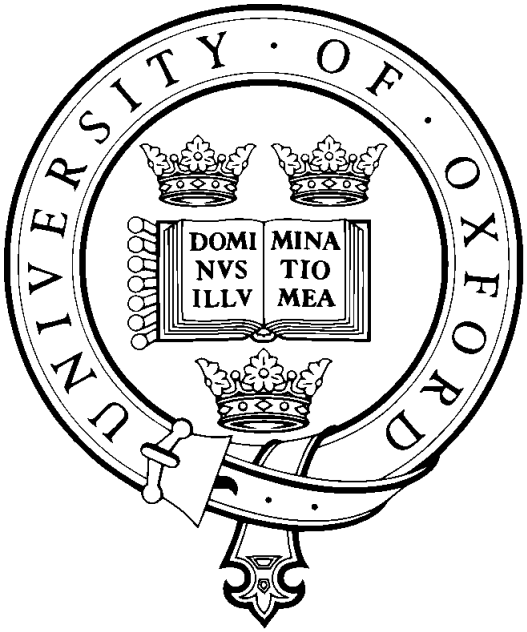
Video and Audio Analysis for the Detection of Obstructive Sleep Apnoea



Elnaz Gederi

Wolfson College
University of Oxford

Supervised by
Professor Lionel Tarassenko
Professor Gari Clifford

Submitted: Trinity Term 2017

This thesis is submitted to the Department of Engineering Science, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own research.

This DPhil was supported by the RCUK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation).

Elnaz Gederi
Wolfson College
University of Oxford
DPhil in Engineering Science
Trinity Term 2017

Abstract

Video and Audio Analysis for the Detection of Obstructive Sleep Apnoea

Elnaz Gederi — Wolfson College

Doctor of Philosophy — Trinity Term 2017

Obstructive sleep apnoea (OSA) is a common sleep disorder characterised by serious sleep fragmentation due to repeated breathing pauses (OSA events) followed by brief awakenings. OSA is under-diagnosed with many consequences that may be life threatening. The standard screening test, polysomnography (PSG), requires an overnight stay in a clinic and the attachment of several on-body sensors which may be uncomfortable for some patients. Large-scale screening for OSA is limited to the availability of equipment and sleep specialists. The manual review of hours of PSG recordings is also cumbersome and costly. The increasing prevalence of OSA and the long waiting times for screening and treatment highlights a need for contact-less home monitoring solutions.

The research presented in this thesis aims to (i) investigate whether automatic processing of video recordings of sleep can be used to differentiate patients with OSA who more urgently require treatment and are more likely to respond to the treatment, and (ii) investigate the possibility of deriving respiratory information (respiratory signal, respiratory rate) from video recordings of sleep and scoring individual OSA events using the derived respiratory information.

This thesis presents the analysis of PSG recordings from 259 patients with suspected OSA and 30 normal volunteers. First, PSG video recordings are used to identify moderate-to-severe OSA patients based on the pattern of their overnight gross-body movements. Then, PSG audio recordings are used to improve the identification of moderate-to-severe OSA patients based on their pattern of snoring, cessation of snoring, and choking sounds. Using the features derived from patients' gross-body movements and audio recordings, moderate-to-severe OSA patients are differentiated from mild-to-no OSA patients with an accuracy of 85.5%, sensitivity of 76.4%, and specificity of 91.1%. Next, a respiratory signal (VDR signal) is derived from the subtle respiratory movements in the PSG video recordings. Patients' respiratory rate (RR) per minute is estimated from the VDR signals. The mean absolute error for the derived RR is 0.82 breaths per minute (bpm) for mild-to-no OSA patients and 0.98 bpm for moderate-to-severe OSA patients. Finally, individual OSA events were scored in the VDR signals and an OSA severity index was calculated. The scored OSA events had an $F_1 score$ of 76.0%, sensitivity of 81.0%, and positive predictive value of 73.8% when compared to the reference OSA events.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors, Professor Lionel Tarassenko and Professor Gari Clifford for their invaluable support, attentive supervision, and encouragement throughout my DPhil. I am deeply grateful to have worked with such inspiring people.

I would like to thank Professor John Stradling for his tremendous clinical advice and for listening to our developing research. I would also like to acknowledge the staff at the Respiratory Medicine group at the Churchill Hospital for their help with collecting the data for this research.

I was extremely fortunate to work amongst talented DPhil students and inspiring researchers at the IBME particularly the Biomedical Signal Processing and Intelligent Patient Monitoring groups. Thank you for your warm friendship, assistance, and creating such an enjoyable environment to be a part of.

I am grateful to the CDT director Professor Alison Noble, all the CDT students, and staff, particularly Jo Armitage, for their advice and support through out my DPhil years. I would like to acknowledge the administrating staff at Wolfson college for their assistance and guidance.

I would like to thank a remarkable individual in my life, my grandmother, who taught me strength and patience. I am grateful to my parents, my brother Arian, and my sister Paniz, for their continuous love and encouragement. Thank you for always being there for me and giving me the best I could have asked for. Finally, words cannot express how grateful I am to my husband, Mehdi, and my lovely daughter Hannah. This thesis would have been impossible without your understanding, patience, and support.

Contents

List of Figures **v**

List of Tables **viii**

1 Background **1**
1.1 Introduction . . . 1
1.2 Sleep physiology . . . 2
1.3 Sleep disorders . . . 6
1.4 Sleep apnoea . . . 8
1.4.1 OSA Physiology . . . 10
1.4.2 Treatment . . . 12
1.5 Aims and overview of this thesis . . . 14
1.5.1 Overview of thesis chapters . . . 15

2 OSA screening **18**
2.1 Sleep questionnaires . . . 18
2.2 Polysomnography . . . 20
2.2.1 Full polysomnography . . . 20
2.2.2 Portable/home monitoring . . . 23
2.3 Oximetry . . . 25
2.4 Actigraphy . . . 29
2.5 Video and audio recording . . . 32
2.6 Sleep scoring . . . 33
2.6.1 Inclusion of arousals in the definition of hypopnoea . . . 35
2.6.2 Automated sleep scoring . . . 36

3 Data **38**
3.1 Hospital patients . . . 38
3.1.1 Overview . . . 38
3.1.2 Room setting . . . 39

3.2 Normal volunteers . . . 40
3.2.1 Overview . . . 40
3.2.2 Room setting . . . 40
3.3 PSG system . . . 41
3.3.1 PSG channels . . . 41
3.4 Data groups . . . 45
3.5 OSA severity based on patient demographics . . . 48

4 Using video derived activity to detect moderate to severe OSA 50
4.1 Introduction . . . 50
4.2 Background: analysis of sleep activity . . . 52
4.2.1 Analysis of sleep activity using actigraphy devices . . . 52
4.2.2 Analysis of sleep activity using video recordings . . . 54
4.3 Materials and methods . . . 56
4.3.1 Data . . . 56
4.3.2 Activity estimation from video data . . . 57
4.3.3 Feature extraction . . . 58
4.3.3.1 Multi-scale entropy analysis . . . 58
4.3.3.2 Inter-Event Interval metrics . . . 60
4.3.4 Classification . . . 61
4.3.4.1 Linear Discriminant Analysis . . . 61
4.3.4.2 Random Forests . . . 66
4.3.5 Data analysis protocol . . . 67
4.4 Results . . . 73
4.4.1 LDA classifier . . . 73
4.4.2 RF classifier . . . 75
4.5 Discussion . . . 75

5 Combination of video derived activity and audio to detect moderate to severe OSA 78
5.1 Introduction . . . 78
5.2 Materials and methods . . . 79
5.2.1 Data . . . 79
5.2.2 Data preprocessing . . . 79
5.2.3 Feature extraction . . . 80
5.2.4 Data analysis protocol . . . 82
5.3 Results . . . 86
5.3.1 LDA classifier . . . 86

5.3.2 RF classifier . . . 88
5.4 Discussion . . . 89

6 Respiratory signal estimation from video recordings 92
6.1 Introduction . . . 92
6.2 Materials and methods . . . 95
6.2.1 Data . . . 95
6.2.2 Extracting respiratory signal from video . . . 96
6.2.3 Non-respiratory movement . . . 102
6.2.4 Assessment of the VDR signals . . . 104
6.2.4.1 Changes in respiratory rate . . . 104
6.2.4.2 Changes in breath amplitude . . . 109
6.3 Results . . . 115
6.3.1 Tuning the model . . . 115
6.3.2 Respiratory rate calculations . . . 116
6.3.3 Breath amplitude to baseline ratios . . . 119
6.4 Discussion . . . 119

7 Detection of apnoeic events using video derived respiration 126
7.1 Introduction . . . 126
7.2 Data . . . 130
7.3 The ground truth for the evaluation of OSA events . . . 131
7.3.1 Oxygen desaturations . . . 131
7.3.2 Apnoeas and hypopnoeas . . . 132
7.4 VDR apnoea and hypopnoea thresholds . . . 133
7.5 Refining event detection using audio . . . 137
7.6 Identifying OSA events for DS1 patients . . . 140
7.7 Comparison of the VDR and reference events . . . 145
7.8 OSA severity threshold . . . 149
7.9 Discussion . . . 152

8 Conclusions and future work 156
8.1 Summary of the contributions of this thesis . . . 157
8.2 Overall conclusions . . . 161
8.3 Limitations of thesis . . . 164
8.4 Future work . . . 165
8.4.1 Inclusion of arousals . . . 165
8.4.2 Identifying events in audio recordings . . . 165

8.4.3 Long-term monitoring of patients 166

Appendix A Sleep questionnaires **167**

A.1 Epworth Sleepiness Scale (ESS) 167

A.2 STOP-BANG questionnaire 168

A.3 Berlin Questionnaire 169

A.3.1 Scoring Berlin Questionnaire 172

Appendix B Classification performance metrics **174**

Appendix C List of Publications **176**

Bibliography **178**

List of Figures

1.1 Sleep publications from 2000 to 2016. 2
1.2 Example EEG patterns for different sleep stages. 4
1.3 Sleep stages across a night. 5
1.4 OSA physiology. 11
1.5 The worldwide map of OSA prevalence. 12
1.6 CPAP device. 14

2.1 Example full and portable PSG systems. 21
2.2 Finger pulse oximeter. 26
2.3 Example PPG signal. 27
2.4 Pulse oximeter pulsatile and non-pusatile absorption. 27
2.5 Light absorbtion for oxyhaemoglobin and reduced haemoglobin. . . . 28

3.1 Snapshots of the sleep clinic rooms. 39
3.2 Snapshot of a patient video. 40
3.3 Snapshot of a normal volunteer video. 41
3.4 A typical set-up of the VISI-3 PSG system. 42
3.5 A snapshot of the VISI-3 PSG software. 44
3.6 Valid PSG respiratory signals versus invalid PSG respiratory signals. 47
3.7 Histograms comparing patient demographics. 49

4.1 30-minute segments of VDA signals from four patients with different OSA severities. 51
4.2 The VDA signal from four patients with different severity levels of OSA. 62
4.3 IEI scores for different OSA sveverity groups in DS1. 63
4.4 A random forest with three decision trees. 67
4.5 Original VDA signal and coarse-grained VDA signal of a severe OSA patient. 69
4.6 Comparison of SE values between moderate-to-severe OSA and mild-to-no OSA patients. 70

4.7 Comparison of SE values between moderate-to-severe OSA and mild-to-no patients for higher m values. 71
4.8 12 VDA classficiation features for example patients. 72
4.9 Data analysis steps. 73
4.10 Boxplots of LDA probabilities on the test data. 74
4.11 Boxplots of RF probabilities on the test data. 76

5.1 Snoring and apnoeic events in audio recordings. 81
5.2 Boxplots showing SE values of moderate-to-severe OSA patients and mild-to-no OSA patients. 83
5.3 Original 2 Hz audio signal and coarse-grained signals of a severe OSA patient. 84
5.4 16 audio classficiation features for example patients. 85
5.5 Boxplots of LDA probabilities on the test data. 87
5.6 RF classification performance using MSE features calculated from audio recordings. 89

6.1 Example sequence of frames. 98
6.2 Tracking a specific pixel on the torso. 99
6.3 Respiratory signals derived from example video frames. 100
6.4 Example 60-second segment of VDR and AEFF signal. 103
6.5 A 240-second segment of a VDR signal and the reference AEFF signal in DS2. 104
6.6 A 60-second window of VDR from a patient in DS2 with the reference respiratory signals. 108
6.7 Normal, out of phase, and paradoxical respiratory movements. 109
6.8 Apnoeic episode with reduced respiratory amplitude. 110
6.9 Calculation of the breath amplitude in the RIPSum signals. 112
6.10 Calculation of the breath amplitude in the VDR signals. 112
6.11 An example section of breath amplitudes, normal breathing baseline, and breath amplitude ratios to normal breathing baseline of a severe OSA patient. 114
6.12 Parameter tuning for the derivation of the VDR signals. 116
6.13 The RR calculations for a mild OSA patient. 117
6.14 The RR calculations for a moderate OSA patient in DS2. 117
6.15 Paired comparison of VDR derived RRs with reference RRs. 118
6.16 Comparison of the changes in breath amplitude. 120

6.17 Comparison of changes from baseline value in the RIPSum and VDR signals. 121

7.1 Apnoeic and hypopnoeic events and desaturations of a severe OSA patient. 133
7.2 Comparison of AHI against ODI. 134
7.3 Comparison of VRDI against AHI. 136
7.4 Comparison of average VRDI against average AHI for $DS2_{Train}$ and $DS2_{Test}$. 137
7.5 False positive OSA events removed using audio. 141
7.6 Comparison of VRDI, refined VRDI, and AHI for DS2 patients. . . . 142
7.7 Comparison of average VRDI, average refined VRDI, and average AHI for $DS2_{Train}$ and $DS2_{Test}$. 142
7.8 Comparison of VRDI, refined VRDI, and ODI for DS1 patients. . . . 143
7.9 Comparison of VRDI, refined VRDI, and ODI for DS1 patients based on OSA severity. 144
7.10 Comparison of average VRDI, average refined VRDI, and average ODI for DS1 patients. 144
7.11 A few examples of TP and FN OSA events for a patient in DS2. . . . 146
7.12 A few examples of TP and FN OSA events for a patient in DS1. . . . 148
7.13 The TH1 and TH2 thresholds to determine OAS severity. 151

8.1 Percentage of patients in each OSA severity correctly classified using the features derived in Chapters 5, 6, and 7. 163
8.2 Percentage of patients in each OSA severity categorised as OSA, no OSA, or possible OSA using the combination of results from Chapters 5 and 7. 164

List of Tables

1.1 Changes to physiology during REM and NREM sleep. 6
1.2 The waiting time for OSA diagnosis and treatment. 8
1.3 Symptoms of OSA 10
1.4 Definition of severity of OSA. 11

2.1 Types of sleep monitoring systems. 25
2.2 Respiratory event definitions in the AASM manuals from 1999 to 2012 34
2.3 The AASM recommended and alternative signals for respiratory analysis. 35

3.1 Details of channels recorded during PSG studies. 43
3.2 Number of patients in each OSA severity group. 45
3.3 Other sleep fragmentation related conditions in DS1 patients. 46
3.4 Details of patients in DS2. 48
3.5 Details of normal volunteers. 48
3.6 Using demographics to identify moderate-to-severe OSA. 48

4.1 Summary of SE parameters used for VDA signals. 69
4.2 LDA classifier performance in differentiating moderate-to-severe OSA patients and the mild-to-no OSA patients. 74
4.3 Performance of the RF classifier. 75

5.1 Summary of SE parameters used for down-sampled audio signals. . . 85
5.2 LDA classification performance using MSE features calculated from audio recordings. 87
5.3 RF classification performance using MSE features calculated from audio recordings. 88

6.1 The absolute error between reference and VDR RR estimates for the DS2 mild-to-no OSA patients and DS3 volunteers. 119
6.2 The absolute error between reference and VDR RR estimates for the DS2 moderate-to-severe OSA patients. 119

7.1 Details of $DS2_{Train}$ and $DS2_{Test}$. 135
7.3 VRDI threshold to differentiate moderate-to-severe OSA patients from mild-to-no OSA patients. 151

8.1 Summary of the out of sample classification results of Chapters 5 and 7. 162

Acronyms

AASM American Academy of Sleep Medicine.

Ac Accuracy.

AE Absolute Error.

AEFF Abdominal Respiratory Effort.

AHI Apnoea Hypopnoea Index.

AR Autoregressive.

AUC Area Under the Curve.

BANG Body mass index, Age, Neck circumference and Gender.

BMI Body Mass Index.

bpm breaths per minute.

CPAP Continuous Positive Airway Pressure.

CSA Central Sleep Apnoea.

ECG Electrocardiography.

EEG Electroencephalography.

EMG Electromyogram.

EOG Electrooculogram.

ESS Epworth Sleepiness Scale.

FN False Negative.

FP False Positive.

IEI Inter-Event Interval.

LDA Linear Discriminant Analysis.

LTP Local Ternary Patterns.

MEMS Microelectromechanical systems.

MSE Multi-Scale Entropy.

NAF Nasal Airflow.

NP Nasal Pressure.

NPV Negative Predictive Value.

NREM Non-Rapid Eye Movement.

ODI Oxygen Desaturation Index.

OSA Obstructive Sleep Apnoea.

PAP Positive Airway Pressure.

PLMD Periodic Limb Movement Disorder.

PLMS Periodic Limb Movement of Sleep.

PPG Photoplethysmogram.

PPV Positive Predictive Value.

PSG Polysomnography.

PTT Pulse Transit Time.

PVDF Polyvinylidene fluoride.

RDI Respiratory Disturbance Index.

REM Rapid Eye Movement.

RF Random Forests.

RGB Red-Green-Blue.

RIP Respiratory Inductance Plethysmography.

RLS Restless Leg Syndrome.

RR Respiratory Rate.

SBD Sleep-related Breathing Disorders.

SE Sample Entropy.

Se Sensitivity.

Sp Specificity.

STOP Snoring, Tiredness, Observed apnoea, Blood pressure.

TEFF Thoracic Respiratory Effort.

TN True Negative.

TP True Positive.

TV Tidal Volume.

UA Upper Airway.

VA Video Actigraphy.

VDA Video Derived Activity.

VDR Video Derived Respiratory.

VRDI Video Respiratory Disturbance Index.

WA Wrist-worn Actigraphy.

Chapter 1

Background

1.1 Introduction

Sleep is something everyone needs but many of us do not get enough. Most people have experienced sleepless night(s) throughout their life and are familiar with fatigue the day after. For some people, this lack of sleep is either due to poor sleep hygiene[1] or a busy lifestyle. For some others, it is due to a sleep disorder, a failure in the complex interactions of the physiological system of sleep and wakefulness.

Inadequate and disrupted sleep is often associated with poor vigilance and memory, premature mortality, cardiovascular disease, hypertension, inflammation, obesity, diabetes, and psychiatric disorders such as anxiety and depression [2, 3]. The effect of 17 hours of continuous wakefulness in decreased performance is similar to a blood alcohol level of 0.05% which in many countries is the legal limit for driving [4].

Sleep deprivation and disorders affect many people worldwide. Insomnia, which is

[1] A set of behavioural and environmental recommendations intended to promote healthy sleep. A few examples of these recommendations are to avoid caffeine, exercise regularly, eliminate noise from the sleeping environment, and maintain a regular sleep schedule [1].

the most common sleep disorder, is reported by 30% of adults and chronic insomnia by 10%. Given the serious effects of lack of sleep in everyday life and the rising number of people with sleep disorders, it is not surprising that the number of publications on sleep has substantially increased in the last few years (Figure 1.1) [4].

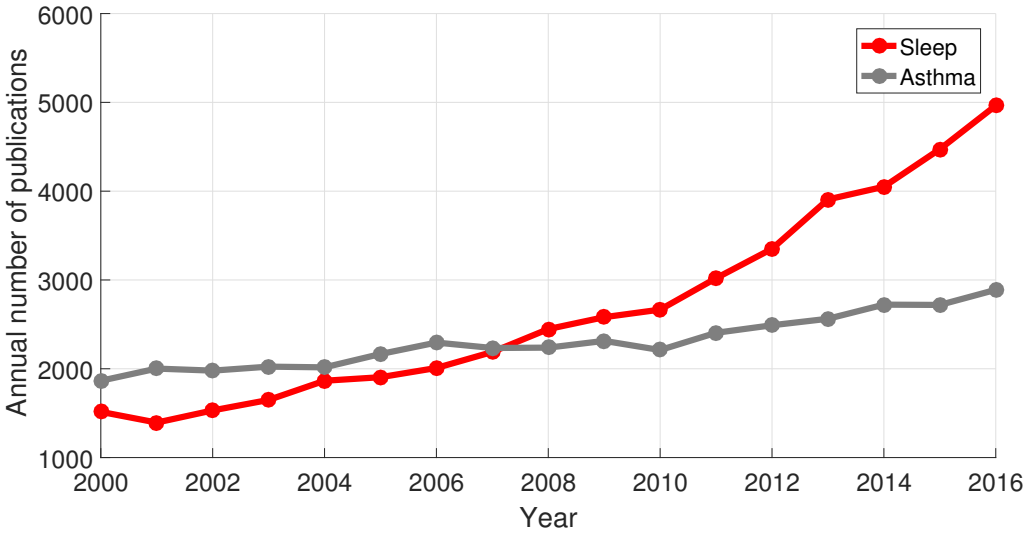


Figure 1.1: Number of publications on sleep from 2000 to 2016 compared to the publications on asthma during the same period (PubMed search for words 'sleep' and 'asthma' in the title in April 2017).

This chapter includes a brief description of the physiology of sleep as well as the classification of sleep disorders. Sleep apnoea and Obstructive Sleep Apnoea (OSA) are then discussed in more detail in terms of physiology, diagnosis and treatment followed by a description of the aims and overview of this thesis.

1.2 Sleep physiology

Normal human sleep comprises two states of Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM) sleep. These two states alternate cyclically during sleep and are very different in characteristics. REM sleep, often called 'active sleep',

was discovered by Aserinsky and Kleitman in 1953 [5]. REM sleep typically includes low-amplitude, high-frequency Electroencephalography (EEG) waves as well as eye movements. The EEG of REM sleep closely resembles the EEG of alert waking. Usually, when people wake up from REM sleep, they report that they had been having vivid dreams. On the other hand, NREM, also called 'quiet sleep', includes variably synchronous cortical EEG activity, as well as minimal psychological activity. People report dreaming far less frequently when awakened from NREM sleep compared to REM sleep. During REM sleep muscles in the arms and legs are temporarily paralysed [6].

In the mid-1930s, Loomis *et al.* [7] provided the earliest detailed description of sleep stages. The authors described five stages of sleep but did not distinguish REM sleep. Sleep staging criteria were standardised by Rechtschaffen & Kales (R&K rules) in 1968 [8]. Based on R&K rules, sleep was divided into 7 stages of wake, stage 1 (NREM), stage 2 (NREM), stage 3 (NREM), stage 4 (NREM), REM, and movement time. Later, the American Academy of Sleep Medicine (AASM) combined stages 3 and 4 and the current AASM guidelines divide NREM sleep into three stages, N1, N2, and N3 [9]. Each stage has different characteristics of EEG patterns, eye movements, and muscle tone. During stage N3 which is the deepest sleep, the EEG waveform becomes slower and more synchronised, and the eyes remain still [2]. Figure 1.2 shows typical EEG patterns for the different sleep stages.

The pattern of sleep in normal human sleeping on a regular schedule starts with NREM sleep and progresses through deeper stages of NREM sleep until the first episode of REM sleep begins around 80 to 100 minutes later. Then NREM and REM

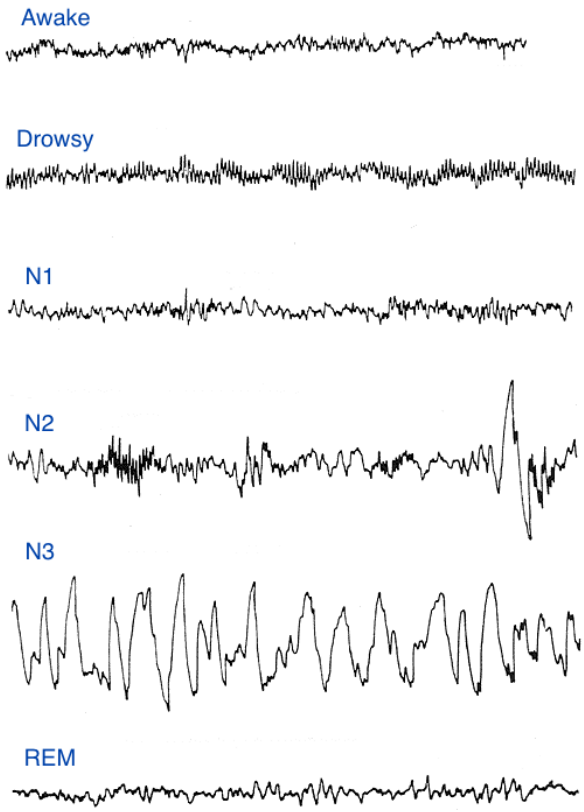


Figure 1.2: Example EEG patterns of different sleep stages (image is adapted from http://sleepdisorders.sleepfoundation.org/chapter-1-normal-sleep/stages-of-human-sleep/)

sleep cycle with a period of approximately 90 minutes. The N3 stage dominates the NREM portion of the initial cycles, and the duration of REM sleep increases across the night [6].

Sleep pattern and duration of sleep cycles change with age. Newborns have a short sleep cycle and enter REM sleep before NREM sleep. At birth, REM sleep is around 50% of total sleep but as the brain matures during the first two years of life it declines to around 20−25%. NREM sleep slow-waves are not present at birth but gradually emerge during the first two years of life. Slow-wave sleep (N3) decreases

across adolescence and continues a slower decline into old age. The proportion of REM sleep to total sleep is normally around 20−25% from childhood into old age [6]. Figure 1.3 shows the stages of sleep in a normal young adult in a single night [6].

Factors such as phase changes in the circadian rhythm, medication, ambient temperature, and sleep disorders alter the normal sleep pattern. There are many other changes that happen in physiology during sleep. Table 1.1 summarises these physiological changes during REM and NREM sleep [10].

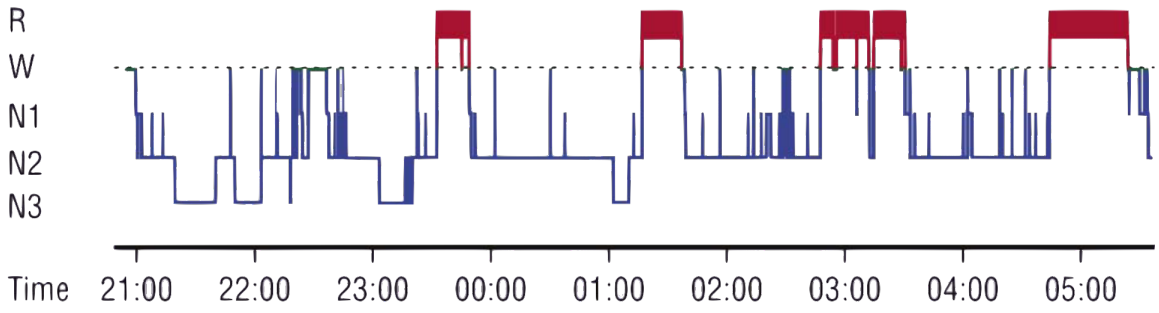


Figure 1.3: The sleep stages across a night in a normal young adult (image adapted from [11]).

In normal people, breathing becomes more rapid and shallow in all stages of sleep than during wakefulness. The lowest mean tidal volume[2] occurs in REM sleep, when the depth of breathing is highly variable [12]. These changes result in lower expired volumes in NREM sleep than during wakefulness, with a further appreciable reduction in ventilation during REM sleep.

Ventilation and respiratory flow show less effective adaptive responses during sleep. The cough reflex, which is the normal reaction to irritants in the airway, is suppressed during REM and NREM sleep. The arousal response to respiratory resistance, for example resistance to breathing in or out, is lowest in N3 stage of sleep. The reduced

[2]Tidal Volume (TV) is the volume of air inspired or expired during each quiet breathing cycle [2].

Table 1.1: Summary of changes to physiology during REM and NREM sleep [10].

Physiological Process	NREM	REM
Brain activity	Decreased from wakefulness	Increased in motor and sensory areas, other areas are similar to NREM
Heart rate	Slower from wakefulness	Varies and increased compared to NREM
Blood pressure	Lower than wakefulness	Varies and increased (up to 30%) from NREM
Sympathetic nerve activity	Decreased from wakefulness	Increased significantly from wakefulness
Muscle tone	Similar to wakefulness	No muscle tone
Blood flow to brain	Decreased from wakefulness	Increased from NREM in some regions of the brain
Respiration	Decreased from wakefulness	Varies and increased from NREM, also may show brief stoppages
Airway resistance	Increased from wakefulness	Varies and increased from wakefulness
Body temperature	Regulated at a lower temperature than wakefulness	Not regulated and temperature will be determined by the local environment

rib cage movement and increased Upper Airway (UA) resistance, during REM sleep, is due to the loss of tone in the UA muscles and the intercostals [10].

1.3 Sleep disorders

The AASM classifies sleep disorders into the following main categories [13]:

Insomnia Insomnias are defined by a repeated difficulty with initiation, duration, consolidation, or quality of sleep that happens despite the individual having enough time and opportunity for sleep. Insomnias result in some form of daytime impairment.

Sleep-related Breathing Disorders (SBD) As the name suggests, SBD are characterised by disordered respiration during sleep. SBD are divided into four main categories of OSA syndrome, Central Sleep Apnoea (CSA) syndrome, sleep-related hypoventilation[3] disorders, and sleep-related hypoxaemia[4] disorders.

Central disorders of hypersomnolence In these disorders, the primary complaint is daytime sleepiness which is not due to disturbed nocturnal sleep or misaligned circadian rhythms (i.e. not due to a circadian rhythm sleep disorder, SBD, or other cause of disturbed night-time sleep).

Circadian rhythm sleep-wake disorders These kind of sleep disorders have shared underlying chronophysiologic basis. The main characteristic of these disorders is a recurrent misalignment between the patient's sleep pattern and the pattern that is desired, needed, or societally expected.

Parasomnias Parasomnias are undesirable physical or experiential events during sleep including abnormal sleep-related movements, behaviours, emotions, perceptions, dreaming, and autonomic nervous system functioning.

Sleep-related movement disorders These disorders are defined by stereotyped movements that disturb sleep. Sleep disorders such as Periodic Limb Movement Disorder (PLMD) and Restless Leg Syndrome (RLS) are examples of this category.

[3]Breathing at an abnormally shallow and slow rate [14].
[4]Reduction of the oxygen concentration in the arterial blood [14].

Other sleep disorders These are sleep disorders that are difficult to fit into any of the categories above such as a sleep disturbance that is caused by a disturbing environmental factor.

The diagnosis of sleep disorders is based on clinical symptoms and the 'gold standard' in-laboratory screening test called Polysomnography (PSG). PSG usually involves an overnight stay in a sleep lab and the attachment of several sensors to the body. PSG has been proven to be accurate with a low failure rate since specialised equipment and trained technical staff are required to record the signals [15]. PSG is expensive, limited by the number of beds available in the sleep lab and the number of trained sleep specialists available in the area. The waiting time for OSA diagnosis and treatment with Continuous Positive Airway Pressure (CPAP) (see Section 1.4.2 below) in five countries is presented in Table 1.2 [16].

Table 1.2: The waiting time for OSA diagnosis and treatment with CPAP in five countries [16] (populations are 2001 estimates).

Country	Population (millions)	# of sleep labs	Waiting time (months)
UK	58	84	7−60
Belgium	10	150	2
Australia	19	65	3−16
US	280	1292	2−10
Canada	31	440	4−36

1.4 Sleep apnoea

The discovery of sleep apnoea in 1965, reported independently by Gastaut *et al.* in France [17] and Jung and Kuhlo in Germany [18], was one of the most important events in the history of sleep disorders medicine. In 1976, a detailed paper by Guilleminault

et al. [19] defined apnoea as the cessation of airflow at the nose and mouth which lasts at least 10 seconds and sleep apnoea syndrome as at least 30 apnoeic episodes over seven hours of overnight sleep. These definitions have been refined over the years. Apnoea is now defined as complete or near complete cessation of airflow (>= 90% reduction in airflow) for at least 10 seconds. Hypopnoea is defined as partial reduction in airflow (>= 30% reduction in airflow) for at least 10 seconds associated with >= 3% reduction in blood oxygen saturation or an arousal. The common duration of apnoeic events is between 20 to 40 seconds and the maximum duration of the events is considered to be around 2 minutes [20].

Three types of apnoea have been defined: obstructive, central, and mixed. Apnoeas are obstructive if there is continued inspiratory effort, central if there is no inspiratory effort, and mixed if inspiratory effort is absent at the beginning, but the effort resumes before the resumption of airflow. There are two types of sleep apnoea syndrome: OSA syndrome in which the patient suffers from repeated obstructive apnoeas and is the more common form, and CSA syndrome in which the patient suffers from repeated central apnoeas. It is possible for a patient to suffer from a mixture of both syndromes. OSA can present with different symptoms as listed in Table 1.3.

There are a few metrics used to determine the severity of sleep apnoea. The main and most commonly used metric is the Apnoea Hypopnoea Index (AHI) which is the number of apnoeas and hypopnoeas per hour of sleep [22]. The second metric is the Oxygen Desaturation Index (ODI), which is the number of oxygen desaturation events per hour of sleep [23]. The third metric, Respiratory Disturbance Index (RDI),

[5]Excessive urination at night.
[6]Involuntary urination.

Table 1.3: Symptoms of OSA [21].

Typical	Snoring Unrefreshing sleep Daytime sleepiness
Frequent	Nocturnal choking Witness apnoeas Nocturia[5] Morning headaches
Less common	Reduced libido Enuresis[6]

includes the number of respiratory-effort related arousals as well as apnoeas and hypopnoeas per hour of sleep. This means a person's RDI can be higher than his or her AHI [22].

1.4.1 OSA Physiology

Normally, the upper part of the throat remains open during sleep allowing air to reach the lungs. However, in patients with OSA, as muscles relax during sleep the collapsible soft tissue that constitutes the wall around the throat obstructs breathing causing what is called an apnoeic event (Figure 1.4). As there is no air flowing into the lungs, the arterial oxygen level drops and carbon dioxide level rises. There are increasingly negative pressure swings in the thorax. Blood pressure initially drops and then drifts upwards during the event. Eventually, the patient awakens with a surge of sympathetic nervous system activity leading to a spike in heart rate and blood pressure and resumption of breathing [15]. The first breath after apnoea includes a deep breath often accompanied with a choking/snorting sound [22]. OSA causes sleep fragmentation and repetitive hypoxaemia leading to adverse consequences such as increase in

the risk of motor vehicle accidents, reduced work performance, hypertension, stroke, heart disease and diabetes [15, 24]. The severity of the disorder based on the AHI is normally grouped into three categories of mild, moderate, and severe (Table 1.4) [25].

Table 1.4: Definition of severity of OSA [25].

OSA severity	AHI criteria
Mild	5⩽AHI<15
Moderate	15⩽AHI<30
Severe	30⩽AHI

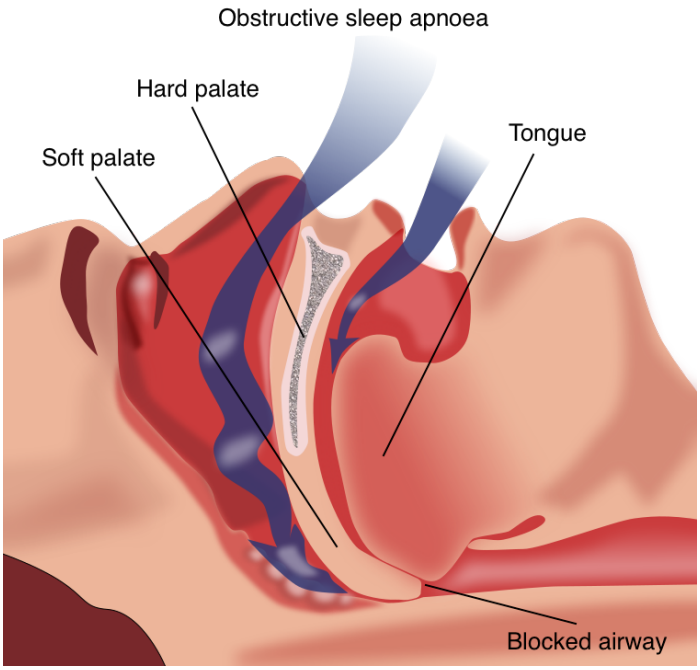


Figure 1.4: OSA occurs when the muscles that support the soft tissues in the throat, such as the tongue and soft palate, relax during sleep causing the airway to narrow or close and partial or complete cessation of breathing. Image adapted from https://en.wikipedia.org/wiki/Obstructive_sleep_apnea.

Figure 1.5 shows a global map of the prevalence of OSA [26]. Even in countries where OSA is a largely recognised condition, a significant part of the symptomatic population are undiagnosed. In the USA, a survey in 1997 on 4925 individuals indicated that as many as 82% of the men and 93% of the women who were likely

to have moderate-to-severe OSA had not yet been diagnosed [27]. OSA is strongly linked to obesity [28] and the increasing prevalence of obesity in the world [29] suggests a future increase in the prevalence of OSA. Understanding the pathogenesis and improving access to the diagnosis of OSA is essential in developing therapeutic strategies and lessening the health burden of OSA [28].

The economic burden of OSA is substantial, accounting for billions of dollars annually. Nearly 20% of the European population suffer from sleep-related breathing disorders with more than 85% of them left undiagnosed due to the screening cost, lack of patient awareness, and limited sleep-monitoring facilities [30].

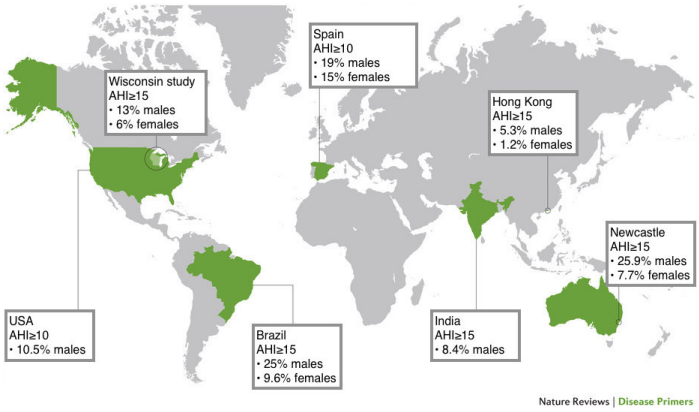


Figure 1.5: The worldwide map of OSA prevalence. The AHI threshold used to determine OSA is specified for each case (adapted from [26]).

1.4.2 Treatment

OSA, once correctly diagnosed, can be managed effectively through a range of the following treatments [31, 32]:

Diet and lifestyle Such as weight loss, avoiding tobacco and alcohol use, avoiding sleep deprivation, and avoiding sleeping in a supine position.

Pharmacological treatment Drug treatments aimed at stimulating neurotransmitters which contract the UA dilator muscles in an effort to maintain UA patency, however, these drugs have had limited success. Also, the use of benzodiazepines should be avoided, and the use of narcotics should be minimised as they may significantly worsen UA dysfunction during sleep [33, 34, 35].

Therapeutic devices Oral appliances such as mandibular advancement devices or tongue trusses which hold the lower jaw and tongue forward. On average, only 52% of patients treated with oral appliances had some success in controlling OSA, therefore the efficacy of these devices is questionable. Effects on sleepiness and quality of life were demonstrated but improvement in other neurocognitive outcomes were not consistent [32].

Surgery Examples are nasal reconstruction to improve normal respiration, tonsillectomy and adenoidectomy usually used for children with OSA in order to enlarge the nasal inferior turbinates [36].

Assistive devices Includes nasal Positive Airway Pressure (PAP) devices such as a CPAP device, which is a device like an oxygen mask worn over the mouth and/or nose forcing pressurised air down the airway thereby keeping it open. CPAP (Figure 1.6), is the most commonly used treatment for symptomatic sleep apnoea and is extremely effective when used correctly. However, approximately 30−35% of patients are intolerant or non-compliant due to side effects such as skin abrasions, bruising, chaffing from the mask, nasal congestion or dryness, and abdominal cramping [31].

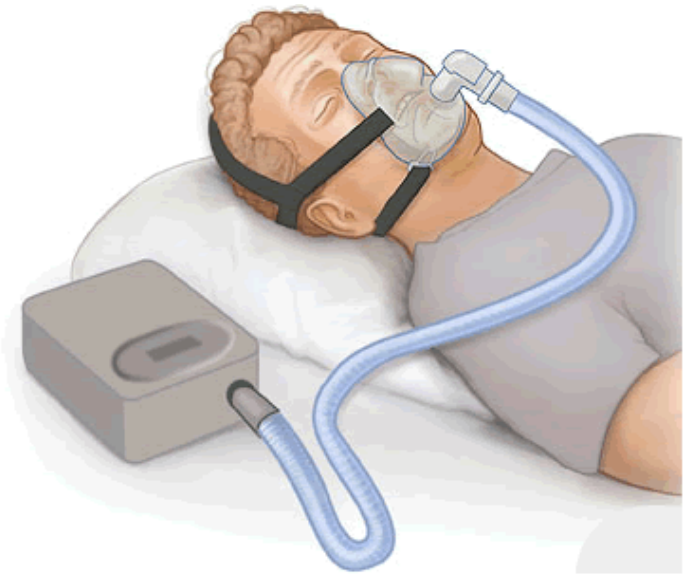


Figure 1.6: A CPAP device includes a flow generator and a face mask. (Taken from http://en.wikipedia.org/wiki/Positive_airway_pressure).

Electrical stimulation Another treatment for OSA patients is the electrical stimulation of the lingual[7] musculature [37, 38].

1.5 Aims and overview of this thesis

The increase in the number of OSA patients and in the awareness of OSA may only increase the waiting times for diagnosis and treatment of the disorder. This has created a need for alternative screening solutions and has led to a rise in home/portable screening solutions for OSA in recent years. Home screening solutions, whether used for clinical diagnosis of OSA or prioritising candidates who more urgently require screening and treatment, can decrease the waiting time. However, large-scale screening for OSA, even with home screening systems (see section 2.2.2) is not common because of the limited availability of equipment. For long-term monitoring of OSA,

[7]Relating to, situated close to, or resembling the tongue (lingua) [14]

the attachment of sensors to the body imposes an additional limitation with compliance and comfort issues. For years, cameras have been part of standard full PSG systems and video recordings of sleep have been manually reviewed for OSA diagnosis. Night-vision cameras, similar to those used in PSG systems, are becoming increasingly cheaper with better resolution[8]. Additionally, there has been a significant rise in the popularity of smartphones around the world, accompanied by a reduction in their cost, improvements in their cameras (particularly in low light conditions), and improvements in the computational power of their on-board processors. Whether a smartphone camera or a stand-alone night-vision camera is used, video recording provides important information about an individual's sleep.

The overall aims of the research presented in this thesis are: (i) to investigate whether automatic processing of video recordings of sleep can be used to identify moderate-to-severe OSA patients who more urgently require treatment and are more likely to respond to treatment, (ii) to investigate the possibility of deriving respiratory information (respiratory signal, respiratory rate) from video recordings, and (iii) to investigate the possibility of automatic identification of individual OSA events[9] using the respiratory information derived from video recordings.

1.5.1 Overview of thesis chapters

This thesis is comprised of 8 chapters. In this chapter, we have presented a general overview of sleep, sleep physiology and disorders with a special focus on sleep apnoea.

[8] A small night-vision camera with a resolution of 1280×720 pixels and a frame rate of 30 frames per second can be purchased for under £15 (March 2017).

[9] The term 'OSA events' in this thesis refers to the combination of obstructive apnoeic and hypopnoeic events.

