 2005. doi: 10.1088/0967-3334/26/4/007. URL <http://dx.doi.org/10.1088/0967-3334/26/4/007>.
- W. Chen, X. Zhu, T. Nemoto, Y. Kanemitsu, K. Kitamura, and K. Yamakoshi. Unconstrained detection of respiration rhythm and pulse rate with one under-pillow sensor during sleep. *Medical & Biological Engineering & Computing*, 43(2):p306 – 312, 2005. ISSN 01400118. URL <http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=19097553&site=ehost-live>.
- Wenxi Chen, Xin Zhu, Tetsu Nemoto, Kei-Ichiro Kitamura, Kayo Sugitani, and Daming Wei. Unconstrained monitoring of long-term heart and breath rates during sleep. *Physiol Meas*, 29(2):N1–10, Feb 2008. doi: 10.1088/0967-3334/29/2/N01. URL <http://dx.doi.org/10.1088/0967-3334/29/2/N01>.
- Ronald D Chervin. Epworth sleepiness scale? *Sleep Medicine*, 4:175–176, 2003.
- Ronald D Chervin and Christian Guilleminault. Assessment of sleepiness in clinical practice. *Nature Medicine*, 1(12):1252–1253, 1995.
- Byung Hun Choi, Gih Sung Chung, Jin-Seong Lee, Do-Un Jeong, and Kwang Suk Park. Slow-wave sleep estimation on a load-cell-installed bed: a non-constrained method. *Physiol Meas*, 30(11):1163–1170, Nov 2009. doi: 10.1088/0967-3334/30/11/002. URL <http://dx.doi.org/10.1088/0967-3334/30/11/002>.
- J. M. Choi, B. O. Kim, B. S. Hwang, R. H. Sohn, and K. S. Park. Unobtrusive body movement monitoring during sleep using infrared motion detector and zigbee protocol. In *Proc. 3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors*, pages 32–33, 2006. doi: 10.1109/ISSMDBS.2006.360090.
- Jung Han Choi and Dong Kyun Kim. A remote compact sensor for the real-time monitoring of human heartbeat and respiration rate. *Biomedical Circuits and Systems, IEEE Transactions on*, 3(3):181 –188, june 2009. ISSN 1932-4545. doi: 10.1109/TBCAS.2009.2019628.
- S Chokroverty. Phasic tongue movements in human rapid eye-movement sleep. *Neurology*, 30(6):665–668, 1980.
- SS Choudhary and SR Choudhary. Sleep effects on breathing and respiratory diseases. *Lung India*, 24(4):117–122, Oct-Dec 2009.
- P. Chow, G. Nagendra, J. Abisheganaden, and Y. T. Wang. Respiratory monitoring using an air-mattress system. *Physiol Meas*, 21(3):345–354, Aug 2000.

- Peter Clarenbach. Parkinson's disease and sleep. *J neurol*, 247(Suppl 4):IV/20–IV/23, September 2000.
- JA Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 70:37–46, 1960.
- R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin. Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5):461–469, Oct 1992.
- Nancy A. Collop. Polysomnography. In T. Lee-Chiong, editor, *Sleep: A Comprehensive Handbook*, chapter Polysomnography, pages 973–976. John Wiley & Sons, Inc., 2006.
- C. L. Comella. Sleep disturbances and excessive daytime sleepiness in parkinson disease: an overview. In P. Riederer, H. Reichmann, M. B. H. Youdim, and M. Gerlach, editors, *Parkinson's Disease and Related Disorders*, volume 70 of *Journal of Neural Transmission. Supplementa*, pages 349–355. Springer Vienna, 2006. ISBN 978-3-211-45295-0.
- CL Comella, TM Nardine, NJ Diederich, and GT Stebbins. Sleep-related violence, injury, and rem sleep behavior disorder in parkinson's disease. *Neurology*, 51(2):526–529, 1998. URL <http://www.neurology.org/cgi/content/abstract/51/2/526>.
- Sunny Consolvo, David W. McDonald, and James A. Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the 27th international conference on Human factors in computing systems, CHI '09*, pages 405–414, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518766. URL <http://doi.acm.org/10.1145/1518701.1518766>.
- Michael Coyle. Ambulatory cardiopulmonary data capture. In *2nd Annual International IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine & Biology*, pages 297–300, 2002.
- K Crowley. Sleep and sleep disorders in older adults. *Neuropsychology Reviews*, 21: 41–53, 2011.
- C. A. Czeisler, M. Dumont, J. F. Duffy, J. D. Steinberg, G. S. Richardson, E. N. Brown, R. Snchez, C. D. Ros, and J. M. Ronda. Association of sleep-wake habits in older people with changes in output of circadian pacemaker. *Lancet*, 340(8825):933–936, Oct 1992.
- A. C. da Rosa and T. Paiva. Automatic detection of k-complexes: validation in normals and dysthymic patients. *Sleep*, 16(3):239–248, Apr 1993.

- H Danker-Hopfe, P Anderer, J Zeitlhofer, M Boeck, H Dorn, G Gruber, E Heller, E Loretz, D Moser, S Parapatics, B Saletu, A Schmidt, and G Dorffner. Interrater reliability for sleep scoring according to the rechtschaffen & kales and the new aasm standard. *Journal of sleep Research*, 18:74, 2009.
- H. Davis, P. A. Davis, A. L. Loomis, E. N. Harvey, and G. Hobart. Changes in human brain potentials during the onset of sleep. *Science*, 86(2237):448–450, Nov 1937. doi: 10.1126/science.86.2237.448. URL <http://dx.doi.org/10.1126/science.86.2237.448>.
- Philip de Chazal, Emer O’Hare, Niall Fox, and Conor Heneghan. Assessment of sleep/wake patterns using a non-contact biomotion sensor. *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 514–517, aug. 2008. ISSN 1557-170X. doi: 10.1109/IEMBS.2008.4649203.
- Philip de Chazal, Conor Heneghan, and Walter T McNicholas. Multimodal detection of sleep apnoea using electrocardiogram and oximetry signals. *Philos Transact A Math Phys Eng Sci*, 367(1887):369–389, Jan 2009. doi: 10.1098/rsta.2008.0156. URL <http://dx.doi.org/10.1098/rsta.2008.0156>.
- Philip De Chazal, Niall Fox, Emer O’Hare, Conor Heneghan, Alberto Zaffaroni, Patricia Boyle, Stephanie Smith, Caroline O’Connell, and Walter T. McNicholas. Sleep/wake measurement using a non-contact biomotion sensor. *J Sleep Res*, 20(2):356–366, Jun 2011. doi: 10.1111/j.1365-2869.2010.00876.x. URL <http://dx.doi.org/10.1111/j.1365-2869.2010.00876.x>.
- T Deboer. Technologies of sleep research. *Cell Mol Life Sci*, 64:1227–1235, 2007.
- W. Dement and N. Kleitman. The relation of eye movements during sleep to dream activity: an objective method for the study of dreaming. *J Exp Psychol*, 53(5):339–346, May 1957.
- George Demiris, Debra Parker Oliver, Jarod Giger, Marjorie Skubic, and Marilyn Rantz. Older adults’ privacy considerations for vision based recognition methods of eldercare applications. *Technol Health Care*, 17(1):41–48, 2009. doi: 10.3233/THC-2009-0530. URL <http://dx.doi.org/10.3233/THC-2009-0530>.
- Department of Economic and Social Affairs, Population Division, United Nations. World population ageing 2009. Technical Report ESA/P/WP/212, United Nations, 2009.
- Mary Amanda Dew, Carolyn C. Hoch, Daniel J. Buysse, Timothy H. Monk, Amy E. Begley, Patricia R. Houck, Martica Hall, David J. Kupfer, and Charles F Reynolds, 3rd. Healthy older adults’ sleep predicts all-cause mortality at 4 to 19 years of follow-up. *Psychosom Med*, 65(1):63–73, 2003.

- Robert F. Dickerson, Eugenia I. Gorlin, and John A. Stankovic. Empath: a continuous remote emotional health monitoring system for depressive illness. In *Proceedings of the 2nd Conference on Wireless Health*, WH '11, pages 5:1–5:10, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0982-0. doi: 10.1145/2077546.2077552. URL <http://doi.acm.org/10.1145/2077546.2077552>.
- Karl Doghramji, Merrill M. Mitler, R. Bart Sangal, Colin Shapiro, Sheila Taylor, Joyce Walsleben, Cynthia Belisle, Milton K. Erman, Rosa Hayduk, Rima Hosn, Edward B. OMalley, JoAnne M. Sangal, Sharon L. Schutte, and James M. Youakim. A normative study of the maintenance of wakefulness test (mwt). *Electroencephalography and clinical Neurophysiology*, 103:554–562, 1997.
- G. Dorfman Furman, Z. Shinar, A. Baharav, and S. Akselrod. Electrocardiogram derived respiration during sleep. In *Proc. Computers in Cardiology*, pages 351–354, 2005.
- Jonas Dorn. distributionplot. Internet, July 2008. URL [www.mathworks.com/matlabcentral/fileexchange/23661-distribution-plot](http://www.mathworks.com/matlabcentral/fileexchange/23661-distribution-plot).
- Maciek Drejak. Sleep cycle alarm clock. <http://www.lexwarelabs.com/sleepcycle/>, Last Accessed on February, 2010 2010. URL <http://www.lexwarelabs.com/sleepcycle/>.
- Antonio Eblen-Zajjur. A simple ballistocardiographic system for a medical cardiovascular physiology course. *Adv Physiol Educ*, 27(1-4):224–229, Dec 2003.
- Farideh Ebrahimi, Mohammad Mikaeili, Edson Estrada, and Homer Nazeran. Automatic sleep stage classification based on eeg signals by using neural networks and wavelet packet coefficients. In *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, pages 1151–1154, 20–25 Aug. 2008. doi: 10.1109/IEMBS.2008.4649365.
- Cindy Ehlers and David Kupfer. Slow-wave sleep: do young adult men and women age differently? *Journal of Sleep Research*, 6(3):211–215, 1997. doi: 10.1046/j.1365-2869.1997.00041.x. URL <http://www.blackwell-synergy.com/doi/abs/10.1046/j.1365-2869.1997.00041.x>.
- European Union. *Demography Report 2010: Older, more numerous and diverse Europeans*. Publications Office of the European Union, Luxembourg, 2011.
- SE Fabregas, J Johnstone, and JR Shambroom. Performance of a wireless dry sensor system in automatically monitoring sleep and wakefulness. *Sleep*, 32 (Suppl.):A388. Abstract 1189, 2009.

- SA Factor, T McAlarney, JR Sanchez-Ramos, and WJ Weiner WJ. Sleep disorders and sleep effect in parkinson's disease. *Mov Disord*, 5(4):280–285, 1990.
- I. Feinberg. Changes in sleep cycle patterns with age. *J. Psychiat. Res.*, 10(3-4):283–306, October 1974.
- WF Flanigan. *The Sleeping Brain*, chapter Behavioral states and electroencephalograms of reptiles. Brain Information Service/Brain Research Institute, Los Angeles: UCLA, 1972.
- A Flexer, G Gruber, and G Dorffner. A reliable probabilistic sleep stage based on a single eeg signal. *Artificial Intelligence in Medicine*, 33:199–207, 2005.
- D. Foley, S. Ancoli-Israel, P. Britz, and J. Walsh. Sleep disturbances and chronic disease in older adults: Results of the 2003 national sleep foundation sleep in america survey. *J. Psychosom. Res.*, 56:497–502, 2004.
- DJ Foley, AA Monjan, SL Brown, EM Simonsick, RB Wallace, and DG Blazer. Sleep complaints among elderly persons: An epidemiologic study of three communities. *Sleep*, 18:425–432, 1995.
- V. Foo Siang Fook, Kng Poh Leong, E. Hao Jian Zhong, M. Jayachandran, Aung Aung Phyo Wai, J. Biswas, Lin Wei Si, and P. Yap. Non-intrusive respiratory monitoring system using fiber bragg grating sensor. *e-health Networking, Applications and Services, 2008. HealthCom 2008. 10th International Conference on*, pages 160 –164, july 2008. doi: 10.1109/HEALTH.2008.4600128.
- Niall A Fox, Conor Heneghan, Monica Gonzalez, Redmond B Shouldice, and Philip de Chazal. An evaluation of a non-contact biomotion sensor with actimetry. *Conf Proc IEEE Eng Med Biol Soc*, 2007:2664–2668, 2007. doi: 10.1109/IEMBS.2007.4352877. URL <http://dx.doi.org/10.1109/IEMBS.2007.4352877>.
- M.G. Frank. *Sleep: A Comprehensive Handbook*, chapter The Function of Sleep, pages 45–48. John Wiley & Sons, Inc, 2006.
- T. Freedom. Sleep and parkinson's disease. *Dis Mon.*, 53(5):275–290, May 2007.
- J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel. Disaggregated end-use energy sensing for the smart grid. *Pervasive Computing, IEEE*, 10(1):28 –39, jan.-march 2011. ISSN 1536-1268. doi: 10.1109/MPRV.2010.74.
- K. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183–192, 1989.

- MD Gjerstad, D Aarsland D, and JP Larsen. Development of daytime somnolence over time in parkinson's disease. *Neurology*, 58(10):1544–1546, May 2002.
- Robert M. Goldberg and Irving Beiber. A portable rem-detecting machine. *IEEE Trans Biomed Eng*, 26(9):513–516, 1979. ISSN 0018-9294. doi: 10.1109/TBME.1979.326433.
- A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, June 13 2000. Circulation Electronic Pages: <http://circ.ahajournals.org/cgi/content/full/101/23/e215>.
- Daniel J. Gottlieb, Coralyn W. Whitney, William H. Bonekat, Conrad Iber, Gary D. James, Michael Lebowitz, F. Javier Nieto, and Carl E. Rosenberg. Relation of sleepiness to respiratory disturbance index: The sleep heart health study. *Am. J. Respir. Crit. Care Med.*, 159(2):502–507, February 1999.
- B.R. Greene, A.O. Donovan, R. Romero-Ortuno, L. Cogan, C. Ni Scanail, and R.A. Kenny. Quantitative falls risk assessment using the timed up and go test. *Biomedical Engineering, IEEE Transactions on*, 57(12):2918–2926, dec. 2010. ISSN 0018-9294. doi: 10.1109/TBME.2010.2083659.
- May L. Griebel and Linda K. Moyer. Pediatric polysomnography. In T. Lee-Chiong, editor, *Sleep: a Comprehensive Handbook*, chapter Pediatric Polysomnography, pages 3–9. John Wiley & Sons, Inc, 2006.
- S Gudmundsson, TP Runarsson, and S Sigurdsson. Automatic sleep staging using support vector machines with posterior probability estimates. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, 2005.
- M.T. Hagan and M.B. Menhaj. Training feedforward networks with the marquardt algorithm. *IEEE trans Neural Networks*, 6(6):989–993, Nov 1994.
- Svenka Happe. Excessive daytime sleepiness and sleep disturbances in patients with neurological diseases. *Drugs*, 63(24):2725–2737, 2003. doi: 0012-6667/03/0024-2725.
- T. Harada, T. Sato, and T. Mori. Estimation of bed-ridden human's gross and slight movement based on pressure sensors distribution bed. In *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, volume 4, pages 3795–3800 vol.4, 2002.