In chapter 2, sleep apnoea screening is discussed. The chapter starts with the description of a few common sleep questionnaires that are used for pre-screening of OSA patients followed by details of full PSG systems and portable PSG systems. Next, pulse oximeters (for oximetry) and actigraphs (for measuring activity) which are better tolerated by the patients and are more commonly used alone, as a single channel tool for screening sleep disorders, are discussed. Video and audio as two contact-less modalities for monitoring OSA patients are briefly presented with more detailed discussions in the following chapters. Finally, the potential benefits of automated screening of OSA and sleep events are discussed.

Chapter 3 provides an overview of the data used in the investigations described in this thesis. Patient demographics as well as the hardware and software used for sleep screening and analysis are discussed followed by algorithms used for identifying OSA and oxygen desaturation events.

Research described in Chapter 4 focuses on the use of gross body sleep activity, which is derived from video recordings to differentiate moderate-to-severe OSA patients from mild-to-no OSA patients. The regularity of the sleep-wake patterns in the activity signals derived from analysis of video recordings is used for the classification of patients.

Chapter 5 investigates whether the addition of features derived from audio data (embedded in the video recordings) to the features derived from video data improves the differentiation of moderate-to-severe OSA patients from mild-to-no OSA patients. The regularity of patterns of snoring followed by silence and then snorting/choking sounds is used for the classification of patients.

Chapter 6 examines the possibility of detecting subtle respiratory movements in the video and extracting a respiratory signal from the video recording. The cyclical pattern of respiratory movements is used to derive the respiratory signal. The respiratory rate and changes in respiratory volume in the derived signal are validated against reference respiratory signals in the PSG recordings.

In Chapter 7, the respiratory signals derived from video recordings are analysed to identify OSA events and derive an OSA severity score comparable with the AHI and ODI. The identified OSA events in the video-derived respiratory signal are compared with the events identified in the reference PSG signals.

Finally, in Chapter 8, the contributions of this thesis are reviewed, and the limitations and future work are discussed.

Chapter 2

OSA screening

This chapter describes the techniques used for the screening and diagnosis of OSA. First, sleep questionnaires which are simple methods for first stage screening of OSA are discussed. This is followed by the description of PSG, which is the standard technique for screening OSA in sleep clinics. Portable and simplified screening methods are then discussed followed by the criteria used for the identification and scoring of OSA events in the PSG data.

2.1 Sleep questionnaires

A number of questionnaires have been designed for the pre-screening of OSA patients. Some of the commonly used questionnaires are listed below:

Epworth Sleepiness Scale (ESS) ESS was developed by Murray Johns at the Epworth Hospital in Melbourne, Australia [39] and it is the most widely used questionnaire for evaluating sleepiness. ESS asks patients to score their expectation of

dozing in differing circumstances. Dozing can be scored as none (0), slight (1), moderate (2), or high (3) for eight hypothetical situations (for details of the questionnaire see Appendix A). Based on studies by Johns [40], Hirshkowitz *et al.* [41], and Sharafkhaneh and Hirshkowitz [42], ESS scores from 0−8 are classified as normal, 9−12 are classified as mild, 13−16 are classified as moderate, and greater than 16 are classified as severe [2].

STOP-BANG The STOP-BANG questionnaire which stands for Snoring, Tiredness, Observed apnoea, Blood pressure (STOP) Body mass index, Age, Neck circumference and Gender (BANG) was developed by Chung *et al.* [43] to estimate the risk of OSA. Details of the questions asked in the STOP part of the questionnaire are provided in Appendix A. The BANG information is preferably assessed by a clinical assistant or may be provided by the patient. Answering yes to three or more items will put the person in the high OSA risk category and answering yes to fewer than three items will put the person in the low OSA risk category.

Berlin Questionnaire The Berlin Questionnaire was designed by a number of US and German pulmonary and primary care specialists at a conference in Berlin in 1996 [44]. The questions were chosen based on the factors or behaviours that across studies were consistently associated with the presence of SBD. The questionnaire includes one introductory question, four snoring-related questions, three questions about daytime sleepiness, and one question about history of high blood pressure. Patients are asked to provide age, weight, height, sex, neck circumference, and ethnicity. Details of the questionnaire are provided in Appendix A.

High risk in category 1 of questions was defined as persistent symptoms (> 3 to 4 times per week) in two or more questions about snoring. High risk in category 2 of questions was defined as persistent (> 3 to 4 times per week) daytime sleepiness, drowsy driving, or both. High risk in category 3 of questions was defined as a history of high blood pressure or a body mass index more than 30 $\frac{kg}{m^2}$. High risk for sleep apnoea was defined as at least two symptom categories. Patients with no persistent symptoms or qualifying for one symptom category only were regarded as being at lower risk.

2.2 Polysomnography

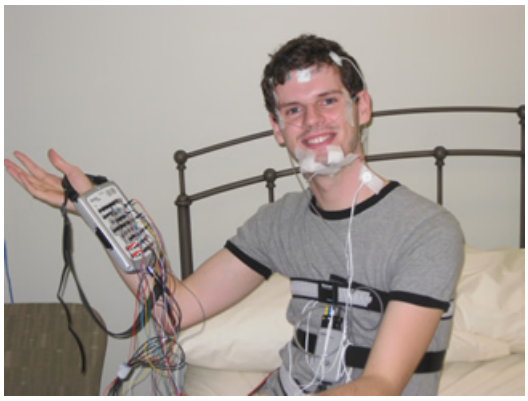
2.2.1 Full polysomnography

The signals gathered through PSG relate to the function of different parts of the body and their sleep and wakefulness-related interaction. PSG systems are normally installed in a quiet room with a bed which is in darkness during the period of study. A nurse attaches various sensors to the patient for the study. An infra-red light source and an infra-red camera provide the visual recording of the patient overnight and a microphone provides the acoustic monitoring. Figures 2.1a and 2.1b show a full PSG device and a participant connected to the PSG system. Below are the list of signals that are usually acquired during a PSG recording [45, 8].

- Neurophysiological monitoring: Based on international standards, at least 4 neurophysiological channels are required; one EEG, two Electrooculogram (EOG), and one Electromyogram (EMG) channels. EEG is the method of recording the



(a)



(b)



(c)



(d)

Figure 2.1: (a) Embla N7000 full PSG system (http://stowood.com/product/embla-n7000/). (b) Participant wearing the Embla N7000 PSG system (http://psychology.msu.edu/SleepLab/Lab.aspx). (c) Stowood Visi Black portable PSG system (http://stowood.com/product/visi-black-series/). (d) Participant wearing the Stowood Visi Black portable PSG system.

brain's electrical activity from surface electrodes. The EOG and EMG are used for detection of REM sleep as REM sleep is usually characterised by suppression of muscle tone (detected by EMG) and rapid eye movements (detected by EOG).

- Respiratory monitoring: Respiratory monitoring includes measurement of oxygenation, respiratory centre drive, ventilation, and recording of sounds related to respiration.

  - Oxygenation: Oxygen saturation is one of the key parameters used in assessing sleep-breathing disorders and is usually measured by a pulse oximeter.

  - Ventilation: Thermistors are used at the mouth or nose to track the presence or absence of airflow by changes in the temperature. Other methods of measuring ventilation are recording pressure derived from a nasal cannula or carbon dioxide partial pressure in the expired air (capnography). However, the amplitude of these ventilation monitoring signals cannot be used to estimate the airflow quantitatively. A more detailed and semi-quantitative technique to measure ventilation is Respiratory Inductance Plethysmography (RIP) which includes the attachment of two strain gauges to the ribcage and abdomen of the patient to analyse the breathing patterns and the coordination of ribcage-abdominal motion.

  - Respiratory centre drive: During obstructive apnoeic events the inspiratory effort continues. Therefore, ribcage-abdominal motion recorded by RIP

or diaphragmatic EMG recordings are used to detect thoraco-abdominal motion and distinguish between obstructive and central sleep apnoea. Alternatively, pleural pressure swings with an esophageal catheter are used to detect increases in negative intra-thoracic pressures (due to continued respiratory effort) and distinguish between obstructive and central sleep apnoea.

  - Sound recording: Microphones are used to record and assess snoring.

- Cardio-circulatory monitoring: Three-lead Electrocardiography (ECG) is used to monitor and record the heart rhythm. In some cases a finger Photoplethysmogram (PPG) or a blood pressure measurement device may be used to measure arterial blood pressure level.

- Body position monitoring: UA obstruction and snoring are affected by gravitation and accelerometers are used to record the patient's body position.

- Video recording: Continuous video monitoring is used to record the patient's detailed physical and acoustic activity during sleep. Video recordings are particularly used for the analysis of parasomnias and nocturnal seizure disorders but can also be used to visually verify apnoeic events.

2.2.2 Portable/home monitoring

In order to overcome the cost and limited availability of in-laboratory PSG studies, there has been a growing interest in alternative approaches to diagnosis such as home PSG systems. Randomised controlled trials have demonstrated that for some patients

without complications such as lung disease, heart failure, or neuromuscular disease, home monitoring systems perform as well as in-laboratory diagnosis and monitoring [46]. Therefore, careful design of home monitoring systems can provide timely and cost-effective management of obstructive sleep apnoea for some groups of patients. There is a wide range of portable monitoring systems recording as many signals as with attended PSG or only one signal such as oxygen saturation. Since the majority of portable monitoring systems do not have EEG and EMG recordings, it is not possible to accurately derive information such as sleep stages, total sleep time, and arousals. Therefore, OSA events are quantified per hour of monitoring rather than per hour of sleep. This may cause underestimation of AHI for patients with poor sleep. For example, for a patient who has been sleeping for 50% of the night, an AHI of 4 events per hour will actually correspond to 8 events per hour when sleep time is taken into account. Additionally, since arousals cannot be scored on most home devices, AHI may underestimate hypopnoeas with a terminating arousal. This is particularly a problem for younger patients who have normal cardiopulmonary function. These patients are less likely to experience oxygen desaturations of above 3% with reductions in airflow but still may experience a very fragmented sleep [47].

In 1994, the AASM classified sleep screenings into four categories based on the number of signals recorded [48]. Types 2, 3, and 4 can be used for ambulatory home monitoring. Type 2 portable monitors record the same montage of signals recorded on an in-laboratory PSGs. Therefore, type 2 monitors are rarely used for home screening as technical staff have to travel to the patient's home to set up and collect the system and data. Figures 2.1c and 2.1d show a portable type 3 PSG device and a participant

connected to the system. Table 2.1 presents the details of each type of monitoring system.

Table 2.1: Types of sleep monitoring systems [48].

Type	Number of signals	Personnel	Description
Type 1	minimum 7	In constant attendance	PSG in a sleep laboratory Minimum signals: EEG, EOG, chin EMG, ECG, air flow, respiratory effort, oxygen saturation
Type 2	Minimum 7	Not in attendance	Portable PSG Minimum signals (same as type 1): EEG, EOG, chin EMG, ECG, air flow, respiratory effort, oxygen saturation
Type 3	Minimum 4	Attended or unattended	Portable cardiopulmonary screening Minimum signals: ECG or heart rate, oxygen saturation, at least 2 channels of respiratory movement, or respiratory movement and air flow
Type 4	1 or 2	Not in attendance	Usually pulse oximetry or airflow

2.3 Oximetry

The arterial oxygen saturation or SpO_2 is a measure of the oxygen content in the blood. SpO_2 is reported as the percentage of haemoglobin molecules that are bound to oxygen $[HbO_2]$ (oxyhaemoglobin) to the total haemoglobin $[Hb]$ (oxyhaemoglobin and deoxyhaemoglobin) as follows [49]:

$$SpO_2 = \frac{[HbO_2]}{[HbO_2] + [HHb]} \times 100 \tag{2.1}$$

where $[HHb]$ is the number of unbound haemoglobin (deoxyhaemoglobin) molecules. The normal range of SpO_2 at sea level is 95-100% with values under 90% considered

to be low (hypoxemia) [50].

Oxyhaemoglobin desaturation is an important clinical consequence of SBD. Direct measurement of blood oxygen is possible with a fixed arterial catheter during sleep but it is highly invasive. Pulse oximetry based on Photoplethysmography[1] has become the standard non-invasive technique for continuous monitoring of oxyhaemoglobin desaturations during sleep and is commonly used for home sleep monitoring and portable PSG systems.

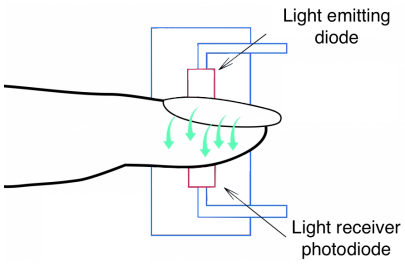


Figure 2.2: Finger pulse oximeter with the light transmitter and receiver on each side of the finger (adapted from [52]).

Most pulse oximeters have a red and infra-red wavelength light transmitter (light emitting diode or LED) on one side of a pulsating arterial vascular area (usually the index finger) and a light receiver placed on the other side. This technique is called transmission pulse oximetry (Figure 2.2). In reflectance pulse oximetry, which is more rarely used, the light transmitter and receiver are placed on the same surface. Figure 2.3 shows an example of pulse oximetry (PPG) waveform.

The transmitted light is absorbed by venous, capillary, and arterial blood as well as other tissues in the finger. The light absorption can be divided into pulsatile absorption from the movement of arterial blood and non-pulsatile absorption from arterial, venous, capillary, and other tissues such as muscle, fat and bone. These

[1]Photoplethysmography is an optical measurement technique that can be used to detect blood volume changes in the microvascular bed of tissue [51]

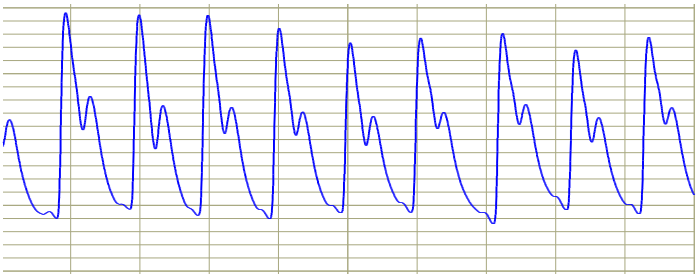


Figure 2.3: Example pulse oximetry signal.

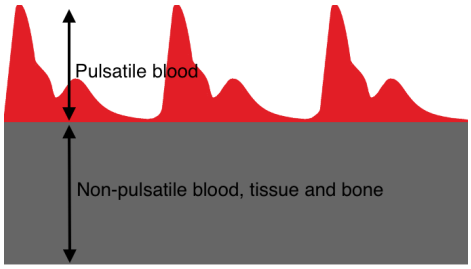


Figure 2.4: Pulse oximeter pulsatile and non-pusatile absorption.

two types of light absorption cause an AC component and a DC component in the pulse oximeter signal (Figure 2.4). The AC component corresponds to the pulsatile absorption and is produced by movement of arterial blood due to the beating of the heart, therefore can be used to extract heart rate. The DC component is related to the non-pulsatile absorption.

Figure 2.5 shows the extinction coefficient[2] of oxyhaemoglobin and reduced haemoglobin at different wavelengths. It can be seen in the figure that the extinction coefficient of oxyhaemoglobin and reduced haemoglobin at 660nm (red) and 940nm (infra-red) wavelengths is different. Oxygen saturation or the relative concentrations of $[HbO_2]$ and $[HHb]$ are calculated by the comparison of absorbances of oxyhaemoglobin and reduced haemoglobin at these two wavelengths.

The main use of pulse oximetry in sleep studies is the measurement of SpO_2 either

[2]A measure of the extent by which the intensity of a beam of light is reduced by passing through a substance [53].

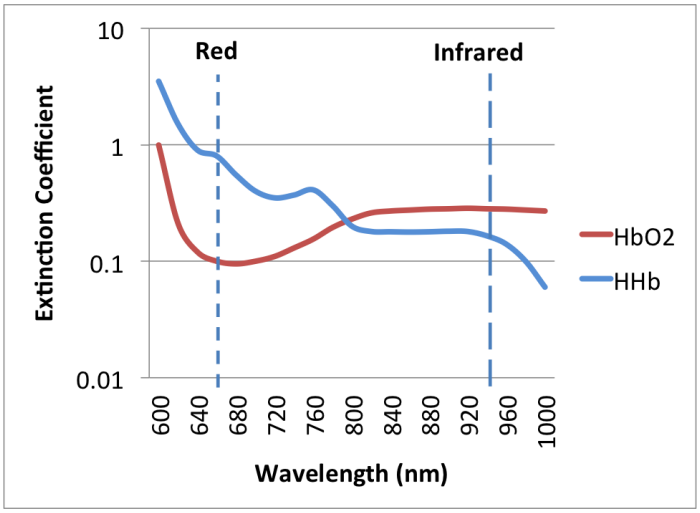


Figure 2.5: Light absorbtion for oxyhaemoglobin and reduced haemoglobin at different wavelengths.

for sleep apnoea alarm systems or for OSA diagnosis. In apnoea alarm systems usually an alarm is triggered when the SpO_2 falls below a predefined value, or when it drops by a certain amount from the normal SpO_2 baseline. Acceptable SpO_2 levels may vary with the type of patient; target values ranging from 85-95% have been considered acceptable for preterm infants [54], while desaturations of more than 2-5% have been considered indicative of OSA in adults [55].

An important limitation of pulse oximetry monitors is the high rate of false alarms, produced by motion artefacts and poor sensor contact [56]. False alarm rates between 70% and 80% have been reported in the literature [57, 58], mainly due to movement artefacts.

Most pulse oximeters deal with this problem by averaging SpO_2 measures to provide a smooth output and to reduce the impact of artefacts. Usual averaging time windows range from 5-12 seconds [59]. Furthermore, several manufacturers have included motion-tolerant algorithms in their systems [57, 59]. Since these algorithms are proprietary, the details of such technologies are not generally available, and their

performance is not well documented.

Various quantitative indices have been derived from overnight pulse oximetry for the diagnosis of OSA. One of them is the ODI which is the average number of oxygen desaturations per hour of sleep. In order to mirror the definition of an abnormal AHI, cut-off points for an abnormal ODI have been proposed (either 5, 10 or 15 desaturations per hour), but there is little evidence that one definition has greater validity than the others[60].

Many pulse oximeters can also track pulse pressure (an index of pulsation at the oximetry sensor site). When both pulse oximetry and ECG recordings are available, Pulse Transit Time (PTT)[3] can be calculated.

Negative pleural pressure[4] provokes a blood pressure drop and therefore lengthens PTT leading to the development of techniques to calculate blood pressure from PTT. The changes in pleural pressure during obstructive apnoea correlates with rising oscillations in PTT amplitude. It has been suggested that PTT might provide a means to differentiate obstructive from central apnoea as the rising oscillations do not occur during central apnoea [2].

2.4 Actigraphy

Actigraphy devices are typically small watch-like devices worn on the wrist to record movements, although they can also be placed on the ankle or trunk. The advantage

[3]Usually measured as the time interval from ECG R wave (peak of ventricular depolarisation) to the pulse recorded in the periphery (commonly the finger) [61].

[4]Pleura is the covering of the lungs (visceral pleura) and of the inner surface of the chest wall (parietal pleura); the two layers enclose a potential space, the pleural cavity. Pleural pressure is the pressure between the visceral pleura and the parietal pleura in the pleural cavity; it is also called intrapleural or intrathoracic pressure [14].

of these devices over other methods of monitoring sleep apnoea such as the pulse oximeter or PSG is that the signal can be continuously recorded over long periods of time, even for weeks or months. This is partly due to battery life (or the ease of recharging by the patient) and partly due to the relatively high levels of compliance as the device's similarity to a normal watch makes it familiar to the patients [62].

Actigraphy hardware provides different range of sensors such as Microelectromechanical systems (MEMS) sensors for 3D acceleration measurement, light, body temperature, humidity, and noise level sensors. 3D actigraphs allow for a very precise automated classification of activity [63].

Older generations of actigraphs on the market usually offer non-directional measurement of acceleration in arbitrary units (counts) rather than g (gravity of Earth). These devices suffer from technological limitations such as limited amount of memory, low sampling rates (below 0.1 Hz) and non-linearity of acceleration measurements. Activity in such devices is usually measured in epochs of several seconds or minutes. Activity data are often post-processed and available as: (i) zero-crossing rate which is counting the number of times the activity reaches zero per epoch, (ii) time-above-threshold which is the amount of time per epoch that the activity is above some defined threshold, or periodic integration which is calculating the area under the curve for each epoch [62].

The primary use of actigraphy has been the assessment of sleep-wake patterns, estimation of total sleep time, and assessment of sleep fragmentation [64]. These analyses are based on the observation that there is less movement during sleep and more movement during wakefulness. In addition to gaining information about one

night's sleep, data collected over long periods can be used to determine circadian rhythm cycles.

Actigraphic analysis depends on where the device is placed on a human body. For sleep and circadian rhythm analysis, the non-dominant wrist is usually the preferred location for the actigraph [65]. However, no significant difference in analysis results is reported between dominant and non-dominant wrists [66, 67].

If the orientation of the person is important to be monitored, such as for OSA patients, the hip or waist of the person may be the preferred location. A correlation between severity of sleep apnoeic events and body position has been shown to exist [68]. Studies on the effect of body posture on OSA have shown that the severity of sleep apnoea increases when sleeping in the supine position [69, 70, 71]. The patient's sleeping position (left side, right side, prone, supine) can therefore be used in combination with other signals for diagnosis. For example, Yoshida [72] found that body position may be useful to predict whether therapeutic devices or a treatment would be effective for that patient. Therefore, apart from using actigraphy as a stand-alone device for sleep analysis, many simplified or full PSG systems have a body position sensor incorporated.

Van Kesteren *et al.* [73] studied the effects of trunk and head position on the AHI in OSA. In order to differentiate the effect of the trunk supine position and head supine position, they used two position sensors, one placed on the mid-forehead and the other placed on the trunk of the subject. The AHI was found to be not position dependent in 41.2% of cases, dependent on the supine position (based on the trunk sensor alone) in 52.3% of cases, and supine position dependent (based on the head

sensor alone) in 6.5% of cases. In 46.2% of the trunk supine position-dependent group, head position was of considerable influence on the AHI (AHI was more than five times higher when the head was also in the supine position compared to when the head was turned to the side). The authors therefore suggest that for patients with suspected OSA two position sensors placed on the head and trunk should be considered for sleep recordings.

2.5 Video and audio recording

Nowadays, synchronised video and audio recording are common features in PSG systems. Unlike the other PSG sensors, since the camera or microphone do not need to be attached to the subject, the recording of audio/video does not disturb the natural sleep environment.

Audio recordings can be used to identify snoring, normal breathing, or obstructive events. Snoring is considered to be one of the most common symptoms of OSA occurring in 70% to 95% of patients [74]. Similar to the production of speech, snoring can be seen as the conversion of an air-stream to audible sound which is modified by the UA. OSA snoring may be differentiated from non-OSA snoring by using speech analysis techniques to analyse the variation in sound intensity during repeated apnoeas and hypopnoeas [74].

In addition to snoring sound analysis, in most OSA patients there is a reduction in snoring sounds during apnoeic/hypopnoeic events followed by a gasping/snorting sound and resumption of snoring during awakening [47, 75].

Video recordings can be used to evaluate parasomnias and seizures and also to

assess SBD. PSG video recordings help with interpreting events that are hard to interpret using only physiological recordings. For example, observations of repeated forward movements of the jaw, arching of the neck, or closing of a gaping mouth, assist in the confirmation of SBD events. Additionally, the snorting sounds in video recordings can be used to investigate whether a small dip in oxygen saturation is an artefact or a real OSA event [23].

Video recordings are particularly important when monitoring children as their PSG recordings are sometimes hard to interpret, and when monitoring people who may not have oxygen desaturations with their OSA events [23]. Simplified PSG systems which include video recording with or without recordings of physiological signals have been proposed for low cost portable/home sleep screening [76].

2.6 Sleep scoring

Traditionally, sleep technicians manually score in-laboratory PSG recordings by analysing hours of data. The scoring of sleep events includes scoring the EEG, cardio-respiratory signals, and movement recordings. The EEG recordings are analysed to identify different sleep stages, arousals, and sleep-wake periods. Movement recordings are analysed to detect sleep-related movement disorders. The respiratory signals are analysed along with pulse oximetry and other recordings such as the ECG to identify respiratory events and cardiac abnormalities [77].

Historically, in 1968, Rechtschaffen & Kales provided one of the earliest standardised manuals for the scoring of sleep stages using EEG recordings. After nearly three decades, the AASM produced a consensus report in 1999, better known as the *Chicago*

Criteria, to improve the standard definition and measurement of SBD in clinical research [78]. In 2004, the AASM appointed a committee to design a manual for the revision of sleep scoring rules, covering not only the scoring of EEG sleep stages but also the scoring of arousals, OSA events, sleep related movement disorders, and cardiac abnormalities [77]. In 2007, the AASM published an update to the 2004 manual introducing key changes to the scoring of EEG recordings [9]. Five years later, in the AASM manual published in 2012, some of the issues associated with scoring respiratory events were addressed and updated [79]. Table 2.2 compares the definitions of OSA events between different manuals. Table 2.3 details the recommended and alternative signals in the AASM 2012 manual for the analysis of respiratory events [79].

Table 2.2: Respiratory event definitions in the AASM manuals (1999−2012) [79].

Manual	Apnoea definition	Hypopnoea definition
AASM 1999 (Chicago criteria)	Apnoea-hypopnoea: >50% airflow reduction or a lesser airflow reduction with associated >3% oxygen desaturation or arousal for ⩾10 s	
AASM 2007	⩾90% airflow reduction for ⩾10 s	Recommended: ⩾30% airflow reduction and ⩾4% desaturation. Alternative: have ⩾50% airflow reduction and ⩾3% desaturation or arousal
AASM 2012	⩾90% airflow reduction for ⩾10 seconds	⩾30% airflow reduction and ⩾3% desaturation or arousal

Table 2.3: The recommended and alternative signals in the AASM 2012 manual for the analysis of respiratory events [79].

Respiratory sensor classification	
Recommended: The sensor that should be used for scoring a respiratory event. Alternative: Sensors that can be used for scoring a respiratory event if the recommended sensor is not functioning or the signal is not reliable.	
Respiratory event	**Sensor**
Apnoea	Recommended: Oronasal thermal airflow sensor Alternative: Nasal pressure transducer, RIPSum[5], RIPflow[6] PVDFSum[7] (acceptable in adults), End-tidal PCO_2 (acceptable in children)
Hypopnoea	Recommended: Nasal pressure transducer (with or without square root transformation) Alternative: Oronasal thermal airflow sensor, RIPSum, RIPflow, Dual thoraco-abdominal RIP belts, PVDFSum (acceptable in adults)

Sleep scoring is still evolving and with the publication of new updates to the scoring rules, there have been notable changes in the analysis of PSG data with clinical impacts. The changes in the definition of hypopnoea can result in large variances in the AHI values obtained with different definitions. Even the use of the recommended and alternative rules within the same manual has been reported to have a significant influence in the analysis outcome [80, 81, 82, 83, 84].

2.6.1 Inclusion of arousals in the definition of hypopnoea

One of the differences in the definitions of hypopnoea published the manuals in Table 2.2 is the inclusion of arousals. The latest manual defines hypopnoea as $\geqslant$ 30% airflow reduction and $\geqslant$ 3% desaturation or arousal for $\geqslant$ 10 seconds. However, the inclusion of arousals is still a matter of controversy [79]. Critics of the inclusion of

[5]The sum of the thoraco-abdominal RIP belt signals.
[6]The time derivative of the RIPSum signal.
[7]The sum of Polyvinylidene fluoride (PVDF) thoraco-abdominal respiratory belt signals.

arousals firstly point to the fact that most studies have not found a link between the severity of arousals and adverse cardiovascular issues. Secondly, opponents argue that the scoring of arousals is less reliable, hence it reduces the reliability of the scored events. Thirdly, they argue that the inclusion of arousals is not applicable to simplified PSG screening or home screening. On the other hand, the advocates of the inclusion of arousals cite evidence that fragmentation of sleep, even without associated oxygen desaturation, can lead to severe daytime sleepiness. Additionally, patients with milder OSA syndromes who may benefit from treatment might not be detected if arousals are not included in the AHI scoring. The supporters of inclusion of arousals also point out that without arousals the sensitivity of full PSG studies in the detection of 'significant' respiratory events is reduced [79].

2.6.2 Automated sleep scoring

With the development of automated in-laboratory or home/portable PSG systems, there is less need for manual scoring of events. These automated scoring systems sometimes output more than an AHI score and provide detailed marked sleep events throughout the recordings. The automated algorithms are reasonably accurate, however review of the original data may be necessary, particularly in the presence of artefacts. One of the advantages of automated scoring algorithms is the reproducibility of the scoring versus human scoring. Secondly, human fatigue or distraction does not play a role in the scoring. Thirdly and most importantly, automated scoring is fast, does not require human resources, and is very cost-effective, allowing a larger number of patients to be screened.

On the other hand, human experience is valuable in manual scoring as artefacts that present like a sleep event to the automated algorithm will be disregarded by the sleep specialist. Loss of employment may be a natural concern to sleep technicians, however their experience may be directed to other areas such as assisting patients and helping with adherence to treatment. Automated algorithms are increasingly becoming part of PSG systems in the laboratories and semi-automated PSG scoring is likely to become the standard in the future [47].

One difficulty with the automated scoring of sleep events is how to define a normal breathing baseline for a recording. Apnoeas, hypopnoeas, and oxygen desaturations are defined as percentage reductions from a baseline level. However, the baseline is not clearly defined in the guidelines for scoring sleep. Additionally, during periods of cyclic apnoeas and hypopnoeas, there is no clear baseline level. Therefore, manual and automated scoring will vary simply because of using different definitions for the baseline [47]. For the same reason, this difference in scoring may exists for manual scoring by different sleep technicians and manual scoring by different algorithms. There is a need for clear definitions of baseline when scoring sleep events.

Chapter 3

Data

This chapter focuses on the data used in this thesis. Two main sources of data one from hospital patients and one from healthy volunteers were collected. These are described below in detail.

3.1 Hospital patients

3.1.1 Overview

This dataset includes in-laboratory overnight PSG recordings collected from three rooms at the sleep lab of the Respiratory Medicine Group, Churchill Hospital, Oxford, UK. Overall, data from 259 patients was collected over a period of two years (from 2011 to 2013). Recordings consist of a set of physiological signals such as ECG, pulse oximetry, pulse rate, sound, respiration rate, and a video file for periods of 7 to 8 hours. Other patient information such as age, gender, height, weight, and neck circumference as well as consultant diagnosis was collected from the patients'

paper records. All patients were referred to the hospital with suspected OSA. From the 259 patients, 151 were diagnosed with different severity levels of OSA. However, 108 patients were diagnosed as normal or as not suffering from OSA. Patients' OSA severities were assessed by the sleep specialists using the patients' ODI automatically generated by the PSG software in addition to physiological signals, history, and demographics information. The ODI used were based on oxygen desaturations of $\geqslant 4\%$ from baseline.

3.1.2 Room setting

The patients in the sleep lab were provided with bed sheets (patterned or plain) and some patients were under the bed sheets during the night. The infra-red camera and the microphone were positioned around 2 to 3 meters away from one of the top corners of the bed. The distance and position of the cameras varied slightly in each room. Figure 3.1 shows snapshots of video captured from sleep rooms in which the data was collected. Figure 3.2 shows an example snapshot of a patient video recorded from one of the rooms.



(a) (b) (c)

Figure 3.1: Snapshots of video captured from the three sleep rooms from which patient data was collected.



Figure 3.2: Snapshot of a video of a patient sleeping on the right side and partially covered with a bed sheet. The head of the patient is masked to preserve anonymity.

3.2 Normal volunteers

3.2.1 Overview

This dataset includes full overnight PSG recordings for 32 normal participants. The participants were mainly PhD students from the Institute of Biomedical Engineering, University of Oxford, Oxford, UK. All data was recorded in a dedicated bedroom with PSG equipment over a period of 3 months. Recordings consist of a set of physiological signals such as ECG, pulse oximetry, pulse rate, sound, respiration rate, and a video file for periods of 7 to 8 hours. Other information such as age, gender, height, weight, and neck circumference was collected on the night of study. The height, weight, and neck circumference of each subject were measured by a member of the research team.

3.2.2 Room setting

The subjects in the study were provided with a thick patterned duvet and most subjects were under the duvet during the night. The infra-red camera and the microphone were positioned a few meters away from one of the top corners of the bed. Figure 3.3

shows an example snapshot from a video recorded from one of the normal volunteers.



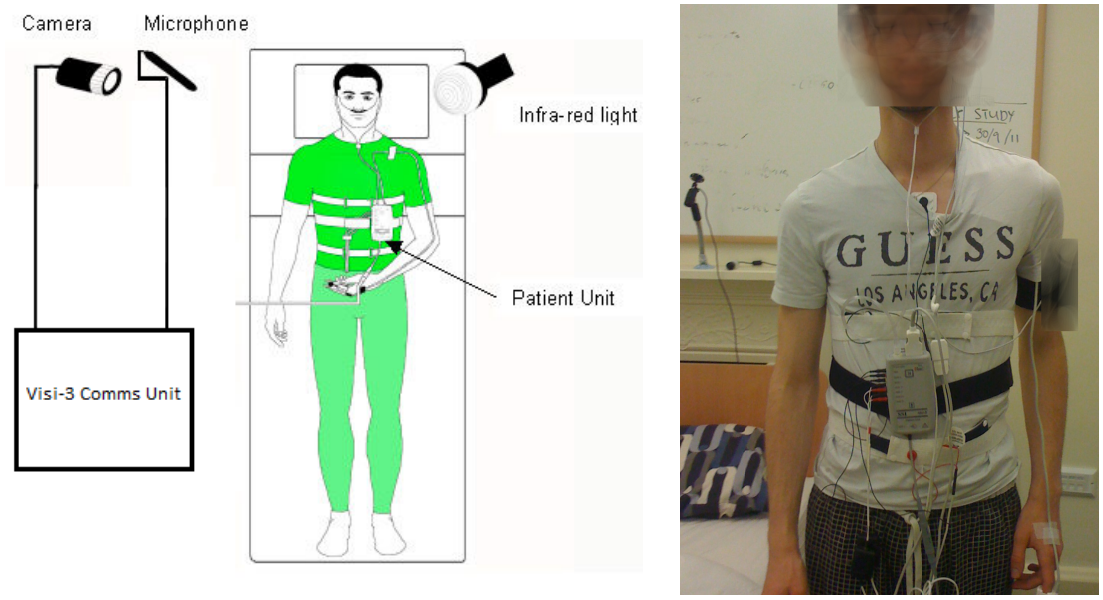
Figure 3.3: Snapshot of a video from a normal subject almost completely covered with a duvet.

3.3 PSG system

The PSG system used in the sleep clinic at the hospital and also for recording normal volunteers was the VISI-3 system from Stowood Scientific Instruments Ltd., Oxford, UK. Figure 3.4a shows the typical set-up of the VISI-3 system, including the sensors acquiring the physiological signals, the microphone, an infra-red light, and the camera. Figure 3.4b shows one of the normal volunteers wearing the sensors for the overnight PSG study.

3.3.1 PSG channels

The PSG system includes several sensors for recording different types of physiological parameters. There is also an infra-red camera, an infra-red light, and a microphone. Table 3.1 gives the details of each channel recorded during a PSG study. Figure 3.5



(a) (b)

Figure 3.4: (a) A typical set-up of the VISI-3 PSG system (Stowood Scientific Instruments Ltd., Oxford, UK) in the sleep lab showing the hardware acquiring the physiological signals, the microphone, an infra-red light, and the camera. Image adapted from VISI-3 hardware documentation. (b) A normal volunteer wearing the PSG sensors for overnight study.

shows a snapshot of the PSG VISI-3 software used to view and analyse the PSG recordings.

Note that in general, the EEG channels are not routinely connected to the patients who are referred to the sleep lab of the Churchill hospital (Oxford, UK) with suspected OSA. Therefore, as can be seen in Table 3.1, the PSG set-up used for both the hospital patients and the normal volunteers in this thesis does not include EEG recordings. Since the EEG recordings are the standard signals used for the detection of arousals, the hypopnoea events reported in this thesis are based only on the criterion of a desaturation of $\geqslant$3% and do not include arousals.

Table 3.1: Details of channels recorded during PSG studies.

Parameter	Description	Recording source	Sampling frequency (Hz)
Sound	Sound recording	Microphone	44k
ECG	Electrocardiography	ECG	256
AEFF	Abdominal respiratory effort	Abdominal respiratory inductance plethysmography (RIP) band	256
TEFF	Thoracic respiratory effort	Thoracic RIP band	256
PLETH	PPG	Pulse oximeter	256
NAF	Nasal Airflow	Nasal airflow pressure sensor	128
Body Position	Supine,prone,left,right	Accelerometer placed on chest	32
Video	Infra-red video recording	Infra-red camera	25
PTT	Pulse transit time	ECG and pulse oximeter	5
R-R	R wave (R to R) interval	ECG	5
PR	Pulse Rate	Pulse oximeter	1
SpO_2	Peripheral oxygen saturation saturation	Pulse oximeter	1

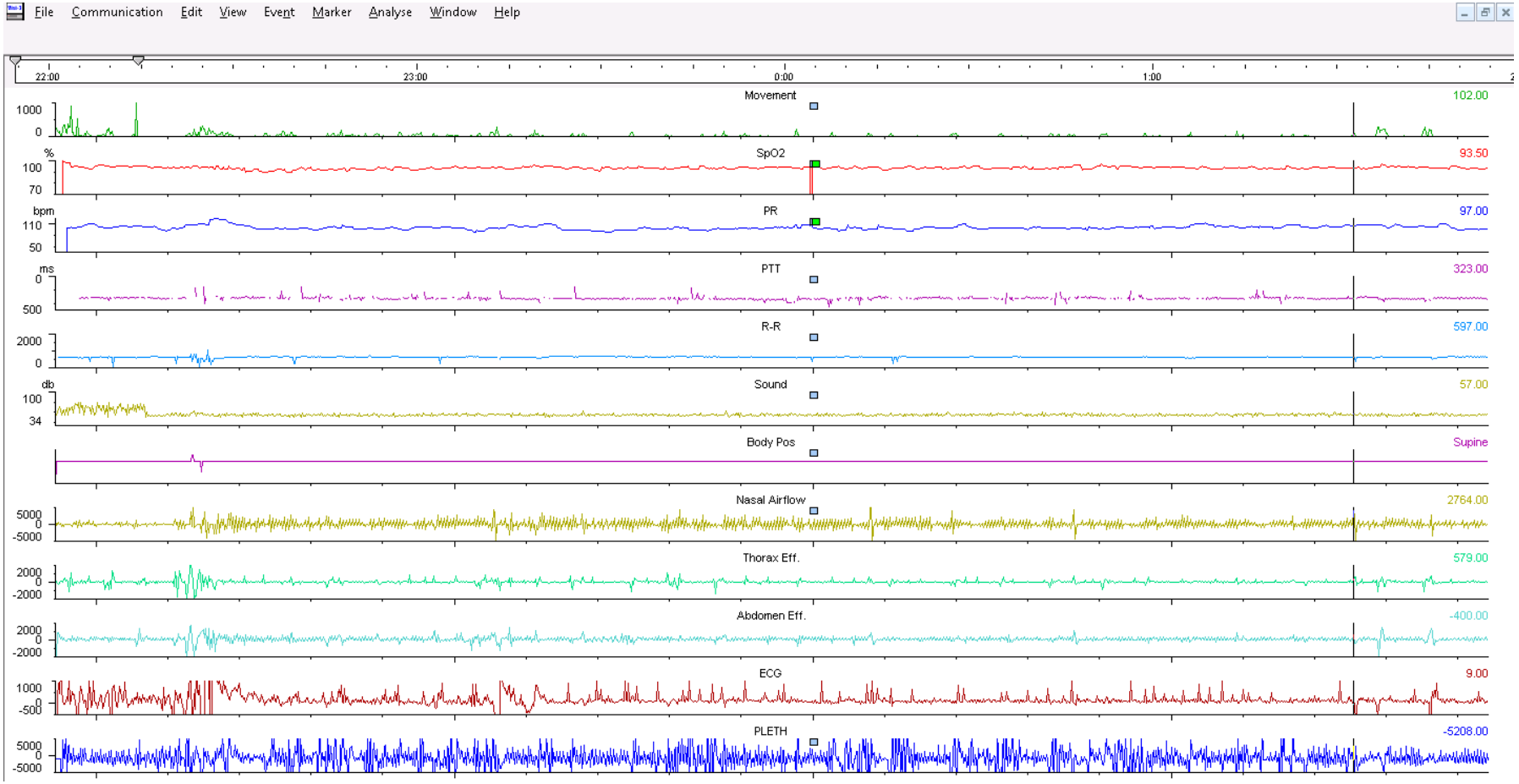


Figure 3.5: A snapshot of the VISI-3 PSG software.

3.4 Data groups

Since the number of signals acquired in each of the three sleep rooms at the hospital varied, the data collected from the hospital was divided into two groups (DS1 and DS2) as described below. In general, patients in two of the sleep rooms were connected to fewer sensors and the respiratory signals including the Abdominal Respiratory Effort (AEFF), Thoracic Respiratory Effort (TEFF), and Nasal Airflow (NAF) signals were not recorded for these patients.

Dataset 1 (DS1) This dataset includes patient data from hospital sleep lab. Out of the 259 patients, 12 had their data removed because of invalid oxygen saturation or audio signals. Details of the remaining 247 DS1 patients can be seen in Table 3.2. 225 patients in DS1 do not have respiratory AEFF, TEFF, and NAF recordings. Although all patients in DS1 were referred to the hospital with suspected OSA, some were diagnosed with other sleep disturbances. Table 3.3 lists some of the other conditions associated with sleep fragmentation that were mentioned in the DS1 patient records.

Table 3.2: Number of patients in each OSA severity group (for the definition of severity groups see chapter 1).

Group	no. of Patients	Sex (m/f)	Age (yrs, $\mu \pm \sigma$)	BMI ($\frac{kg}{m^2}$, $\mu \pm \sigma$)	Neck Circumference (cm, $\mu \pm \sigma$)
Non-OSA	102	85m, 17f	53.6±12.4	29.3±6.8	41.6±3.7
Mild OSA	56	45m, 11f	58.2±11.5	31.0±8.2	41.6±4.0
Moderate OSA	47	38m, 9f	58.2±11.4	34.0±8.5	44.0±4.8
Severe OSA	42	33m, 9f	53.8±9.9	38.3±8.6	45.8±4.3
Mild/No OSA	158	130m, 28f	54.8±12.2	29.8±7.3	41.6±3.8
Moderate/Severe OSA	89	71m, 18f	56.1±10.9	36.0±8.8	44.9±4.6
Total	247	201m, 46f	55.3±11.7	32.1±8.4	42.8±4.4

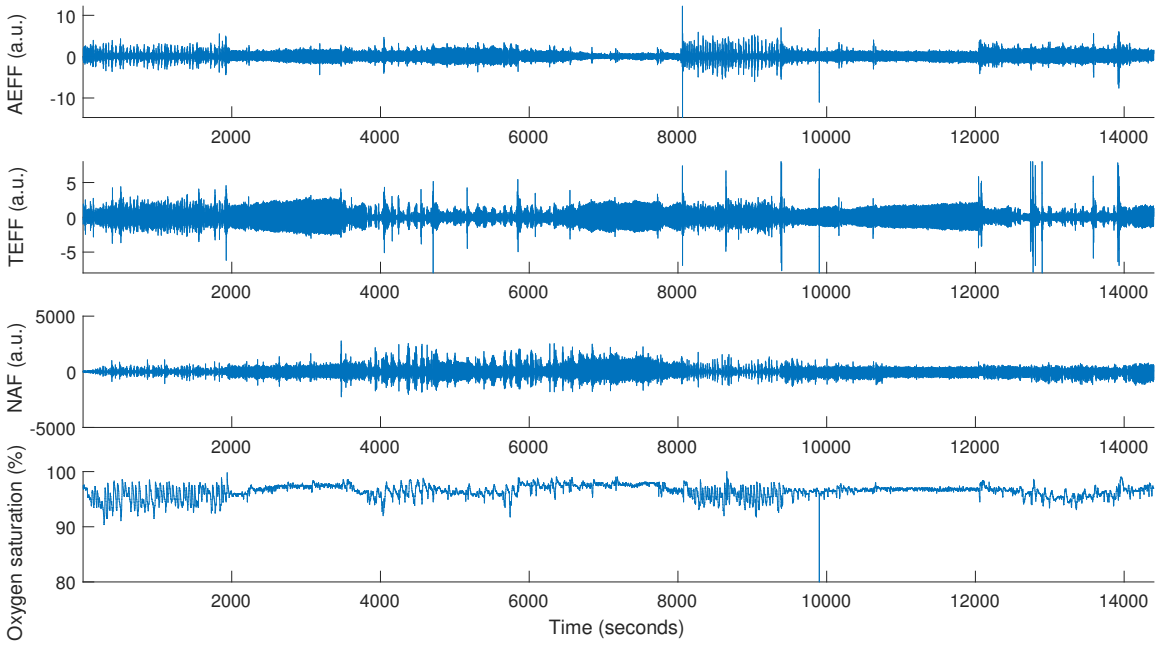
Table 3.3: Some of the other conditions associated with sleep fragmentation mentioned in DS1 patient records.