- Amr A. Hassaan and Ahmed A. Morsy. Adaptive hybrid system for automatic sleep staging. In *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, pages 1631–1634, 20–25 Aug. 2008. doi: 10.1109/IEMBS.2008.4649486.
- T Hastie, R Tibshirani, and JH Friedman. *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- PJ Hauri and J Wisbey. Wrist actigraphy in insomnia. *Sleep*, 15:293–301, 1992.
- J Hedner, G Pillar, SD Pittman, D Zou, L Grote, and DP White. A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *sleep*, 27:1560–1566, 2004.
- Jan Hedner, David P. White, Atul Malhotra, Sarah Herscovici, Stephen D. Pittman, Ding Zou, Ludger Grote, and Giora Pillar. Sleep staging based on autonomic signals: a multi-center validation study. *J Clin Sleep Med*, 7(3):301–306, Jun 2011. doi: 10.5664/JCSM.1078. URL <http://dx.doi.org/10.5664/JCSM.1078>.
- C. Hernandez-Espinosa and M. Fernandez-Redondo. Multilayer feedforward weight initialization. In *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on*, volume 1, pages 166 –170 vol.1, 2001. doi: 10.1109/IJCNN.2001.939011.
- Sarah Herscovici, Avivit Peer, Surik Papyan, and Peretz Lavie. Detecting rem sleep from the finger: an automatic rem sleep algorithm based on peripheral arterial tone (pat) and actigraphy. *Physiological Measurement*, 28:129–140, 2007. doi: doi:10.1088/0967-3334/28/2/002. URL [http://www.iop.org/EJ/article/0967-3334/28/2/002/pm7\\_2\\_002.pdf?request-id=b79e3e94-9996-452e-ba2e-dabbabda0bc3](http://www.iop.org/EJ/article/0967-3334/28/2/002/pm7_2_002.pdf?request-id=b79e3e94-9996-452e-ba2e-dabbabda0bc3).
- J. Hislop, S. Arber, R. Meadows, and S. Venn. Narratives of the night: The use of audio diaries in researching sleep. *Sociological Research Online*, 10(4), 2005.
- B Hjorth. *CEAN - Computerised EEG Analysis*, chapter Time domain descriptors and their relation to a particular model for generation of EEG activity, pages 3–8. Gustav Fischer Verlag, 1975.
- M Hnatiuc, G Caruntu, and V Sarbu. The fuzzy system classifier using an intelligent mattress. In Paul Schiopu, editor, *Proc. of SPIE*, volume 7297 of *Advanced topics in optoelectronics, microelectronics and nanotechnologies IV*, 2009.
- E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, and W. C. Dement. Quantification of sleepiness: A new approach. *Psychophysiology*, 10(4):431–436, 1973.

- M. Holtzman, A. Arcelus, R. Goubran, and F. Knoefel. Breathing signal fusion in pressure sensor arrays. In *Medical Measurements and Applications, 2008. MeMeA 2008. IEEE International Workshop on*, pages 71–76, may 2008a. doi: 10.1109/MEMEA.2008.4543001.
- M. Holtzman, A. Arcelus, I. Veledar, R. Goubran, H. Sveistrup, and P. Guitard. Force estimation with a non-uniform pressure sensor array. In *Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE*, pages 1974–1979, may 2008b. doi: 10.1109/IMTC.2008.4547372.
- M. Holtzman, D. Townsend, R. Goubran, and F. Knoefel. Validation of pressure sensors for physiological monitoring in home environments. In *Medical Measurements and Applications Proceedings (MeMeA), 2010 IEEE International Workshop on*, pages 38–42, 30 2010-may 1 2010. doi: 10.1109/MEMEA.2010.5480218.
- M. Holtzman, R. Goubran, and F. Knoefel. Maximal ratio combining for respiratory effort extraction from pressure sensor arrays. In *Proc. (MeMeA) Workshop IEEE Int Medical Measurements and Applications*, pages 88–92, 2011. doi: 10.1109/MeMeA.2011.5966674.
- T Hori, Y Sugita, E Koga, S Shirakawa, K Inoue, S Uchida, H Kuwahara, M Kousaka, T Kobawashi, Y Tsuji, M Terashima, K Fukuda, and N Fukuda. Proposed supplements and amendments to a manual of standardized terminology, techniques and scoring system for sleep stages of human subjects, the rechtschaffen & kales (1968) standard. *Psychiatry and Clinical Neurosciences*, 55:305–310, 2001.
- C. C. Hsia, K. J. Liou, A. W. Aung, V. Foo, W. Huang, and J. Biswas. Analysis and comparison of sleeping posture classification methods using pressure sensitive bed system. *Conf Proc IEEE Eng Med Biol Soc*, 1:6131–6134, 2009. doi: 10.1109/IEMBS.2009.5334694. URL <http://dx.doi.org/10.1109/IEMBS.2009.5334694>.
- Eero Huupponen, Germn Gmez-Herrero, Antti Saastamoinen, Alpo Vrri, Joel Hasan, and Sari-Leena Himanen. Development and comparison of four sleep spindle detection methods. *Artif Intell Med*, 40(3):157–170, Jul 2007. doi: 10.1016/j.artmed.2007.04.003. URL <http://dx.doi.org/10.1016/j.artmed.2007.04.003>.
- C Iber, S Ancoli-Israel, AL Chesson, and S Quan. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*. American Academy of Sleep Medicine, Westchester, IL, 1st edition edition, 2007. URL <http://www.aasmnet.org/Store/ProductDetails.aspx?pid=176>.
- Information Unit, An Roinn Slainte (Department of Health), Ireland. *Long-stay Activity Statistics*. 2010.

- A Iranzo, JL Molinuevo, J Santamaria, M Serradell, MJ Marti, Valdeoriola, and E Tolosa. Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: a descriptive study. *Lancet Neurology*, 5(7):572–577, July 2006.
- B. Irie and S. Miyake. Capabilities of three-layered perceptrons. In *IEEE International Conference on Neural Networks*, volume 1, 1988.
- B. H. Jansen and K. Shankar. Sleep staging with movement-related signals. *Int J Biomed Comput*, 32(3-4):289–297, May 1993.
- B. H. Jansen, B. H. Larson, and K. Shankar. Monitoring of the ballistocardiogram with the static charge sensitive bed. *IEEE Trans Biomed Eng*, 38(8):748–751, Aug 1991. doi: 10.1109/10.83586. URL <http://dx.doi.org/10.1109/10.83586>.
- G. Jean-Louis, H. vonGizycki, F. Zizi, A. Spielman, P. Hauri, and H. Taub. The actigraph data analysis software: I. a novel approach to scoring and interpreting sleep-wake activity. *Percept Motor Skills*, 85(1):207–216, 1997.
- G. Jean-Louis, D.F. Kripke, R.J. Cole, J.D. Assmus, and R.D. Langer. Sleep detection with an accelerometer actigraph: comparisons with polysomnography. *Physio Behav*, 72:21–28, 2001.
- T. Joachims. *Making large-scale SVM learning practical. Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1999.
- M.W. Johns. A new method for measuring daytime sleepiness: the epworth sleepiness scale. *Sleep*, 14(6):540–545, 1991.
- M.W. Johns. Reliability and factor analysis of the epworth sleepiness scale. *Sleep*, 15(4):376–381, 1992.
- I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics, Aberdeen, 2nd edition edition, 2004.
- M.H. Jones, R. Goubran, and F. Knoefel. Reliable respiratory rate estimation from a bed pressure array. *Conf Proc IEEE Eng Med Biol Soc*, 1:6410–6413, 2006a. doi: 10.1109/IEMBS.2006.260164. URL <http://dx.doi.org/10.1109/IEMBS.2006.260164>.
- M.H. Jones, R. Goubran, and F. Knoefel. Identifying movement onset times for a bed-based pressure sensor array. In *2006 IEEE International Workshop on Medical Measurement and Applications*, pages 111–114, 2006b.
- K. Kaida, M. Takahashia, T. kerstedtb, A. Nakataa, Y. Otsukaa, T. Haratania, and K. Fukasawaa. Validation of the karolinska sleepiness scale against performance and eeg variables. *Clinical Neurophysiology*, 117(7):1574–1581, July 2006.

- E.S. Katz, M.G. Greene, K.A. Carlson, P. Galster, G.M. Loughlin, J.L. Carroll, and C.L. Marcus. Night-to-night variability of polysomnography in children with suspected obstructive sleep apnea. *J Pediatr*, 140:589–594, 2002.
- C. Kaufmann, R. Wehrle, T.C. Wetter, F. Holsboer, D.P. Auer, T. Pollmacher, and M. Czisch. Brain activation and hypothalamic functional connectivity during human non-rapid eye movement sleep: an eeg/fmri study. *Brain*, 129:655–667, 2006.
- K. Kaye, P. E. Hesla, and O. Rsj. The actiocylographic monitor of sleep. *Sleep*, 2(2): 253–260, 1979.
- A.H. Khandoker, M. Palaniswami, and C.K. Karmakar. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ecg recordings. *IEEE Trans Inf Technol Biomed*, 13(1):37–48, Jan 2009a. doi: 10.1109/TITB.2008.2004495. URL <http://dx.doi.org/10.1109/TITB.2008.2004495>.
- Ahsan H Khandoker, Jayavardhana Gubbi, and Marimuthu Palaniswami. Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings. *IEEE Trans Inf Technol Biomed*, 13(6):1057–1067, Nov 2009b. doi: 10.1109/TITB.2009.2031639. URL <http://dx.doi.org/10.1109/TITB.2009.2031639>.
- Ahsan H. Khandoker, Chandan K. Karmakar, and Marimuthu Palaniswami. Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings. *Computers in Biology and Medicine*, 39(1):88 – 96, 2009c. ISSN 0010-4825. doi: DOI:10.1016/j.combiomed.2008.11.003. URL <http://www.sciencedirect.com/science/article/B6T5N-4VCH3J4-1/2/4cdcacf807bd723e67e9ca57ccb17261>.
- Ahsan H Khandoker, Chandan K Karmakar, and Marimuthu Palaniswami. Automated recognition of patients with obstructive sleep apnoea using wavelet-based features of electrocardiogram recordings. *Comput Biol Med*, 39(1):88–96, Jan 2009d. doi: 10.1016/j.combiomed.2008.11.003. URL <http://dx.doi.org/10.1016/j.combiomed.2008.11.003>.
- Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(12):273 – 324, 1997. ISSN 0004-3702. doi: 10.1016/S0004-3702(97)00043-X. URL <http://www.sciencedirect.com/science/article/pii/S000437029700043X>. jce:title;Relevancej/ce:title.
- I. Korhonen, T. Iivainen, R. Lappalainen, T. Tuomisto, T. Kbi, V. Pentikinen, M. Tuomisto, and V. Turjanmaa. Terva: system for long-term monitoring of wellness

- at home. *Telemed J E Health*, 7(1):61–72, 2001. doi: 10.1089/153056201300093958. URL <http://dx.doi.org/10.1089/153056201300093958>.
- J. M. Kortelainen and J. Virkkala. Fft averaging of multichannel bcg signals from bed mattress sensor to improve estimation of heart beat interval. In *Proc. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2007*, pages 6685–6688, 2007. doi: 10.1109/IEMBS.2007.4353894.
- J.M. Kortelainen, M.O. Mendez, A.M. Bianchi, M. Matteucci, and S. Cerutti. Sleep staging based on signals acquired through bed sensor. *IEEE Trans Inf Technol Biomed*, 14(3):776–785, May 2010. doi: 10.1109/TITB.2010.2044797. URL <http://dx.doi.org/10.1109/TITB.2010.2044797>.
- C. A. Kushida, A. Chang, C. Gadkary, C. Guilleminault, O. Carrillo, and W. C. Dement. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*, 2(5):389–396, Sep 2001.
- H. Kuwahara, H. Higashi, Y. Mizuki, S.M.M. Tanaka, and K. Inanaga. Automatic real-time analysis of human sleep stages by an interval histogram method. *Electroencephalography and Clinical Neurophysiology*, 70:220–229, 1988.
- N.M. Al Lawati, S.R. Patel, and N.T. Ayas. Epidemiology, risk factors, and consequences of obstructive sleep apnea and short sleep duration. *Progress in Cardiovascular Diseases*, 51(4):285–293, 2009.
- TL Lee-Chiong, editor. *Sleep: A Comprehensive Handbook*. John Wiley & Sons, Inc, 2006.
- Andrew S P. Lim, Lei Yu, Madalena D. Costa, Sue E. Leurgans, Aron S. Buchman, David A. Bennett, and Clifford B. Saper. Increased fragmentation of rest-activity patterns is associated with a characteristic pattern of cognitive impairment in older individuals. *Sleep*, 35(5):633–640, 2012. doi: 10.5665/sleep.1820. URL <http://dx.doi.org/10.5665/sleep.1820>.
- Y.G. Lim, K.K. Kim, and K.S. Park. Ecg recording on a bed during sleep without direct skin-contact. *IEEE Journal of Biomedical Engineering*, 54(4):718–725, 2007. doi: 10.1109/TBME.2006.889194.
- M Littner, C.A. Kushida, W.M. Anderson, D. Bailey, R.B. Berry, D.G. Davila, M. Hirshkowitz, S. Kapen, M. Kramer, D. Loubé, M. Wise, and S.F. Johnson; Standards of Practice Committee of the American Academy of Sleep Medicine. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep*, 26(3):337–341, May 2003.

- S.W. Lockley. Circadian rhythms: Influence of light in humans. In Larry R. Squire, editor, *Encyclopedia of Neuroscience*, pages 971 – 988. Academic Press, Oxford, 2009. ISBN 978-0-08-045046-9. doi: DOI:10.1016/B978-008045046-9.01619-3. URL <http://www.sciencedirect.com/science/article/B98GH-4TVBCX5-W4/2/b71cf4c4f09ad760344f6a8cc2610622>.
- S.W. Lockley, J.W. Cronin, E.E. Evans, B.E. Cade, C.J. Lee, C.P. Landrigan, J.M. Rothschild, J.T. Katz, C.M. Lilly, P.H. Stone, D. Aeschbach, and C.A. Czeisler. Effect of reducing interns' weekly work hours on sleep and attentional failures. *N Engl J Med*, 351(18):1829–1837, 2004. doi: 10.1056/NEJMoa041404. URL <http://content.nejm.org/cgi/content/abstract/351/18/1829>.
- A. L. Loomis, E. N. Harvey, and G. Hobart. Further observations on the potential rhythms of the cerebral cortex during sleep. *Science*, 82(2122):198–200, Aug 1935a. doi: 10.1126/science.82.2122.198. URL <http://dx.doi.org/10.1126/science.82.2122.198>.
- A. L. Loomis, E. N. Harvey, and G. Hobart. Potential rhythms of the cerebral cortex during sleep. *Science*, 81(2111):597–598, Jun 1935b. doi: 10.1126/science.81.2111.597. URL <http://dx.doi.org/10.1126/science.81.2111.597>.
- J. Ltjnen, I. Korhonen, K. Hirvonen, S. Eskelinen, M. Myllymki, and M. Partinen. Automatic sleep-wake and nap analysis with a new wrist worn online activity monitoring device vivago wristcare. *Sleep*, 26(1):86–90, Feb 2003.
- V.M. Lubecke and O. Boric-Lubecke. Wireless technologies in sleep monitoring. In *Radio and Wireless Symposium, 2009. RWS '09. IEEE*, pages 135 –138, jan. 2009. doi: 10.1109/RWS.2009.4957303.
- David C Mack, James T Patrie, Robin A Felder, Paul M Suratt, and Majd Alwan. Sleep assessment using a passive ballistocardiography-based system: Preliminary validation. *Conf Proc IEEE Eng Med Biol Soc*, 1:4319–4322, 2009a. doi: 10.1109/IEMBS.2009.5333805. URL <http://dx.doi.org/10.1109/IEMBS.2009.5333805>.
- David C Mack, James T Patrie, Paul M Suratt, Robin A Felder, and Majd A Alwan. Development and preliminary validation of heart rate and breathing rate detection using a passive, ballistocardiography-based sleep monitoring system. *IEEE Trans Inf Technol Biomed*, 13(1):111–120, Jan 2009b. doi: 10.1109/TITB.2008.2007194. URL <http://dx.doi.org/10.1109/TITB.2008.2007194>.
- D.C. Mack, M. Alwan, B. Turner, P. Suratt, and R.A. Felder. A passive and portable system for monitoring heart rate and detecting sleep apnea and arousals: Preliminary validation. pages 51 –54, april 2006. doi: 10.1109/DDHH.2006.1624795.