Condition	no. of Patients
CSA	5
Parasomnia	1
RLS	1
(Severe) COPD[1]	2
Depression	4
OSH[2]	1
UARS[3]	4
PLMD	9
Restlessness	1
ADHD[4] (woke severely)	1
Cheyne-Stokes respiration	1

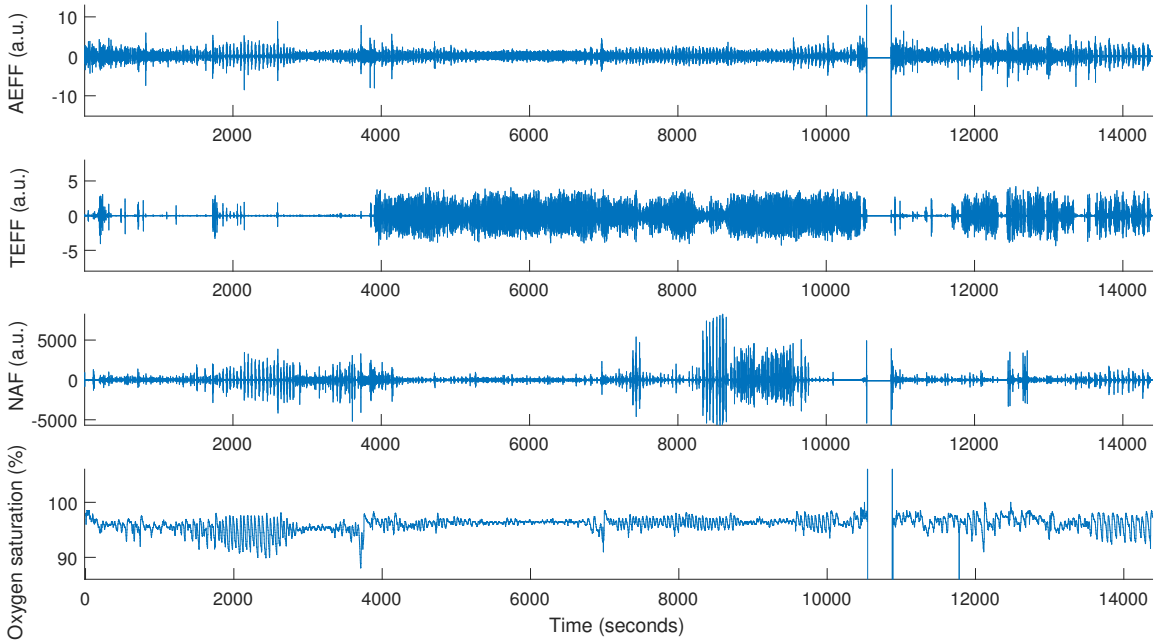
Dataset 2 (DS2) This dataset is a subset of DS1 which includes data from 30 patients from the hospital sleep lab with AEFF, TEFF, and NAF recordings. Of the 30 patients in DS2, 8 had their data removed as the RIP signals (AEFF or TEFF signals) of these patients had unusual changes in the amplitude of the signal as shown in Figure 3.6, possibly due to the displacement of one of the RIP elastic bands. Amongst the 8 removed patients, one was diagnosed with moderate OSA, one with mild OSA and the remaining 6 were diagnosed with no OSA. The details of patients in this group are given in Table 3.4.

Dataset 3 (DS3) This dataset includes all data from normal volunteers. The details of subjects in this groups are presented in Table 3.5.

[1]Chronic obstructive pulmonary disease
[2]Obesity hypoventilation syndrome
[3]Upper airway resistance syndrome
[4]Attention deficit hyperactivity disorder



(a)



(b)

Figure 3.6: a. Normal quality AEFF, TEFF, NAF and oxygen saturation signals from a patient in DS2. b. AEFF, TEFF, NAF and oxygen saturation signals from a patient in DS2 with the TEFF signal having unexplained large increases. (a.u.) = Arbitrary unit.

Table 3.4: Details of patients in DS2 (for the definition of OSA severity groups see Chapter 3).

Group	no. of Patients	Sex (m/f)	Age (yrs, $\mu \pm \sigma$)	BMI ($\frac{kg}{m^2}$, $\mu \pm \sigma$)	Neck Circumference (cm, $\mu \pm \sigma$)
Non-OSA	10	5m, 5f	53.4±8.5	36.4±9.9	41.8±4.7
Mild OSA	3	1m, 2f	44.3±7.5	40.3±14.8	38.9±0.7
Moderate OSA	3	3m, 0f	62.0±14.0	40.2±14.5	46.5±3.6
Severe OSA	6	6m, 0f	50.3±9.3	45.5±14.8	49.1±5.5
Mild/No OSA	13	6m, 7f	51.3±8.9	37.34±10.7	41.1±4.3
Moderate/Severe OSA	9	9m, 0f	54.2±11.7	43.7±14.0	48.2±4.9
Total	22	15m, 7f	52.5±10.0	40.0±12.3	44.0±5.7

Table 3.5: Details of normal volunteers.

no. of subjects	Sex (m/f)	Age (yrs, $\mu \pm \sigma$)	BMI ($\frac{kg}{m^2}$, $\mu \pm \sigma$)	Neck Circumference (cm, $\mu \pm \sigma$)
32	19m, 13f	25.1±3.2	24.2±3.6	35.7±3.4

3.5 OSA severity based on patient demographics

As used in the STOP-BANG questionnaire, patient demographics including Body Mass Index (BMI), age, neck size, and gender can be used for a preliminary assessment of the patient's risk of OSA. Figure 3.7 compares the demographics of moderate-to-severe OSA patients versus mild-to-no OSA patients. Table 3.6 shows the ability of patient demographics to differentiate moderate-to-severe OSA from mild-to-no OSA patients in DS1 using the STOP-BANG thresholds.

Table 3.6: The use of patient demographics to differentiate between moderate-to-severe OSA patients and mild-to-no OSA patients in DS1. Ac = Accuracy, Se = Sensitivity, Sp = Specificity, PPV = Positive predictive value, NPV = Negative predictive value, AUC = Area under the curve (for the definitions of classification performance metrics see Appendix B).

Criteria	Threshold	Ac(%)	Se(%)	Sp(%)	PPV(%)	NPV(%)	AUC(%)
BMI	35	70.8	43.8	86.0	63.9	73.1	64.9
Neck size	40	72.4	43.8	88.6	68.4	73.6	66.2
Age	50	52.6	40.4	59.5	36.0	63.9	50.0
Gender	Male	40.0	79.8	17.7	35.3	60.8	48.7

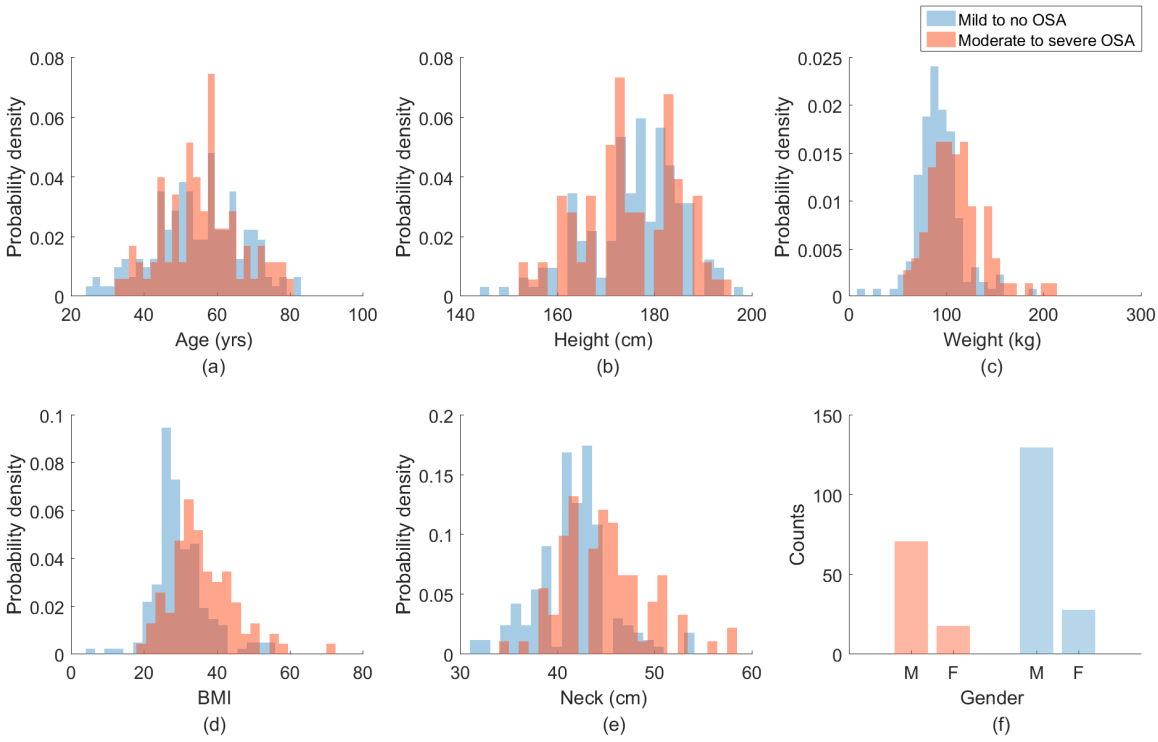


Figure 3.7: Normalised histograms comparing demographics of moderate-to-severe OSA patients versus mild-to-no OSA patients. Histograms (a) to (e) are normalised so that sum of the area under the bins equals 1.

Chapter 4

Using video derived activity to detect moderate to severe OSA

4.1 Introduction

This chapter investigates the possibility of using whole body activity derived from overnight video recordings of sleep to identify moderate-to-severe OSA patients. Patients with moderate-to-severe OSA are defined as suffering from more than 15 episodes of apnoea or hypopnoea per hour of sleep. During each apnoeic episode, there is limited or no body movement apart from small respiratory efforts occurring around the torso. The awakening following each apnoeic episode causes larger movements in other parts of the body, particularly the head while the patient is gasping for air. A series of apnoea and awakenings cause a cyclical pattern of no-movement followed by a surge in the activity which is the focus of the analyses in this chapter. The prominence of this cyclical pattern in moderate-to-severe OSA patients is illustrated in Figure 4.1

which includes a 30-minute segment of sleep activity from four patients with OSA of different severity levels. The extraction of the Video Derived Activity (VDA) signals is based on frame differencing and is explained in section 4.3.2 in more detail. As the figure shows, patients with mild OSA, particularly mild supine OSA (OSA only occurring in supine position), behave similarly to non-OSA patients for the majority of their sleep.

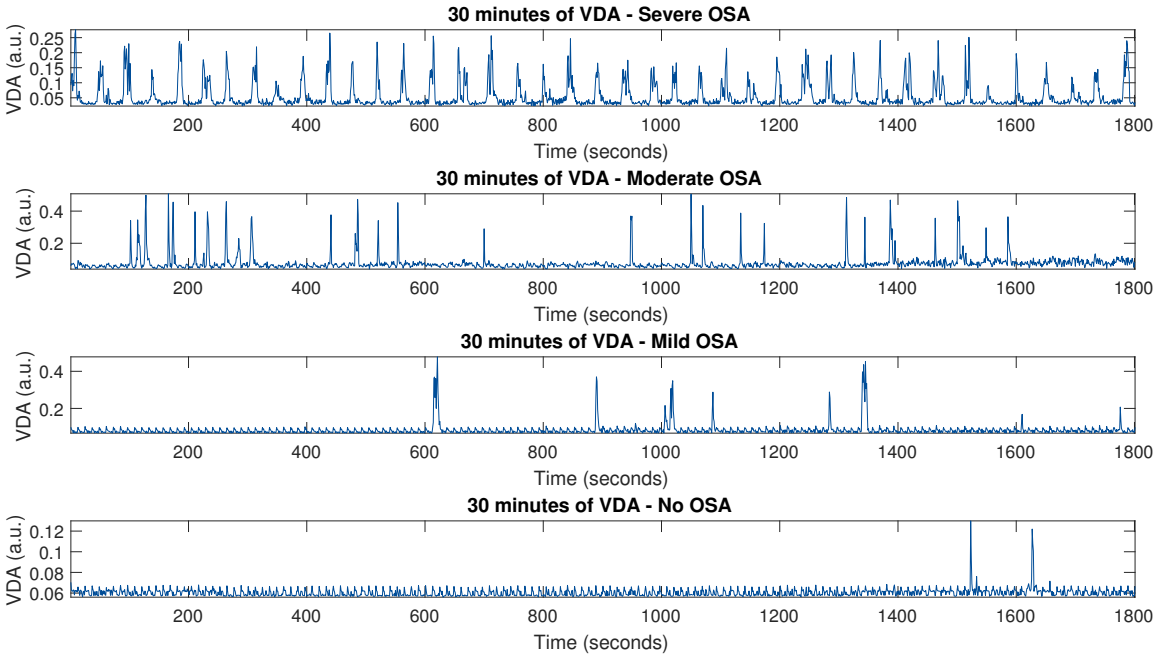


Figure 4.1: Example 30-minute segments of Video Derived Activity (VDA) signals from four patients with different OSA severities. The figure shows a distinct repetitive pattern in the sleep activity of moderate or severe OSA patients due to regular apnoea and awakenings. Patients with mild OSA may behave similarly to non-OSA patients for the majority of their sleep. (a.u.) = Arbitrary unit.

4.2 Background: analysis of sleep activity

4.2.1 Analysis of sleep activity using actigraphy devices

Over the years, recordings of activity have been used in the field of sleep medicine to monitor sleep/wake periods and play a central role in the assessment of quality of sleep [64]. The Sensitivity (Se) of actigraphy in differentiating wakefulness from sleep can be as high as 95%, but due to the absence of movement during wakefulness in bed, Specificity (Sp) as low as 35% has been reported in certain patient populations (for the definitions of Se and Sp see Appendix B). Activity has also been used to study the effect of body position on sleep apnoea [68, 69, 70, 71, 73, 85] and assess whether the effectiveness of therapeutic devices was influenced by body posture [72]. Another important area of actigraphy usage is the analysis of circadian rhythm abnormalities [62], often linked with psychiatric and neurodegenerative diseases [86, 87].

Additionally, actigraphy has been used to detect Periodic Limb Movement of Sleep (PLMS) and OSA in patients. Sforza *et al.* [88] used a device specifically tailored to detect limb movements. A total of 43 patients (33 males, 10 females; age = 57.6±3.7), who were referred for insomnia and/or excessive daytime sleepiness, underwent one or two nights of PSG with simultaneous bilateral recording of limb activity. The authors found that actigraphy-PLMS correlated highly with PSG-PLMS ($r = 0.87$) and that actigraphy-PLMS had Se = 88% and Sp = 76% for detecting PLMS index > 10. King *et al.* [89] fixed an actigraphy device to the big toe of five patients with known PLMS. For a PLMS index > 25, the Actiwatch had 100% Se and 97% Sp.

Sadeh *et al.* [90, 91] found that actigraphy is not considered a valid tool for

assessing sleep-disordered breathing. Elbaz *et al.* [92] analysed 20 subjects (15 males, 5 females; age = 52±15; BMI = 28±5) with suspected OSA using an actigraph worn on the non-dominant wrist as well as computerised PSG. The authors found that AHI calculated with the total actigraphy-derived sleep time (instead of the time at which the patient went to bed) improved the validity of the AHI estimate based on simple respiratory polygraphy[1] (Se went from 50% to 88% while Negative Predictive Value (NPV) increased from 75% to 92.5% - for the definition of NPV see Appendix B).

Hedner *et al.* [93] developed an automated algorithm for sleep/wake assessment in normal and OSA patients. They studied 228 patients (190 OSA, 38 normal; age = 48.8±14.0; 163 male, 65 female; BMI = 29.4±6.3) who underwent PSG while wearing wrist actigraphs. First, for each 30-second epoch of the activity recording, they determined wake versus sleep states using an adaptive threshold for an "energy" value calculated for the epoch. Next, if an all-pole based linear predictive model detected periodicity in an epoch (such as is expected during periodic apnoeic events), the epoch was marked as corresponding to sleep overruling the decision from the first step. The results, grouped by the severity of OSA, indicated a relatively high agreement between their detected sleep/wake states and the PSG scoring (86% agreement for normals decreasing to 79% for severe OSA patients). The Se decreased from 90% for normals to 85% for severe OSA patients while the Sp remained around 70% for all patient groups.

Despite the widespread use of actigraphy for sleep assessment, there is no standard in actigraphic sleep-wake scoring rules comparable to the R&K rules [94]. Sleepwake

[1]Respiratory polygraphy is defined as recording of respiratory parameters (airflow, thoracic and abdominal movements, and pulse oximetry) but not of neurophysiologic sleep parameters [92].

scoring algorithms for newer devices may not have been developed yet or may need to be validated against PSG standard scores. As a result of differences in hardware, most actigraphs implement their own sleep detection algorithms, and manufacturer differences in data sampling frequency, processing and analysis make it difficult to compare actigraphic studies [65]. However, the availability and simplicity of actigraphic analysis tools make it an attractive choice for wearable sleep monitoring scenarios.

4.2.2 Analysis of sleep activity using video recordings

Video recordings have been widely used to correlate PSG signals with patients' behaviour, particularly respiratory and body movements during sleep [95, 76, 96, 97]. Simplified PSG systems which include video recording with or without recordings of physiological signals have been proposed for low-cost portable/home sleep screening [76]. Sivan *et al.* [76] studied the use of home video recording of sleep to improve patient selection for PSG studies. In 58 children suspected of having OSA, they compared the results of overnight PSG studies with 30-minute video recording of sleep during periods of snoring and breathing difficulty at home. The manual PSG-based diagnoses were highly correlated with the video-derived results, with an overall agreement of 84%, Se of 94% and Sp of 68%.

Wang *et al.* [98] described an automated video monitoring algorithm tailored to recover the posture of a sleeping person where body parts may be covered and partially visible. 32 video clips were filmed for eight subjects with different environment settings such as different illuminations, camera angles, and various occlusion levels (occlusion here refers to body parts being partially covered by bedding or hands). The results

were compared with a general method proposed by Ramanan *et al.* [99] to identify and track the posture of individuals in video recordings. The authors report that their method outperforms the general method in detecting head, torso, and lower body (recognition rates: head: 0.89 vs. 0.66; torso: 0.95 vs. 0.59; right upper leg: 0.75 vs. 0.50; left upper leg: 0.71 vs. 0.38). The authors believe the analysis of the activity of the detected body parts will have diagnostic value for sleep apnoea.

Liao *et al.* [100] presented a near-infra-red video-based system to estimate sleep/wake status. They studied 10 video recordings of subjects with simultaneous PSG studies and actigraph recordings (only available for eight of the 10 studies). Their method, based on thresholded frame differencing and a motion history image technique [101], showed that the average recognition accuracy for video recordings (92.13%) was slightly higher than recognition accuracy using actigraphy (91.24%). Liao *et al.* [102] also proposed a home sleep monitoring system focused on handling variations in brightness and light. Different background modelling techniques such as consecutive frame differences, Gaussian mixture models, local binary patterns, and Local Ternary Patterns (LTP) were investigated to handle the changes in the environment's light levels. The study of three videos with 1800 sleep/wake episodes showed that LTP had the highest accuracy for detecting sleep/wake states (around 57% of wake and 98% of sleep states in uniform and varying light conditions).

Heinrich *et al.* [103] compared home-recorded Wrist-worn Actigraphy (WA) and Video Actigraphy (VA) extracted using recursive search motion estimation. Study subjects included a first group of 5 adult subjects (4 male, 1 female; age between 24 to 55, no self-reported sleep disturbance) wearing a wrist-worn actigraph and using

a video recording system, and a second group of four subjects who underwent PSG at a sleep lab wearing a wrist-worn actigraph. The results for the first group showed 61% agreement between WA and VA, with WA failing to recognise 20.2% of all body part movements indicating that VA may be more sensitive. For the second group, a sleep efficiency parameter (defined as the ratio of total sleep time over total time in bed, expressed in %) calculated from WA and VA had reasonable agreement with the reference PSG (5.8% VA error vs. 4.5% WA error).

Despite a significant increase in the number of publications on the use of video recordings for monitoring sleep in recent years, attempts to analyse the video data from real patients to detect and score sleep disturbances are limited. The aim of this chapter is to use PSG video recordings of suspected OSA patients to provide a severity-based patient classification which facilitates referral for screening and treatment of OSA.

4.3 Materials and methods

4.3.1 Data

The video dataset used in this chapter is DS1 (see Section 3.4 in Chapter 3) which includes in-laboratory PSG recordings of 247 patients who were referred to the sleep lab of the Churchill Hospital (Oxford, UK) with suspected OSA. The PSG recordings include a number of physiological signals as well as audio and video recordings. The videos in DS1 were recorded in grey scale in MPEG2 format with a resolution of 352×288 pixels, and a frame rate of 25 fps.

4.3.2 Activity estimation from video data

The video data was analysed from 2 to 6 hours into the recording (total 4 hours) to increase the likelihood of the patient being asleep during the study period. An investigation of 4 hours of sleep data would be sufficient to detect apnoea in patients according to clinical experts (J. R. Stradling, personal communication, September 2010). The activity patterns of interest and specific to OSA patients are episodes of no movement or little respiratory efforts (apnoea) lasting more than 10 seconds followed by large movements and gasps for breath (awakenings). Both types of activity have frequencies similar or below the average respiratory rate (6 to 36 breaths per minute (bpm) or 0.1Hz to 0.6Hz). Therefore, the frame rate of the videos could be reduced from 25Hz to 1Hz. This process was carried out using *ffmpeg* software (https://ffmpeg.org) with option '-r' which removes the frames from the original video to achieve the desired frame rate.

The activity signals were then extracted from the down-sampled video recordings by taking the mean of the absolute difference between every two consecutive frames as described below.

Let's denote the frame at time t as fr_t and each pixel in fr_t as $p_{x,y,t}$ where x and y represent the location of the pixel in the frame with $p_{0,0,t}$ being the top left corner. The absolute difference between the pixel intensities of two consecutive video frames (δ) was calculated as follows:

$$\delta_{x,y,t} = |p_{x,y,t} - p_{x,y,t-1}|, \quad \forall t > 1 \tag{4.1}$$

As most movement took place in a small area of the frames (around the head

and torso) with the rest of the frames being still, at each t, only the δ values above 90th percentile were used (90th percentile of 352×288 pixels). This threshold was determined by investigating the approximate ratio of the area that the patient's upper body part was covering in the frame with respect to the whole frame area. As Figure 3.1 shows, the top half of the bed occupies a small area of the frames. The VDA signal was calculated as:

$$VDA(t) = \frac{1}{M} \sum_{y} \sum_{x} \delta_{x,y,t} \times \Theta(\delta_{x,y,t} - P_t^{90}), \tag{4.2}$$

where P_t^{90} is the 90th percentile of the $\delta_{x,y,t}$ values, Θ is the Heaviside step function[2], and M is the total number of $\delta_{x,y,t} > P_t^{90}$.

4.3.3 Feature extraction

In order to analyse the correlation between a patient's sleep activity and the severity of OSA, quantifiable measurements of activity patterns are required. Here, multi-scale entropy analysis as well as three Inter-Event Interval (IEI) metrics were used for the analysis of activity patterns, which are described below.

4.3.3.1 Multi-scale entropy analysis

Multi-Scale Entropy (MSE) [104] provides a way of describing the complexity of a time series and has been used for identifying rhythms in physiological time-series. This motivates our choice of MSE for sleep activity analysis. Given a time series y with length D, the MSE process involves first generating a number of "coarse-grained"

[2]Step function taking a value of zero for negative arguments and a value of one for positive arguments.

time series (x^τ) by averaging y over non-overlapping windows of length τ (assumes there is no aliasing issue):

$$x_i^\tau = \frac{1}{\tau}\sum_{t=1}^{\tau} y_{[(i-1)\tau+t]}, \quad 1 \leqslant i \leqslant D/\tau \tag{4.3}$$

followed by calculating Sample Entropy (SE) [105] for each coarse-grained time series.

SE is a method of estimating the regularity of a finite-length time series being low in regular time series and high in irregular time series. If $\{x_1, ..., x_N\}$ represents a time series of length N and $\mathbf{x}_m(i) = \{x_i, ..., x_{(i+m)-1}\}$, $1 \leqslant i \leqslant N - m$ represent patterns of length m, $n_i^m(r)$ is defined as the number of vectors $\mathbf{x}_m(j)$ that have a distance smaller than r with $\mathbf{x}_m(i)$. Here, the distance between two vectors is defined as the maximum absolute difference between the components of the two vectors. $U_i^m(r) = n_i^m(r)/(N-m)$ is the probability that the distance between vector $\mathbf{x}_m(i)$ and any other vector $\mathbf{x}_m(j)$ is smaller than r. $U^m(r) = 1/(N-m)\sum_{i=1}^{N-m} U_i^m(r)$ is the probability of any two vectors of length m being within distance r of each other. Richman and Moorman [105] define SE as the following:

$$\hat{H}_{SE}(m, r, N) = -ln\frac{U^{m+1}(r)}{U^m(r)}. \tag{4.4}$$

$\hat{H}_{SE}(m, r, N)$ calculates the negative natural logarithm of the conditional probability that if two patterns of length m are within distance r of each other, they will still be within distance r of each other when one more point is added. If the probability is high that m points as well as $m + 1$ points will be within distance r of each

other, then SE will be low indicating a signal that is regular. However, a signal that has different probabilities for m and $m+1$ points will have a high SE, indicating a more complex signal. It is expected that the VDA signal of mild-to-no OSA patients will have a low SE value because of having few peaks and being mainly flat, while the VDA signal for moderate-to-severe OSA patients will have a high SE value because of the irregular pattern and the large number of peaks in the signal.

4.3.3.2 Inter-Event Interval metrics

Based on the clinical definition of OSA and severity of OSA, three IEI metrics were defined using the peaks[3] in the VDA signal. Let's denote the location of the peaks in the VDA signal as $\{l_1, ..., l_K\}$ where K is the number of peaks in the signal. The first metric, IEI_f, was defined as the number of peaks ($=K$) in the whole VDA signal as a measure of sleep fragmentation. Frequent apnoea and awakenings in moderate-to-severe OSA sufferers are expected to cause a large number of peaks in the VDA signal and a higher IEI_f score compared to the number of peaks for mild-to-no OSA patients.

Additionally, as higher severity of OSA is directly related to a higher number of apnoeic events per hour of sleep [84], the intervals between the peaks in the VDA signal may give information about the severity of OSA. The intervals between the peaks are calculated as below:

$$d_i = (l_i - l_{i-1}), \forall i > 1 \tag{4.5}$$

[3]A point in the signal is defined as a peak (local maximum) if it is larger than its two neighbouring points. The peaks have to be at least 10 seconds apart which is the minimum duration of an apnoeic or hypopnoeic event.

IEI_m was defined as the mean of the intervals between peaks in the activity signal:

$$IEI_m = \frac{1}{K-1}\sum_{i=1}^{K-1} d_i \tag{4.6}$$

IEI_v was defined as the variance of the intervals between peaks in the activity signal:

$$IEI_v = \frac{1}{K-1}\sum_{i=1}^{K-1} (d_i - \mu_d)^2 \tag{4.7}$$

where $\mu_d = IEI_m$ is the mean of the intervals. Moderate-to-severe OSA patients are expected to have lower IEI_v and IEI_m as the interval between apnoeas is shorter in these patients comparing to mild-to-no OSA patients.

Figure 4.2 shows examples of the peaks in the VDA signal from four patients with different severity levels of OSA.

Figure 4.3 compares IEI scores between different OSA severity groups in DS1.

4.3.4 Classification

4.3.4.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) has been used in a variety of applications such as image retrieval and face recognition. Classical LDA projects the data onto a lower-dimensional vector space such that the ratio of the between-class distance to the within-class distance is maximised, thus achieving maximum discrimination. The optimal projection can be calculated by computing the eigen-decomposition of the covariance matrices [106].

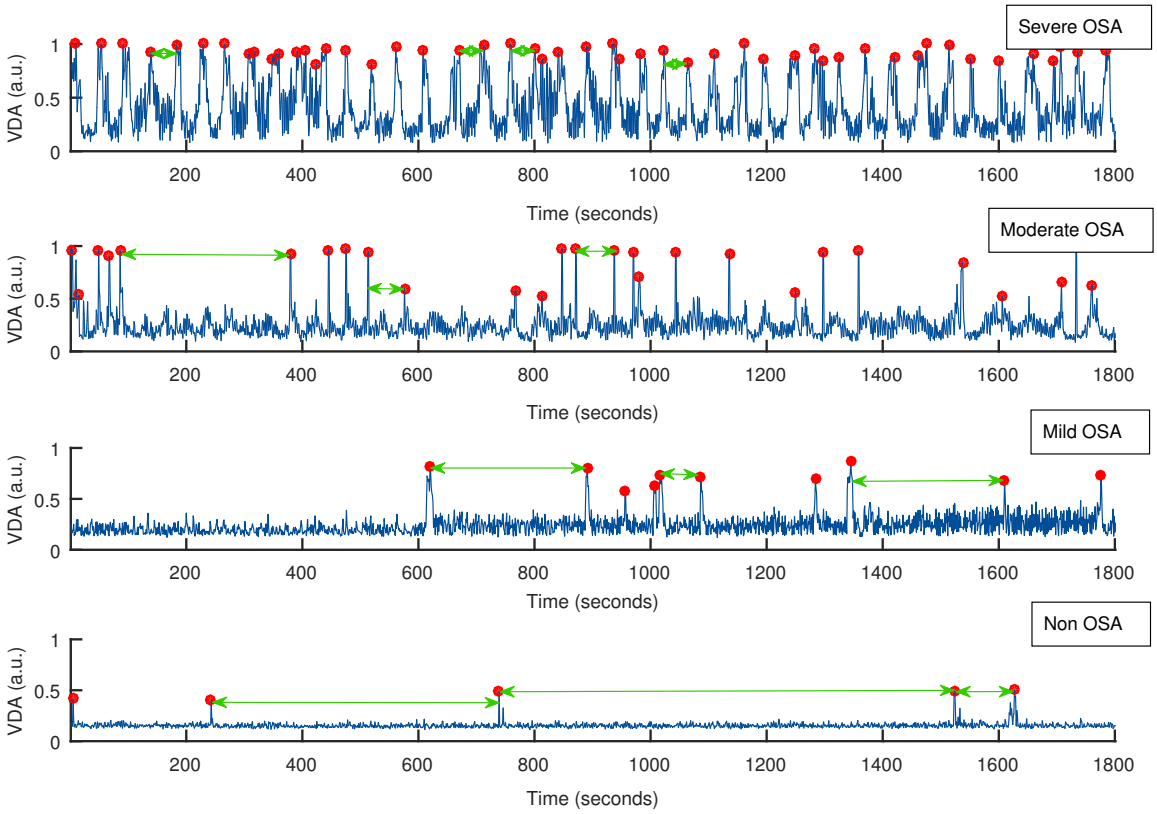


Figure 4.2: The VDA signal from four patients with different severity levels of OSA and the peaks (red dots) for which the IEI metrics were calculated. The green arrows show the distances between a few example peaks. (a.u.) = Arbitrary unit.

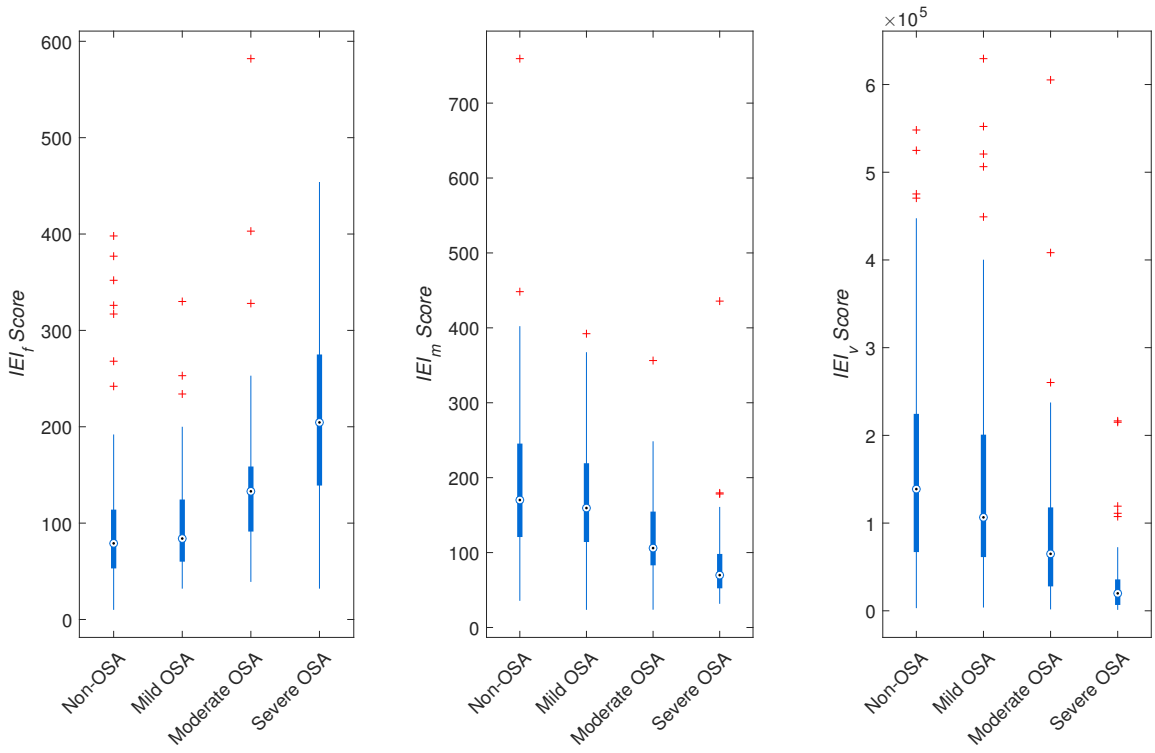


Figure 4.3: IEI scores for different OSA sveverity groups in DS1. Filled circles indicate medians, boxes represent the lower quartile and upper quartile ranges, whiskers indicate the range between sample minimum and maximum, and red pluses are outliers.

If there are C classification groups, Bayes' Rule minimises the total error of classification by assigning the object to group i which has the highest conditional probability where $P(i|\mathbf{x}) > P(j|\mathbf{x}), \forall j, i$ [107]. Using Bayes Theorem we have:

$$P(i|\mathbf{x}) = \frac{P(\mathbf{x}|i).P(i)}{\sum_{\forall k} P(\mathbf{x}|k).P(k)} \tag{4.8}$$

where $P(i)$ is the unconditional probability of class i and $P(\mathbf{x}|i)$ is the probability of $\mathbf{x}$ given that it came from class i. Therefore, based on Bayes' Rule, an object is assigned to i if,

$$\frac{P(\mathbf{x}|i).P(i)}{\sum_{\forall k} P(\mathbf{x}|k).P(k)} > \frac{P(\mathbf{x}|j).P(j)}{\sum_{\forall k} P(\mathbf{x}|k).P(k)}, \forall j \neq i \tag{4.9}$$

The denominators are positive and the same, therefore they can be cancelled out resulting in:

$$P(\mathbf{x}|i).P(i) > P(\mathbf{x}|j).P(j), \forall j \neq i. \tag{4.10}$$

It is then assumed that the data is drawn from a multivariate normal distribution whose probability distribution function is given by:

$$P(\mathbf{x}|i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathbf{\Sigma}_i|^{\frac{1}{2}}} exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)) \tag{4.11}$$

where $\boldsymbol{\mu}_i$ is the vector mean and $\mathbf{\Sigma}_i$ is the covariance matrix of group i, and d is the number of variables. Inputting the distribution formula into Bayes' Rule, cancelling factors of $(2\pi)^{\frac{d}{2}}$, taking the natural logarithm of both sides, and multiplying by -2 gives:

$$
\begin{aligned}
ln(|\boldsymbol{\Sigma}_i|) - 2ln(P(i)) + (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) < \\
ln(|\boldsymbol{\Sigma}_j|) - 2ln(P(j)) + (\mathbf{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j), \forall i \neq j.
\end{aligned} \tag{4.12}
$$

Assuming all covariance matrices are equal, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_j$, allows for further simplification. Writing $(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$ as $\mathbf{x}\boldsymbol{\Sigma}^{-1}\mathbf{x}^T - 2\boldsymbol{\mu}_i\boldsymbol{\Sigma}^{-1}\mathbf{x}^T + \boldsymbol{\mu}_i\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i^T$ means that terms can be cancelled, and multiplying both sides of the inequality by $-\frac{1}{2}$ gives rise to:

$$
f_i(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i - \frac{1}{2} \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_i + ln(P(i)) \tag{4.13}
$$

which is the linear discriminant function for the ith class. An object with measurement $\mathbf{x}$ is assigned to group i if:

$$
f_i(\mathbf{x}) > f_j(\mathbf{x}), \forall i \neq j. \tag{4.14}
$$

The reason for choosing LDA as the first classification algorithm was because it can be easily computed, works well in practice, generalises well for large data, and has no hyper-parameters to tune.

It should be noted that LDA frequently achieves good performances on the tasks of face and object recognition, even though the assumptions of common covariance matrix among groups and normality are often violated [108, 109].

4.3.4.2 Random Forests

Random Forests (RF) are a type of ensemble classifier based on decision trees and bagging. In bagging, each tree is independently constructed using a bootstrap sample of the dataset and successive trees do not depend on earlier trees. In the end, a simple majority vote is taken for prediction. The RF proposed by Breiman [110] add an additional layer of randomness to bagging. As well as using a different bootstrap sample of the data to construct each tree, RF change how the classification or regression trees are constructed. While in standard trees each node is split using the best split among all variables, in RF each node is split using the best between a subset of variables randomly chosen at that node. Having only two main parameters to tune (i.e. the number of variables in the random subset at each node, $mtry$, and the number of trees, b, in the forest) makes RF easy to train, particularly as results are not usually very sensitive to the changes to these two parameters.

An overview of the RF algorithm is as follows:

- 1 - For b=1 to B
  - a - Draw a bootstrap sample Z of size N from the training data
  - b - Grow an unpruned random forest classification or regression tree T_b on Z by repeating the following steps for each terminal node:
    - i - Select $mtry$ variables at random from the p total variables
    - ii - Select the best variable and split-point among the $mtry$ variables
    - iii - Split the node into two child nodes
- 2 - Output the ensemble of trees $\{T_b\}_{b=1}^{B}$

In the case of regression, the prediction at point x will be:

$$f_{RF}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{4.15}$$

For classification, if we assume that $C_b(x)$ is the class prediction of point x for the bth RF tree,

$$C_B(x) = majority\ vote\ of\ \{C_b(x)\}_{b=1}^{B} \tag{4.16}$$

Figure 4.4 shows a simple example of a RF with three decision trees.

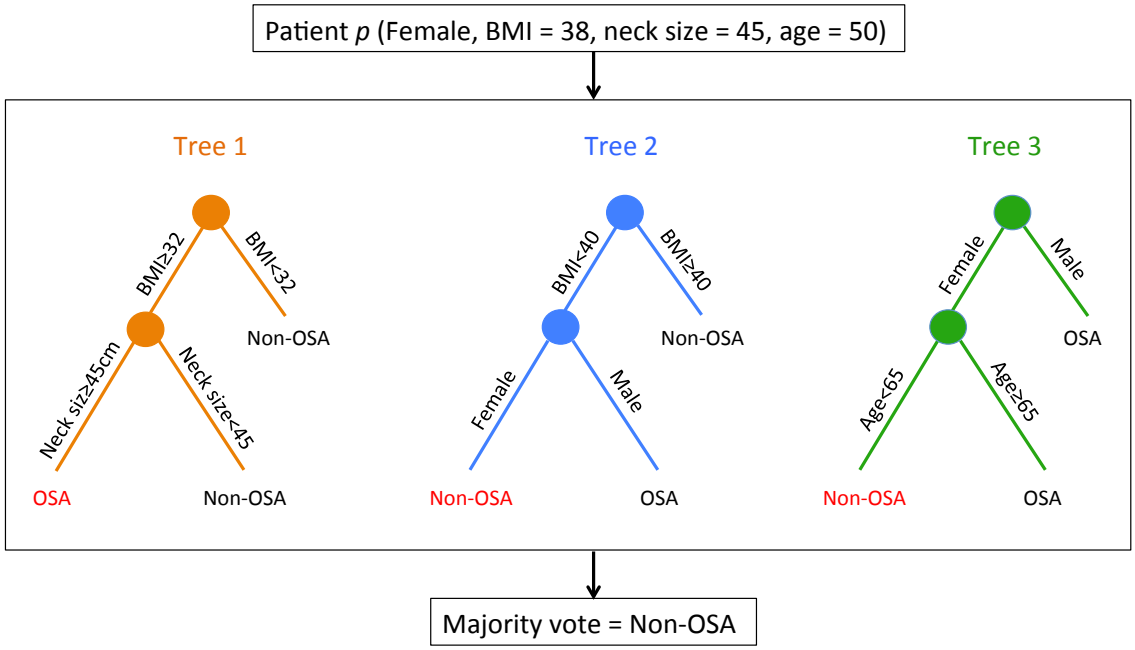


Figure 4.4: A random forest with three decision trees. The text in red is the vote of each tree.

4.3.5 Data analysis protocol

In order to derive the features for the classification of moderate-to-severe OSA patients from mild-to-no OSA patients, initially SE values were investigated for τ =

$\{1, 2, ..., 256\}$, $m = \{1, 2, ..., 8\}$ and $r = \{0.1, 0.15, 0.2, 0.25\}$ (calculated for the VDA signals). As it is illustrated in Figure 4.6, τ values that are too close produce similar SE distributions for moderate-to-severe and mild-to-no OSA patients. Additionally, Figure 4.5 shows that the coarse-grained VDA signal of a severe OSA patient seems to become too smooth (flat) for higher values of τ losing many of the awakening-related peaks. This results in relatively similar SE values for moderate-to-severe OSA and mild-to-no OSA patients. As it can be seen in figures 4.6 and 4.7, $\tau > 40$ for $m = 1$, $\tau > 30$ for $m = 2$, and $\tau > 20$ for $m = 3$, made little difference in the SE values for the moderate-to-severe OSA and mild-to-no OSA groups. Varying τ for $m > 3$ also made little difference to the SE values. Therefore, for generating the MSE features, SE was calculated for parameters $m = 1$, $r = 0.1$, $\tau = \{5, 10, 20, 30, 40\}$, $m = 2$, $r = 0.15$, $\tau = \{5, 10, 20, 30\}$, and $m = 3$, $r = 0.15$, $\tau = \{5, 10, 20\}$ slightly relaxing the similarity criterion for $m = 2$ and $m = 3$ (12 features in total). Table 4.1 presents a summary of the SE parameters used to calculate the 12 classification features and Figure 4.8 shows the differences in the values of these features for three example severe OSA, moderate OSA, and non-OSA patients.