- Tony Malim and Ann Birch. *Introductory Psychology*. Palgrave Macmillan, 2000.
- A. Mamelak and J. A. Hobson. Nightcap: a home-based sleep monitoring system. *Sleep*, 12(2):157–166, Apr 1989.
- K. Manabe, T. Matsui, M. Yamaya, T. Sato-Nakagawa, N. Okamura, H. Arai, and H. Sasaki. Sleep patterns and mortality among elderly patients in a geriatric hospital. *Gerontology*, 46(6):318–322,, 2000.
- D. Marquardt. An algorithm for least squares estimation of non-linear parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441, June 1963.
- E. Mattila, I. Korhonen, J. Merilahti, A. Nummela, M. Myllymaki, and H. Rusko. A concept for personal wellness management based on activity monitoring. pages 32–36, 30 2008-feb. 1 2008. doi: 10.1109/PCTHEALTH.2008.4571020.
- W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bussetin of Mathematical Biophysics*, 5(5):115–133, 1943.
- N. McGrogan, E. Braithwaite, and L. Tarassenko. Biosleep: a comprehensive sleep analysis system. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 2, pages 1608–1611 vol.2, 2001.
- J.E. McHugh, A.M. Casey, and B.A. Lawlor. Psychosocial correlates of aspects of sleep quality in community-dwelling irish older adults. *Aging & Mental Health*, 15(6):749–755, 2011.
- K. McMorro and W. Roeger. The economic consequences of ageing populations: A comparison of the eu, us and japan. Internet, [http : //ec.europa.eu/economy\\_finance/publications/publication11151\\_en.pdf](http://ec.europa.eu/economy_finance/publications/publication11151_en.pdf), May 2012.
- M.O. Mendez, A.M. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel. Sleep apnea screening by autoregressive models from a single ecg lead. *IEEE Trans Biomed Eng*, 56(12):2838–2850, Dec 2009. doi: 10.1109/TBME.2009.2029563. URL <http://dx.doi.org/10.1109/TBME.2009.2029563>.
- M.O. Mendez, J. Corthout, S. Van Huffel, M. Matteucci, T. Penzel, S. Cerutti, and A.M. Bianchi. Automatic screening of obstructive sleep apnea from the ecg based on empirical mode decomposition and wavelet analysis. *Physiol Meas*, 31(3):273–289, Mar 2010. doi: 10.1088/0967-3334/31/3/001. URL <http://dx.doi.org/10.1088/0967-3334/31/3/001>.

- Juho Merilahti, Ari Saarinen, Juha Parkka, Kari Antila, Elina Mattila, and Ilkka Korhonen. Long-term subjective and objective sleep analysis of total sleep time and sleep quality in real life settings. *Conf Proc IEEE Eng Med Biol Soc*, 2007:5202–5205, 2007. doi: 10.1109/IEMBS.2007.4353514. URL <http://dx.doi.org/10.1109/IEMBS.2007.4353514>.
- HA Middelkoop, E van Dam, and DA Smilde van den Doel. 45-hour multiple site actimetry in 20 healthy subjects: relations between body and limb movements and the effects of circadian sleep-wakefulness. *Sleep-Wake research in the Netherlands*, pages 681–684, 1995.
- L.E. Miles and W. C. Dement. Sleep and aging. *Sleep*, 3(3):119–220, 1980.
- Michael S Miletin and Patrick J Hanly. Measurement properties of the epworth sleepiness scale. *Sleep Medicine*, 4:195–199, 2003.
- M.L. Minsky and S.A. Papert. *Perceptrons*. MIT Press, Cambridge, MA, USA, 1969.
- MM Mitler. Methods of testing for sleepiness. *Behavioral Medicine*, 21:171–183, 1996.
- T Monk, D Buysse, and LR Rose. Wrist actigraphic measures of sleep in space. *Sleep*, 22(7):948–954, 1999.
- Timothy H. Monk, Charles F. Reynolds III, David J. Kupfer, Daniel J. Buysse, Patricia A. Coble, Amy J. Hayes, Mary Ann Machen, Susan R. Petrie, and Angela M. Ritenour. The pittsburgh sleep diary. *J Sleep Res*, 3:111–120, 1994.
- C. Morgenstern, R. Jan, M. Schwaibold, and W. Randerath. Automatic classification of inspiratory flow limitation assessed non-invasively during sleep. *Conf Proc IEEE Eng Med Biol Soc*, 2008:1132–1135, 2008. doi: 10.1109/IEMBS.2008.4649360. URL <http://dx.doi.org/10.1109/IEMBS.2008.4649360>.
- D. J. Mullaney, D. F. Kripke, and S. Messin. Wrist-actigraphic estimation of sleep time. *Sleep*, 3(1):83–92, 1980.
- K Nakajima, Y Matsumoto, and T Tamura. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiological Measurement*, 22:N21–N28, 2001.
- Grainne O’Donoghue, Niall Fox, Conor Heneghan, and Deirdre Hurley. Objective and subjective assessment of sleep in chronic low back pain patients compared with healthy age and gender matched controls: a pilot study. *BMC Musculoskeletal Disorders*, 10(1):122, 2009. ISSN 1471-2474. doi: 10.1186/1471-2474-10-122. URL <http://www.biomedcentral.com/1471-2474/10/122>.

- WG Oerlemans and AW de Weerd. The prevalence of sleep disorders in patients with parkinson's disease. a self-reported, community-based survey. *Sleep Medicine*, 3(2): 147–149, March 2002.
- MA Ohayon, MA Carskadon, C Guilminault, and MV Vitielo. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan. *Sleep*, 27:7, 2004.
- Shima Okada, Yuko Ohno, Goyahan, Kumi Kato-Nishimura, Ikuko Mohri, and Masako Taniike. Examination of non-restrictive and non-invasive sleep evaluation technique for children using difference images. *Conf Proc IEEE Eng Med Biol Soc*, 2008:3483–3487, 2008. doi: 10.1109/IEMBS.2008.4649956. URL <http://dx.doi.org/10.1109/IEMBS.2008.4649956>.
- A Oksenberg, C Gordon, E Arons, and L Sazbon. Phasic activities of rapid eye movement sleep in vegetative state patients. *Sleep*, 24(6):703–706, September 2001.
- J Orem. Neuronal mechanisms of respiration in rem sleep. *Sleep*, 3(3-4):251–267, 1980.
- P Paavilainen, I Korhonen, J Ljtinen, L Cluitmans, M Jylh, A Srel, and M Partinen. Circadian activity rhythm in demented and non-demented nursing-home residents measured by telemetric actigraphy. *J Sleep Res*, 14(1):61–68, Mar 2005a.
- P. Paavilainen, I. Korhonen, and Partinen M. Telemetric activity monitoring as an indicator of long-term changes in health and well-being of older people. *Gerontechnology*, 4(2):77–85, 2005b.
- Jiapu Pan and Willis J. Tompkins. A real-time qrs detection algorithm. *Biomedical Engineering, IEEE Transactions on*, BME-32(3):230–236, March 1985. ISSN 0018-9294. doi: 10.1109/TBME.1985.325532.
- Jean Paquet, Anna Kawinska, and Julie Carrier. Wake detection capacity of actigraphy during sleep. *Sleep*, 30(10):1362–1369, 2007.
- James Pardey, Stephen Roberts, Lionel Tarassenko, and John Stradling. A new approach to the analysis of the human sleep/wakefulness continuum. *Journal of Sleep Research*, 5(4):201–210, 1996. doi: 10.1111/j.1365-2869.1996.00201.x. URL <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2869.1996.00201.x>.
- H-J Park, J-S Oh, D-U Jeong, and K-S Park. Automated sleep stage scoring using hybrid rule- and case-based reasoning. *Computers and Biomedical Research*, 33((20)): 330–349, October 2000. URL <http://www.ingentaconnect.com/content/ap/co/2000/00000033/00000005/art01549>.