Table 4.1: A summary of the parameters used for calculating SE from VDA signals to derive the 12 classification features.

Feature no.	m	r	τ
1	1	0.1	5
2	1	0.1	10
3	1	0.1	20
4	1	0.1	30
5	1	0.1	40
6	2	0.15	5
7	2	0.15	10
8	2	0.15	20
9	2	0.15	30
10	3	0.15	5
11	3	0.15	10
12	3	0.15	20

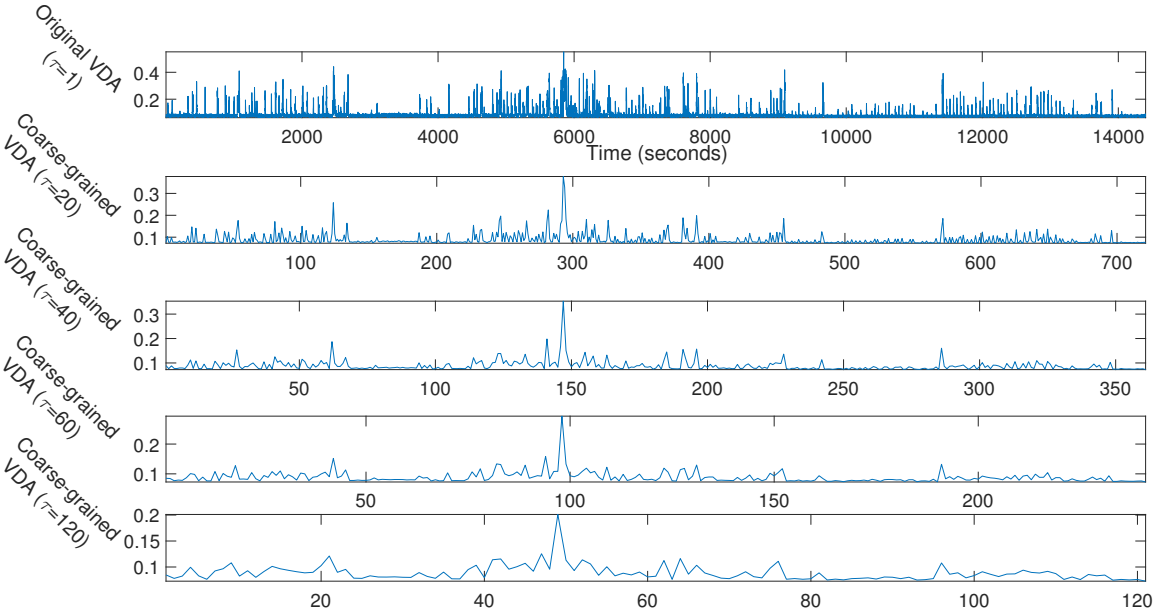


Figure 4.5: Original VDA signal and coarse-grained VDA signal of a severe OSA patient for $\tau = \{20, 40, 60, 120\}$. Every point on X axis of the original VDA signal represents one second, while for the coarse-grained signals every point on the X axis represents τ seconds.

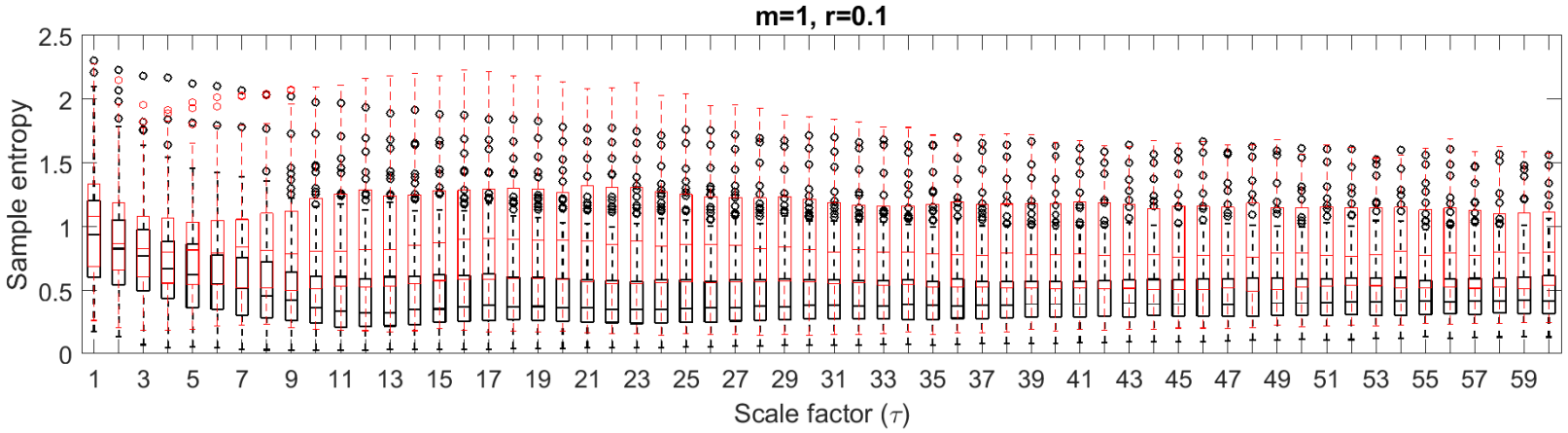


Figure 4.6: Comparison of SE values calculated from the VDA signals between the moderate-to-severe OSA (red boxes) and mild-to-no OSA (black boxes) patients (calculated for $m = 1$, $r = 0.1$, and $\tau = \{1, ..., 60\}$).

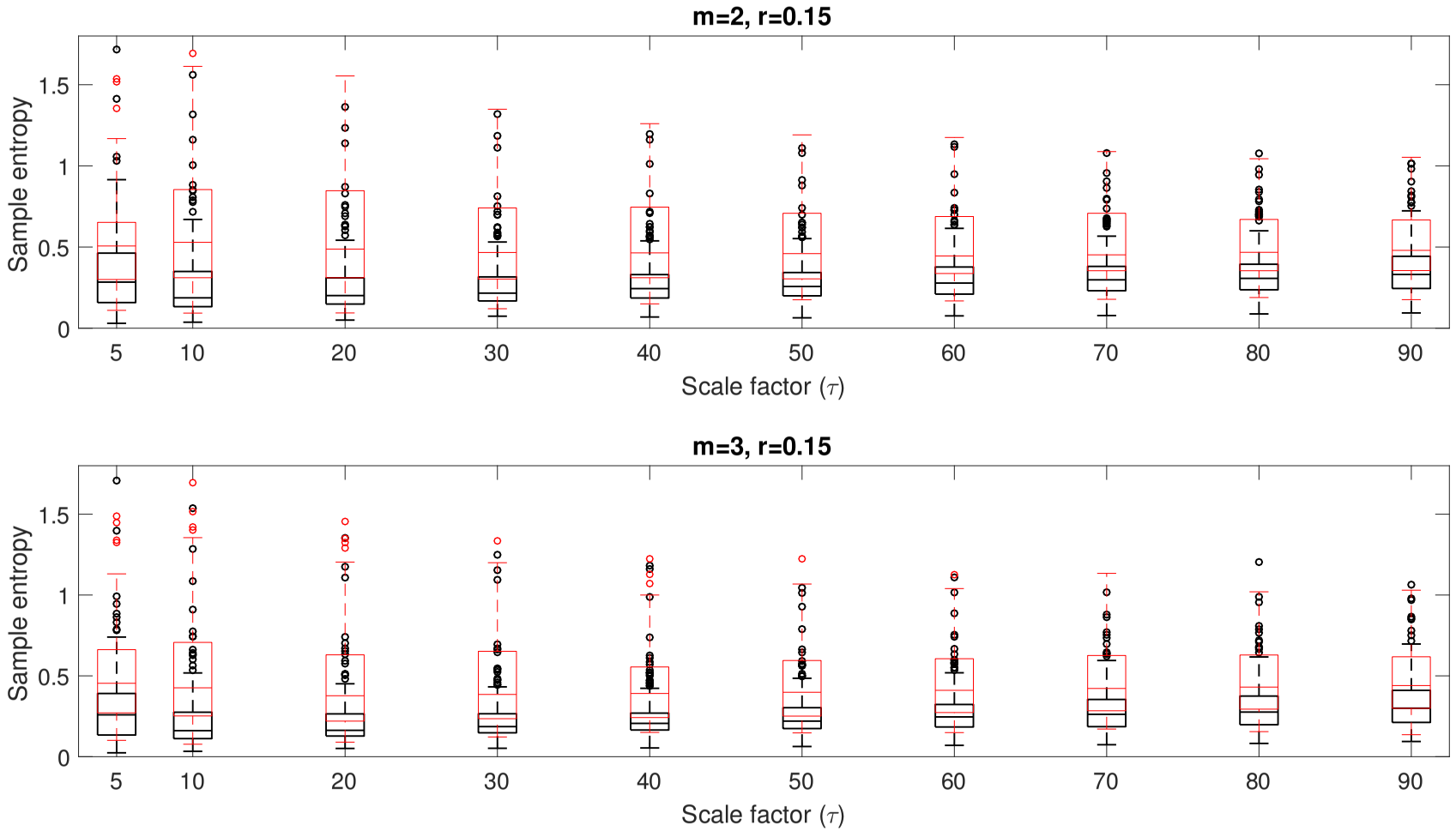


Figure 4.7: Comparison of SE values calculated from the VDA signals between the moderate-to-severe OSA (red boxes) and mild-to-no OSA (black boxes) patients (calculated for $m = 2$ and $m = 3$, $r = 0.15$, and $\tau = \{5, 10, 20, ..., 90\}$).

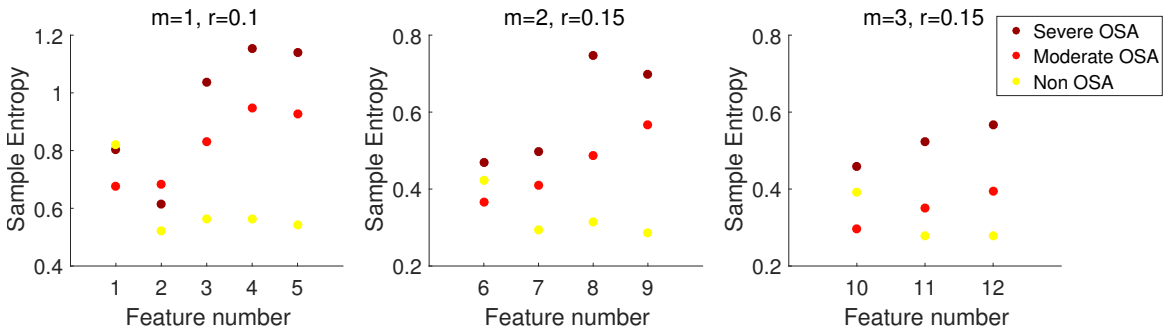


Figure 4.8: The differences in the values of the 12 selected features calculated from the VDA signals for three example severe OSA, moderate OSA, and non-OSA patients.

The LDA and RF classifiers described in 4.3.4 were used to classify moderate-to-severe OSA patients vs. mild-to-no OSA patients. In addition to the MSE and IEI features explained in section 4.3.3, patient BMI and neck size were also used as features to investigate whether any improvements in the classification performance could be achieved. The OSA severity assigned to the patients by the clinicians at the Respiratory Medicine Group (Churchill Hospital, Oxford, UK) was used as the ground truth.

The data for the 247 observations (patients) was divided into a training set (X_{Train}) and a test set (X_{Test}) using 5-fold cross validation (X_{Train} includes 80% of the data and X_{Test} includes 20% of the data). At each cross validation round, four of the five data partitions were used as X_{Train} to train the classifier and the remaining partition was used as X_{Test} to test the model. This 5-fold cross validation process was repeated 50 times each time selecting different partitions of the data. The mean and standard deviation of the classification performance for all X_{Test} partitions were used to assess the overall performance of the classifier. Figure 4.9 presents an overview of the steps of data analysis. Final patient categories were determined by taking the majority vote of the classifiers over the 50 iterations (e.g. if in more than 25 iterations a patient

was categorised as having moderate-to-severe OSA this category was chosen for the patient and the mild-to-no OSA category was chosen otherwise).

With respect to the choice of tunable parameters for the RF classifier, i.e. the number of trees and $mtry$, no improvement to the performance was achieved when the parameters were modified from their default values (500 trees and $mtry = \sqrt{numberof features}$). The number of trees was set to $\{20, 50, 150, 300, 500, 1000\}$ and $mtry$ was modified from 1 to the total number of features in increments of 1.

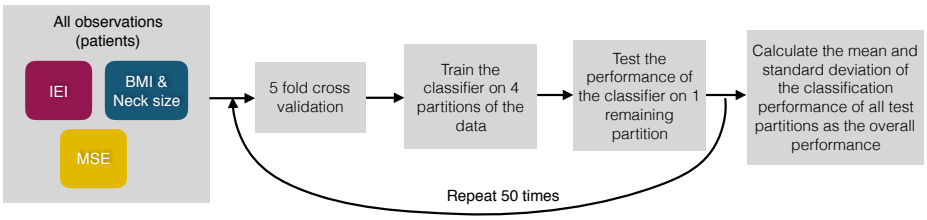


Figure 4.9: The steps of data analysis using the LDA and RF classifiers.

4.4 Results

4.4.1 LDA classifier

The LDA classification results are presented in Table 4.2. The bottom row in the table is the use of the Oxygen Desaturation Index (ODI) with a threshold of $\geqslant 15$ events/h for moderate-to-severe OSA to differentiate the same groups of patients and is provided for reference.

Figure 4.10 presents boxplots of LDA output probabilities on the test data (grouped by OSA severity) over all 5-fold cross validation rounds.

Table 4.2: The performance of the LDA classifier in differentiating between moderate-to-severe OSA patients from mild-to-no OSA patients. The classifier uses MSE and IEI features calculated from the VDA signals (MSE_{Vid} and IEI_{Vid} respectively), and the combination of each set of features with patient demographics (BMI and neck size). The performance of ODI in the same patient groups is presented at the end for comparison (using the threshold of ODI $\geqslant$ 15 events/h for moderate-to-severe OSA). For the definitions of Accuracy (Ac), Positive Predictive Value (PPV), and Area Under the Curve (AUC) see Appendix B.

Features	Ac(%)	Se(%)	Sp(%)	PPV(%)	NPV(%)	AUC(%)
MSE_{Vid} Train	83.8±0.5	66.9±0.9	93.3±0.4	84.9±0.9	83.4±0.4	80.1±0.6
MSE_{Vid} Test	81.4±5.0	61.7±11.6	92.5±4.8	82.9±9.8	81.3±4.8	77.1±6.1
IEI_{Vid} Train	74.8±0.6	69.0±1.0	78.0±1.1	64.0±1.0	81.7±0.5	73.5±0.6
IEI_{Vid} Test	74.3±5.8	68.3±12.2	77.7±7.7	63.9±8.6	81.6±5.9	73.0±6.3
MSE_{Vid} & IEI_{Vid} Train	84.3±0.5	68.5±1.0	93.1±0.5	85.0±1.0	84.0±0.4	80.8±0.6
MSE_{Vid} & IEI_{Vid} Test	80.8±5.1	61.5±11.9	91.7±5.1	81.4±9.8	81.1±5.0	76.6±6.2
BMI & $Necksize$ Train	68.6±1.4	59.9±0.9	73.5±2.5	56.3±1.7	76.5±0.6	66.7±1.0
BMI & $Necksize$ Test	68.4±6.3	59.7±12.5	73.4±9.3	56.5±9.3	76.6±5.7	66.5±6.5
MSE_{Vid} & BMI & $Necksize$ Train	85.6±0.4	71.4±1.0	93.6±0.5	86.4±1.0	85.3±0.4	82.5±0.5
MSE_{Vid} & BMI & $Necksize$ Test	82.4±5.0	66.4±11.1	91.3±5.2	81.8±9.5	83.0±4.9	78.9±5.9
IEI_{Vid} & BMI & $Necksize$ Train	75.8±0.6	71.0±0.9	78.5±0.7	65.0±0.8	82.8±0.5	74.7±0.7
IEI_{Vid} & BMI & $Necksize$ Test	74.7±6.2	70.1±11.8	77.3±7.4	63.9±8.8	82.4±6.2	73.7±6.7
MSE_{Vid} & IEI_{Vid} & BMI & $Necksize$ Train	85.8±0.5	72.6±0.9	93.3±0.6	86.0±1.1	85.8±0.5	83.0±0.6
MSE_{Vid} & IEI_{Vid} & BMI & $Necksize$ Test	82.1±5.1	67.1±10.9	90.5±5.4	80.6±9.3	83.2±4.9	78.8±5.9
ODI (15 events/h threshold) Train	93.5±0.3	96.6±0.3	91.8±0.5	86.9±0.7	98.0±0.2	94.2±0.3
ODI (15 events/h threshold) Test	93.5±3.4	96.6±4.1	91.8±4.9	87.3±6.7	98.0±2.3	94.2±3.1

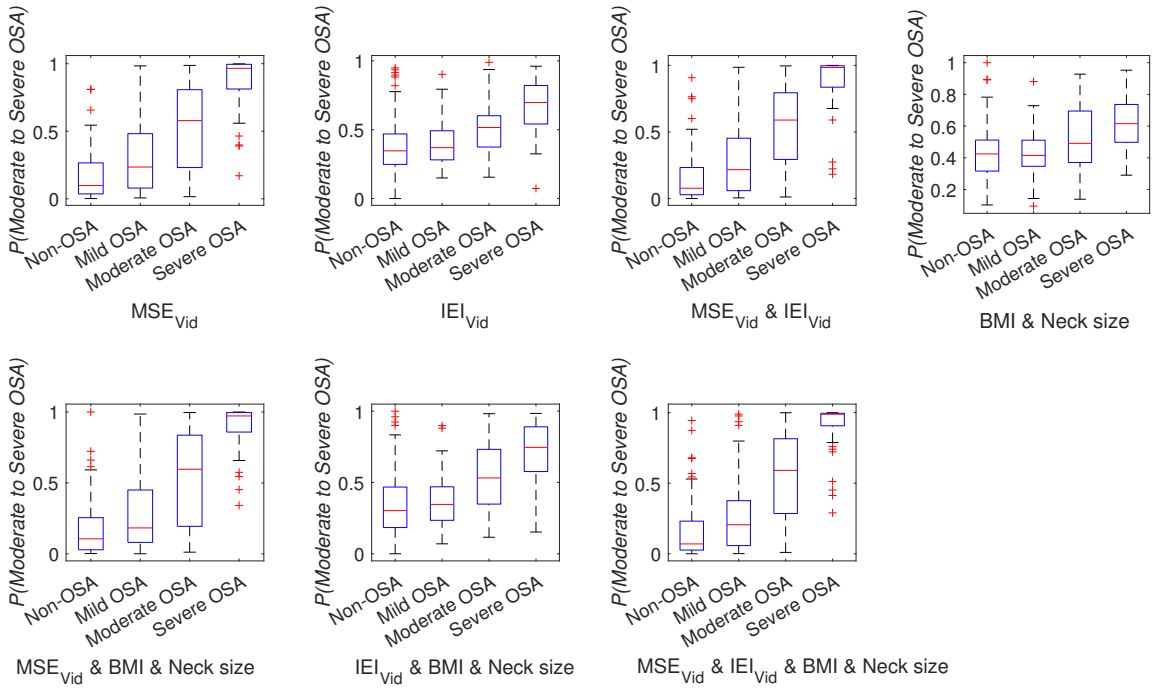


Figure 4.10: Boxplots of LDA probabilities on the test data grouped by severities over all 5-fold cross validation steps.

Table 4.3: The performance of the RF classifier differentiating moderate-to-severe OSA patients from mild-to-no OSA patients. The classifier uses MSE and IEI features calculated from the VDA signals (MSE_{Vid} and IEI_{Vid} respectively), and the combination of both set of features with demographics (BMI and neck size). The RF parameters $mtry_{best}$ and $ntree_{best}$ were selected in the validation step and used to build the final classifier. The performance of ODI on differentiating the same patient groups is presented at the end for comparison (using the threshold of ODI $\geqslant$ 15 events/h for moderate-to-severe OSA).

Features	Ac(%)	Se(%)	Sp(%)	PPV(%)	NPV(%)	AUC(%)
MSE_{Vid} Test	74.9±5.6	60.0±11.7	83.3±7.6	67.9±10.2	78.9±5.1	71.7±6.0
IEI_{Vid} Test	71.2±6.0	55.5±11.8	80.2±7.7	61.9±10.7	76.3±5.3	67.8±6.5
MSE_{Vid} & IEI_{Vid} Test	74.5±5.6	61.5±11.3	81.7±7.1	66.1±9.7	79.3±5.2	71.6±6.1
BMI & $Necksize$ Test	62.4±5.7	45.5±10.9	72.1±7.8	48.1±8.9	70.2±5.0	58.8±5.9
MSE_{Vid} & BMI & $Necksize$ Test	76.0±5.2	61.8±11.5	83.9±6.6	69.0±9.2	79.8±4.9	72.9±5.9
IEI_{Vid} & BMI & $Necksize$ Test	74.2±5.5	58.3±12.6	83.2±7.0	66.8±9.9	78.3±5.5	70.8±6.2
MSE_{Vid} & IEI_{Vid} & BMI & $Necksize$ Test	77.2±5.5	63.3±11.7	85.0±6.3	70.9±9.4	80.7±5.2	74.1±6.2
ODI (15 events/h threshold) Test	93.5±3.4	96.6±4.1	91.8±4.9	87.3±6.7	98.0±2.3	94.2±3.1

4.4.2 RF classifier

The RF classification results are presented in Table 4.3. The bottom row in the table is the use of the ODI with a threshold of $\geqslant$ 15 events/h for moderate-to-severe OSA to differentiate the same groups of patients and is provided for reference. As RF produce fully grown trees on the training data, the training error which is usually very small (near zero) is uninformative and hence this error is not included.

Figure 4.11 presents boxplots of RF output probabilities on the test data (grouped by OSA severity) over all 5-fold cross validation rounds.

4.5 Discussion

The investigation into the use of video-derived activity for differentiating moderate-to-severe OSA from mild-to-no OSA patients demonstrates that video recordings have the potential to be used as a first step screening tool. The classifiers were evaluated

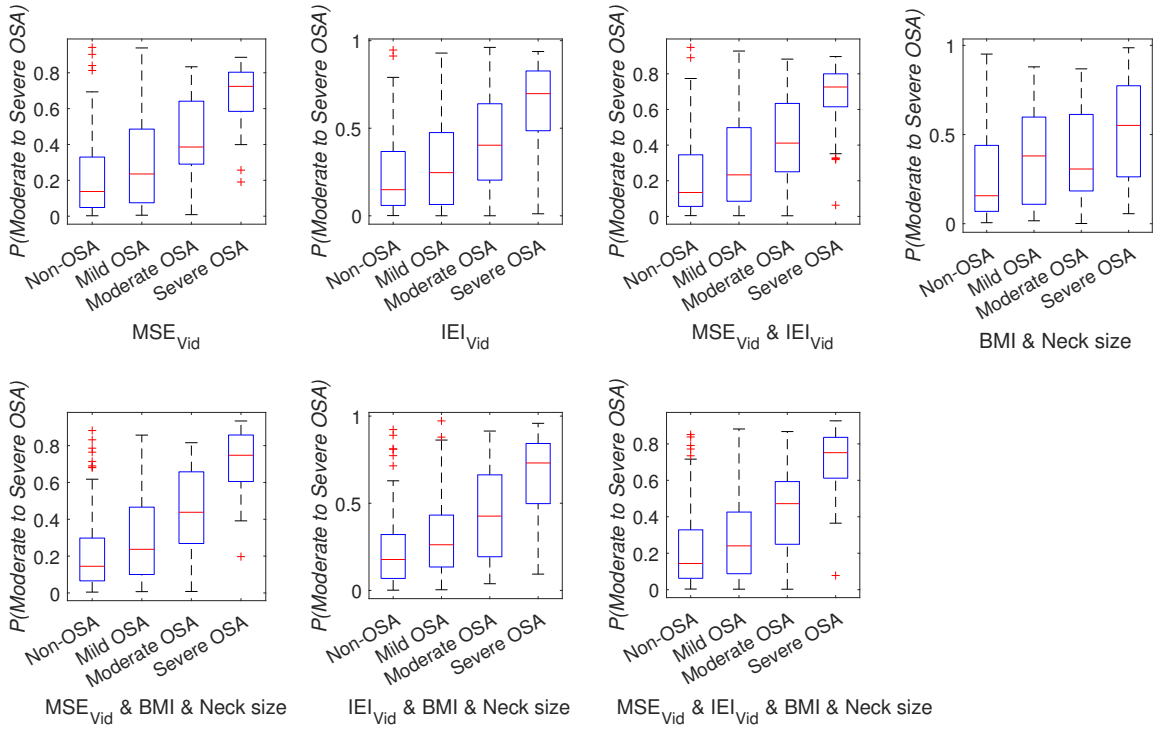


Figure 4.11: Boxplots of RF probabilities on the test data grouped by severities over all 5-fold cross validation steps.

on real PSG video data from patients referred to the Churchill Hospital (Oxford, UK) with suspected OSA. Using MSE features extracted from full-body video activity with an LDA classifier, moderate-to-severe OSA patients were differentiated from mild-to-no OSA patients with an accuracy of approximately 81%, sensitivity of 62%, and specificity of 93% (second row of Table 4.2). When patient BMI and neck size were added to the MSE features, the accuracy and sensitivity increased to 82% and 66% respectively but the specificity decreased to 91%. Adding the IEI features to MSE features did not result in any improved performance. This may be due to the fact that both types of features contain similar information about the VDA signal and are not independent of each other.

Using the IEI features alone with the LDA classifier, resulted in 74% accuracy on the test set. The addition of patient BMI and neck size to the IEI features did not

result in any noticeable improvement to classification accuracy.

For the MSE features, increasing τ to values higher than 40 made little difference in the SE distributions of moderate-to-severe OSA patients and mild-to-no OSA patients. This may be due to the fact that the average duration of apnoeic events is between 20 to 40 seconds [20] and a coarse-grained representation of the signals using τ values above the average event duration smooths the awakening-related peaks.

LDA had a higher classification performance than the RF classifier. This may be due to the fact that LDA builds a model better able to classify the test data compared to the RF classifier without over-fitting the training data.

For the misclassified moderate-to-severe OSA patients, the main reason for misclassification tended to be having a borderline moderate OSA, an ODI of lower than 15, or OSA with no awakening movements (patients resumed breathing with no movement after apnoea). For mild-to-no OSA patients, having other sleep complications such as parasomnia, Cheyne-Stokes respiration[4], obstructive snoring, or ODI of higher than 15 seem to the main reasons for misclassification.

Finally, using the ability of the ODI to differentiate between moderate-to-severe OSA and mild-to-no OSA patients as the reference for assessing the classifiers may not be a fair test. This is because the ODI values have been used by the clinicians for their diagnosis. Furthermore, the patients were not monitored in their natural sleep environment which may have caused excessive disturbed sleep and abnormal patterns in their VDA signal in comparison to a normal night.

[4]A form of breathing with a cyclical variation in the rate and/or depth. In Cheyne-Stokes respiration, the breathing becomes slower and/or shallower or until breathing stops for several seconds before the rate and/or depth increases to a peak and then decreases again. It occurs when the sensitivity of the respiratory centres in the brain is impaired [14].

Chapter 5

Combination of video derived activity and audio to detect moderate to severe OSA

5.1 Introduction

In this chapter, we extend the work of the previous chapter and investigate using the combination of audio and video recordings of sleep to improve the differentiation between moderate-to-severe OSA patients and mild-to-no OSA patients. Since most cameras are equipped with a microphone and include both sound and video recording, the addition of information from audio to improve the accuracy of OSA analyses in the previous chapter is a natural next step. Audio recordings include important apnoea related features such as the patient's snoring, deep breaths, and choking sounds [47]. The use of audio recordings alone for the detection of OSA has been studied previously

by Roebuck [111]. Roebuck's findings show that MSE analysis of audio recordings outperform other methods of audio analysis in differentiating between moderate-to-severe OSA and mild-to-no OSA patients. The work in this chapter is based on the MSE analysis in the previous chapter and Roebuck's approach. Using MSE analysis, OSA related features are extracted from audio recordings and used in combination with features extracted from VDA signals to detect moderate-to-severe OSA.

5.2 Materials and methods

5.2.1 Data

The video data used in this chapter is DS1 (see Section 3.4) which includes in-laboratory PSG recordings of patients who were referred to the sleep clinic of Churchill Hospital (Oxford, UK) with suspected OSA. The PSG recordings include a number of physiological signals as well as audio and video recordings. The videos in DS1 were recorded in grey scale in MPEG2 format with a resolution of 352×288 pixels, and a frame rate of 25 fps. The audio embedded in the video recordings was sampled at a frequency of 44 kHz.

5.2.2 Data preprocessing

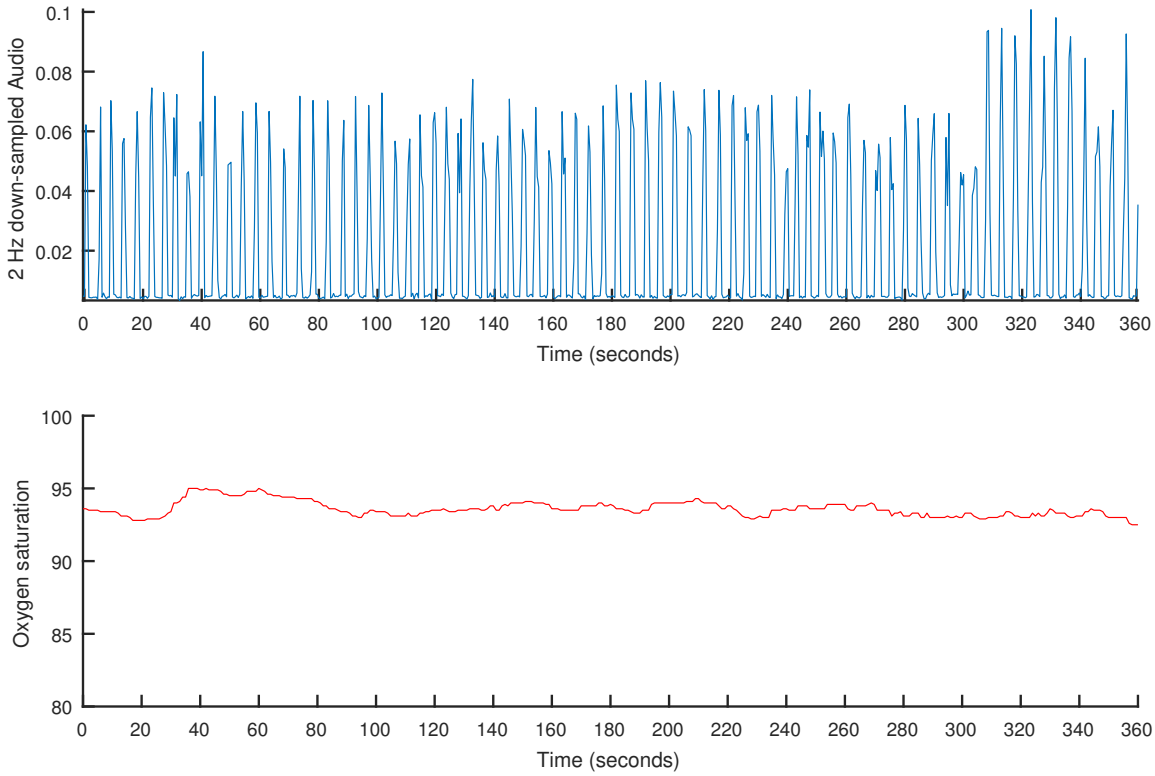
The data was analysed from 2 hours to 6 hours into the recording to increase the likelihood of the patient being asleep during the study period. The audio recordings were extracted from video files to flac format[1] with a 44 kHz sampling frequency.

[1]FLAC (Free Lossless Audio Codec) is a lossless audio format. The audio is compressed in FLAC format without any loss in quality [112].

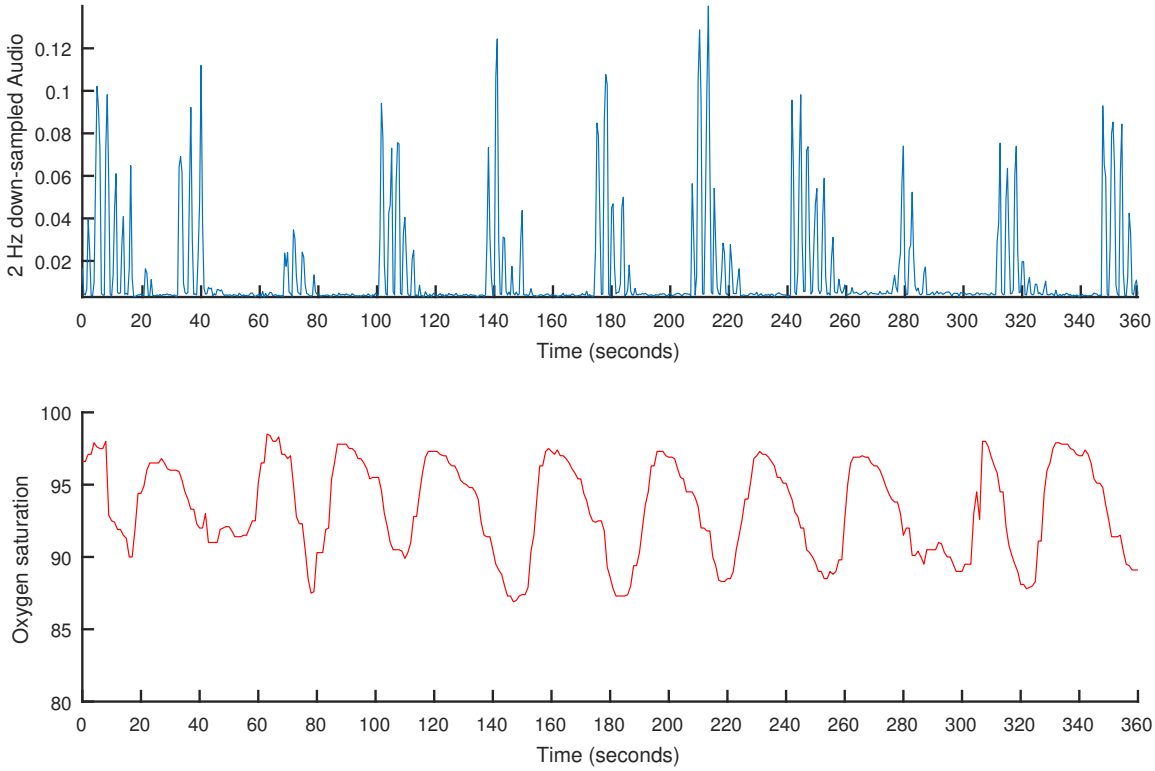
Subsequently, the audio signals were down-sampled to 2 Hz by taking the maximum of every 0.5 second of the signal. Since awakening choking sounds may happen in a sub-second period, using the maximum value instead of the usual averaging ensures that large fast peaks are included in the down-sampled signal. Additionally, as snoring mainly happens during inhalation, it has a frequency similar to respiration. The 2 Hz sampling frequency is sufficiently high to keep the fastest respiratory/snoring frequency which is considered to be 0.6 Hz (36 bpm) here.

5.2.3 Feature extraction

The audio recordings of OSA sufferers include three apnoea related patterns; (i) a majority of OSA patients snore during their sleep, therefore outside periods of apnoea, the audio signal, as illustrated in Figure 5.1a, has a repetitive pattern with the same frequency as breathing; (ii) as the airway is completely obstructed during apnoea and partially obstructed during hypopnoea, the amplitude of the audio signal becomes close to zero or is reduced during OSA events; (iii) the first breath after apnoea includes a deep breath often accompanied with a choking/snort sound which causes a spike in the audio signal. With an increase in the severity of OSA, the number of OSA events per hour of sleep is increased. Therefore, the reductions in the amplitude of the signal followed by the awakening spikes become a regular pattern in the audio recordings of the moderate-to-severe OSA patients. An example of a series of apnoeic events and awakenings is presented in Figure 5.1b. In order to derive features indicative of OSA from the audio recordings, multi-scale entropy analysis was used as a measure of regularity for time series data.



(a) A period of snoring which causes a repetitive pattern in the 2 Hz down-sampled audio signal.



(b) A series of apnoeic events presented with a reduced amplitude followed by awakenings causing a spike in the 2 Hz down-sampled audio signal.

Figure 5.1: Snoring and apnoeic events in audio recordings.

5.2.4 Data analysis protocol

In order to derive the features for the classification of moderate-to-severe OSA patients from mild-to-no OSA patients, initially SE values were calculated for the 2 Hz audio signals for $\tau = \{1, 2, ..., 256\}$, $m = \{1, 2, ..., 8\}$ and $r = \{0.1, 0.15, 0.2, 0.25\}$. As it is illustrated in Figure 5.2, for τ from 1 to 10, the change in SE values between moderate-to-severe OSA group and mild-to-no OSA group is noticeable. However, τ values above 10 that are too close produce similar SE distributions for the moderate-to-severe and mild-to-no OSA patient groups.

Figure 5.3 shows that the coarse-grained audio signal of a severe OSA patient becomes smooth (flat) for higher values of τ, losing patterns of snoring for $\tau > 5$ and losing many of the awakening-related peaks for $\tau > 40$.

As shown in Figure 5.2, for $m = 2$ only $\tau = 1$, $\tau = 2$, and around $\tau = 25$ generate SE values with a good separation between moderate-to-severe OSA and mild-to-no OSA groups. Varying τ for $m \geqslant 3$ made little difference in SE values and did not produce a better separation between the two patient groups. Therefore, to generate the MSE features, SE was calculated for parameters $m = 1$, $r = 0.1$, $\tau = \{1, 2, ..., 9\}$ and $\tau = \{10, 20, 30, 40\}$, $m = 2$, $r = 0.15$, $\tau = \{1, 2, 25\}$ slightly relaxing the similarity criterion for $m = 2$ (16 features in total). Table 5.1 presents a summary of the SE parameters used to calculate the 16 classification features and Figure 5.4 shows the differences in the values of these 16 features for four example severe OSA, moderate OSA, snorer non-OSA, and non-snorer non-OSA patients.

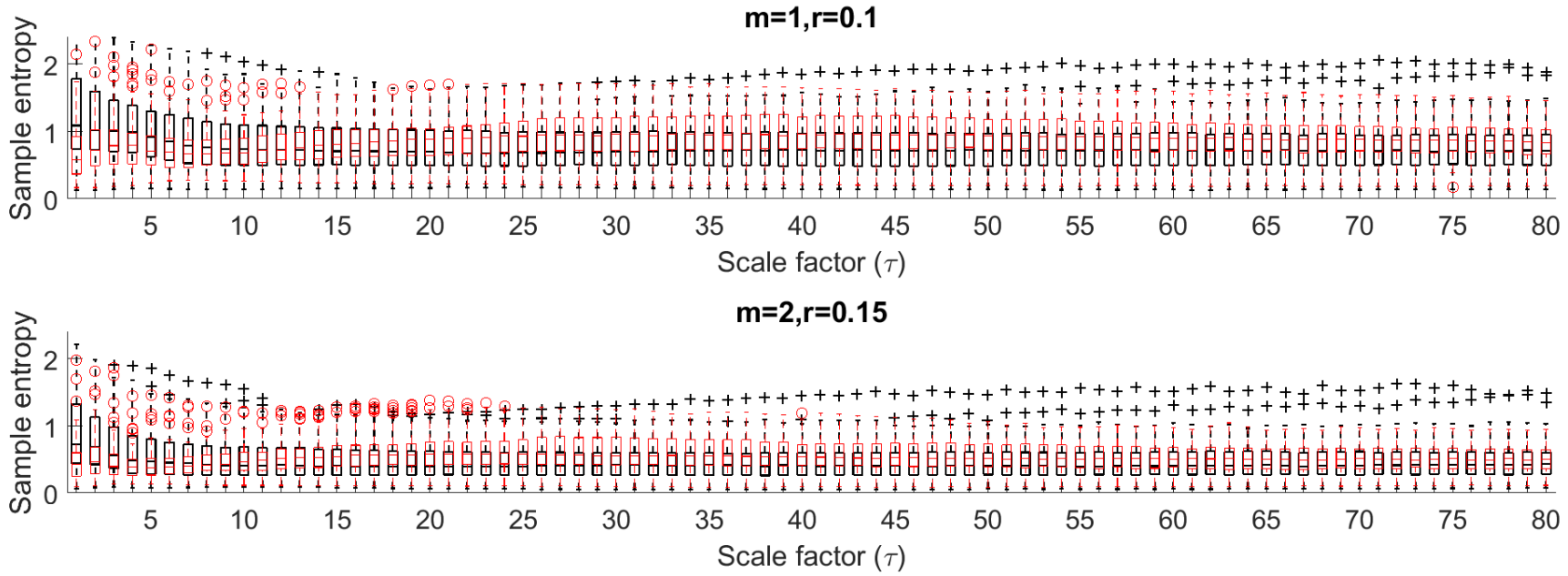


Figure 5.2: Boxplots showing SE values (calculated for the 2 Hz down-sampled audio signals) for moderate-to-severe OSA patients (red boxes) and mild-to-no OSA patients (black boxes) for values of $\{m = 1, r = 0.1\}$, $\{m = 2, r = 0.15\}$, and τ between 1 to 60.

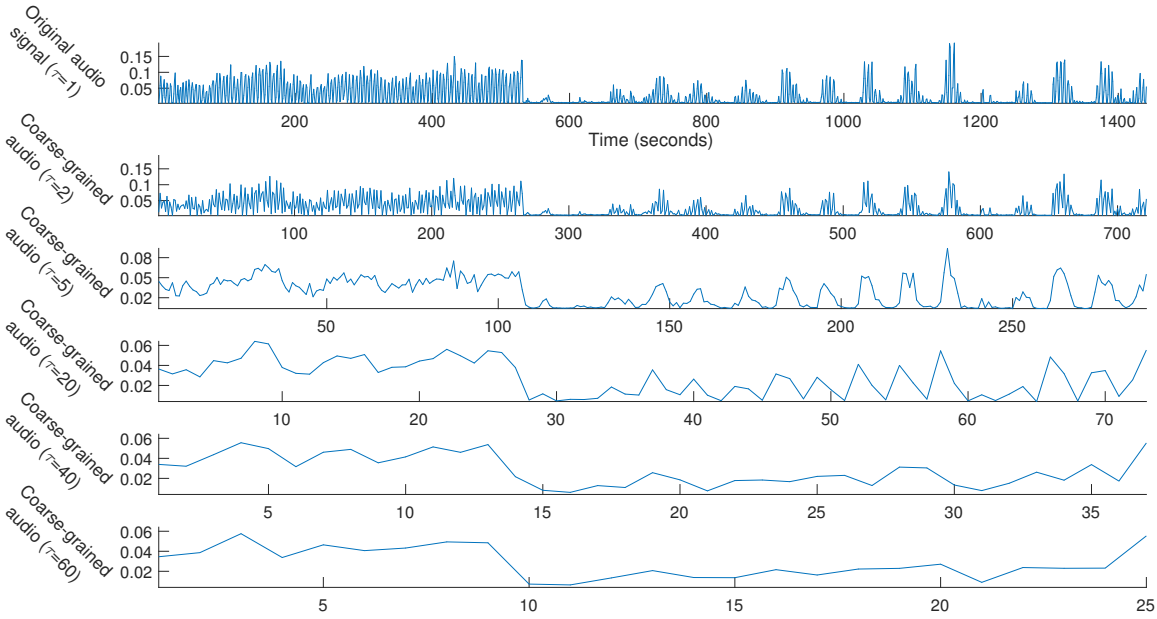


Figure 5.3: Original 2 Hz down-sampled audio signal and the corresponding coarse-grained signals of a severe OSA patient for $\tau = \{2, 5, 20, 40, 60\}$. Every point on X axis of the original 2 Hz down-sampled audio signal represents 0.5 seconds, while for the coarse-grained signals every point on the X axis represents $\frac{\tau}{2}$ seconds.