- Sairam Parthasarathy, Marypat Fitzgerald, James L. Goodwin, Mark Unruh, Stefano Guerra, and Stuart F. Quan. Nocturia, sleep-disordered breathing, and cardiovascular morbidity in a community-based cohort. *PLoS One*, 7(2):e30969, 2012. doi: 10.1371/journal.pone.0030969. URL <http://dx.doi.org/10.1371/journal.pone.0030969>.
- Ya-Ti Peng, Ching-Yung Lin, Ming-Ting Sun, and C.A. Landis. Multimodality sensor system for long-term sleep quality monitoring. *Biomedical Circuits and Systems, IEEE Transactions on*, 1(3):217–227, sept. 2007. ISSN 1932-4545. doi: 10.1109/TBCAS.2007.914481.
- Giora Pillar, Amir Bar, Arie Shlitner, Robert Schnall, Jacob Shefy, and Peretz Lavie. Autonomic arousal index: an automated detection based on peripheral arterial tonometry. *Sleep*, 25(5):543–549, Aug 2002.
- Giora Pillar, Amir Bar, Michal Betito, Robert P Schnall, Itsik Dvir, Jacob Sheffy, and Peretz Lavie. An automatic ambulatory device for detection of aasm defined arousals from sleep: the wp100. *Sleep Med*, 4(3):207–212, May 2003.
- SD Pittman, MM MacDonald, RB Fogel, A Malhotra, K Todros, B Levy, AB Geva, and DP White. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep*, 27(7):1394–1403, Nov 2004.
- CP Pollak and D Perlick. Sleep problems and institutionalization of the elderly. *Neurology*, 4:204–210, 1991.
- CP Pollak, WW Tryon, H Nagaraja, and Roger Dzwonczyk. How accurately does wrist actigraphy identify the states of sleep and wakefulness. *Sleep*, 24(8):957–965, 2001.
- P. Prakash, A. Johnston, B. Honari, and S.F. McLoone. Optimal wafer site selection using forward selection component analysis. In *Advanced Semiconductor Manufacturing Conference (ASMC), 2012 23rd Annual SEMI*, pages 91–96, may 2012. doi: 10.1109/ASMC.2012.6212875.
- W.H. Press and G.B. Rybicki. Fast algorithm for spectral analysis of unevenly sampled data. *Astrophysical Journal Part 1*, 338:277–280, March 1989. doi: 10.1086/167197.
- P. Prinz, W. Obrist, and H. Wang. Sleep patterns in healthy elderly subjects: Individual differences as related to other neurological variables. *Sleep Res.*, 4(132), 1975.
- PN Prinz, LH Larsen, KE Moe, EM Dulberg, and MV Vitiello. C stage, automated sleep scoring: Development and comparison with human sleep scoring for healthy older men and women. *Sleep*, 17(8):711–717, 1994.

- Anil Natesan Rama, S. Charles Cho, and Clete A. Kushida. Normal human sleep. In T. Lee-Chiong, editor, *Sleep: a Comprehensive Handbook*, chapter Normal Human Sleep, pages 3–9. John Wiley & Sons, Inc, 2006.
- M. J. Rantz, M. Skubic, S. J. Miller, and J. Krampe. Using technology to enhance aging in place. *Lect Notes Comput Sci*, 5120:169–176, 2008. doi: 10.1007/978-3-540-69916-3\_20. URL [http://dx.doi.org/10.1007/978-3-540-69916-3\\_20](http://dx.doi.org/10.1007/978-3-540-69916-3_20).
- Marilyn J Rantz, Marjorie Skubic, and Steven J Miller. Using sensor technology to augment traditional healthcare. *Conf Proc IEEE Eng Med Biol Soc*, 1:6159–6162, 2009. doi: 10.1109/IEMBS.2009.5334587. URL <http://dx.doi.org/10.1109/IEMBS.2009.5334587>.
- A. Rechtschaffen and A. Kales. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *NIH Publication No. 204, US Government Printing Office, Washington, DC.,* 1968.
- S Redline, HL Kirchner, SE Quan, DJ Gottlieb, V Kapur, and A Newman. The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture. *Arch Intern Med*, 164(4):406–418, February 23 2004.
- DP Redmond and FW Hegge. The design of human activity monitors. In LE Scheving, F Halberg, and CF Ehret, editors, *Chronobiology and Chronobiological Engineering*, pages 202–215. Martinus Nijhoff Publishers, 1987.
- C. Richard and R. Lengelle. Joint time and time-frequency optimal detection of k-complexes in sleep eeg. *Comput Biomed Res*, 31(3):209–229, Jun 1998.
- Rixt F. Riemersma-van der Lek, Dick F. Swaab, Jos Twisk, Elly M. Hol, Witte J. G. Hoogendijk, and Eus J. W. Van Someren. Effect of bright light and melatonin on cognitive and noncognitive function in elderly residents of group care facilities. *JAMA: The Journal of the American Medical Association*, 299(22):2642–2655, 2008. doi: 10.1001/jama.299.22.2642. URL <http://jama.ama-assn.org/content/299/22/2642.abstract>.
- Ann E. Rogers, Claire C. Caruso, and Michael S Aldrich. Reliability of sleep diaries for assessment of sleep/wake patterns. *Nursing Research*, 42(6):368–371, 1993.
- F. Rosenblatt. *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books, New York, USA, 1962.
- M. Rosenblatt, editor. *The Quefrequency Alanalysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking*, volume Proceedings of

- the Symposium on Time Series Analysis; Ch. 15, pp 209–243 of *Proceedings of the Symposium on Time Series Analysis*, New York, 1963. Wiley.
- W. H. Rosenblatt. Ballistocardiography: concepts of its applicability in the office practice of cardiology. *Dis Chest*, 32(4):400–412, Oct 1957.
- Leon Rosenthal. Physiologic processes during sleep. In T Lee-Chiong, editor, *Sleep: A Comprehensive Handbook*, chapter Physiologic processes during sleep, pages 19–23. John Wiley & Sons, Inc, 2006.
- S. Rostig, JW Kantelhardt, T Penzel, W Cassel, JH Peter, C Vogelmeier, HF Becker, and A Jerrentrup. Nonrandom variability of respiration during sleep in healthy humans. *Sleep*, 28(4):411, 2005.
- T Roth and S Ancoli-Israel. Daytime consequences and correlates of insomnia in the united states: results of the 1991 national sleep foundation survey ii. *Sleep*, 22:S354–S358, 1999.
- T Roth and TA Roehrs. Etiologies and sequelae of excessive day- time sleepiness. *Clinical Therapy*, 18:52–576, 1996.
- T. Roth, DB Rye, LD Borchert, C Bartlett, DL Bliwise, C Cantor, JM Gorell, JP Hubble, B Musch, CW Olanow, C Pollak, MB Stern, and RL Watts. Assessment of sleepiness and unintended sleep in parkinson’s disease patients taking dopamine agonists. *Sleep Med*, 4(4):275–280, July 2003.
- D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning representations by back-propagating errors. *Letters to Nature*, 323:533–536, 1986.
- DB Rye. Parkinson’s disease and rls: the dopaminergic bridge. *Sleep Med*, 5(3):317–328, May 2004.
- A. Sadeh. The role and validity of actigraphy in sleep medicine: an update. *Sleep Medicine Reviews*, 15(4):259–264, August 2011.
- A Sadeh, KM Sharkey, and MA Carskadon. Activity-based sleep/wake identification: An empirical test of methodological issues. *Sleep*, 17:201–207, 1994.
- Avi Sadeh, Peter J Hauri, Daniel F Kripke, and Peretz Lavie. The role of acigraphy in the evaluation of sleep disorders. *Sleep*, 18(4):288–302, 1995.
- A. Sarela, I. Korhonen, J. Lotjonen, M. Sola, and M. Myllymaki. Ist vivago reg; - an intelligent social and remote wellness monitoring system for the elderly. pages 362 – 365, april 2003. doi: 10.1109/ITAB.2003.1222554.