Once the MSE features were calculated from the audio recordings, the LDA and RF classifiers introduced in Section 4.3.4 were used to classify moderate-to-severe OSA patients versus mild-to-no OSA patients. The audio MSE features were then combined with video MSE features from Chapter 4 as well as patient BMI and neck size to investigate whether any improvements in the classification performance could be achieved.

As with Chapter 4, the data for the 247 observations (patients) was divided into training sets (X_{Train}) and test sets (X_{Test}) using 5-fold cross validation. At each cross validation round, four of the five data partitions were used as X_{Train} to train the classifier and the remaining partition was used as X_{Test} to test the model. This 5-fold cross validation process was repeated 50 times, each time selecting different partitions of the data. The mean and standard deviation of the classification performance on

Table 5.1: A summary of the parameters used for calculating SE from down-sampled audio signals to derive the 16 classification features.

Feature no.	m	r	τ
1	1	0.1	1
2	1	0.1	2
3	1	0.1	3
4	1	0.1	4
5	1	0.1	5
6	1	0.1	6
7	1	0.1	7
8	1	0.1	8
9	1	0.1	9
10	1	0.1	10
11	1	0.1	20
12	1	0.1	30
13	1	0.1	40
14	2	0.15	1
15	2	0.15	2
16	2	0.15	25

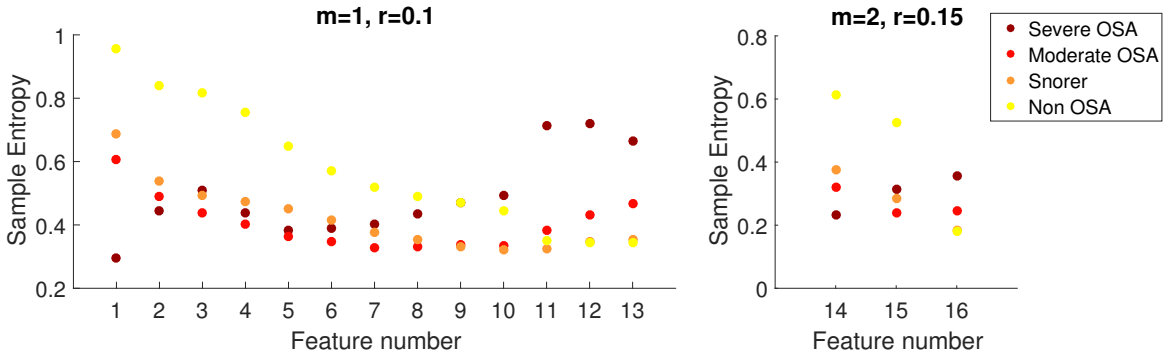


Figure 5.4: The differences in the values of the 16 selected features calculated from audio recordings for four example severe OSA, moderate OSA, snorer non-OSA, and non-snorer non-OSA patients.

all the X_{Test} partitions were used to assess the overall performance of the classifier. Final patient categories were determined by taking the majority vote of the classifiers over the 50 iterations.

With regards to the choice of tunable parameters for the RF classifier, i.e. the number of trees and $mtry$, no improvement to the performance was achieved when the parameters were varied from their default values (500 trees and $mtry = \sqrt{numberoffeatures}$). The number of trees was set to $\{20, 50, 150, 300, 500, 1000\}$ and $mtry$ was modified from 1 to the total number of features in increments of 1.

5.3 Results

In this section the performance of the proposed classifiers is assessed on real patient data. As mentioned earlier, the classifiers used the SE values calculated for the down-sampled audio signals to classify moderate-to-severe OSA patients and mild-to-no OSA patients. The OSA severity assigned to the patients by the physicians at the Respiratory Medicine Group (Churchill hospital, Oxford, UK) was used as the ground truth.

5.3.1 LDA classifier

The LDA classification results are presented in Table 5.2. The bottom row in the table is the performance of ODI (using the threshold of $\geqslant 15$ events/h for moderate-to-severe OSA) to differentiate the same groups of patients and is provided for reference.

Table 5.2: LDA classification performance using MSE features calculated from audio recordings (MSE_{Aud}), and the combination of audio MSE features, with VDA MSE features (MSE_{Vid}, derived in Chapter 4) and patient demographics (BMI and neck size). The classification performance of ODI is presented at the end for comparison.

Features	Ac(%)	Se(%)	Sp(%)	PPV(%)	NPV(%)	AUC(%)
MSE_{Aud} Train	83.3±0.5	73.7±0.8	88.7±0.6	78.6±0.9	85.7±0.4	81.2±0.5
MSE_{Aud} Test	80.0±5.3	70.5±10.7	85.5±6.6	73.9±9.2	83.9±5.4	78.0±5.8
MSE_{Vid} Train	83.8±0.5	66.9±0.9	93.3±0.4	84.9±0.9	83.4±0.4	80.1±0.6
MSE_{Vid} Test	81.4±5.0	61.7±11.6	92.5±4.8	82.9±9.8	81.3±4.8	77.1±6.1
MSE_{Aud} & MSE_{Vid} Train	89.0±0.5	80.5±0.9	93.8±0.5	88.0±0.9	89.5±0.5	87.2±0.6
MSE_{Aud} & MSE_{Vid} Test	84.5±4.8	74.3±10.0	90.2±5.0	81.5±8.1	86.3±4.9	82.2±5.5
MSE_{Aud} & BMI & $Necksize$ Train	85.4±0.5	76.4±0.9	90.4±0.6	81.8±1.0	87.2±0.4	83.4±0.6
MSE_{Aud} & BMI & $Necksize$ Test	81.7±5.0	73.4±10.4	86.5±5.9	75.9±8.4	85.4±5.3	79.9±5.6
MSE_{Vid} & BMI & $Necksize$ Train	85.6±0.4	71.4±1.0	93.6±0.5	86.4±1.0	85.3±0.4	82.5±0.5
MSE_{Vid} & BMI & $Necksize$ Test	82.4±5.0	66.4±11.1	91.3±5.2	81.8±9.5	83.0±4.9	78.9±5.9
MSE_{Aud} & MSE_{Vid} & BMI & $Necksize$ Train	89.8±0.4	81.3±0.9	94.7±0.4	89.6±0.7	90.0±0.5	88.0±0.5
MSE_{Aud} & MSE_{Vid} & BMI & $Necksize$ Test	85.2±4.5	75.1±9.8	90.9±4.7	82.8±7.8	86.8±4.8	83.0±5.2
ODI (15 events/h threshold) Train	93.5±0.3	96.6±0.3	91.8±0.5	86.9±0.7	98.0±0.2	94.2±0.3
ODI (15 events/h threshold) Test	93.5±3.4	96.6±4.1	91.8±4.9	87.3±6.7	98.0±2.3	94.2±3.1

Figure 5.5 presents boxplots of LDA output probabilities on the test data (grouped by OSA severity) over all 5-fold cross validation rounds.

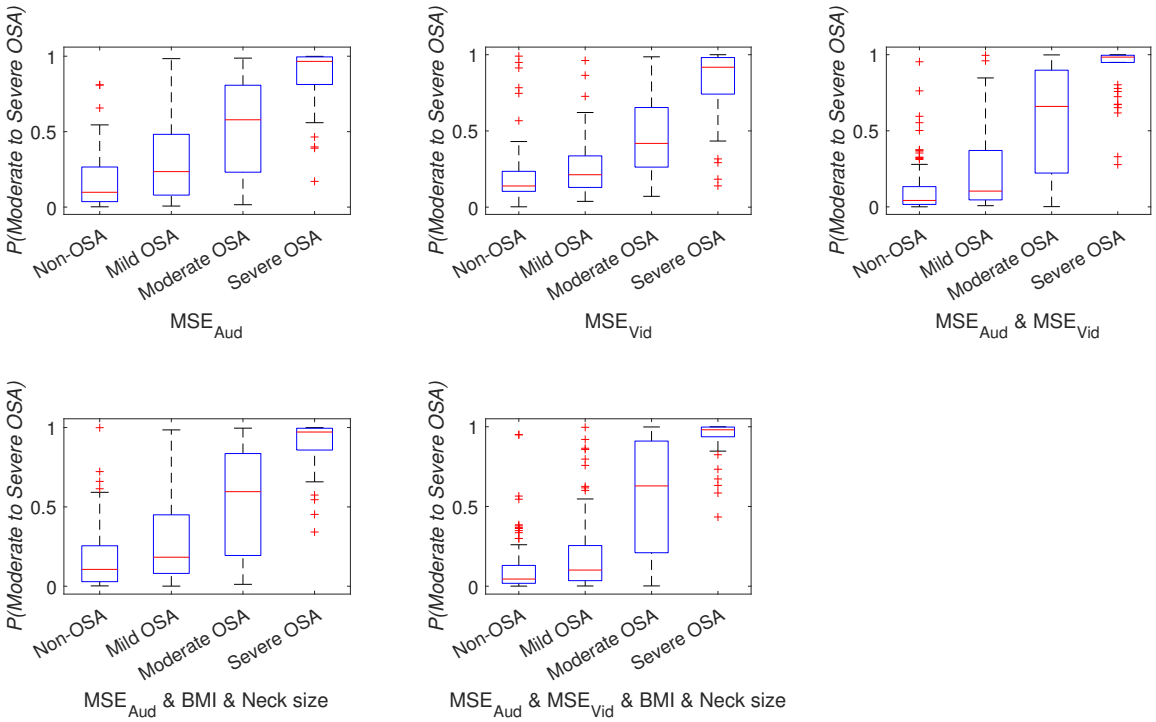


Figure 5.5: Boxplots of LDA probabilities on the test data grouped by severities over all 5-fold cross validation steps.

5.3.2 RF classifier

The RF classification results are presented in Table 5.3. The bottom row in the table is the performance of ODI (using the threshold of $\geqslant$ 15 events/h for moderate-to-severe OSA) to differentiate the same groups of patients and is provided for reference. As RF produce fully-grown trees on the training data, the training error is usually very small and not reported here.

Table 5.3: RF classification performance using MSE features calculated from audio recordings (MSE_{Aud}), and the combination of audio MSE features, with VDA MSE features (MSE_{Vid}, derived in Chapter 4) and patient demographics (BMI and neck size). The classification performance of ODI is presented at the end for comparison.

Features	Ac(%)	Se(%)	Sp(%)	PPV(%)	NPV(%)	AUC(%)
MSE_{Aud} Test	78.3±5.9	63.1±12.3	87.0±6.8	74.2±10.7	80.9±6.1	75.1±6.6
MSE_{Vid} Test	74.9±5.6	60.0±11.7	83.3±7.6	67.9±10.2	78.9±5.1	71.7±6.0
MSE_{Aud} & MSE_{Vid} Test	82.0±5.4	68.6±10.6	89.6±5.9	79.6±10.0	83.6±5.2	79.1±6.1
MSE_{Aud} & BMI & $Necksize$ Test	79.5±5.2	64.8±11.2	88.0±6.5	76.1±9.9	81.8±5.5	76.4±5.8
MSE_{Vid} & BMI & $Necksize$ Test	76.0±5.2	61.8±11.5	83.9±6.6	69.0±9.2	79.8±4.9	72.9±5.9
MSE_{Aud} & MSE_{Vid} & BMI & $Necksize$ Test	82.9±5.2	70.1±10.0	90.1±5.8	80.7±9.8	84.3±5.0	80.1±5.7
ODI (15 events/h threshold) Test	93.5±3.4	96.6±4.1	91.8±4.9	87.3±6.7	98.0±2.3	94.2±3.1

Figure 5.6 presents boxplots of RF output probabilities on the test data (grouped by OSA severity) over all 5-fold cross validation rounds.

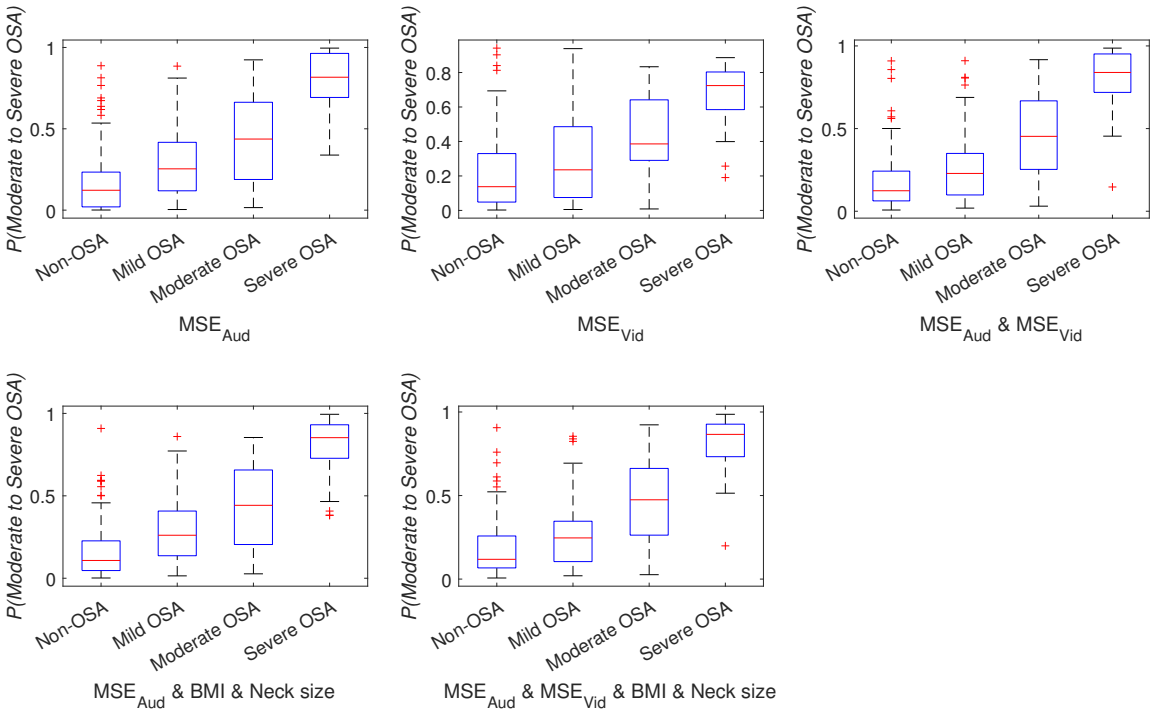


Figure 5.6: Boxplots of RF probabilities on the test data grouped by severities over all 5-fold cross validation steps.

5.4 Discussion

This chapter investigated combining OSA-related features derived from video recordings with features derived from audio recordings by using both sets of features together for the classification of moderate-to-severe OSA patients from mild-to-no OSA patients. In particular, the regularity of sleep-wake periods from the VDA signals was combined with the regularity of snoring, silence due to apnoea, and choking and resumption of snoring due to awakening from audio recordings. The combination of video features and audio features increased the classification accuracy by above 3% (Ac of 84.5%) compared to when video activity features or audio features are used on their own. When patient BMI and neck size were added to the video activity and audio features, the accuracy improved to 85.2%.

The LDA classifier achieved a better performance on the combination of video activity and audio features than the RF classifier. This is in agreement with the results reported in Chapter 4 and may be due to the fact that the LDA classifier builds a model with a better performance on the test data compared to the RF classifier which may over-fit the train data.

The classification accuracy, despite being lower compared to the accuracy of ODI derived from pulse oximetry, is achieved using a contact-less device without restrictions on the recording environment, bedding, clothing, etc. Even the patients in the dataset with no OSA were referred to the clinic with suspected OSA and some were diagnosed with other sleep disturbances making them difficult to differentiate from OSA patients. The specificity is slightly above that obtained with ODI, however, sensitivity is lower compared to ODI. The lower number of patients in the moderate-to-severe OSA group versus the number of patients in the mild-to-no OSA group may contribute to the lower sensitivity.

As the boxplots of the sample entropy values for audio recordings show in Figure 5.2, when the audio signal is averaged over τ values below 5, the repetitive snoring pattern causes low SE values for moderate-to-severe OSA patients compared to mild-to-no OSA patients. As the value of τ increases to above average respiratory rate frequencies (12-18 bpm [113] or $\tau > 10$), the coarse-grained signal loses the pattern of snoring and the periods of apnoea/awakenings become dominant in the signal. Apnoea and awakening peaks in the signal cause a higher SE value for moderate-to-severe patients compared to mild-to-no OSA patients. Increasing τ to values higher than 80 decreased the separation between SE distributions of moderate-to-severe OSA

patients and mild-to-no OSA patients. As explained in section 4.5, this may be because a coarse-grained representation for the signals using values above the average duration of apnoeic events, which is around 40 seconds, smooths the awakening-related peaks.

Overall, as the audio recordings contain information about snoring as well as apnoea and awakening sounds and the VDA signals include information about apnoea and awakenings movements, the two types of signals have complementary information. For example, in cases where snoring stops but resumes after a while, the pattern in the audio signal might resemble an apnoea-like silence followed by an awakening spike. In such cases, an awakening-related movement is not visible in the VDA signal and the addition of VDA features to the audio features results in the whole period not being mistaken for an apnoeic event by the classifier.

Chapter 6

Respiratory signal estimation from video recordings

6.1 Introduction

Respiratory activity leading to the right amount of oxygen and carbon dioxide being exchanged is essential for healthy life. While acute respiratory failure can become life-threatening within a few minutes, chronic respiratory problems can lead to serious long-term consequences that affect the quality of life. Monitoring respiratory activity is therefore important in diagnosing these failures and can be life-saving [114].

The apparent need for respiratory monitoring devices and methodologies has led to the development of different equipment to measure respiratory variables. Folke *et al.* provide an extensive review and comparison of non-invasive methods of monitoring respiratory activity [114]. From a clinical perspective, respiratory activity includes Respiratory Rate (RR) and depth as well as quantifiable measures of the gas exchange

taking place. The authors classify the measurement methods into the following three categories based on their detection approaches: (i) movement, volume, and tissue composition detection, (ii) airflow sensing, (iii) and blood gas measurement.

Although in many cases, respiratory monitoring involves measuring several variables for the correct diagnosis and treatment, many respiratory-related health issues can benefit from simple monitoring of respiratory rate (RR) and respiratory tidal volume (TV). Routine measurement of RR on a general ward in hospital is performed by manually counting chest movements over a period of approximately 30 seconds, which is time-consuming and prone to errors [115, 116]. For patients who are already connected to RIP bands, oronasal airflow thermistors, PPG, or ECG, RR can be measured using these on-body sensors. However, there is still a real need for reliable non-contact respiratory measurement systems for other groups of patients who are not connected to monitors. An example of such groups of patients are infants and adults who are required to be monitored in their sleep and can hugely benefit from a non-disturbing monitoring environment.

The use of video cameras as non-contact health-monitoring devices has become increasingly popular in recent years. Different research groups have demonstrated the potential use of colour video recordings in the estimation of blood oxygen saturation [117] and the measurement of heart rate [117, 118]. Thermal video recordings have been used by several groups for the detection of respiratory-induced changes in the temperature around the nose and mouth area [119, 120, 121, 122, 123, 124]. The issue with using thermal cameras is that they are costly and not a particularly feasible solution for home monitoring.

Daylight video cameras have been investigated for extracting RR and respiratory waveform. Frigola *et al.* [125], Makkapati *et al.* [126], and Li *et al.* [127] proposed using optical flow to extract the respiratory signal. Tan *et al.* [128] used a webcam and the thresholded frame differencing technique to generate the respiratory signal. The method was tested on a normal subject sitting in front of the camera, wearing different clothes and positioned at different distances to the camera as well as on a paediatric patient. The authors report a drop in performance for plain clothes and distances above 0.5m. Poh *et al.* [129] extracted RR from videos recorded with a webcam. The RR was calculated from the power spectrum of the heart rate variability signal derived by applying independent component analysis to the changes in the Red-Green-Blue (RGB) values. Tarassenko *et al.* [117] presented a non-contact vital-sign monitoring method in a clinical environment to measure several physiological parameters including RR. The signals were extracted by tracking the changes in RGB colour intensities in an exposed area of forehead and an Autoregressive (AR) model was used to find the dominant respiratory frequency. Koolen *et al.* [130] used the Eulerian video magnification technique [131] to enhance respiratory movements and used an optical flow algorithm to extract the movements from videos of 7 infants who underwent PSG. Tveit *et al.* [132] measured the RR of two infants and two adults using a phase-based method. They used an activity detection method to identify when the camera is blocked by a carer or the baby is removed.

Apart from daytime monitoring, night-vision cameras are particularly useful for monitoring respiration during sleep. Light projection systems that enhance respiratory movements have been proposed by Aoki *et al.* [133] and Martinez et al. [134].

Bartula et al. *et al.* [135] used a monochrome day/night camera to extract the respiration by calculating the cross-correlation of 1D vertical projection of consecutive images. Heinrich *et al.* [136] computed a reference expiratory pause frame and extracted the respiration waveform by calculating the cross-correlation of the reference frame with other video frames.

While the growing interest in the vision-based monitoring of respiration is promising, there is little evidence of the validation of these state-of-the-art methods against real patient data, particularly patients with sleep disorders. The objective of this chapter is to advance the new area of vision-based respiration monitoring by developing a real-time technique which (i) works with low quality off-the-shelf cameras to monitor patients over long periods of time, (ii) is validated against real patient data recorded in general PSG room settings.

6.2 Materials and methods

In this section details of the datasets used for the study, the technique used to extract respiratory signals from video recordings, and the assessment of the derived signals are given.

6.2.1 Data

Two sources of data have been used for this study; (i) DS2 which includes PSG recordings from 22 patients who were referred to the sleep lab at the Churchill Hospital (Oxford, UK) with suspected OSA. PSG recordings include infra-red video recordings in grey scale in MPEG2 format with a 352×288 resolution and 25 fps frame rate. The

patients in the sleep lab were provided with bed sheets (patterned or plain) and some patients were under the bed sheets during the night. (ii) DS3 which includes full overnight PSG recordings from 32 normal participants including infra-red video recordings in grey scale in MPEG2 format, with a 720×576 resolution and 25 fps frame rate. In the PSG system the recording of the physiological signals (e.g. the RIP signals) started a with a delay from the video recording. The amount of delay is recorded in the log files of each PSG recording. Therefore, the video recordings were synchronised with the rest of the physiological recordings using the delay stated in the log files.

The subjects in the study were provided with a thick patterned duvet and most subjects were under the duvet during the night. The PSG recordings of both DS2 and DS3 include three respiratory related signals, nasal airflow (NAF) and thoracic and abdominal RIP bands (TEFF and AEFF respectively). The data is used from 2 hours to 6 hours into the recordings to increase the likelihood of the patients/subjects being asleep during the period of study (for details of these datasets see Section 3.4 of Chapter 3).

6.2.2 Extracting respiratory signal from video

In this section we explain the methodology that was developed to derive a respiratory signal from overnight video recordings. The technique is based on the repetitive movement of breathing, which is most pronounced in the chest and abdominal areas.

Since the body moves through a similar trajectory during inhalation and exhalation, the video frames which show the body at specific points of the inhale/exhale

cycle are similar. This is illustrated in Figures 6.1 and 6.2. Figure 6.1 shows a sequence of frames which are captured at 0.5 second intervals from a demonstration video. The frames show the repetitive, although subtle, movement of the torso during two cycles of breathing. The red dots in the figure are a specific point on the torso which is selected manually for demonstration. Figure 6.2 presents the changes to the X and Y position of the red dot during breathing. As can be seen in 6.2b, the selected point on the torso is at relatively similar positions during the corresponding stages of inhalation or exhalation.

In Figure 6.2b, if point E is the trough of exhalation, points D and F, which are 0.5 second before and after E respectively, are at relatively similar positions. Similarly, points C and G, which are 1 second before and after E respectively, are also at relatively similar positions. In a generalised form, if the trough of exhale happens at frame at 4 seconds, all the frame pixels at 3.5 seconds and all the frame pixels at 4.5 seconds are relatively similar. Additionally, all the frame pixels at 3 seconds and all the frame pixels at 5 seconds are relatively similar.

On the other hand, for point C which is midway between inhalation and exhalation, the positions of points B and D which are 0.5 second before and after C are very different. The difference increases for points A and E which are 1 second before and after C. If we generalise this to all pixels in the frames, the minimum difference between frames which are at equivalent intervals before and after a frame at time t ($f(t)$) occurs when t is at the peak of inhalation or at the trough of exhalation. The maximum difference between frames which are at equivalent intervals before and after $f(t)$ occurs when t is at the mid-point between inhalation and exhalation.

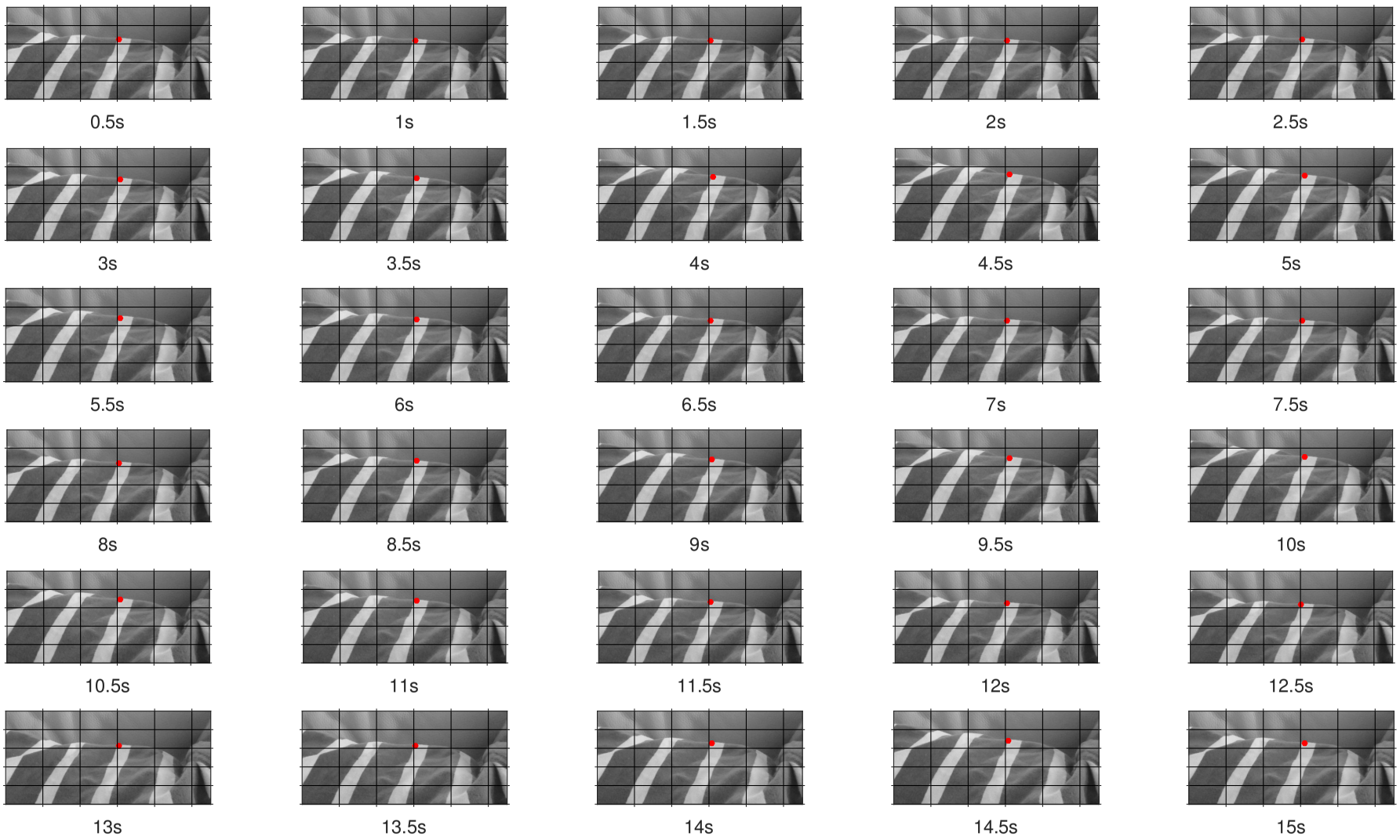
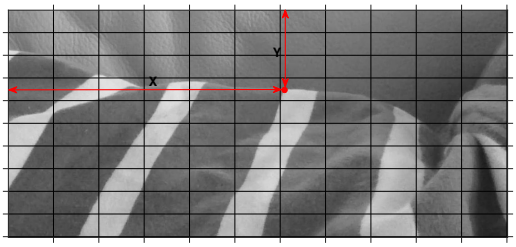
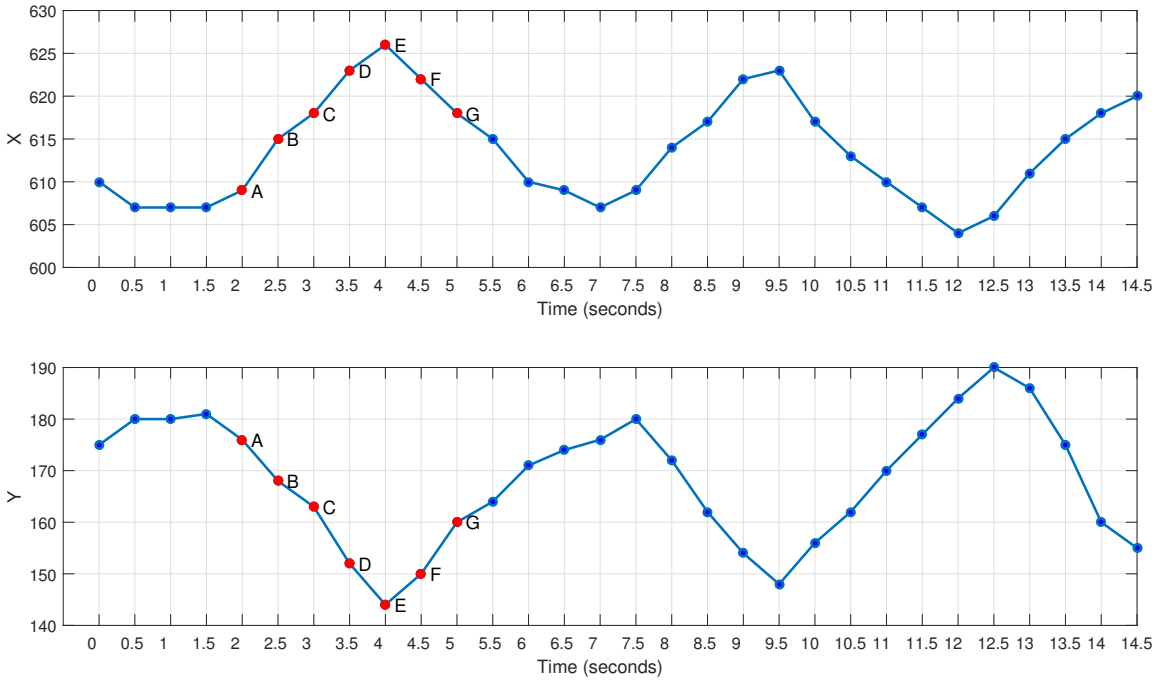


Figure 6.1: A sequence of frames which are captured at 0.5 second intervals from a demonstration video. The red dots track the position of a specific pixel on the torso which is manually selected in each frame.



(a)



(b)

Figure 6.2: (a) A specific pixel on the torso which is manually selected in each frame. (b) The changes in the X and Y positions of the pixel during two respiratory cycles.

Therefore, using the example frames in Figure 6.1, a respiratory signal can be generated by calculating the mean absolute difference between $f(t-0.5)$ and $f(t+0.5)$, where $t = 0.5, ..., 14$ seconds. The generated respiratory signal is shown in the middle plot of Figure 6.3. The top plot of the same figure is the Y position of the red dot in frames in Figure 6.1 to show the respiration cycle.

Since there might be slight differences between the depth and speed of respiratory cycles, if a sequence of frames is used as opposed to one frame, a smoother respiratory

signal will be generated. In other words, the respiratory signal is calculated by taking the overall mean of the absolute difference between $f(t-0.5)$ and $f(t+0.5)$, $f(t-1)$ and $f(t+1)$, and $f(t+1.5)$ and $f(t-1.5)$, where $t = 1.5, ..., 13$ seconds. As can be seen in the figure, the generated respiratory signal has a frequency twice that of the true RR since it has two troughs and two peaks in each respiratory cycle. The generated respiratory signal using a sequence of frames is presented in the bottom plot of Figure 6.3.

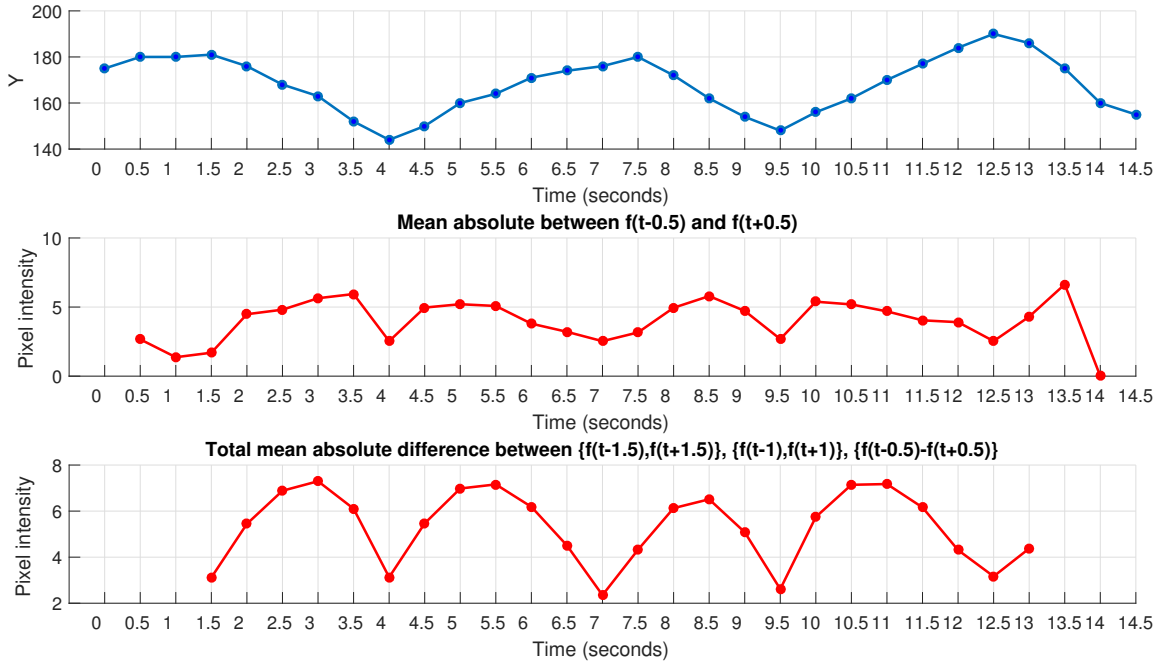


Figure 6.3: Top plot: The Y position of the red dot in frames in Figure 6.1 to show the respiratory cycle. Middle plot: The generated respiratory signal by calculating the mean absolute difference between $f(t-0.5)$, and $f(t+0.5)$, where $t = 0.5, ..., 14$ seconds and $f(t)$ denotes a frame at time t. Bottom plot: The generated respiratory signal by taking the overall mean of the absolute difference between $f(t-0.5)$ and $f(t+0.5)$, $f(t-1)$ and $f(t+1)$, and $f(t+1.5)$, and $f(t-1.5)$, where $t = 1.5, ..., 13$ seconds.

The same explanation is now presented in mathematical terms. Let us denote a frame with dimensions $W \times H$ at time t as $f(t)$ and each pixel in $f(t)$ as $p_{x,y}(t)$ where x and y represent the location of the pixel in the frame with $p_{0,0}(t)$ being the top left

corner. Note that t, from now on, is the sampling time of the video frames. Therefore, if the frame rate of the video is denoted with fs frames per second, $f(t)$ and $f(t+1)$ are two consecutive frames 1/fs seconds apart. The difference between two frames at times t and t' is defined as the absolute difference of the pixels at the same position:

$$f(t) - f(t') = \frac{1}{W \times H} \sum_{y=1}^{H} \sum_{x=1}^{W} |p_{x,y}(t) - p_{x,y}(t')| \tag{6.1}$$

We define a sequence of frames starting at time t with length N as:

$$seq(t)_N = (f(t), f(t+1), ..., f(t+N)) \tag{6.2}$$

and a sequence starting at time t with length $-N$ as:

$$seq(t)_{-N} = (f(t), f(t-1), ..., f(t-N)). \tag{6.3}$$

The difference between two sequences of frames with length N starting at time t and t' is defined as the mean of the difference between the frames at the same position in the sequence:

$$seq(t)_N - seq(t')_N = \frac{1}{N}(f(t) - f(t'), ..., f(t+N) - f(t'+N)) \tag{6.4}$$

In ideal conditions, in a respiration cycle, the following equation holds at time T for the the peak of inhalation and the trough of exhalation:

$$seq(T)_N - seq(T)_{-N} = 0 \tag{6.5}$$

In reality, however, factors such as noise, changes in the background light conditions, and other body movements, cause the frames occurring at the same interval before and after each inhalation and exhalation to be slightly different from each other, in which case the above condition of equality (Eq.6.5) does not hold.

The Video Derived Respiratory (VDR) signal can be calculated from:

$$VDR(t) = seq(t)_N - seq(t)_{-N} \tag{6.6}$$

The optimum value for the sequence length, N, will be derived in the results section of this chapter.

As explained earlier for the example of Figure 6.1, there are two local minima in the VDR signal for each respiratory cycle, one for the inhalation peak one for the exhalation trough. Therefore, the dominant frequency of the signal is twice the respiratory frequency. Figure 6.4 shows a 60-second segment of the respiratory signal derived from a patient's video in DS2 and the reference AEFF signal. Both signals are down-sampled to 4 Hz. AEFF is band-pass filtered with a respiratory passband from 0.1 Hz to 0.6 Hz and VDR is band-pass filtered with a passband from 0.2 Hz to 1.2 Hz. The AEFF signal is overlaid on the VDR signal for the comparison of the shape of the signals and the positions of the peaks and troughs.

6.2.3 Non-respiratory movement

As previously mentioned, the methods described above assume that the subject is relatively still. However, this may not be the case in practice and the subject may move several times during the night. While the measurement of the respiratory signal

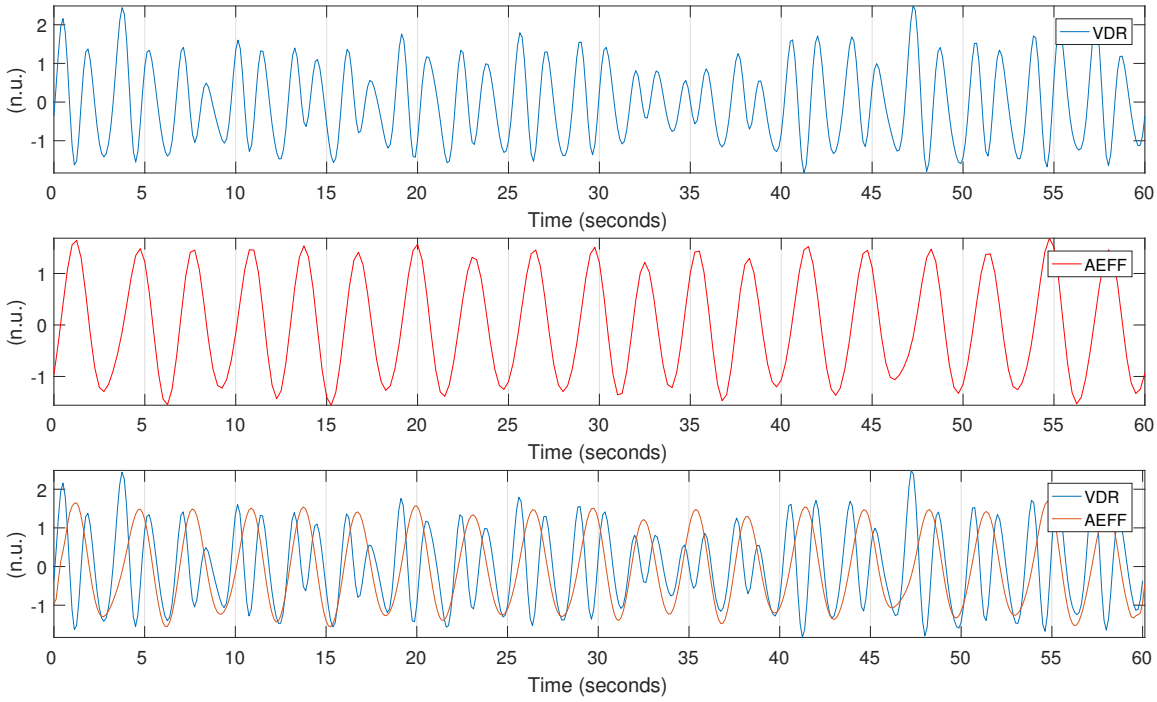


Figure 6.4: A 60-second segment of a VDR signal from a patient in DS2 and the reference AEFF signal. The signals are overlaid for the comparison of the shape of the signals and the positions of the peaks and troughs.

is a challenge when the subject is moving, this method gives a good estimate of the respiratory signal as soon as the subject becomes relatively still again. This is because at each point of time, the current frame and its neighbouring frames are used in the algorithm as opposed to deriving a reference frame which requires identifying episodes of movement and updating the reference frame after each of these episodes.

The quick recovery after movement is particularly important in detecting apnoeic events as moderate-to-severe OSA patients have numerous awakening-related movements throughout the night. Figure 6.5 shows the effect of movement and the changes in the amplitude of the VDR and the reference AEFF signal.

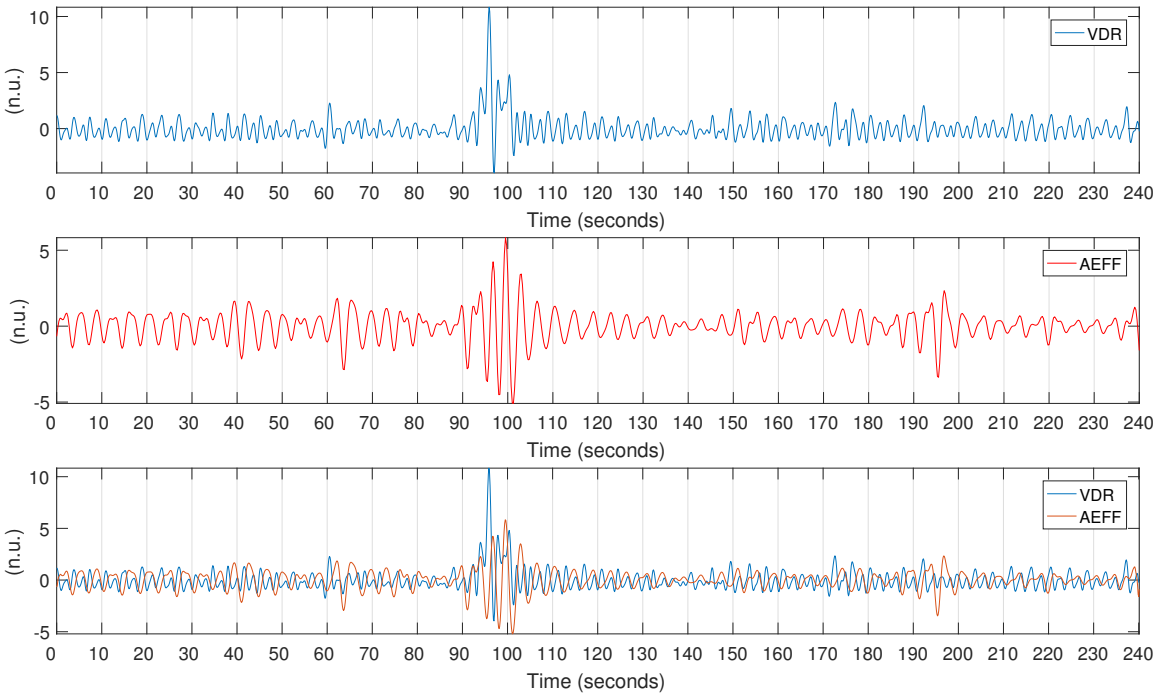


Figure 6.5: A 240-second segment of the respiratory signal derived from a video in DS2 and the reference AEFF signal. The signals are overlaid to compare the effect of movement (happening around 100 seconds) on the VDR and AEFF signals. (n.u.) = Normalised unit.

6.2.4 Assessment of the VDR signals

6.2.4.1 Changes in respiratory rate

Respiratory rate calculation using an autoregressive model In order to derive the RR, the dominant frequency of the respiratory signal needs to be calculated. Using an autoregressive (AR) model is a robust method for calculating the dominant frequency, which has previously been applied to EEG [137], ECG [138, 139, 140, 141, 142], and most recently to PPG [143, 144] signals to extract the respiratory rate.

The AR model assumes that the desired signal is the linear combination of its past values plus an error term with zero mean:

$$x(n) = -\sum_{i=1}^{I} a_i x(n-i) + e(n) \tag{6.7}$$

where a_i are the parameters of the model. $e(n)$ and $x(n)$ can be considered as the input and outputs of a system with the transfer function H as:

$$H(z) = \frac{X(z)}{E(z)} = \frac{1}{\sum_{i=0}^{I} a_i z^{-i}} \tag{6.8}$$

where $X(z)$ and $E(z)$ are the z-transform of $x(n)$ and $e(n)$ respectively. $H(z)$ can be re-written as:

$$H(z) = \frac{z^I}{(z-z_1)(z-z_2)\dots(z-z_I)} \tag{6.9}$$

The denominator has I roots each representing a pole in the transfer function, $H(z)$. The poles appear as complex conjugate pairs, each representing a frequency of the output signal. The poles with the larger magnitude represent the frequencies with the higher magnitude peaks, which is essentially the frequencies intended to be identified. The resonant frequency of each peak, f is given by the phase angle, θ of the corresponding pole:

$$f = \frac{\theta}{2\pi\Delta t} \tag{6.10}$$

where Δt is the sampling interval (in our case is the inverse of video frame rate). Based on the method explained in Orphanidou *et al.* [142], the objective is to: (i) fit a polynomial function to the signal, (ii) find the poles within the respiration frequency

range, (iii) pick the pole with the largest magnitude and all other poles with magnitude of larger than 95% of the largest pole, and (iv) and calculate the RR from the pole with the lowest frequency amongst the candidate poles found in step (iii).

Since the respiratory frequencies (between 0.1 Hz-0.6 Hz or 6-36 bpm) are small compared to the original sampling frequency of the VDR and the reference respiratory signals (25 Hz VDR, 128 Hz NAF, and 256 Hz AEFF and TEFF signals), the VDR and reference signals were down-sampled. Down-sampling improved the resolution of the phase angles making it easier to select the dominant respiratory frequency. The VDR signals were down-sampled to 8 Hz and the reference respiratory signals were down-sampled to 4 Hz after applying an anti-aliasing filter.

Additionally, since the VDR and reference signals were relatively noisy, they were band-pass filtered to only keep the normal respiratory frequencies. Three band-pass filters of 0.1 Hz-0.6 Hz, 0.2 Hz-0.6 Hz, 0.1 Hz-0.5 Hz, as suggested by Orphanidou *et al.* [142], were applied to the reference respiratory signals. The VDR signals were filtered using three band-pass filters of 0.2 Hz-1.2 Hz, 0.4 Hz-1.2 Hz, 0.2 Hz-1.0 Hz. The cut-off frequencies are doubled for the VDR signals since the frequency of the VDR signals is twice the respiratory frequency.

When using the AR method to calculate the RR, the parameter to select is the order of the model. This parameter was varied from 4 to 20. High order models were found to divide the respiratory frequency into a few poles with smaller magnitude (over-fitting). A model order of 8 for the reference respiratory signal and a model order of 11 for the VDR signal generated the most stable results and were chosen for this work.

Respiratory rate calculation by counting respiratory cycles As with to the AR method in the previous section, the respiratory signals were first down-sampled to 8 Hz for the VDR signals and 4 Hz for the reference respiratory signals. Then the signals were bandpass-filtered to emphasise the respiratory frequencies. The VDR signals were filtered using a 0.2 Hz-1.2 Hz band-pass filter and the reference respiratory signals were filtered using a 0.1 Hz-0.6 Hz bandpass filter. The RR was then simply determined by counting the troughs in the signal. For reference respiratory signals such as the NAF signal, the interval between every two troughs is one respiratory cycle. The interval between every two troughs is half a respiratory cycle in the VDR signals and therefore the number of troughs was divided by two to calculate the RR.

Respiratory rate calculation protocol DS3 has full PSG recordings of normal subjects and DS2 has full PSG recordings of patients. Both datasets include three reference respiratory signals, the NAF, TEFF, and AEFF signals. The RR was derived from the down-sampled signals using both the AR method and the trough-counting method. The RR was calculated for every 60-second window of the signals moving in steps of 15 seconds.

For the mild-to-no OSA patients in DS2 and the normals subjects in DS3, the ground truth RR was calculated as the mean of the three reference RRs. The RR derived from video was compared with the ground truth by calculating the absolute error (AE) between the two values.

Figure 6.6 illustrates a 60-second window of the VDR signal from a patient in DS2 along with the reference respiratory signals. All signals are down-sampled and band-pass filtered (8 Hz for VDR with a 0.2 Hz-1.2 Hz band-pass filter, and 4 Hz

for the reference signals with a 0.1 Hz-0.6 Hz band-pass filter). The interval between every two troughs marked with red stars indicates the duration of a respiratory cycle for the reference signals and half a respiratory cycle for the VDR.

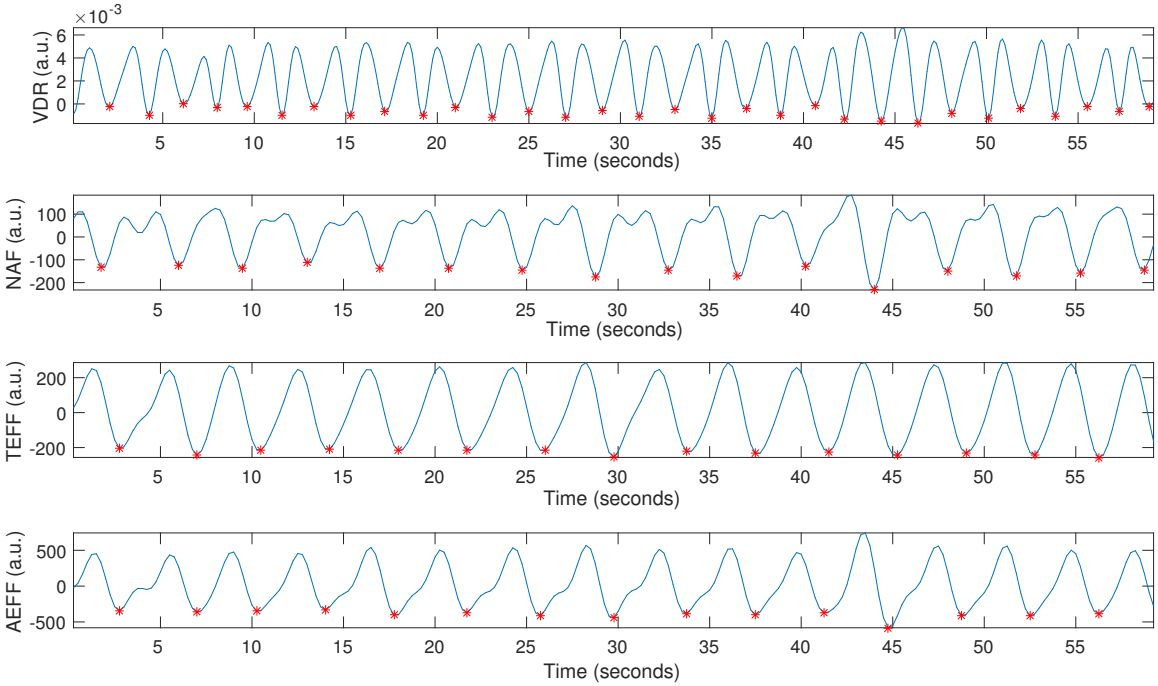


Figure 6.6: A 60-second window of VDR from a patient in DS2 with the reference respiratory signals. All signals are down-sampled and band-pass filtered. The red stars show the troughs in the signals. (a.u.) = Arbitrary unit.

Patients with moderate-to-severe OSA have frequent OSA events during the night. As explained in Chapter 2, during an OSA event the airway is completely or partially obstructed but the respiratory effort continues and the respiratory movements are still present. These respiratory movements, however, may be different in thoracic and abdominal areas. The respiratory movements in the thorax and abdomen may be paradoxical (chest and abdomen move in opposite directions) or there may be a phase shift between them. Figure 6.7 illustrates normal, out of phase, and paradoxical thorax and abdomen movements. The continued respiratory efforts during apnoeic events usually have a reduced amplitude as shown in Figure 6.8.

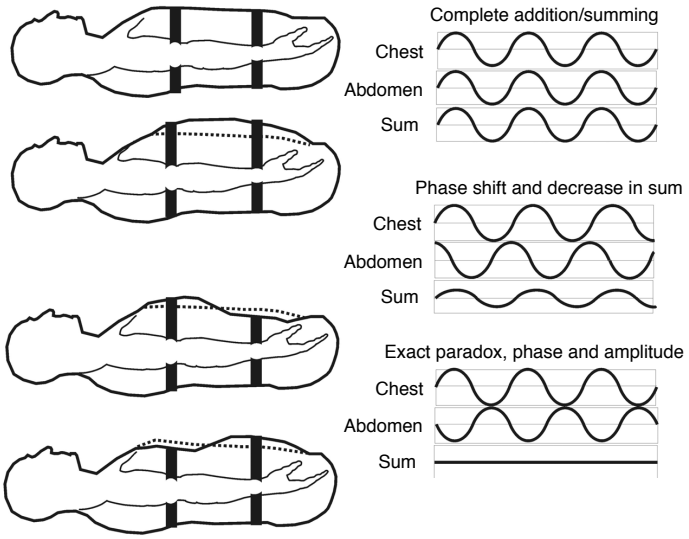


Figure 6.7: Normal, out of phase, and paradoxical respiratory movements in the thoracic and abdominal areas (image adapted from [145]).

Regardless of the amplitude of the respiratory efforts and the type of respiratory movements during apnoea, RR in the VDR signal can still be compared with the reference signals particularly TEFF and AEFF. During an apnoeic event the NAF signal may become flat if the airway is completely obstructed and be unsuitable for RR comparison. For moderate-to-severe OSA patients, the same methods used for mild-to-no OSA patients were used to extract RR. The ground truth RR used for the moderate-to-severe OSA patients was the average of the RR derived from the TEFF and AEFF signals.

6.2.4.2 Changes in breath amplitude

This section investigates whether it would be possible to identify the drops in TV in the VDR signals which is a requirement for the identification of apnoeas and hypopnoeas. According to the AASM manuals, the preferred signal for monitoring changes in TV is the oronasal thermal flow measurement as it has the advantage of being

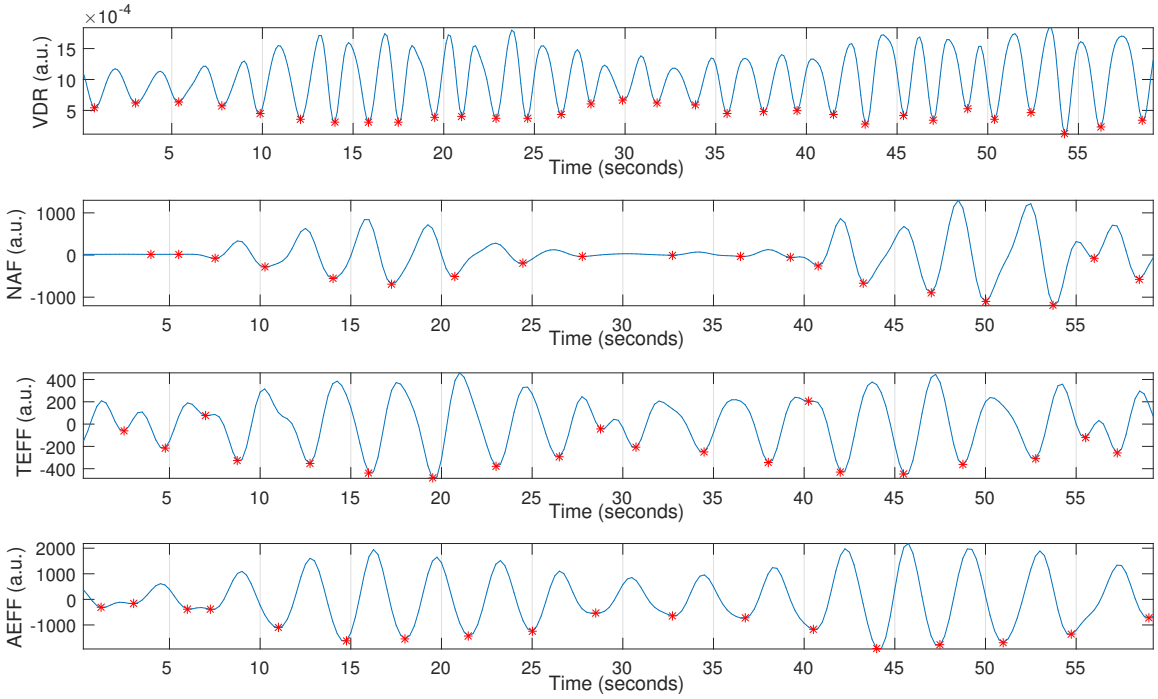


Figure 6.8: Apnoeic episode with reduced respiratory amplitude. The red stars show the troughs in the signals. (a.u.) = Arbitrary unit.

able to detect both nasal and oral airflow. The measurement of nasal airflow alone is not recommended for apnoea detection as the signal may show decreased TV during mouth breathing. The sum of the RIP TEFF and AEFF signals (RIPSum) is the preferred alternative measurement for apnoea detection if the oronasal thermal sensor is not available. For the detection of hypopnoea, the Nasal Pressure (NP) signal is the recommended signal and the RIPSum signal is the preferred alternative [79] (for more details of the AASM recommendations see Chapter 2).

This research uses the RIPSum signal for hypopnoeas as well as apnoeas. The similarity of the RIP band measurements to respiratory measurements from the video and the non-linear characteristics of the NP signal[1] makes hypopnoea detected in the VDR and RIPSum signals more comparable. The breath amplitude of the RIPSum signal provides an estimate of TV when calibrated and if uncalibrated, the changes

[1]The square transformation of the NP signal approximates flow more closely [146].

in the signal show the relative change in TV compared to baseline breathing. During OSA events, the more out of phase the TEFF and AEFF signals become (moving toward paradoxical breathing), the lower the amplitude of the RIPSum will be. The RIPSum signal shows absent or minimal excursions during apnoea and reduced excursions compared to baseline breathing during hypopnoea [146].

In order to analyse the VDR signals' ability to track the changes in the TV, both the RIPSum and VDR signals were first down-sampled to 4 Hz. The RIPSum signal was filtered with a 0.1 Hz−0.6 Hz bandpass filter and the VDR signal with a 0.2 Hz−1.2 Hz bandpass filter. Next, the breath amplitudes were calculated in the VDR signals and the reference RIPSum signals. For the RIPSum signals the breath amplitudes were calculated by first identifying the inhalation peaks and exhalation troughs in the signal and then calculating each breath amplitude as the difference between each inhalation peak and the following exhalation trough (Figure 6.9). For the VDR signals first the peaks and troughs in the signal were identified. Then, the breath amplitudes were calculated as the sum of the difference between every two consecutive peaks from their following troughs (Figure 6.10). The use of a pair of adjacent peaks rather than a single peak was because every peak-to-trough in the VDR signals represents half of a breathing cycle (see section 6.2.2). The breath amplitudes of the RIPSum signal are denoted with a_{RIPSum} and the breath amplitudes of the VDR signals are denoted with a_{VDR}.

Both a_{RIPSum} and a_{VDR} signals were then divided into overlapping 10 second windows, $w(k)$, moving in steps of 1 second. The length of the windows was chosen to be 10 seconds because the minimum duration of apnoeic/hypopnoeic events is defined

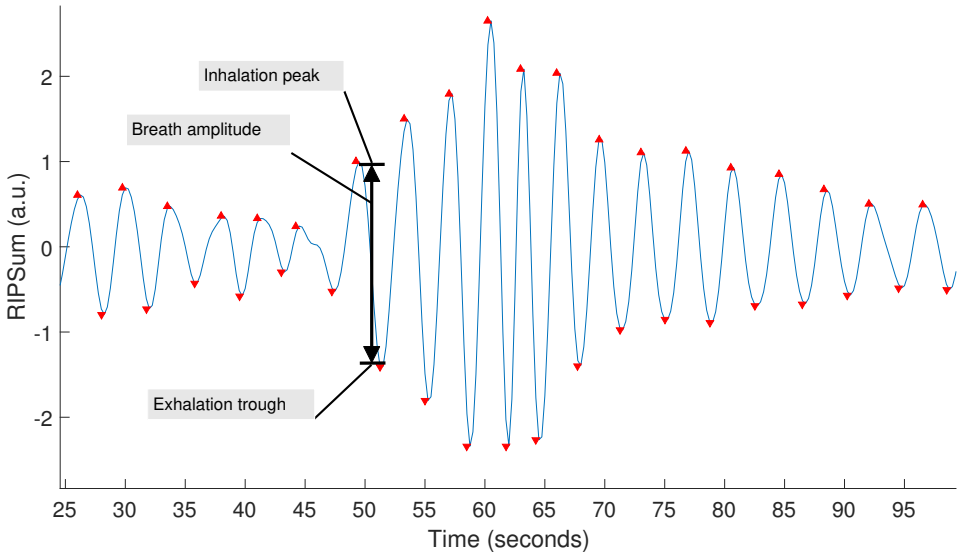


Figure 6.9: The breath amplitude in the RIPSum signal is calculated as the difference between each inhalation peak and the following exhalation trough. (a.u.) = Arbitrary unit.

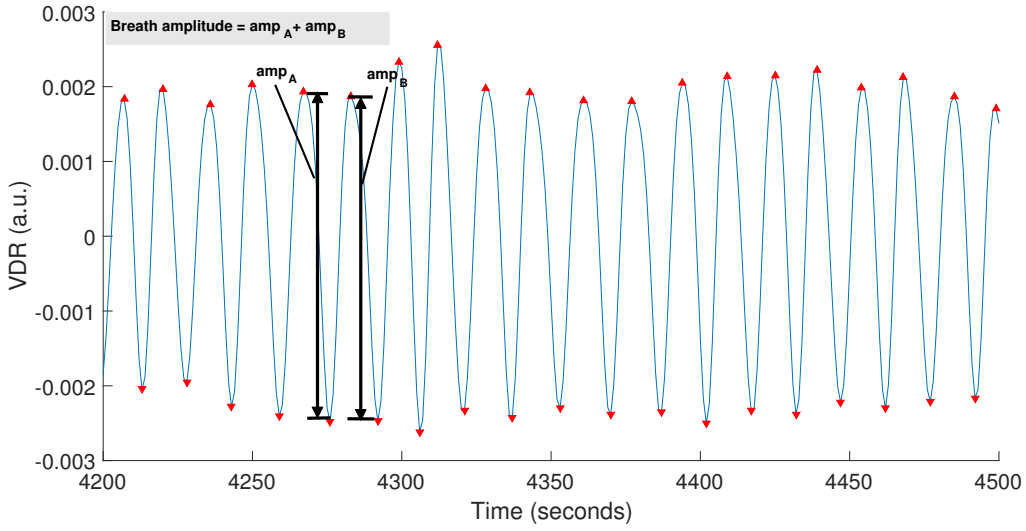


Figure 6.10: The breath amplitudes in the VDR signal is calculated as the sum of the difference between every two consecutive peaks from their following trough. (a.u.) = Arbitrary unit.

to be 10 seconds. During an apnoea or a hypopnoea, the amplitude of the signal must not exceed the corresponding threshold for at least a period of 10 seconds. Therefore, the breath amplitude for each $w(k)$ in the VDR signal, denoted with $wa_{VDR}(k)$, was calculated as the maximum of the breath amplitudes within that $w(k)$. Similarly, the breath amplitude for each $w(k)$ in the RIPSum signal, denoted with $wa_{RIPSum}(k)$, was calculated as the maximum of the breath amplitudes within that $w(k)$.

For the calculation of a baseline for normal breathing the AASM manuals recommend taking the mean of the 3 largest breaths in the previous two minutes of the signal [79]. Based on this guideline a baseline value $b_{RIPSsum}(k)$ was calculated for each $w(k)$ in the RIPSum signal as the mean of the 3 highest breath amplitudes in a_{RIPSum} of the previous 2 minutes. Similarly, a baseline value $b_{VDR}(k)$ was calculated for each $w(k)$ in the VDR signal as the mean of the 3 highest breath amplitudes in a_{VDR} of the previous 2 minutes. In order to remove extreme out-of-range artefacts in both signals, the values outside 5 times standard deviation of the signals were not used to calculate the baselines, wa_{RIPSum}, and wa_{VDR}.

The changes in the breath amplitude in the VDR and RIPSum signals were calculated as the ratio of the breath amplitude of each $w(k)$ to the normal breathing baseline of that $w(k)$ as follows:

$$br_{VDR}(k) = \frac{wa_{VDR}(k)}{b_{VDR}(k)} \tag{6.11}$$

$$br_{RIPSum}(k) = \frac{wa_{RIPSum}(k)}{b_{RIPSum}(k)} \tag{6.12}$$

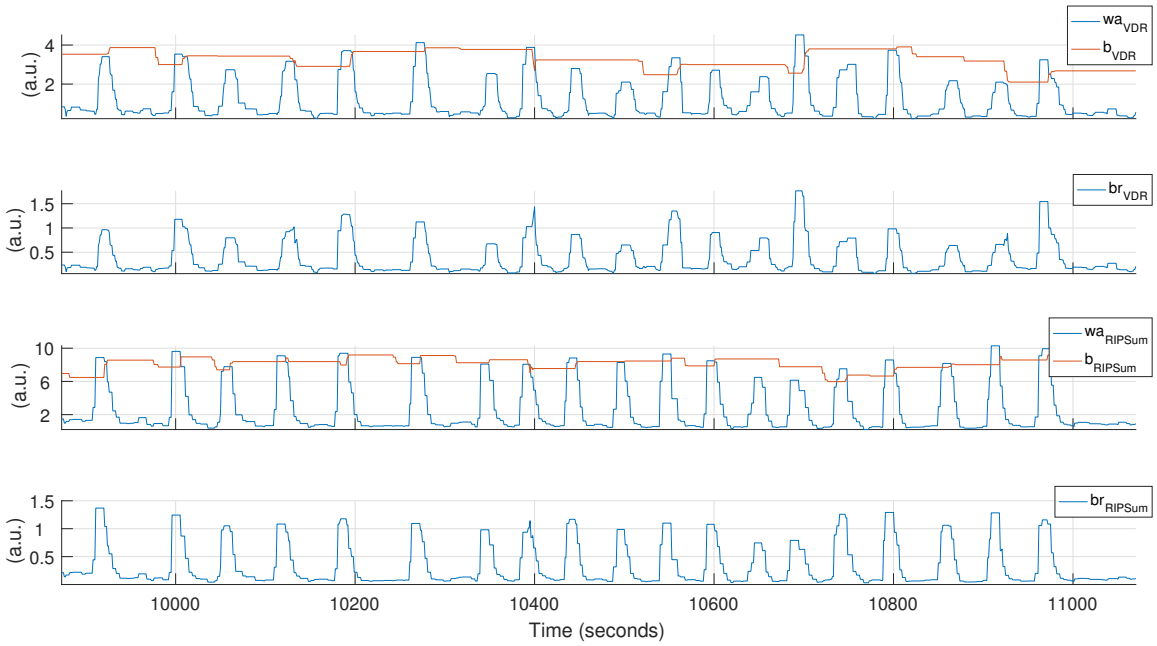


Figure 6.11: An example section of breath amplitudes, normal breathing baseline, and breath amplitude ratios to normal breathing baseline of a severe OSA patient. (a.u.) = Arbitrary unit.

Note that sometimes instead of the ratio of current breath amplitude to baseline breath amplitude (equivalent to br_{VDR} and br_{RIPSum}) the phrases '*drop from baseline*' or '*percentage drop from baseline*' are used. In these cases, drop from baseline is equivalent to $1 - br_{VDR}$ and $1 - br_{RIPSum}$ and percentage drop from baseline is equivalent to $(1 - br_{VDR}) \times 100$ and $(1 - br_{RIPSum}) \times 100$.

Figure 6.11 shows an example section of the wa_{VDR} signal with the corresponding baseline b_{VDR} and breath amplitude ratios br_{VDR} from a severe OSA patient. The figure also presents the wa_{RIPSum} signal and the corresponding baseline b_{RIPSum} and breath amplitude ratios br_{RIPSum}.

6.3 Results

6.3.1 Tuning the model

The method developed above for extracting the VDR signals from the video recordings requires one parameter, which is the length of the sequence N in equation 6.4, to be tuned. To find the optimum value for N, first the reference respiratory signals of mild-to-no OSA patients in DS2 were divided into 60-second windows with an overlap of 15 seconds. Then, the reference RR values were calculated for these 60-second windows (details of RR calculation and the reference value calculation have been given in section 6.2.4.1). From all the 60-second windows, 200 windows were randomly selected from each of the following RR ranges: (a) below 10 bpm (a total of 231 windows), (b) between 10 bpm and 15 bpm (a total of 17847 windows), and (c) above 15 bpm (a total of 856 windows). Next, the VDR signals were derived for the 600 windows (200 each for (a),(b) and (c)) for a range of values of $N = 10, 15, ..., 100$. For each value of N, the RRs were calculated for the derived VDR signals and the error between the VDR derived RRs and reference RRs were compared. As can be seen in Figure 6.12, the Absolute Error (AE) for RR between 10 bpm and 15 bpm is relatively similar for all values of N. The AE for RR above 15 bpm does not noticeably change for values above 30 frames and the AE for RR below 10 bpm does not noticeably change for values above 45 frames. Choosing larger values for N (for example, $N = 60$) despite slightly decreasing the AE leads to processing a larger number of frames and reduction in the processing performance. Therefore, 45 frames was chosen as the optimum value for N.

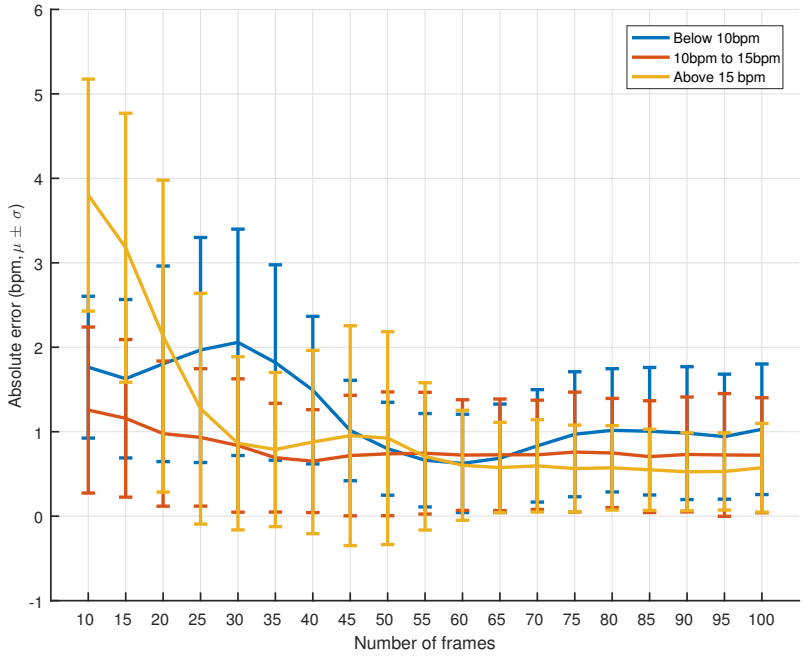


Figure 6.12: The mean and standard deviation of absolute error between RR derived from reference respiratory signals and RR derived from VDR when the length of sequence N was changed from 10 to 100 frames.

6.3.2 Respiratory rate calculations

The extracted RRs for one of the mild OSA patients in DS2 is shown in Figure 6.13 as an example. The bottom plot shows the AE between the VDR derived RR and the reference RR.

RR calculations from one of the moderate OSA patients in DS2 are shown in Figure 6.14 as an example. The bottom plot shows the AE between the VDR derived RR and the reference RR.

The scatter and Bland-Altman plots in Figure 6.15 compare the RR estimates derived for each sliding 60-second window (moving in 15-second steps) of the VDR signals with the corresponding reference RR values. The size of the dots on the plot is proportional to the number of RR estimates taking that value to display the overlapping data points.

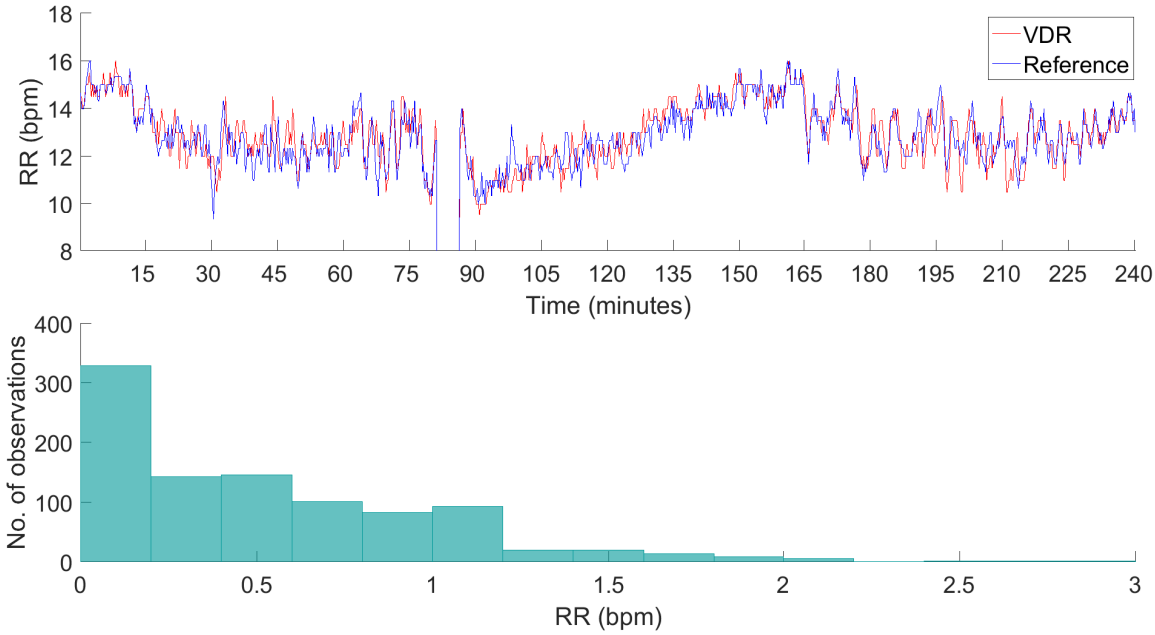


Figure 6.13: The top plot presents RR calculations from the VDR and the reference signals for a mild OSA patient in DS2. The RR is updated every 15 seconds and is calculated for the previous 60 seconds. The bottom plot shows a histogram of the AE between the RR calculated from the VDR and the reference value.

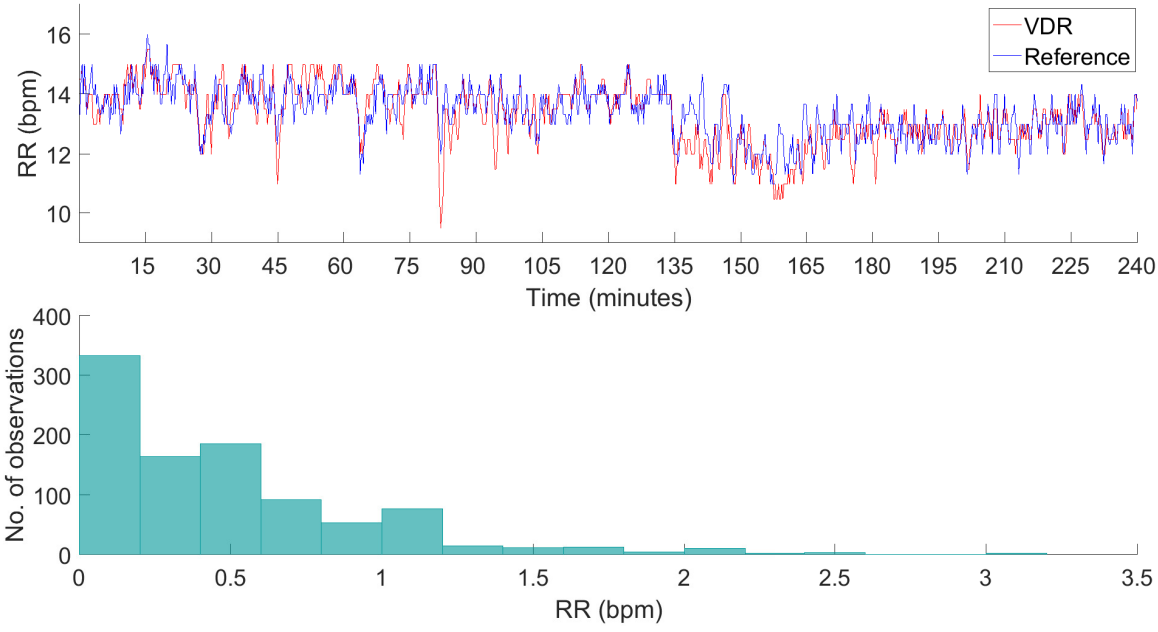
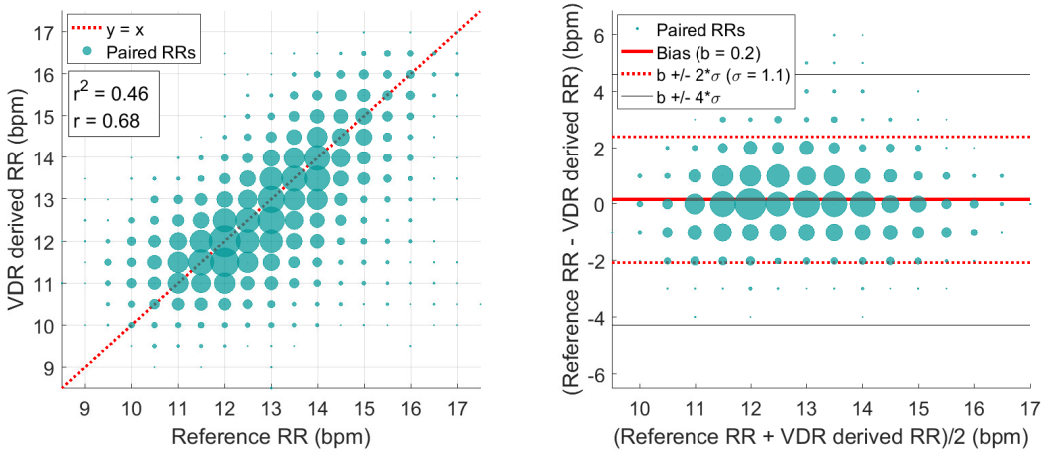
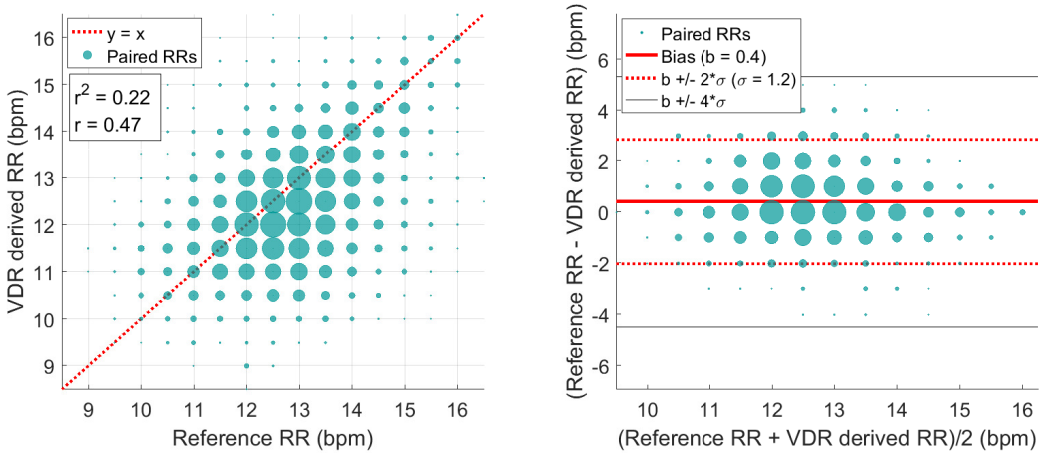


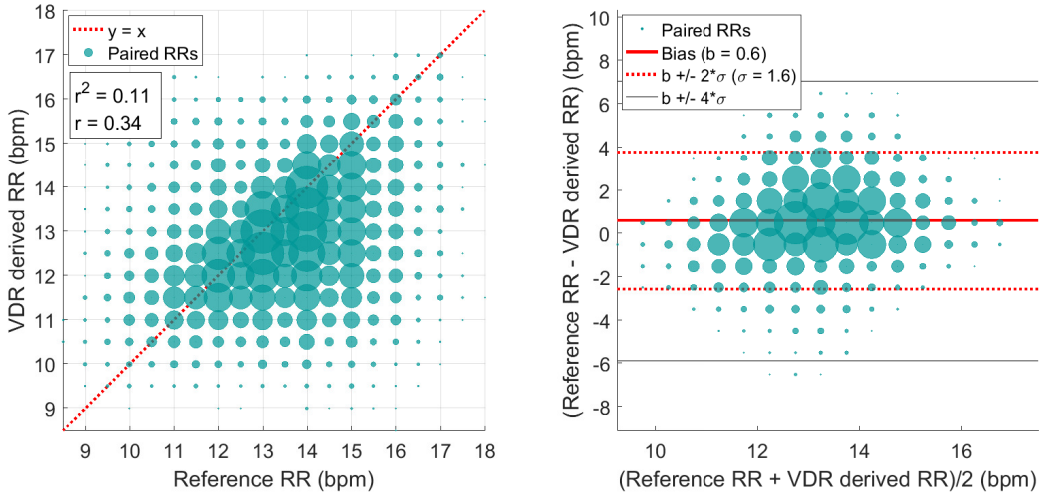
Figure 6.14: The top plot presents RR calculations from the VDR and the reference signals for a moderate OSA patient in DS2. The RR is updated every 15 seconds and is calculated for the previous 60 seconds. The bottom plot shows a histogram of the AE between the RR calculated from the VDR signal and the reference value.



(a) Mild-to-no OSA patients in DS2



(b) Moderate-to-severe OSA patients in DS2



(c) Normal participants in DS3

Figure 6.15: Scatter plots (left side) and Bland-Altman lots (right size) comparing each pair of VDR derived RR and the corresponding reference RR (blue-green dots). The size of the dots on the plot is proportional to the number of RR estimates taking that value to display the overlapping datapoints. The lines show the bias (solid red), 95% confidence interval (around $2 \times \sigma$, dotted red), and outlier thresholds (black).

Table 6.1: The absolute error between RR derived from the VDR signal and the reference signals for the mild-to-no OSA patients in DS2 and for normal volunteers in DS3.

Dataset	Method	AE($\mu \pm \sigma$)
DS3	AR	2.9±2.5
DS3	Trough-counting	1.28±1.15
DS2 (mild-to-no OSA)	AR	2.3±2.4
DS2 (mild-to-no OSA)	Trough-counting	0.82±0.77

Table 6.2: The absolute error between RR derived from the VDR signal and the reference signals for the moderate-to-severe OSA patients in DS2.

Dataset	Method	AE($\mu \pm \sigma$)
DS2 (moderate-to-severe OSA)	AR	2.6±2.2
DS2 (moderate-to-severe OSA)	Trough-counting	0.98±0.84

Table 6.1 shows the AE between RR derived from the VDR signal and the reference signals for mild-to-no OSA patients in DS2 and for normal volunteers in DS3.

Table 6.2 shows the AE between RR derived from the VDR signal and the reference signals for moderate-to-severe OSA patients in DS2.

6.3.3 Breath amplitude to baseline ratios

Figure 6.16 compares the changes in breath amplitude, br_{VDR}, calculated from the VDR signals with the corresponding changes in breath amplitude, br_{RIPSum}, calculated from the RIPSum signal. Figure 6.17 compares the overall histograms of changes in breath amplitude from baseline in the RIPSum signals and VDR signals.

6.4 Discussion

In this chapter a technique for deriving a respiratory signal from infra-red low quality video recordings of sleeping subjects was presented. The proposed method was based

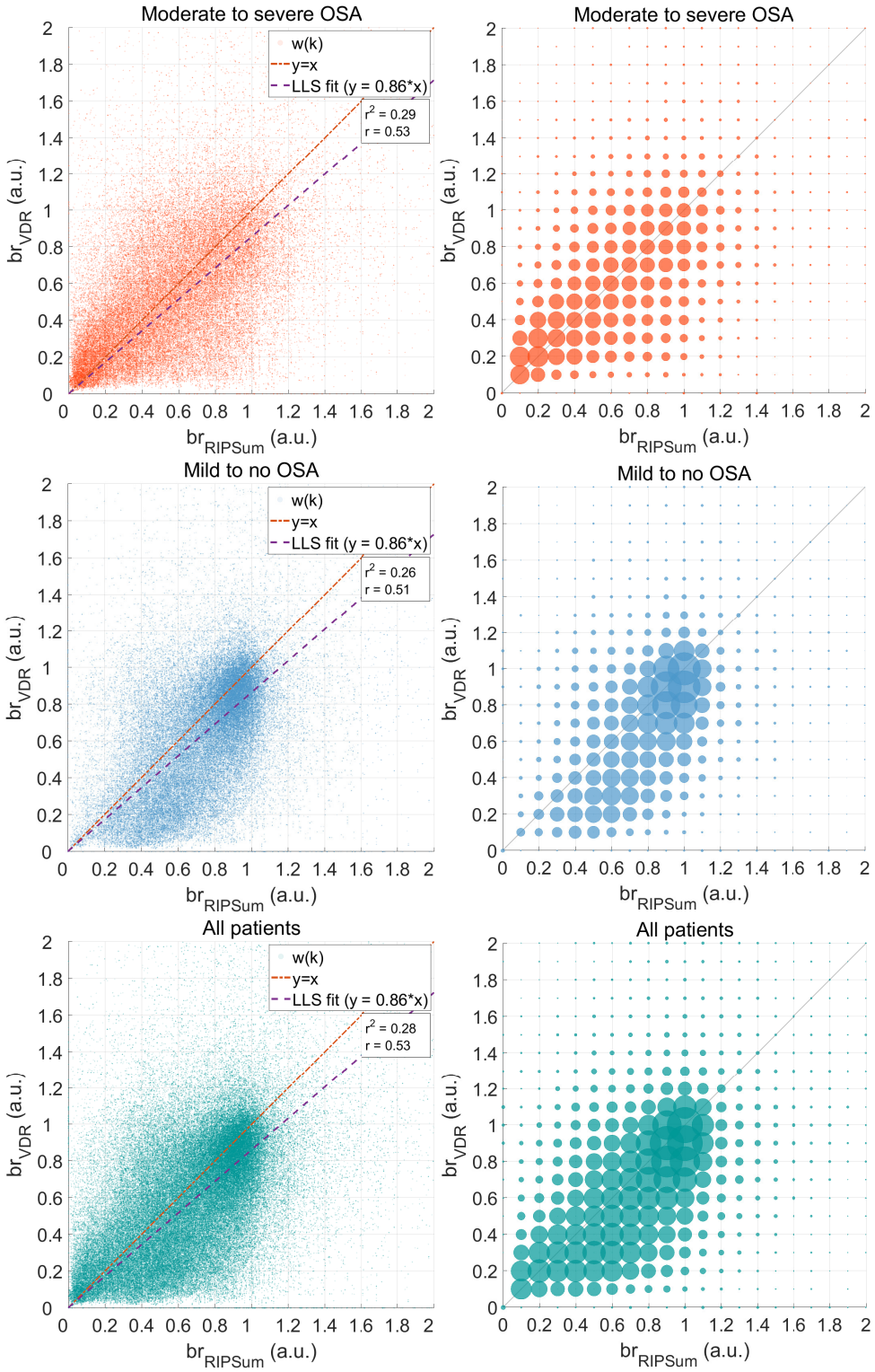


Figure 6.16: The changes in breath amplitude from baseline, br_{VDR}, calculated from the VDR signals compared with the corresponding changes in breath amplitude from baseline, br_{RIPSum}, calculated from the RIPSum signals. The figures on the left side show scatter plots with detailed values. The figures on the right side show scatter plots with values rounded to one decimal point with the size of the markers scaled to the number of points taking that value. br_{VDR} and br_{RIPSum} values above 1.0 are due to deep breaths, e.g. during awakenings, which are larger than baseline breath amplitude. LLS: Linear least squares.