- N Schaltenbrand, R Lengelle, M Toussaint, R Luthringer, G Carelli, A Jacqmin, E Lainey, A Muzet, and JP Macher JP. Sleep stage scoring using the neural network model: comparison between visual and automatic analysis in normal subjects and patients. *Sleep*, 17(1):23–35, 1996.
- CH Schenck, SR Bundlie, and MW Mahowald. Delayed emergence of an parkinsonian disorder in 38 initially diagnosed with idiopathic rapid eye movement sleep behaviour disorder. *Neurology*, 46(6):388–393, February 1996.
- CH Schenck, AL Callies, and MW Mahowald. Increased percentage of slow-wave sleep in rem sleep behaviour disorder (rbd): a reanalysis of previously published data from a controlled study of rbd reported in sleep. *Sleep*, 26(8):1066, December 2003.
- Hartmut Schulz. Rethinking sleep analysis. *Journal of Clinical Sleep Medicine*, 4(2):99, 2008.
- Rosalyn Seeton and Andy Adler. Sensitivity of a single coil electromagnetic sensor for non-contact monitoring of breathing. *Conf Proc IEEE Eng Med Biol Soc*, 2008:518–521, 2008. doi: 10.1109/IEMBS.2008.4649204. URL <http://dx.doi.org/10.1109/IEMBS.2008.4649204>.
- JR Shambroom, J Johnstone, and SE Fabregas. Evaluation of a portable monitor for staging sleep. *Sleep*, 32 (Suppl.):A386. Abstract 1182, 2009.
- JR Shambroom, SE Fabregas, and J Johnstone. Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research*, [epub ahead of print], 2011.
- J. Shawe-Taylor and N Christiani. *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
- J. W. Shin, Y. R. Yoon, and J. C. Principe. Evaluation of body movement during sleep by thermopile using wavelet and neuro-fuzzy. In *Proc. 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 2406–2407, 17–21 Sept. 2003. doi: 10.1109/IEMBS.2003.1280401.
- Jae Hyuk Shin, Young Joon Chee, Do-Un Jeong, and Kwang Suk Park. Nonconstrained sleep monitoring system and algorithms using air-mattress with balancing tube method. *IEEE Trans Inf Technol Biomed*, 14(1):147–156, Jan 2010. doi: 10.1109/TITB.2009.2034011. URL <http://dx.doi.org/10.1109/TITB.2009.2034011>.
- JR Smith and I Karacan. Electroencephalographic sleep stage scoring by an automated hybrid system. *Electroencephalography and Clinical Neurophysiology*, 31(3):231–237, 1971.

- Simon Smith and John Trinder. Detecting insomnia: comparison of four self-report measures of sleep in a young adult population. *Journal of Sleep Research*, 10(3): 229–235, 2001. ISSN 1365-2869. doi: 10.1046/j.1365-2869.2001.00262.x. URL <http://dx.doi.org/10.1046/j.1365-2869.2001.00262.x>.
- L De Souza, AA Benedito-Silva, ML Pires, D Poyares, S Tufik, and HM Calil. Further validation of actigraphy for sleep studies. *Sleep*, 26(1):81–85, 2003.
- W B Spillman, M Mayer, J Bennett, J Gong, K E Meissner, B Davis, R O Claus, A A Muelenaer Jr, and X Xu. A 'smart' bed for non-intrusive monitoring of patient physiological factors. *Measurement Science and Technology*, 15(8):1614, 2004. URL <http://stacks.iop.org/0957-0233/15/i=8/a=032>.
- John M. Stern. *Atlas of EEG Patterns*. Lippincott Williams & Wilkins, US, 2004.
- K Stiasny-Kolster, G Mayer, S Schafer, JC Moller, M Heinzel-Gutenbrunner, and WH Oertel. The rem sleep behaviour disorder screening questionnaire—a new diagnostic instrument. *Mov Disord*, 22(16):2386–2393, December 2007.
- RA Stoohs, HC Blum, L Knack, and C Guilliminault. Non-invasive estimation of esophageal pressure based on intercostal emg monitoring. In *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, CA, USA, September 1-5 2004. IEEE.
- EK Tan, SY Lum SM Fook-Chong, ML Teoh, Y Yih, L Tan, A Tan, and MC Wong. Evaluation of somnolence in parkinson's disease: comparison with age- and sex-matched controls. *Neurology*, 58(3):465–468, February 2002.
- E Tandberg, JP Larsen, and K Karlsen. Excessive daytime sleepiness and sleep benefit in parkinson's disease: a community-based study. *Mov Disord*, 14(6):922–927, November 1999.
- The Report of an American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. *Sleep*, 22:667–689, 1999.
- MJ Thorpy. Sleep disorders in parkinson's disease. *Clin Cornerstone*, 24(6 Suppl 1A): S7–15, 2004.
- D. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel. Effect of windowing on central apnea detection. In *Proc. (MeMeA) Workshop IEEE Int Medical Measurements and Applications*, pages 117–120, 2010. doi: 10.1109/MEMEA.2010.5480221.

- D. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel. Relative thresholding with under-mattress pressure sensors to detect central apnea. *IEEE trans Inst Meas*, 60(10):3281–3289, 2011a. doi: 10.1109/TIM.2011.2123250.
- D. I. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel. Simulated central apnea detection using the pressure variance. In *Proc. Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society EMBC 2009*, pages 3917–3920, 2009a. doi: 10.1109/IEMBS.2009.5333551.
- D. I. Townsend, M. Holtzman, R. Goubran, M. Frize, and F. Knoefel. Measurement of torso movement with delay mapping using an unobtrusive pressure-sensor array. *IEEE Trans Inst Meas*, 60(5):1751–1760, 2011b. doi: 10.1109/TIM.2010.2092872.
- Daphne Townsend, Frank Knoefel, and Rafik Goubran. Privacy versus autonomy: a tradeoff model for smart home monitoring technologies. *Conf Proc IEEE Eng Med Biol Soc*, 2011:4749–4752, 2011c. doi: 10.1109/IEMBS.2011.6091176. URL <http://dx.doi.org/10.1109/IEMBS.2011.6091176>.
- Daphne I Townsend, Rafik Goubran, Monique Frize, and Frank Knoefel. Preliminary results on the effect of sensor position on unobtrusive rollover detection for sleep monitoring in smart homes. *Conf Proc IEEE Eng Med Biol Soc*, 2009:6135–6138, 2009b. doi: 10.1109/IEMBS.2009.5334690. URL <http://dx.doi.org/10.1109/IEMBS.2009.5334690>.
- WW Tryon. *Activity Measurement in Psychology and Medicine*. Plenum, 1991. URL <http://books.google.ie/books?id=cxrsUd-bYpsC&lpg=PA149&ots=a0fvbBgvWr&dq=tryon%20activity%20and%20sleep&pg=PP1#v=onepage&q=tryon%20activity%20and%20sleep&f=false>.
- M. Uchida, Shuxue Ding, Wenxi Chen, T. Nemoto, and Daming Wei. An approach for extractions of pulse and respiration information from pulsatile pressure signals. In *Biomedical Engineering, 2003. IEEE EMBS Asian-Pacific Conference on*, pages 180–181, 20-22 Oct. 2003. doi: 10.1109/APBME.2003.1302643.
- ATM Van De Water, A HOLMES, and DA HURLEY. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography a systematic review. *Journal of Sleep Research*, 20(1pt2):183–200, 2011. ISSN 1365-2869. doi: 10.1111/j.1365-2869.2009.00814.x. URL <http://dx.doi.org/10.1111/j.1365-2869.2009.00814.x>.
- H.F. Machiel Van Der Loos, Nino Ullrich, and Hisato Kobayashi. Development of sensate and robotic bed technologies for vital signs monitoring and sleep quality improvement.