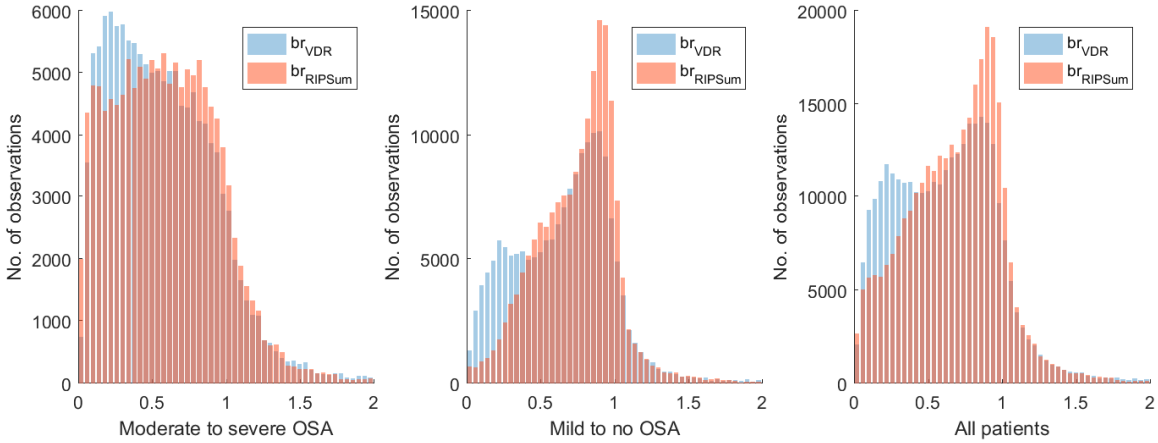


Figure 6.17: Histograms comparing the changes from baseline value in the RIPSum signals (br_{RIPSum}) to changes from baseline value in the VDR signals (br_{VDR}). The br_{VDR} and br_{RIPSum} values above one are due to deep breaths, e.g. during awakenings.

on the repetitive nature of breathing movements with the torso being in relatively the same position during similar stages of inhalation and exhalation. The method was validated on two datasets, one dataset which includes data from 32 normal volunteers and one dataset which includes data from 22 patients with different severity levels of OSA or no OSA. The respiratory signals derived from videos in both datasets have been compared with reference respiratory signals (NAF, TEFF, AEFF) included in the PSG recordings. The sleeping positions vary among subjects and patients and there were no limitations on the pattern, colour, or thickness of bedding and clothing.

If the two methods used to derive the RR are compared, it is found that the trough-counting method has a lower error. The AR model was found to select frequencies for RR that were too high which is in agreement with the findings of Loerup [50]. Loerup found that particularly for RR below 16 bpm (which is often the case for RR during sleep) the AR model is more sensitive to high frequency noise and produces RR estimates that are too high.

The comparison of the RR derived from the video recordings of normal volunteers

in DS3 using the trough counting method with the RR derived from the reference signals showed an AE of 1.28±1.15 bpm ($\mu \pm \sigma$). For the same method, the comparison of the RR derived from the video recordings of the 13 mild-to-no OSA patients in DS2 with the RR derived from the reference signals showed an AE of 0.82±0.77 bpm ($\mu \pm \sigma$). Finally, the comparison of the RR derived from the video recordings of the 9 moderate-to-severe OSA patients in DS2 with the RR derived from the reference signals showed an AE of 0.98±0.84 bpm ($\mu \pm \sigma$). These error rates are noticeably lower compared to typical values of above ±4 bpm for the inter-observer average error rate (error rate between RR measured manually by different nurses and doctors) mentioned in the literature [115, 147]. The RR error rates obtained in this chapter also outperform the error rates reported in the review of different RR estimation algorithms from PPG and ECG signals by Charlton *et al.* [148] (the standard deviation of error in RR estimations reported here is less than half the standard deviation of the error of the algorithms investigated by Charlton *et al.*).

Despite the resolution of the video data recorded at the hospital being nearly 4 times lower than that of the videos recorded from normal subjects[2], the RR error is lower for the mild-to-no OSA patients in DS2. This may be because most subjects in DS3 are covered with a thick duvet during the night while the patients in DS2 were only provided with bed sheets. Figure 6.15c shows that the RR estimates for normal subjects in DS3 are lower than the reference values for many of the cases which also indicates that the thick duvet on the subjects has made the respiratory movements more difficult to detect.

[2]Hospital video recordings had a resolution of 352×288 pixels and normal subject video recording had a resolution of 720×576 pixels.

The low error for the RRs evaluated from hospital patients demonstrates the feasibility of using low-cost low-resolution cameras for respiration monitoring in hospital settings. Remote contact-less sleep monitoring is particularly important as a comfortable environment makes a significant difference in the quality of sleep. The small difference between the error rates for mild-to-no OSA patients and the moderate-to-severe OSA patients in DS2 indicates that it is possible to derive a reliable RR from the VDR signals despite the large number of awakening-related movements in moderate-to-severe OSA patients.

The scatter plots comparing the changes in breath amplitude from baseline between the RIPSum and VDR signals show a large number of points around the equality ($y = x$) line, especially for moderate-to-severe OSA patients. It can be seen in Figure 6.16 that for these patients, the majority of br_{RIPSum} values that are around 90% drop from the baseline lie between 80% to 90% drop from the baseline in br_{VDR}. This slight difference in the drop from baseline breathing could be due to the RIP bands measuring the perimeter of the torso around the chest and abdomen while video recordings measure the expansions and contractions of the whole torso. It could also indicate that a slightly higher threshold is needed to detect episodes of apnoea in the VDR signal. For mild-to-no OSA patients the VDR signal seems to over-estimate the breath amplitude drops from baseline. This is mainly because many non-OSA patients in the DS2 dataset suffer from other sleep disturbances such as PLMD, RLS, and insomnia that cause a severe number of awakenings/movements overnight. These non-SBD related movements have a greater impact on the video recordings which record the whole body than on the RIP signals.

One issue with second-to-second comparison of the br_{RIPSum} and br_{VDR} values is that as explained in Section 6.2.1 in the PSG system there was a delay between the onset of the recording of the physiological signals (e.g. RIP signal) and video recording. Therefore, the video recordings were synchronised with the rest of the physiological signals using the delay stated in the log files of the PSG recordings. The synchronisation of the VDR and RIPSum signals was then verified by taking the cross-correlation between the two signals for each patient. However, the two signals were still sometimes a few seconds out of sync as the difference in the shape of the signals made it difficult to synchronise the signals to the exact second. Therefore, when one point on the br_{RIPSum} signal shows a 90% drop from the baseline, the same amount of drop from the baseline in the br_{VDR} signal may occur a few seconds before or after, causing low correlation coefficients in Figure 6.16.

The histograms presented in Figure 6.17 show that, in particular, for moderate-to-severe OSA patients the distribution of br_{RIPSum} values is close to that of br_{VDR}. The histograms for mild-to-no OSA patients show a greater number of br_{VDR} values between 0 to 0.4 than br_{RIPSum}. This, as explained above, is likely to be related to non-OSA patients suffering from other sleep disturbances with several awakenings overnight.

Overall, the low error rates on the RRs derived from the VDR signals in this chapter are a confirmation of the possibility of remotely and continuously monitoring patients' RR using video cameras. Additionally, using video recordings for RR monitoring resolves some of the issues with the standard methods for RR monitoring (sensors such as RIP bands and oronasal flow cannula) which are restrictive for the

patients and not environmentally friendly if the attachments are disposable.

The results of monitoring breath amplitude from VDR signals show a correlation between the drops in breath amplitude from baseline between the VDR signals and RIPSum signals. This correlation is stronger for moderate-to-severe OSA patients who have frequent episodes of reduced breathing and awakening but go back to sleep immediately after an awakening due to fatigue. br_{VDR} and br_{RIPSum} become less correlated for mild-to-no OSA patients. As many of these patients suffer from other sleep disturbances, they will benefit from further screening to investigate the cause of their sleep fragmentation.

Chapter 7

Detection of apnoeic events using video derived respiration

7.1 Introduction

As previously discussed in chapter 2, the standard for the diagnosis of OSA is overnight PSG in a sleep laboratory. From the PSG studies, the AHI, which is the key measure for quantifying the severity of OSA, is derived. AHI is defined as the number of apnoeas and hypopnoeas per hour of sleep. An obstructive apnoea is defined as $\geqslant$ 90% reduction of airflow for at least 10 seconds with presence of respiratory effort. A hypopnoea is defined as $\geqslant$ 30% reduction of airflow for at least 10 seconds with $\geqslant$ 3% reduction in oxygen saturation or a terminating arousal [2]. The duration of apnoeic/hypopnoeic events normally does not exceed 2 minutes and the events usually last between 20 to 40 seconds [20]. The severity of apnoea is categorised into 4 classes: (a) no OSA (AHI<5), (b) mild (AHI$\geqslant$ 5), (c) moderate (AHI$\geqslant$ 15), and (d) severe

(AHI$\geqslant$ 30) [2].

When calculation of AHI is not possible, the ODI, has been used as an alternative measure for the severity of OSA. ODI is the number of oxygen desaturations per hour of sleep. Desaturations have been defined as $\geqslant 3\%$ or $\geqslant 4\%$ reduction in oxygen saturation. The $\geqslant 3\%$ threshold was used here as it is the latest recommended threshold in the AASM manual for scoring sleep events [79].

In common practice, the PSG recordings are visually analysed by sleep specialists, investigating all the abnormalities present in the respiratory signals as well as the rest of the PSG signals. Considering the length of overnight PSG recordings and the complexity of the apnoea scoring process involved, the manual analysis of PSG studies is cumbersome, costly, and consumes medical resources. This leads to long waiting lists for overnight sleep studies and limited availability of sleep clinics (see Table 1.2). Therefore, a system that completely or partially automates the PSG analysis process may have a significant impact on costs and medical resources.

As a result, the development of automated techniques for the detection of OSA, severity of OSA, and apnoeic events has become an area of research interest in recent years. The proposed methods for automated detection of apnoeic events in the literature mainly use a single or a combination of the physiological signals in the PSG recordings. In a comprehensive review, Alvarez-Estevez *et al.* [20] have classified these approaches into three categories based on the signals they use: i) pulse oximetry, ii) single respiratory signal, iii) multiple respiratory signals with or without other PSG signals.

One of the simplest yet most informative sensors to measure respiratory distur-

bances during sleep is the pulse oximeter. The oxygen saturation signal from pulse oximetry has been used by Burgos *et al.* [149] to develop an on-the-fly classifier to detect abnormalities in 60-second segments of the signal. Gil *et al.* [150] used the PPG signal from pulse oximetry and investigated the relation between the decreases in the amplitude of the PPG signal and apnoeic events in children.

Other groups have used respiratory signals alone without pulse oximetry as a simplified form of apnoea detection. Macey *et al.* [151] developed a model to detect central apnoeic events in infants using an abdominal respiratory signal. Varady *et al.* [152] used artificial neural networks to detect normal breathing, apnoea, and hypopnoea in 16-seconds windows of NAF and RIP signals. Han *et al.* [153] used a threshold on the mean magnitude of the second derivative of the NAF signal to detect apnoeic events. Nakano *et al.* [154] developed an algorithm using power spectral analysis of the NAF signal to detect apnoeic events and calculate an RDI. Nazeran *et al.* [155] used fuzzy logic to detect apnoeic events in the respiratory airflow signal. Houdt *et al.* [156] used the nasal pressure and thoracoabdominal RIP signals to detect apnoeic events by analysing the duration, amplitude and slope of individual breaths.

The use of pulse oximetry signal or respiratory signals alone is not ideal as each of these signals may underestimate or overestimate the number of apnoeic/hypopnoeic events [157]. A number of research groups have used a combination of respiratory signals and pulse oximetry which is closer to the AASM recommendations for scoring sleep events, particularly, if arousals are taken into account for scoring hypopnoeas in addition to oxygen desaturations. Taha *et al.* [158] differentiated apnoeic and hypopnoeic events by amplitude-based analysis of the RIPSum signal and oxygen

saturation. Tian *et al.* [159] used nasal airflow and oxygen saturation signals and artificial neural networks to detect apnoeic events. By performing an amplitude-based analysis on the combination of nasal airflow and PPG signals, Sommermeyer *et al.* [160] attempted to detect and differentiate between obstructive and central apnoeas. The amplitude changes in the PPG signal were used to detect arousals. Maali *et al.* [161], Otero *et al.* [162], and Alvarez-Estevez *et al.* [163] used the thoracoabdominal RIP, nasal airflow, oxygen desaturation signals and fuzzy logic to detect apnoeic events. Al-Angari *et al.* [164] extracted features from the RIP signals, and combined those features with heart rate variability of the ECG signal and oxygen saturation to detect apnoea in 60-second epochs.

Overall, the research on automated detection of apnoeic events using PSG signals shows strong agreement between the detected events and expert manual scoring. Many PSG systems, therefore, provide some degree of automated event analysis, ranging from presenting an AHI score to individually marked events.

Despite the increasing number of publications on using video cameras for patient monitoring, to the authors knowledge only two research groups have used video derived respiratory signals to perform some form of automated apnoea detection on clinical data. Takemura *et al.* [165] proposed a model to differentiate obstructive vs. central apnoeas and evaluated their method on five patients. Their proposed model measures the movement of 100 or more sampling points of the upper-body using a non-contact 3-dimensional position sensor and outputs the sum total as respiratory movement. The obstructive events are differentiated from central events based on the reduction in respiratory effort in abdomen and thorax. Wang *et al.* [166] pro-

posed a system that uses frame differencing and background modelling followed by an adaptive template of normal breathing to differentiate normal vs. apnoeic breathing. Their method was tested on nine 15-minute video clips from eight-hour recordings of three OSA patients.

The work in this chapter focuses on using overnight video recordings of sleep recorded by contact-less infra-red cameras to detect OSA events. In particular, we investigate the possibility of using the VDR signals (as explained in Chapter 6) from 247 patients, to detect individual apnoeic/hypopnoeic events as well as to differentiate moderate-to-severe OSA patients from mild-to-no OSA patients based on the number of detected events.

7.2 Data

In this chapter two sources of data were used for the studies: (i) DS1 which includes PSG recordings of 247 patients who were referred to the sleep lab of Churchill Hospital (Oxford, UK) with suspected OSA. The majority of the PSG studies in DS1 do not include reference respiratory signals (NAF, TEFF, and AEFF). (ii) DS2 which includes 22 of the patients in DS1 whose PSG recordings include reference respiratory related signals. PSG recordings of patients in both DS1 and DS2 include infra-red video recordings in grey scale in MPEG2 format with a 352×288 resolution and 25 fps frame rate. In the PSG system the recording of the physiological signals (e.g. the RIP signals) started a with a delay from the video recording. The amount of delay is recorded in the log files of each PSG recording. Therefore, the video recordings were synchronised with the rest of the physiological recordings using the delay stated in

the log files. The data is used from 2 hour to 6 hour into the recordings to increase the likelihood of patients being asleep during the period of study (for details of these datasets see Section 3.4 of Chapter 3).

7.3 The ground truth for the evaluation of OSA events

Scoring of individual apnoeic events by clinicians or the PSG software was not available for any of the patients in our datasets. The reference OSA events were therefore identified from the sum of thoracic and abdominal RIP signals (the RIPSum signal) and the oxygen saturation signal as described below. These reference events were then used to assess the performance of the proposed method for the identification of OSA events in the VDR signals.

7.3.1 Oxygen desaturations

Desaturations of $\geqslant 3\%$ were detected in the oxygen saturation signal by first identifying all the peaks, pk_{SpO_2}, and troughs, tr_{SpO_2}, in the oxygen saturation signal. For each trough in tr_{SpO_2}, the closest previous peak and next peak to the trough are denoted with pk_{prev} and pk_{next}. If the difference between the trough and pk_{prev} was $\geqslant 3\%$, the period between pk_{prev} and pk_{next} that was $\geqslant 3\%$ below pk_{prev} was identified as a desaturation event. If pk_{next} was not $> 3\%$ from the trough, a re-saturation of $> 2\%$ was considered as end of event. The bottom plot in Figure 7.1 shows a section of the oxygen saturation signal from a severe OSA patient and the identified desaturation

events.

7.3.2 Apnoeas and hypopnoeas

For the detection of the apnoeic and hypopnoeic events in the RIPSum signal, the br_{RIPSum} signal derived in the previous chapter was used. For the detection of apnoeic events, any reduction in br_{RIPSum} to a value $\leqslant 0.1$ [1] (breath amplitudes with more than 90% drop from the baseline $b_{RIPSsum}$) was identified as the start of an apnoeic event with the end of the event set to the first following point $\geqslant 0.1$. For the hypopnoeic events, any decrease in br_{RIPSum} to values $\leqslant 0.7$ [2] (breath amplitudes with more than 30% drop from the baseline b_{RIPSum}) was identified as the potential start of a hypopnoeic event (denoted with e_{Start}) with the potential end of the event set to the first following point $\geqslant 0.7$ (denoted with e_{End}). If an apnoea event existed between e_{Start} and e_{End} or a desaturation of $\geqslant 3\%$ existed between e_{Start} and e_{End}+10 seconds, the event was identified as hypopnoea and otherwise discarded. Note that the addition of 10 seconds to e_{End} is due to the length of $w(k)$ being 10 seconds (see Section 6.2.4.2), therefore, e_{Start} to e_{End} in the br_{RIPSum} signal corresponds to e_{Start} to e_{End}+10 in the oxygen saturation signal. Finally, based on the AASM guidelines [146], hypopnoeic events that included an apnoeic event inside were identified as an apnoeic event instead of hypopnoea. The top plot in Figure 7.1 shows a section of the br_{RIPSum} signal from a severe OSA patient and the identified apnoeic and hypopnoeic events.

As neither of the reference respiratory signals were available for the majority of the

[1] Based on the apnoea threshold defined in the AASM manual for scoring sleep events [79].
[2] Based on the hypopnoea threshold defined in the AASM manual for scoring sleep events [79].

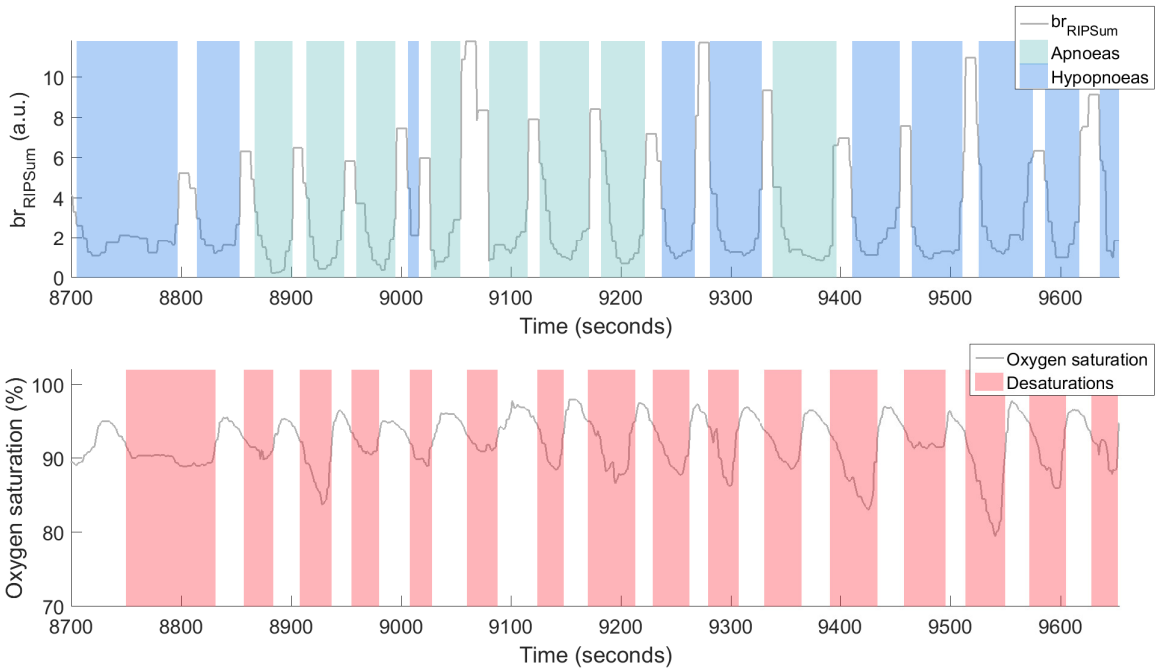


Figure 7.1: Top plot: The br_{RIPSum} and the apnoeic and hypopnoeic events of a severe OSA patient. Bottom plot: the corresponding oxygen saturation signal and the desaturation events. Desaturations are defined as drops of $\geqslant 3\%$ from the baseline.

patients in DS1, the identified OSA events were compared with desaturations. ODI has been shown to have a strong correlation with AHI [167]. Figure 7.2 compares the ODI values and AHI values for the DS2 patients.

7.4 VDR apnoea and hypopnoea thresholds

As discussed at the end of the previous chapter, the RIPSum and VDR signals both measure the changes in the breath amplitude of the person in a slightly different way. While RIPSum measures the expansions and contractions of torso as the perimeter of chest and abdomen, the VDR measurements of the expansions and contractions of torso as explained in the previous chapter are slightly different. As a consequence, the thresholds defined for the RIPSum signal may not be optimum for the VDR signals. Additionally, the standard definition of hypopnoea requires a drop in the amplitude

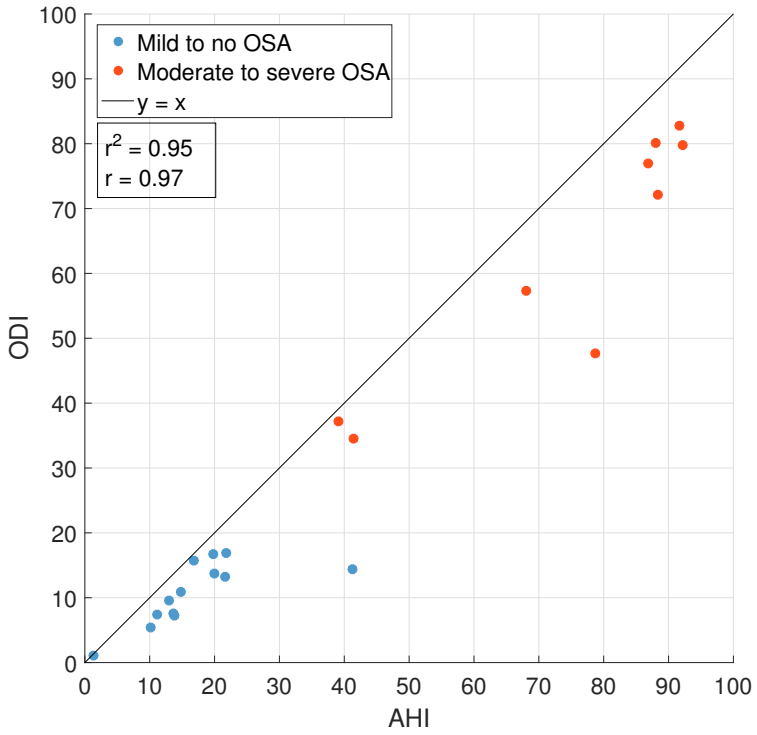


Figure 7.2: Comparison of the number of desaturations per hour (ODI) and number of OSA events per hour (AHI) for the DS2 patients.

of the signal as well as a drop in the oxygen saturation. Since the oxygen saturation measurements are not available when monitoring patients with only a video camera, it is quite likely that a lower threshold for VDR hypopnoea events is required.

To search for apnoea and hypopnoea thresholds tailored for the VDR signals, the 22 patients in DS2 were divided into two groups of training and test denoted with $DS2_{Train}$ and $DS2_{Test}$ here. The patients in $DS2_{Train}$ and $DS2_{Test}$ were randomly selected making sure that approximately equal number of patients from each OSA severity were assigned to each group. Table 7.1 shows the number of patients in each OSA severity group in $DS2_{Train}$ and $DS2_{Test}$.

For finding the threshold for apnoea, a range of thresholds, thr_a, from 0.05 to 0.3 with steps of 0.01, were tested on the $DS2_{Train}$ br_{VDR} signals. The points below thr_a in br_{VDR} were compared against the $DS2_{Train}$ $br_{RIPSum} \leqslant 0.1$. For each point on the

Table 7.1: Number of patients in each OSA severity group in $DS2_{Train}$ and $DS2_{Test}$.

Severity	$DS2_{Train}$	$DS2_{Test}$
Non OSA	5	5
Mild OSA	2	1
Moderate OSA	2	1
Severe OSA	3	3

$br_{RIPSum} \leqslant 0.1$, if br_{VDR} was below thr_a, this was labelled as a true positive. For each point on the $br_{RIPSum} \leqslant 0.1$, if br_{VDR} was above thr_a, this was labelled as a false negative. Finally, for each point on the $br_{RIPSum} > 0.1$, if br_{VDR} was below thr_a, this was labelled as a false positive. The threshold 0.12 was found to provide the highest F_1score and was chosen as the apnoea threshold, thr_a. F_1score was calculated the standard way, as defined in Appendix B.

Similarly, for finding the threshold for hypopnoea, a range of thresholds, thr_h, from 0.5 to 0.8 with steps of 0.01, were tested on the $DS2_{Train}$ br_{VDR} signals. The points below thr_h in br_{VDR} were compared against br_{RIPSum} hypopnoeic events. For each point on the br_{RIPSum} inside a hypopnoeic event, if br_{VDR} was below thr_h, this was labelled as a true positive. For each point on the br_{RIPSum} inside a hypopnoeic event, if br_{VDR} was above thr_h, this was labelled as a false negative. And finally, for each point on the br_{RIPSum} inside a hypopnoeic event, if br_{VDR} was below thr_h, this was labelled as a false positive. The threshold 0.62 was found to provide the highest F_1score and was chosen as the hypopnoea threshold, thr_h. The apnoea and hypopnoea thresholds were then tested on $DS2_{Test}$ br_{VDR} signals.

Note that according to the AASM manuals, hypopnoea events that include an apnoeic event should be identified as apnoea, however, doing so would result in finding wrong thresholds for apnoea and hypopnoea. Therefore, only for finding the thr_a and

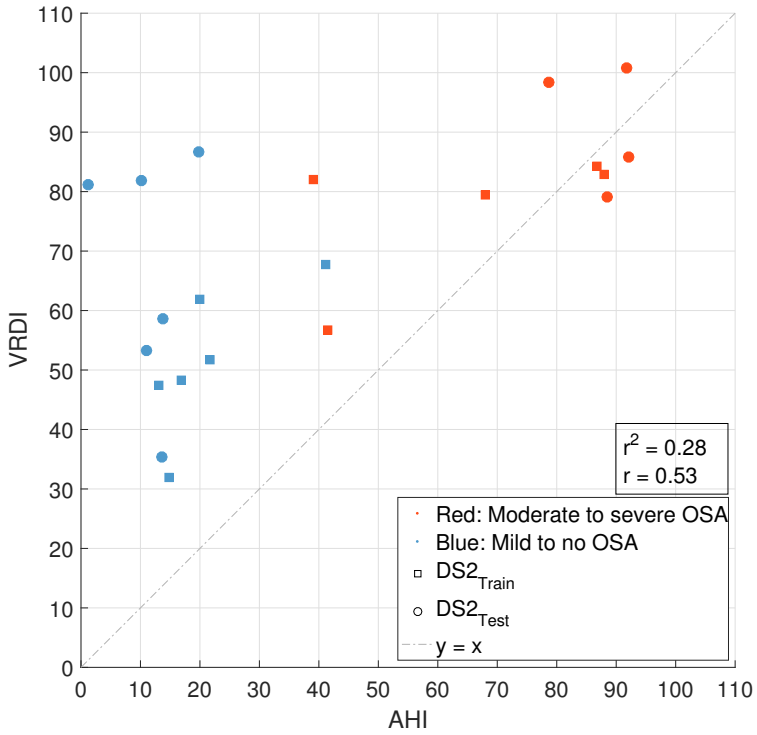


Figure 7.3: Comparison of the number of OSA events per hour identified using br_{VDR} with the events identified using br_{RIPSum} for each individual patient in DS2.

thr_h thresholds, these hypopnoea events were treated as hypopnoea. In all other cases in this chapter, both for VDR signals and RIPSum signals, hypopnoea events including apnoea were regarded as apnoea.

Figure 7.3 compares the number of OSA events per hour identified using br_{VDR} signals for each individual patient in $DS2_{Train}$ and $DS2_{Test}$ with the number of OSA events per hour identified using br_{RIPSum}. The number of OSA events per hour identified using br_{VDR} signals is denoted with Video Respiratory Disturbance Index (VRDI). Figure 7.4 compares the average number of apnoea and/or hypopnoea events per hour identified using br_{VDR} signals with the average number of apnoea and/or hypopnoea events per hour identified using br_{RIPSum} in $DS2_{Train}$ and $DS2_{Test}$.

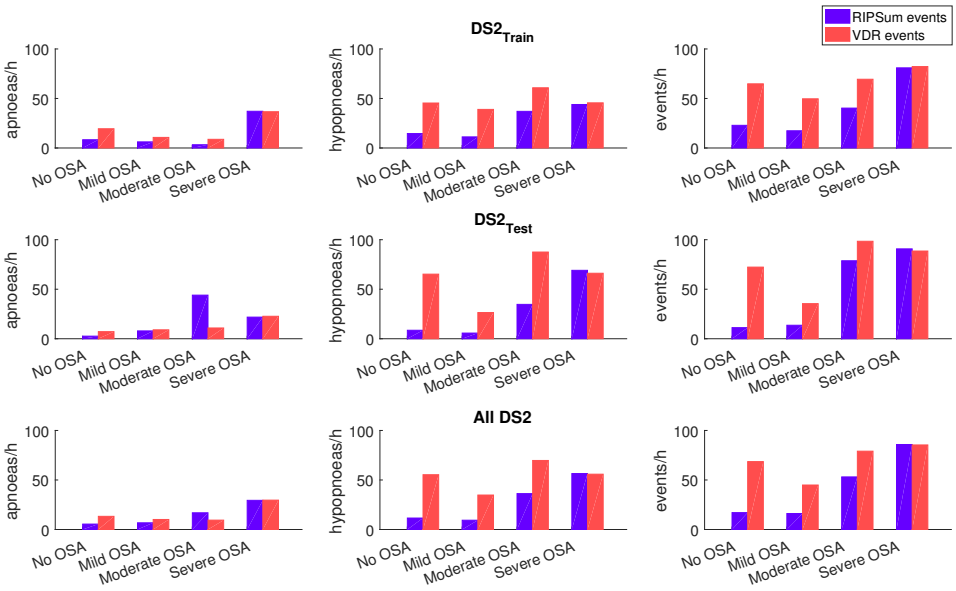


Figure 7.4: Comparison of the average number of apnoea and/or hypopnoea events per hour identified using br_{VDR} signals with the average number of apnoea and/or hypopnoea events per hour identified using br_{RIPSum}.

7.5 Refining event detection using audio

As shown in Figures 7.3 and 7.4, the hypopnoeic events detected using br_{VDR} signals include a number of false positives particularly causing a high number of events per hour for mild-to-no OSA patients. When the RIPSum signal is used to identify the hypopnoeic events, many of the $\geqslant 30\%$ reductions in the breath amplitude are rejected as there is no desaturation associated with them. Without the use of oxygen saturation signal, there is a need for some additional information to filter out the falsely detected hypopnoeic events. One source of data that usually accompanies the video recordings is the embedded audio signal which provides additional information about OSA events. Since most OSA patients snore in their sleep and the snoring stops or is reduced during OSA events, the presence of snoring sound during the identified events may help with identifying the false positives. Additionally, OSA

events usually end with a deep breath, a choking/snorting sound, and resumption of snoring [168, 169] all of which cause a spike in the audio signal. The power of the audio signal is therefore expected to reach a local minimum during OSA events followed by a local maximum caused by the awakening.

To use the audio recordings for the detection of false positive events, the recordings extracted from video data were down-sampled from 44 kHz to 16 Hz. The down-sampled 16 Hz frequency is enough to identify snoring and awakening related choking sounds. The audio signals were then normalised to have unit variance and divided into sliding 10-second windows, $w(k)$, moving with steps of 1 second. If the down-sampled audio signal is denoted with Aud and the 16 Hz sampling frequency is denoted with fs, $w(k)$ is as follows:

$$w(k) = \{Aud_{(k-1)*10*fs+1}, ..., Aud_{k*10*fs}\} \tag{7.1}$$

The maximum energy of each $w(k)$, denoted with $E_a(k)$, was defined as follows:

$$E_a(k) = max\{Aud^2_{(k-1)*10*fs+1}, Aud^2_{(k-1)*10*fs+2}, ..., Aud^2_{k*10*fs}\} \tag{7.2}$$

For each of the hypopnoeic events identified in the previous section, the minimum E_a value within the event was compared to the maximum E_a value just after the event (the awakening).

Let us denote the start and end of the current hypopnoeic event with $e_{Start}(i)$ and $e_{End}(i)$ and the start and end of the next OSA event (apnoea or hypopnoea) with $e_{Start}(i+1)$ and $e_{End}(i+1)$. $R_h(i)$, for the ith hypopnoeic event, was defined as the

ratio between the minimum E_a value within $e_{Start}(i)$ and $e_{End}(i)$ to the maximum E_a value between $e_{End}(i)$ and $e_{Start}(i+1)$:

$$R_h(i) = \frac{min(E_a(e_{Start}(i)), ..., E_a(e_{End}(i)))}{max(E_a(e_{End}(i)), ..., E_a(e_{Start}(i+1)))} \tag{7.3}$$

For moderate and severe OSA patients, the OSA events are usually occurring repeatedly within a short time interval. For mild-to-no OSA patients, however, there might be a long interval between the end of one event to the start of the next event. Since the choking sound should occur right after the end of an OSA event, if the interval between $e_{End}(i)$ and $e_{Start}(i+1)$ was more than 30 seconds, in Eq. 7.3 $e_{Start}(i+1)$ was replaced with $e_{End}(i) + 30$.

During OSA events, R_h is expected to be lower than normal breathing because of the cessation of snoring (small minimum value) followed by a deep breath, choking sound, and resumption of snoring (large maximum value). During normal breathing with no snoring and when the patient is snoring, R_h is expected to be higher.

The threshold for R_h was found using $DS2_{Train}$. If R_h for a hypopnoeic event was below a threshold, thr_{Aud}, the event was kept and removed otherwise. The events that were kept during this process are going to be labelled as 'refined events' from now on. The value of thr_{Aud} was varied from 0.1 to 1 in steps of 0.01. For each value of thr_{Aud} the refined events were compared with the events identified in the RIPSum signals. For each point within an OSA event in the br_{RIPSum}, if the point was identified as OSA event in the refined br_{VDR} events, this was labelled as a true positive. For each point within an OSA event in the br_{RIPSum}, if the point was not identified as OSA event in the refined br_{VDR} events, this was labelled a false negative. For each point

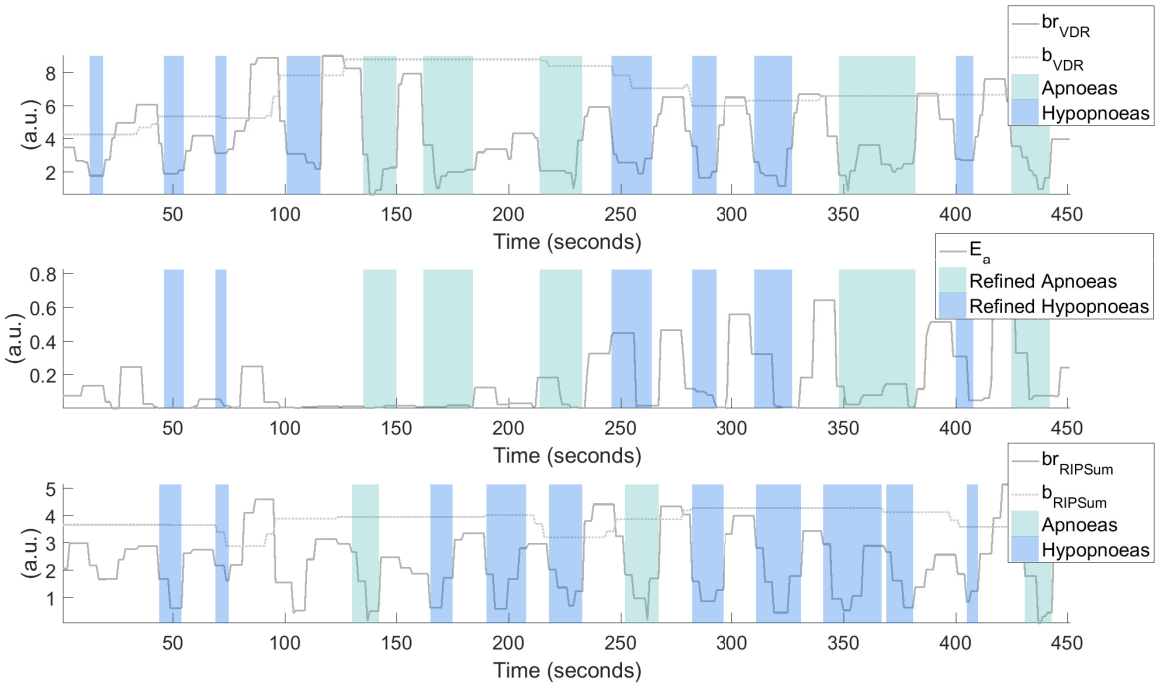
not in an OSA event in the br_{RIPSum}, if the point was identified to be in an OSA event in the refined br_{VDR} events, this was labelled as a false positive. The optimum threshold with the highest F_1score was found to be $thr_{Aud} = 0.39$. Figures 7.5a and 7.5b show examples of false positive events that were removed when audio was used to refine the events.

Figure 7.6 compares the number of OSA events/h identified using br_{VDR}, the number of refined OSA events/h (refined using audio), and the number of OSA events/h identified using br_{RIPSum} for each patient in DS2. Figure 7.7 compares the average number of OSA events/h identified using br_{VDR}, the average number of refined OSA events/h, and the average number of OSA events/h identified using br_{RIPSum} for each OSA severity group.

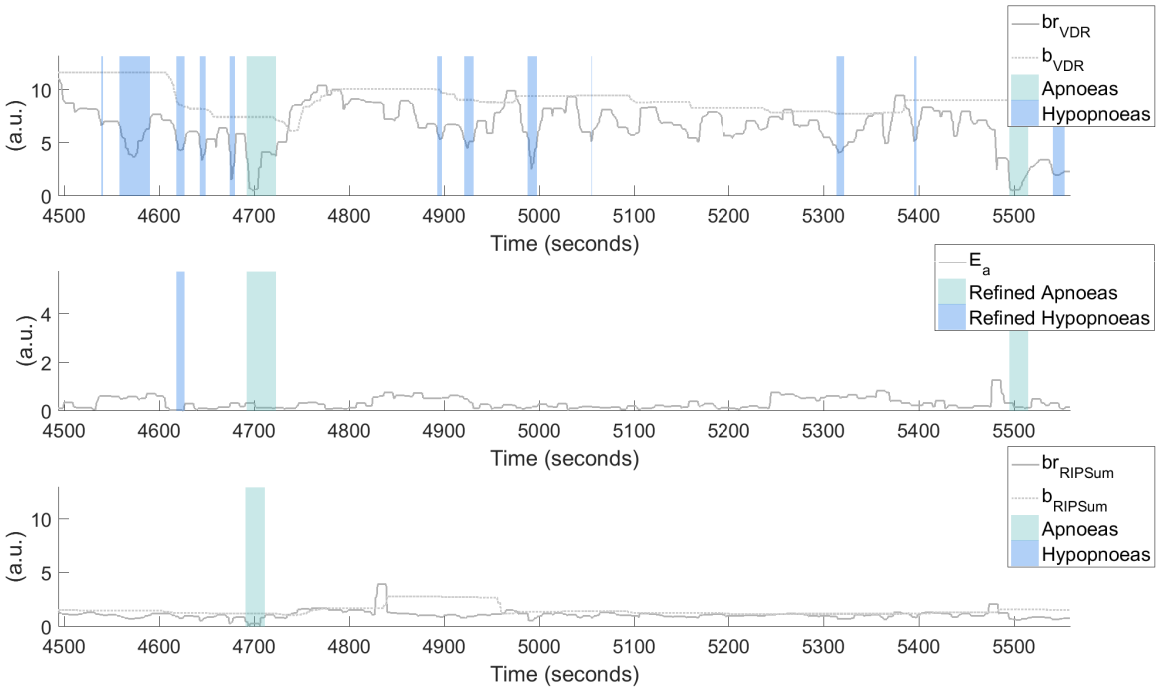
7.6 Identifying OSA events for DS1 patients

In this section we assess the performance of OSA event detection on DS1 dataset with 247 patients. The majority of the patients in DS1 do not have reference respiratory recordings of their sleep due to undergoing reduced more convenient PSG recordings in the sleep clinics at Oxford. In the absence of the respiratory recordings and therefore AHI, the sleep clinicians use the ODI value as an alternative. Using the desaturations as the ground truth for OSA events is not ideal as in some patients desaturations may underestimate or overestimate the number of events. However, as ODI has been shown to have a high correlation with AHI and is used in hospital clinics as an alternative to AHI, we have used the ODI as the reference for DS1.

The same thresholds found in Section 7.4 were first applied to the br_{VDR} signals to



(a)



(b)

Figure 7.5: (a) and (b): Two example figures showing false positive events that were removed using audio. In each figure, the top plot shows the original events identified in the br_{VDR} signal, the middle plot shows the refined events (with removed false positive (FP) events) using audio and the E_a signal, the bottom plot shows the reference events identified in the br_{RIPSum} signal.