- Auton. Robots*, 15(1):67–79, 2003. ISSN 0929-5593. doi: <http://dx.doi.org/10.1023/A:1024444917917>.
- C. van Rijsbergen. *Information Retrieval*. Butterworths, London, 1979.
- EJW Van Someren, RH Lazeron, and BF Vonk. Wrist acceleration and consequences for actigraphic rest-activity registration in young and elderly subjects. *Sleep-Wake Research in the Netherlands*, pages 6123–6125, 1995.
- BV Vaughn and P Gaiallanza. Technical review of polysomnography. *Chest*, 134:1310–1319, 2008.
- MP Villa, S Piro, A Dotta, E Bonci, P Scola, B Paggi, MG Paglietti, F Midulla, and R Ronchetti. Validation of automated sleep analysis in normal children. *Eur Respir J*, 11(2):458–461, 1998. URL <http://erj.ersjournals.com/cgi/content/abstract/11/2/458>. Low correlation on identifying sleep stages using automated analysis, higher correlation on apnoea events (68
- J Virkkala, J Hasan, A Varri, AL Himanen, and K Muller. Automatic sleep stage classification using two-channel electro-oculography. *Journal of Neuroscience Methods*, 166:109–115, 2007.
- MV Vitiello. Sleep in normal aging. In S Ancoli-Israel, editor, *Sleep in the Older Adult*, volume 1, chapter Sleep in Normal Aging, pages 171–176. Elsevier Saunders, 2 edition, 2006.
- MV Vitiello, LH Larsen, and KE Moe. Age-related sleep change: Gender and estrogen effects on the subjective-objective sleep quality relationships of healthy, noncomplaining older men and women. *Journal of Psychosomatic Research*, 56:503–510, 2004.
- L Walsh, SF McLoone, J Behan, and T Dishongh. The deployment of a non-intrusive alternative to sleep/wake wrist actigraphy in a home-based study of the elderly. In *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS 2008*, pages 1687–1690, 20–25 Aug. 2008.
- L Walsh, LK Barger, EE Flynn-Evans, and SW Lockley. Comparison between actigraphy and polysomnography in a real world environment. In *23rd meeting of the Associated Professional Sleep Societies (APSS)*, Seattle, USA, June 6-11 2009a.
- L. Walsh, S. McLoone, J. Behan, T. Dishongh, D. Prendergast, and B. Quigley. Determining sleep metrics and nocturnal activity using an unobtrusive pressure sensing grid. In *Proc. 1st International Conference of the IET Assisted Living Conference 2009*, March 2009b.

- L. Walsh, B.R. Greene, A. Burns, and C.N. Scanail. Ambient assessment of daily activity and gait velocity. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*, pages 418–425, may 2011a.
- L Walsh, E Moloney, and S McLoone. Identification of nocturnal movements during sleep using the non-contact under mattress bed sensor. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 1660–1663, 30 2011-sept. 3 2011b. doi: 10.1109/IEMBS.2011.6090478.
- K. Watanabe, T. Watanabe, H. Watanabe, H. Ando, T. Ishikawa, and K. Kobayashi. 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, 2005. ISSN 0018-9294.
- T. Watanabe and K. Watanabe. Noncontact method for sleep stage estimation. *Biomedical Engineering, IEEE Transactions on*, 51(10):1735–1748, 2004. ISSN 0018-9294.
- J. B. Webster, D. F. Kripke, S. Messin, D. J. Mullaney, and G. Wyborney. An activity-based sleep monitor system for ambulatory use. *Sleep*, 5(4):389–399, 1982.
- Bernard Widrow and Samuel D. Stearns. *Adaptive signal processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1985. ISBN 0-13-004029-0.
- IH Witten, E Frank, L Trigg, M Hall, G Holmes, and SJ Cunningham. Weka: Practical machine learning tools and techniques with java implementations. In *Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems*, page 192196, 1999.
- Amy R. Wolfson, Mary A. Carskadon, Christine Acebo, Ronald Seifer, Gahan Fallone, Susan E. Labyak, and Jennifer L. Martin. Evidence for the validity of a sleep habits survey for adolescents. *Sleep*, 26(2):213–216, 2003.
- World Health Organisation (WHO). Diabetes action now: An initiative of the world health organisation and the international diabetes federation. WHO Library, 2004.
- K Wright, J Johnstone, SE Fabregas, and JR Shambroom. Evaluation of a portable dry sensor-based automatic sleep monitoring system. *Sleep*, 31 (Suppl.):A337. Abstract 1023, 2008a.
- K Wright, J Johnstone, SE Fabregas, and JR Shambroom. Assessment of a wireless dry headband technology for automatic sleep monitoring. *Journal of Sleep Research*, 17 Suppl. 1:230. Abstract P428, 2008b.

- Kenneth P. Wright, Joseph T. Hull, and Charles A. Czesler. Relationship between alertness, performance, and body temperature in humans. *Am J Physiol Regul Integr Comp Physiol*, 283:1370–1377, 2002.
- Alberto Zaffaroni, Philip de Chazal, Conor Heneghan, Patricia Boyle, Patricia Ronayne Mppm, and Walter T McNicholas. Sleepminder: an innovative contact-free device for the estimation of the apnoea-hypopnoea index. *Conf Proc IEEE Eng Med Biol Soc*, 2009:7091–7094, 2009. doi: 10.1109/IEMBS.2009.5332909. URL <http://dx.doi.org/10.1109/IEMBS.2009.5332909>.
- H. Zepelin. *Principles and practice of sleep medicine*, chapter Mammalian Sleep, pages 82–92. Saunders, Philadelphia, 2000.
- H. Zhang, M. Nikolaou, and Y. Peng. Development of a data-driven dynamic model for a plasma etching reactor. *Journal Vac. Sci. Technol.*, 20(3):891–901, 2001.
- X. Zhu, W. Chen, T. Nemoto, Y. Kanemitsu, K.-I. Kitamura, K.-I. Yamakoshi, and D. Wei. Real-time monitoring of respiration rhythm and pulse rate during sleep. *Biomedical Engineering, IEEE Transactions on*, 53(12):2553–2563, Dec. 2006. ISSN 0018-9294. doi: 10.1109/TBME.2006.884641.
- Xin Zhu, Wenxi Chen, T. Nemoto, Y. Kanemitsu, K. Kitamura, and K. Yamakoshi. Accurate determination of respiratory rhythm and pulse rate using an under-pillow sensor based on wavelet transformation. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 5869–5872, Sept. 2005. doi: 10.1109/IEMBS.2005.1615825.
- Xin Zhu, Wenxi Chen, Zunyi Tang, Tetsu Nemoto, and Daming Wei. Automatic home care system for monitoring hr/rr during sleep. *Conf Proc IEEE Eng Med Biol Soc*, 2008:522–525, 2008. doi: 10.1109/IEMBS.2008.4649205. URL <http://dx.doi.org/10.1109/IEMBS.2008.4649205>.