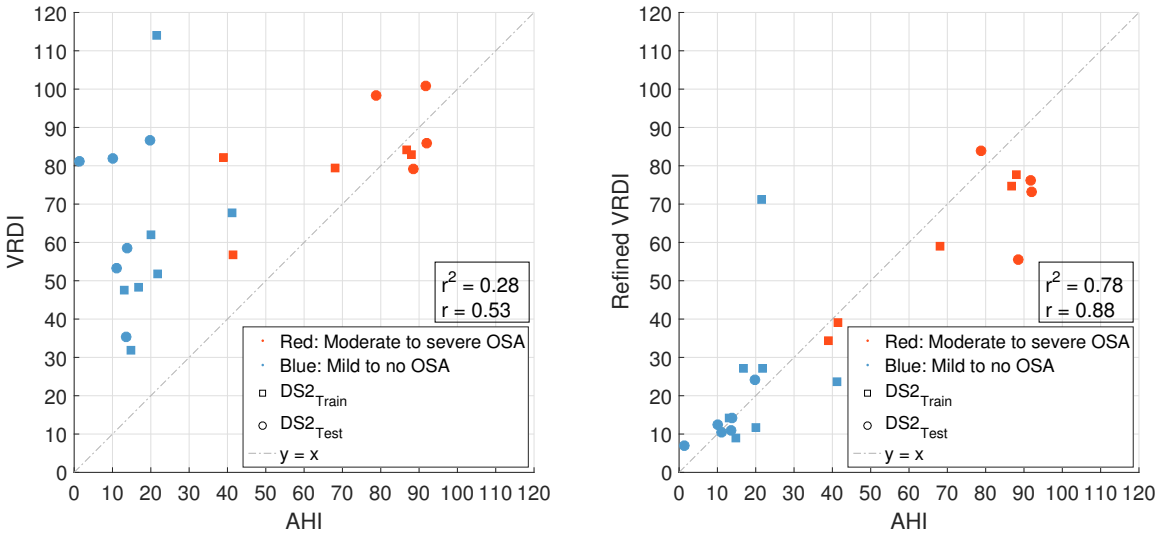


Figure 7.6: Left plot: Comparison of the number of OSA events per hour identified using br_{VDR} and the events identified using br_{RIPSum} for each individual patient in DS2. Right plot: Comparison of the number of refined OSA events per hour identified using br_{VDR} and the events identified using br_{RIPSum} for each individual patient in DS2.

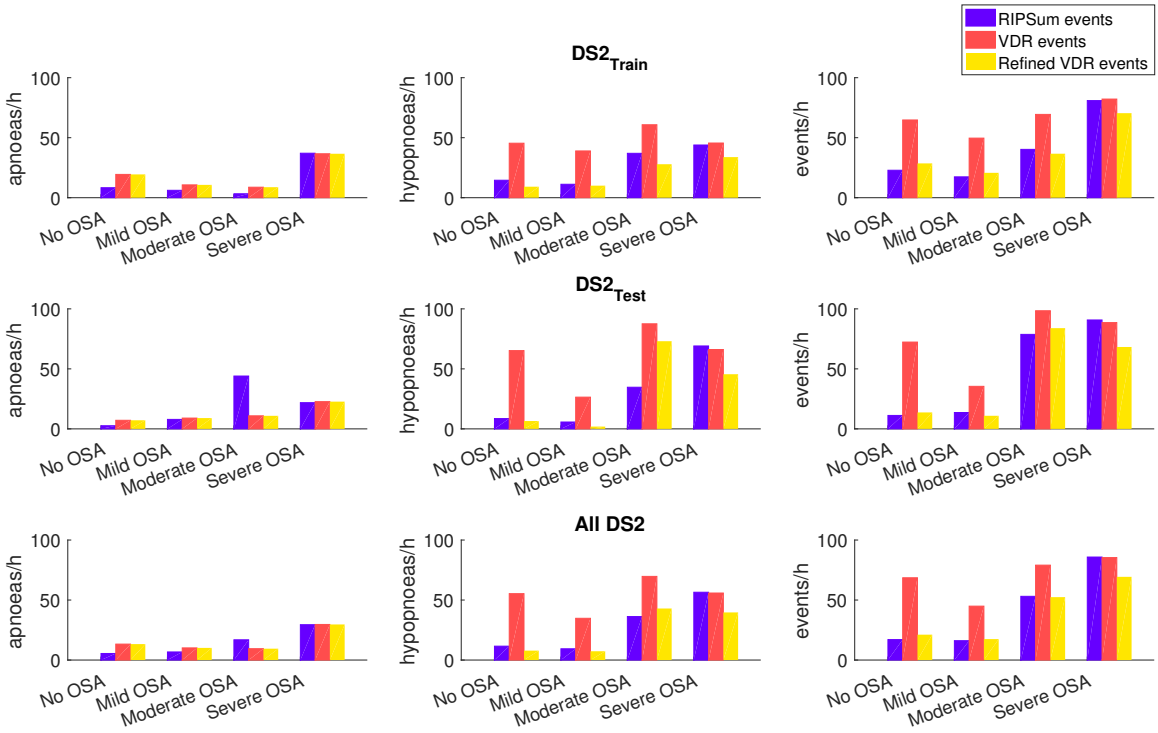


Figure 7.7: Comparison of the average number of apnoea and/or hypopnoea events per hour identified using br_{VDR} signals with the average number of the same events refined using audio and average number of apnoea and/or hypopnoea events per hour identified using br_{RIPSum} for DS2 patients.

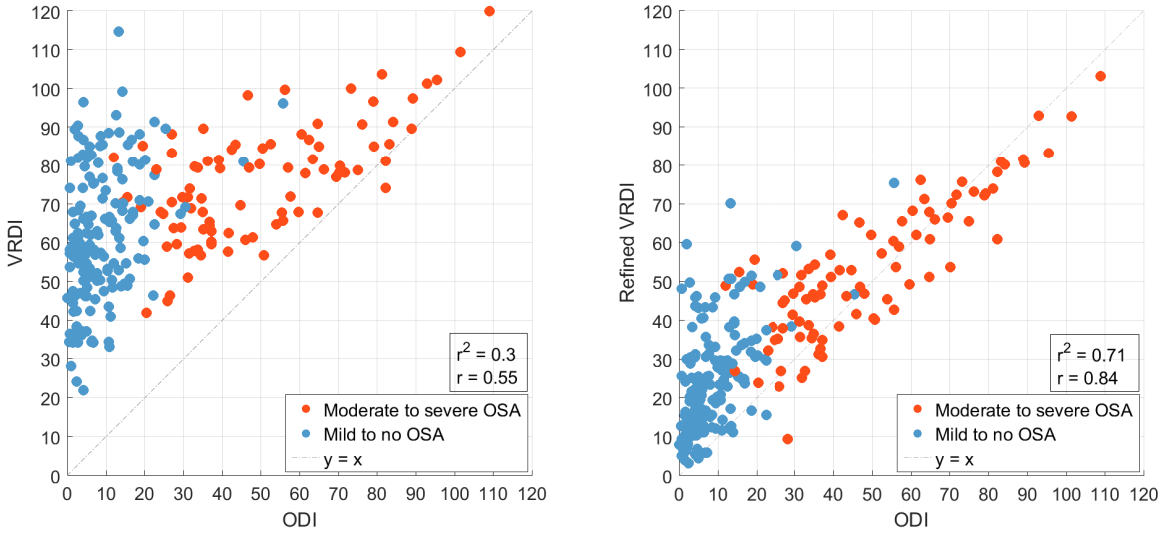


Figure 7.8: Left plot: Comparison of the number of OSA events per hour identified using br_{VDR} and the number of desaturation events per hour for each individual patient in DS1. Right plot: Comparison of the number of refined OSA events per hour identified using br_{VDR} and the number of desaturation events per hour for each individual patient in DS1.

identify the apnoeic and hypopnoeic events of the patients in the DS1 dataset. Then, the events were refined using the audio recordings of the patient with the threshold found in Section 7.5. Figure 7.8 compares the number of OSA events/h identified using br_{VDR}, the number events/h refined using audio, and the number of desaturation/h for each patient in DS1. Figure 7.9 shows the audio refined VRDI values against ODI with the points coloured based on OSA severity assigned to the patients by the clinicians. Figure 7.10 compares the average number of OSA events/h identified using br_{VDR}, the average number of refined events/h, and the average number of desaturations/h for each OSA severity group.

From this point on, VRDI refers to the number of refined VDR events per hour of sleep.

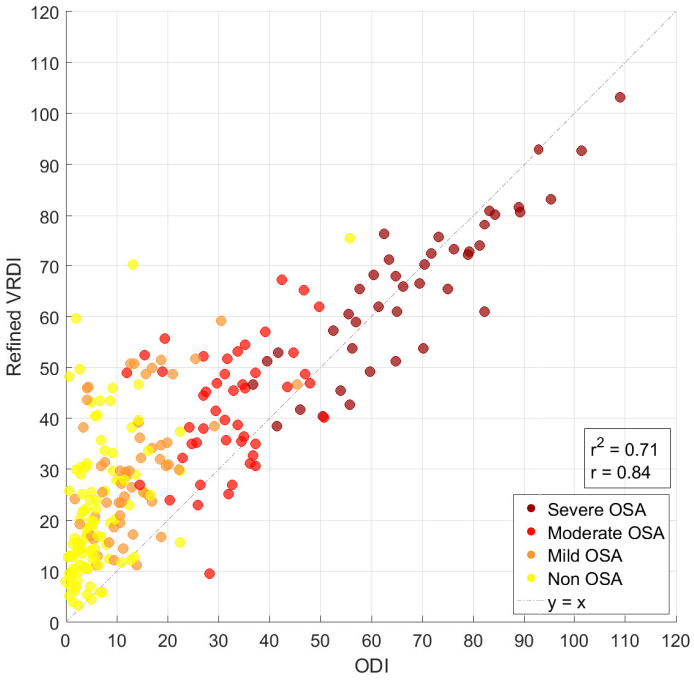


Figure 7.9: Comparison of the number of audio refined OSA events per hour identified using br_{VDR} and the number of desaturation events per hour for each individual patient in DS1. The points on the plots are coloured based on OSA severity assigned to the patients by clinicians.

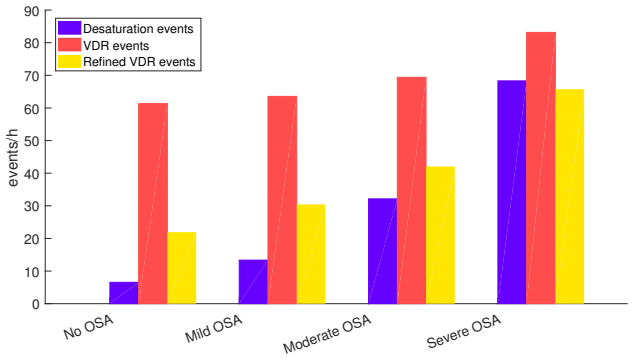


Figure 7.10: Comparison of the average number of events per hour identified using br_{VDR} signals with the average number of events refined using audio and average number of desaturation events per hour for DS1 patients.

7.7 Comparison of the VDR and reference events

In this section the identified events for the DS1 and DS2 datasets are going to be compared on an event-by-event basis. For the DS2 dataset, the events identified in the br_{VDR} signal are compared with the events identified in the reference br_{RIPSum} signal. Since the identified events in the br_{VDR} and br_{RIPSum} signals might have slight different durations or event start and end times, the definitions of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) used for the comparison of the br_{VDR} and br_{RIPSum} events are described below:

- TP: For every event identified in the br_{RIPSum} signal, if within the duration of the event there is an event identified in the br_{VDR} signal, it is labelled a TP.
- FN: For every event identified in the br_{RIPSum} signal, if within the duration of the event there is no event identified in the br_{VDR} signal, it is labelled a FN.
- FP: For every event identified in the br_{VDR} signal, if within the duration of the event there is no event identified in the br_{RIPSum} signal, it is labelled a FP.
- TN: A sliding non-overlapping 30-second window is passed through the br_{RIPSum} and br_{VDR} signals. For every 30-second window of the br_{RIPSum} signal without any OSA events, if there are no events detected in the corresponding 30-second window of the br_{VDR} signal, the widnow is labelled a TN. The 30-second length chosen for the windows is the average duration of apnoeic events [20].

Figure 7.11 shows a few examples of TP and FN events for a patient in DS2 dataset. The identified events are then assessed based on the following metrics, which are

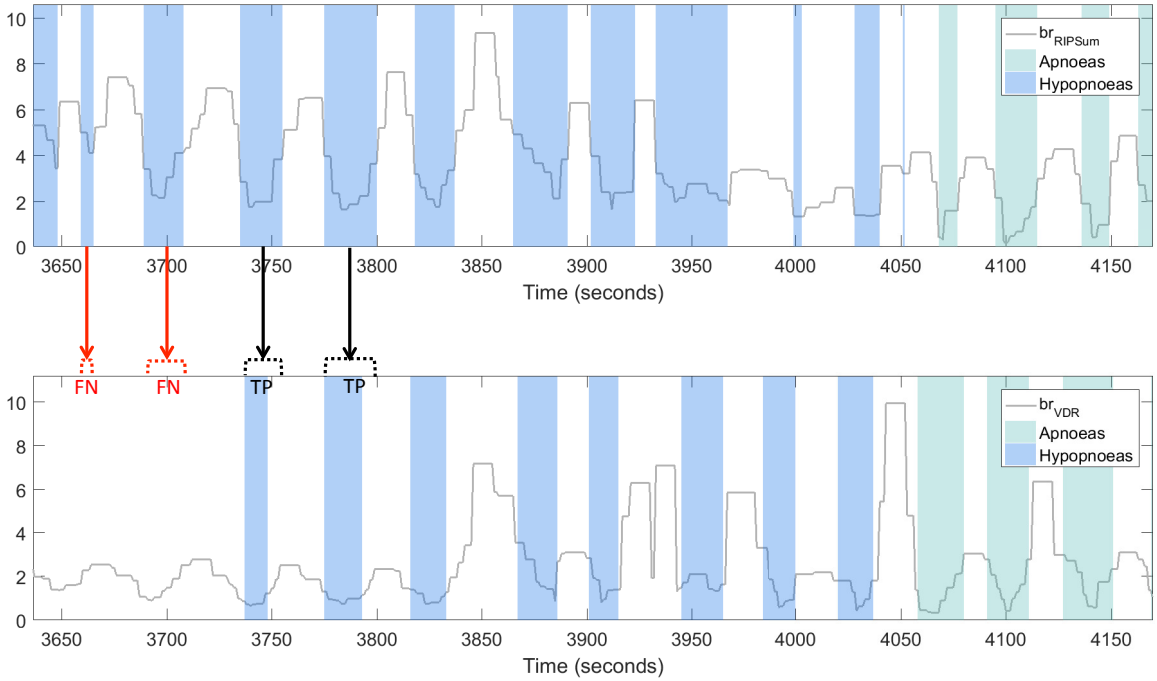


Figure 7.11: A few examples of TP and FN OSA events identified in the br_{VDR} signal for a patient in DS2 dataset. The br_{VDR} events are compared with the OSA events identified in the br_{RIPSum} signal. The dotted brackets show the period corresponding to the start and end of an OSA event in the br_{RIPSum} signal (only the first four events are annotated as an example).

slightly redefined in this section for the event comparison (note that in all other parts of the thesis definitions of these metrics stays the same as the standard definitions defined in Appendix B):

- Se: (Number of TP events)/(Total number of OSA events identified in br_{RIPSum})
- PPV: (Number of TP events)/(Number of TP events + Number of FP events)
- F_1score: $(2 \times Se \times PPV)/(Se + PPV)$
- Sp: (Number of TN 30-second windows)/(Total number of 30-second windows with no OSA events)

Since the RIPSum signals are not available for the majority of the patients in the DS1 dataset, the events identified in the DS1 br_{VDR} signals are compared with the

desaturation events identified in the oxygen saturation signals. The challenge with the comparison of the br_{VDR} OSA events with desaturations is that oxygen desaturations occur with a delay from the onset of OSA events. Although the desaturation delay varies from person to person, average delays of around 30 seconds have been reported in the literature [170]. Therefore, in order to compare the OSA events in the br_{VDR} signal with the desaturations, this delay of 30 seconds is taken into account.

The definitions of TP, TN, FP, and FN used for the comparison of the br_{VDR} events and desaturations are described below:

- TP: For every desaturation event identified in the oxygen saturation signal with the start of the event denoted by e_{Start} and the end of the event denoted by e_{End}, if within e_{Start} - 30 seconds and e_{End} - 30 seconds there is an event identified in the br_{VDR} signal, it is labelled a TP.

- FN: For every desaturation event identified in the oxygen saturation signal with the start of the event denoted by e_{Start} and the end of the event denoted by e_{End}, if within e_{Start} - 30 seconds and e_{End} - 30 seconds there is no event identified in the br_{VDR} signal, it is labelled a FN.

- FP: For every event identified in the br_{VDR} signal with the start of the event denoted by e_{Start} and the end of the event denoted by e_{End}, if within the e_{Start} + 30 seconds and e_{End} + 30 seconds there is no desaturation event identified in the oxygen saturation signal, it is labelled a FP.

- TN: A sliding non-overlapping 30-second window is passed through the oxygen saturation signal. For every 30-second window of the oxygen saturation signal

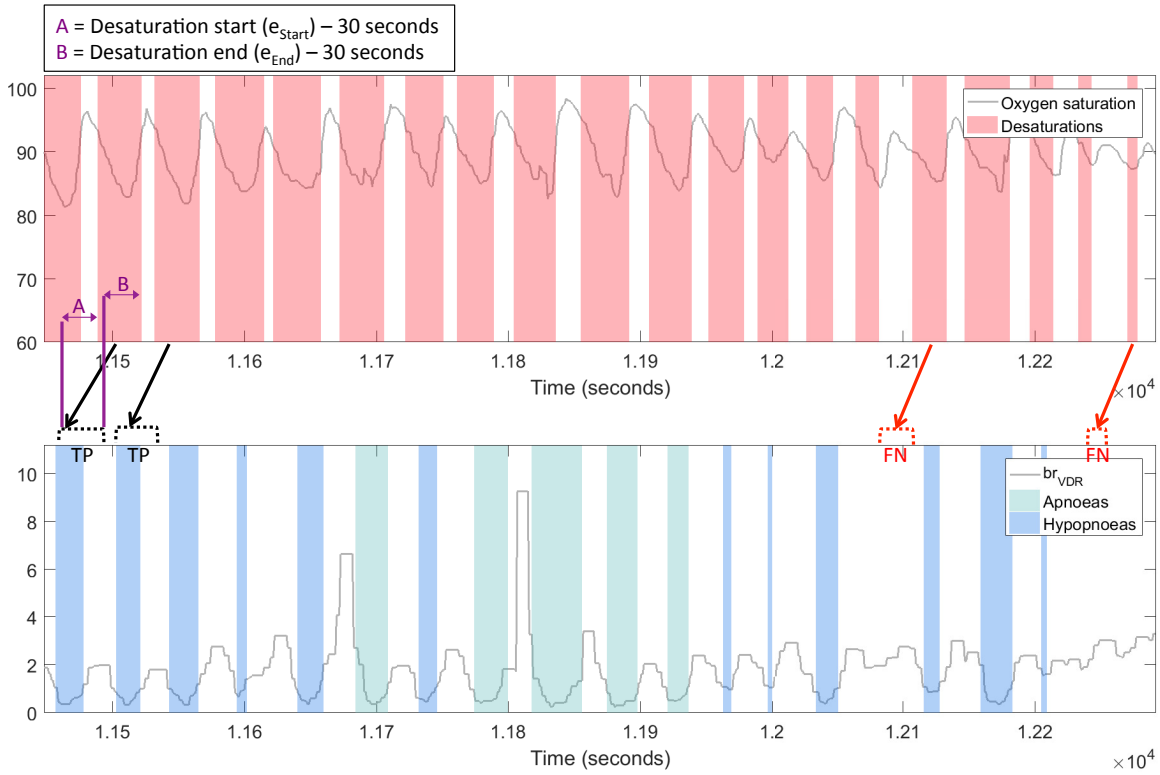


Figure 7.12: A few examples of TP and FN OSA events identified in the br_{VDR} signal for a patient in DS1 dataset. The br_{VDR} events are compared with the desaturation events identified in the oxygen saturation signal. The dotted brackets show the period corresponding to the start and end of the desaturation event shifted backwards by 30 seconds (the period between e_{Sstart} - 30 seconds and e_{End} - 30 seconds). Note that only a few events are annotated in the image as an example.

without any OSA events and with the start of the window denoted by e_{Start} and the end of the window denoted by e_{End}, if there are no events detected in the br_{VDR} signal within e_{Start} - 30 seconds and e_{End} - 30 seconds, the window is labelled a TN. As before, the 30-second length chosen for the windows is the average duration of apnoeic events [20].

Se, PPV, Sp, and F_1score were calculated as for DS2. Figure 7.12 shows the OSA events identified in the br_{VDR} signal for a patient in DS1 dataset.

Table 7.2 lists the results of comparing br_{VDR} and br_{RIPSum} OSA events for DS2 patients with moderate-to-severe OSA. Since patients with mild-to-no OSA have very

Table 7.2: The results of comparing br_{VDR} and br_{RIPSum} OSA events for DS2 patients and the results of comparing br_{VDR} OSA events and the desaturations events for DS1 patients.

Dataset	Reference events	F_1score(%)	Se(%)	PPV(%)	Sp(%)
DS2 (moderate-to-severe OSA)	br_{RIPSum} events	82.3±13.5	76.4±12.9	89.9±16.9	49.3±23.8
DS1 (moderate-to-severe OSA)	Desaturation events	76.0±15.2	81.0±12.1	73.8±19.8	50.9± 22.4

few or no OSA events, for many of the patients it is not possible to calculate Se, PPV, and F_1score and therefore the results are not reported here for mild-to-no OSA patients.

7.8 OSA severity threshold

For classifying the patients into the three OSA severity groups, the AHI thresholds of between 5 to 15 events/h for mild OSA, between 15 to 30 events/h for moderate OSA and above 30 events/h for severe OSA have been defined by the AASM [25]. However, there is a noticeable difference between the number of OSA events detected using AASM criteria recommended in 2007 and AASM criteria recommended in 2012. Specifically, with changing the definition of desaturations accompanying hypopnoea events from a $\geqslant 4\%$ drop from the oxygen desaturation baseline (2007 AASM recommended) to a $\geqslant 3\%$ drop from the oxygen desaturation baseline or an arousal (2012 AASM criteria), it is not surprising that the number of hypopnoeic events per hour has increased.

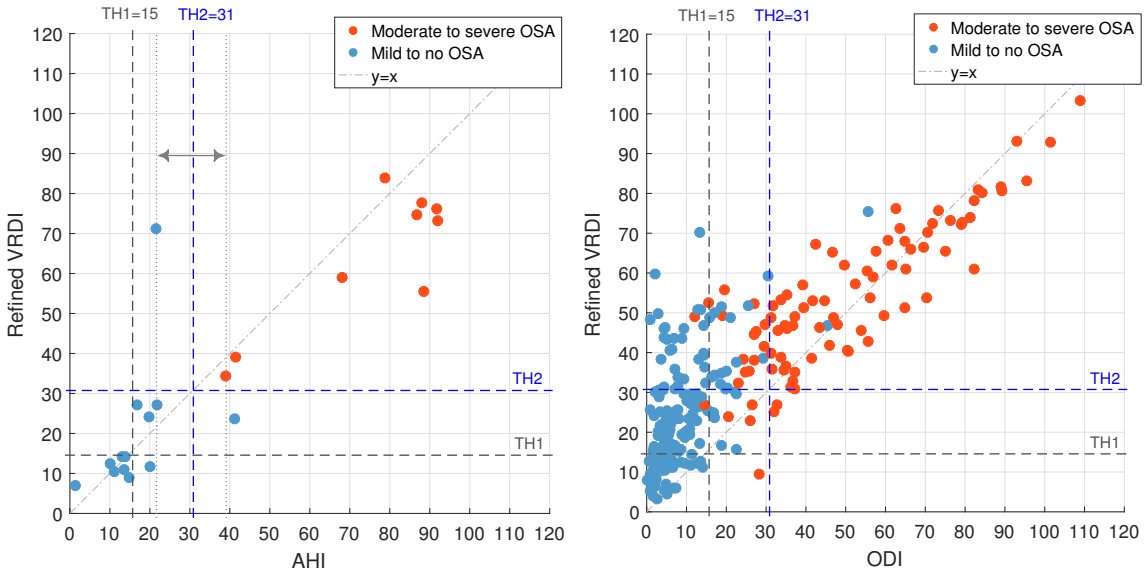
Literature comparing the 2012 AASM criteria for identifying apnoea and hypopnoea with the 2007 AASM criteria reports on average around two folds and up to three folds increase in the AHI values [79, 81, 82, 83]. The number of patients diagnosed with OSA has consequently increased when the 2012 AASM criteria are used. A quote

from an article published from some of the authors of the 2012 AASM manual reads *"It should be noted that using $\geqslant$ 3% instead of $\geqslant$ 4% desaturation requirement for defining hypopnea does increase the AHI substantially, with median AHI in a general community sample being almost twice as great using a 3% as a 4% criterion."* [79].

Since the OSA severity assigned to the patients by the clinicians in our dataset is mainly based on the 2007 AASM criteria, in this section a new threshold is calculated for moderate-to-severe OSA to make the OSA severity derived from the VRDI values comparable to the OSA severity assigned to the patients by the clinicians.

It can be seen in Figure 7.13a that for the patients in the DS2 dataset the AHI values for moderate-to-severe OSA patients are higher than 39 events/h. Additionally the AHI values for mild-to-no OSA patients are lower than 22 events/h. The mid-point separating the highest AHI value for mild-to-no OSA patients and lowest AHI value for moderate-to-severe OSA patients is 31 events/h. As this 31 events/h (TH2) threshold is in agreement with the reported increases in the AHI values using the 2012 AASM criteria, it will be used in this chapter instead of the 15 events/h threshold (TH1) to differentiate between moderate-to-severe OSA patients and mild-to-no OSA patients based on their VRDI values. Figure 7.13b shows the VRDI values of moderate-to-severe OSA and mild-to-no OSA patients in DS1 and the TH1 and TH2 thresholds.

Table 7.3 compares the results of classifying the patients in DS2 and DS1 into moderate-to-severe OSA and mild-to-no OSA using the VRDI values and the TH1 and TH2 thresholds. The ground truth used is the OSA severity assigned to the patients by the clinicians.



(a) DS2 (b) DS1

Figure 7.13: The AHI and VRDI values of moderate-to-severe OSA patients and mild-to-no OSA patients in DS2 [left] and DS1 [right] with respect to the TH1 (15 events/h) and TH2 (31 events/h) thresholds used to differentiate between moderate-to-severe OSA and mild-to-no OSA.

Table 7.3: Comparison of the results of classifying the patients in DS2 and DS1 into moderate-to-severe OSA and mild-to-no OSA using the VRDI values and the TH1 (15 events/h) and TH2 (31 events/h) thresholds.

Dataset	Threshold	Ac(%)	Se (%)	Sp (%)	PPV (%)	NPV (%)	AUC (%)
DS2	TH1 (15 events/h)	77.3	100.0	61.5	64.3	100.0	80.8
DS2	TH2 (31 events/h)	95.5	100.0	92.3	90.0	100.0	96.2
DS1	TH1 (15 events/h)	53.4	98.9	27.9	43.6	97.8	63.4
DS1	TH2 (31 events/h)	79.4	91.0	72.8	65.3	93.5	81.9

7.9 Discussion

This chapter investigated the possibility of detecting apnoeic and hypopnoeic events from the VDR signals and the classification of patients into mild-to-no OSA and moderate-to-severe OSA groups based on the number of events per hour. A similar approach to the standard guidelines for the identification of OSA events from PSG signals was used. An apnoea threshold and a hypopnoea threshold were used to find drops in the breath amplitude from the normal breathing baseline and identify OSA events in the VDR signals. The optimum apnoea and hypopnoea thresholds for the VDR signals were calculated using 50% of the patients in the DS2 dataset as the training test and assessed on the remaining 50% of patients (as the test set).

The threshold found for apnoea was 0.12 (or $\geqslant$ 88% drop from normal breathing baseline) which is very close to the $\leqslant$ 0.1 apnoea threshold defined for the standard PSG signals. For the hypopnoeic events the threshold found was 0.62 (or $\geqslant$38% drop from baseline breathing) which is slightly lower than the $\leqslant$ 0.7 threshold defined for hypopnoeic events. This lower hypopnoea threshold is due to the requirement for the hypopnoeic events in the RIPSum to be accompanied by a $\geqslant$3% reduction in the oxygen saturation.

Even with the more restrictive hypopnoea threshold of $\leqslant$ 0.62 for the VDR signals compared to the standard $\leqslant$ 0.7 for the RIPSum signals, the VDR hypopnoea events included a large number of false positives. When oxygen saturation is used with the RIPSum signals to identify hypopnoea, most of these false positive events are detected and removed. However, since oxygen saturation information is not available when the VDR signals are used alone an alternative source of information is required

to identify the false positive events. The audio data that is embedded in the video recordings were used to detect and remove the false positive hypopnoea events in the VDR signals. The hypopnoea events that were not followed by an awakening related spike in the audio signal were removed.

The refinement of hypopnoeic events using audio resulted in a noticeable improvement in the number of OSA events detected in the VDR signals (VRDI) and an increase in the agreement of VRDI with the AHI value. Overall, for the majority of the patients, the use of the audio signal successfully identifies the falsely-detected hypopnoeas. However, as Figure 7.6 shows, there were a few patients in DS2 for whom the audio refinement for hypopnoeic events was not successful. One patient was described as having violent PLMD related leg movements which caused both awakening-like peaks in the VDR signal and audio signal. This patient had an AHI of around 20 but a VRDI of around 70. Another example were two severe OSA patients who were awake during the study period. They were lying down but turning in bed occasionally causing artefacts in the RIPSum and oxygen saturation signals that resemble hypopnoea. For both of these patients the AHI is higher than the VRDI value.

The audio refinement of the OSA events identified using the DS1 VDR signals also shows a great improvement in filtering out falsely-detected hypopnoeas. As Figure 7.8 shows, VRDI values are overall higher than ODI values particularly for mild-to-no OSA patients. One reason for this may be that all the patients included in the study were referred to the sleep lab with suspected OSA. While not all were diagnosed with OSA many were diagnosed with other sleep disturbances such as PLMD, Cheyne-

Stokes respiration, etc. These other sleep disturbances may cause a pattern similar to OSA in the VDR signals. Another contributor to VRDI values being higher than ODI values may, as shown in Figure 7.2, be that ODI underscores AHI.

One important consideration here is that as explained in section 3.3 arousals were not included in the definition of hypopnoea in this chapter as EEG recordings were not available. While RIP signals and oximetry may not be affected by arousals, the video recordings capture the smallest arousal-related movements. Therefore, VRDI values may in fact be closer to AHI calculations that include arousals as well as oxygen desaturations. Further studies with EEG recordings are required to investigate the possibility of using videos for detecting hypopnoeas that include arousals and/or desaturations.

The individual comparison of OSA events identified in the VDR signals and RIP-Sum signals for moderate-to-severe OSA patients in DS2 showed a good agreement with an $F_1 score$ of 82.3%, Se of 76.4% and PPV of 89.9%. The events identified in the VDR signals of the DS1 dataset were compared with oxygen desaturations as there were no reference respiratory signals available for the majority of the patients. Comparison of the OSA events with the desaturation events is not ideal as desaturations occur with a delay from the onset of OSA events. The desaturation delay was set according to the average desaturation delay reported in the literature (30-seconds) when the events were compared. The results for moderate-to-severe OSA patients in DS1 showed an $F_1 score$ of 76.0%, Se of 81.0%, and PPV of 73.8% between the OSA events and the desaturation events.

There are several reports in the literature comparing the expert manual scoring of

OSA events in the PSG recordings with the automated scoring. However, the majority of these papers compare the performance based on the number of detected events per hour and not individual events [171, 172]. Additionally, many of these papers compare the automated and manual scoring of the same physiological signals (for example RIP and oxygen saturation signals). Some other papers, report the results based on an epoch by epoch comparison [173, 174, 158]. The signals are divided into, 30-second or 60-second non-overlapping epochs and the epochs are usually labelled as either SBD or normal epochs. There are variations in when an epoch is defined to be an SBD epoch between these papers. Taha *et al.* [158] report a Se of 93% and PPV of 97% when automated and manual scoring of RIP and oxygen saturation signals are compared on a 30-second epoch basis. Chazal *et al.* [174] detect SBD events using the ECG signal and report an Ac of 90%, Se of 89%, and PPV of 86% when the events are compared with the manual scoring on a 60-second epoch basis. Comparison of the results obtained here with the results reported in the literature is not a fair test as the methods (e.g. event by event comparison versus epoch by epoch comparison) and signals are not directly comparable.

Overall, this chapter shows that respiratory signals derived from video recordings provide a promising alternative for the scoring of OSA events during sleep. The agreement between the individual OSA events identified in the VDR signals and the reference events shows that detailed examination of OSA events is possible in the VDR signals. Also, the accuracy of 79.4% and sensitivity of 91.0% in differentiating between moderate-to-severe and mild-to-no OSA patients in DS1, indicates that the VRDI score can be used to identify which patients require further OSA screening.

Chapter 8

Conclusions and future work

Currently, large-scale screening for OSA, even with the home screening systems is limited because of the availability of equipment, and long-term monitoring of OSA is challenging because of the requirement to attach sensors to the body. The increasing number of people at risk of OSA [28] and the long waiting times before screening and treatment (see Table 1.2) create the need for contact-less home monitoring solutions. Additionally, as the review of several hours of over-night recordings is cumbersome and costly, the development of an automated sleep scoring system may be a very worthwhile contribution to the screening of OSA.

The overall aims of the research presented in this thesis were (i) to investigate whether automatic processing of video recordings of sleep can be used to identify patients with OSA who more urgently require treatment and are more likely to respond to the treatment (moderate-to-severe OSA patients). (ii) to investigate the possibility of deriving respiratory information (respiratory signal, respiratory rate) from video recordings and of identifying individual OSA events using the derived respiratory

information.

8.1 Summary of the contributions of this thesis

In chapter 2, the current standard methods for the screening of OSA inside sleep laboratories and at home were explained. In addition to the standard PSG methods, the following on-body sensors that are commonly used alone as a single channel screening tool were discussed: actigraphs which are used for the assessment of sleep fragmentation and sleep-wake periods, and pulse oximeters which are used for the identification of SBD events. As even these single on-body sensors may impose limitations for short or long-term monitoring of sleep, the role of video and sound recording that are part of most full-PSG systems was reviewed. Since the aim of this thesis was to investigate contact-less automated screening of OSA, the advantages and disadvantages of automatic analysis of sleep and scoring of OSA events were also presented.

Chapter 3 explained the patient data and normal volunteer data collected for the studies analysed in this thesis. For the 247 hospital patients, patient demographics were collected from paper records of patients over a period of two years. The video recordings included in the full PSG of hospital patients were recorded with the lowest resolution allowed by the system. This was to ensure that the PSG recording could be copied on a DVD to be given to the patient. Even with the lowest resolution, each video recording requires around 5GB of space. To the author's knowledge, there is no published literature on automated analysis of video recordings of sleep or deriving respiratory information from video recordings of sleep using a large number of patients.

Chapter 4 starts the investigations by examining the gross body sleep-wake move-

ments of patients during sleep. The interest here was the numerous repeated cycles of apnoea and awakenings in the sleep of moderate-to-severe OSA patients and whether these sleep-wake patterns can be used to identify significant OSA. The gross body movements of the patients were extracted from the video recordings (VDA signals). As mentioned in Chapter 2, actigraph-recorded activity has been used for the analysis of sleep-wake episodes for many years [64]. However, the limitation of actigraphy devices is that the movements recorded are affected by the location of the sensor on the body. For example, if the sensor is placed on the wrist or the torso, the head movements of the patients, which are of great importance in the analysis of OSA, will be missed. Activity derived from video recordings includes movements from all parts of body and may be superior in detecting OSA-related motion compared to actigraphs. The regularity of the patterns of sleep-wake in the VDA signals was extracted by calculating the sample entropy value of the original signal and coarse-grained VDA signals (original VDA signal averaged over non-overlapping time scales). Finally, sample entropy values of the most relevant time scales were identified and used as features for two classification models. Using these features calculated from the VDA signals, patients with moderate-to-severe OSA were differentiated from mild-to-no OSA patients with an Ac of 81%, Se of 62%, and Sp of 93%. When features derived from VDA signals were combined with the patient's BMI and neck size the classification performance improved to give an Ac of 82%, Se of 66%, and Sp of 91%.

As mentioned in Chapter 3, the sleep audio recordings of OSA patients include a few patterns that are specific to apnoeic/hypopnoeic events. In Chapter 5, the aim was to use the audio recordings embedded in the video data in addition to the

video recordings to differentiate patients suffering from moderate-to-severe OSA from those suffering from mild-to-no OSA. As with OSA sleep-wake activity, the audio signals of OSA patients also include a repetitive pattern of silence (during apnoea) followed by a snort/choking sound or loud breathing and resumption of breathing (awakening). Additionally, most patients with OSA snore during their sleep, therefore during periods of normal breathing a repetitive pattern of snoring can be identified in the audio signal. To analyse the regularity of these patterns, sample entropy was estimated for the original audio signal and for coarse-grained audio signals. Sample entropy values of the most relevant time scales were chosen and used as features for two classification models. Using the features derived from the audio signal alone, moderate-to-severe OSA was differentiated from mild-to-no OSA with an Ac of 80%, Se of 71%, and Sp of 86%. When features calculated from the audio recordings were combined with the features calculated from the VDA signals and patient BMI and neck size, the classification performance improved to an Ac of 85%, Se of 75%, and Sp of 91%.

Clinically, the respiratory signals in the PSG recordings are very useful for the identification of apnoeic/hypopnoeic events and calculation of an AHI to assess the severity of OSA. Therefore, Chapter 6 moved away from analysis of gross body movements and focused on subtle respiratory movements in the trunk of the patients that are visible in the video recordings. An algorithm was designed to generate a respiratory signal from the video recordings based on the similar pattern of the movements of chest and abdomen during each respiratory cycle. To validate the VDR signals, two approaches were taken. First, the patient's respiratory rates (per minute) were

estimated from the VDR signals and assessed against respiratory rates derived from reference respiratory signals included in the PSG studies. Secondly, the changes in the breath amplitude of the VDR signals were assessed against changes in the breath amplitude of the reference respiratory signals. The changes in breath amplitude are important as they are routinely used for detecting episodes of apnoea and hypopnoea.

The use of video recordings for sleep analysis is a relatively new area. To the author's knowledge, no literature has been published on validating respiratory signals from video recordings of sleep on more than a few patients. In fact, most of the papers in the literature on contact-less sleep respiration monitoring report results on healthy volunteers in controlled environments or on a handful of patients for short periods of time [117].

In chapter 7, the VDR signals were analysed to identify episodes of apnoeic or hypopnoeic breathing. Two thresholds were found for apnoea and hypopnoea in the VDR signals. The OSA events were identified by comparing the drops in the breath amplitude from baseline breathing against the apnoea/hypopnoea thresholds. The identified events were assessed against reference OSA events identified using RIPSum and oxygen saturation signals. As the hypopnoea events identified in the VDR signals included a large number of false positives, the audio signals of the video recordings were used to remove these events. The false positive events were detected by searching for a period of low audio energy (during the OSA event) followed by higher audio energy (during awakening) for each event. Refining the VDR events noticeably improved the calculated VRDI value.

The refined events were individually compared with the reference OSA events.

When the DS2 VDR events were compared with the RIPSum events, an $F_1 score$ of 82.3%, Se of 76.4%, and PPV of 89.9% was achieved. When the DS1 VDR events were compared with the oxygen desaturation events, an $F_1 score$ of 76.0%, Se of 81.0%, and PPV of 73.8% was achieved. Finally, with a new threshold of 31 events/h to detect moderate-to-severe OSA, the DS1 patients with moderate-to-severe OSA were differentiated from those with mild-to-no OSA patients with an Ac of 79%, Se of 91%, and Sp of 73% using the VRDI values.

8.2 Overall conclusions

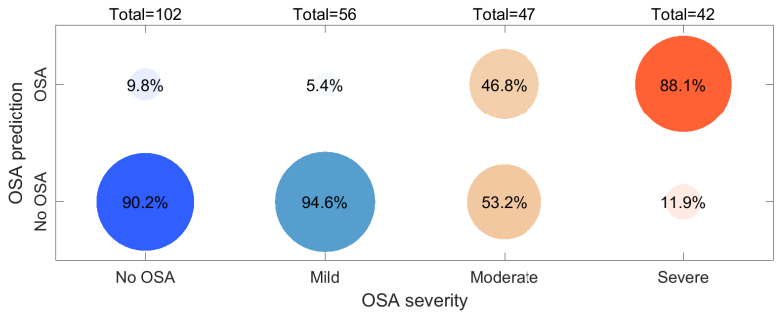
Two types of sleep activity were studied in this thesis: (i) gross body movements to analyse sleep-wake pattern, and (ii) small respiratory movements in the trunk of the patients to analyse individual OSA events. Table 8.1 below summarises the out-of-sample test performance achieved for classifying moderate-to-severe OSA and mild-to-no OSA using each type of activity combined with features derived from an audio channel recorded simultaneously. As can be seen in the table, the classification using gross body movements achieved a high Sp and low Se. On the other hand, the classification using respiratory movements, achieved a high Se and low Sp. Details of the percentage of each OSA severity correctly classified using the features derived in Chapters 5, 6, and 7 are presented in Figure 8.1.

The analysis in chapters 5 and 7 both agreed on the category (moderate-to-severe vs. mild-to-no OSA) assigned to 199 (81%) of the patients. Amongst these 199 patients, 79 have moderate-to-severe OSA (89 % of total moderate-to-severe OSA patients) and 120 have mild-to-no OSA (76% of total mild-to-no OSA patients). If the

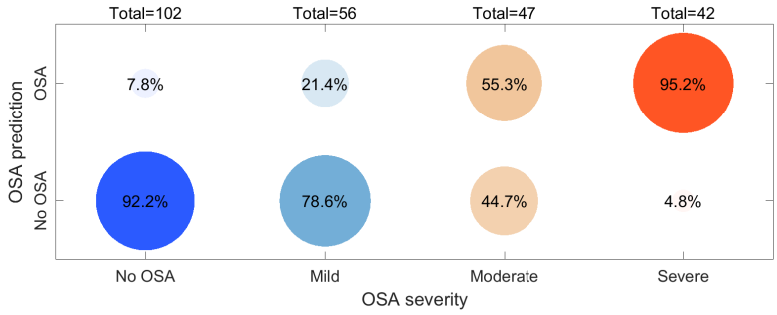
Table 8.1: Summary of the classification results of Chapters 5 and 7 and the results for the patient group both classifiers agree on.

Chapter(s)	Activity type	Ac(%)	Se(%)	Sp(%)	PPV(%)	NPV(%)	AUC(%)	Description
5	Gross body	85.8	76.4	91.1	82.9	87.3	89.3	247 DS1 patients. LDA classifier using MSE_{Vid}, MSE_{Aud}, BMI Neck size
6 & 7	Respiratory	79.4	91.0	72.8	65.3	93.5	81.9	247 DS1 patients. Threshold of $\geqslant$31 events/h on VRDI.
4 to 7	Gross body and respiratory	90.5	90.5	90.4	84.8	94.2	90.4	Using output labels of 199 DS1 patients for which both the LDA classifier above and the VRDI threshold agreed.

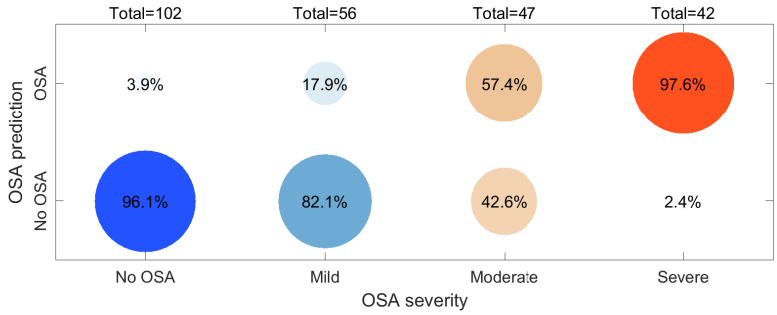
categories assigned to these 199 patients are compared with their true OSA categories, it results in an Ac of 91%, Se of 91%, and Sp of 90% (see Table 8.1). The classifiers disagreed on the category assigned to 48 patients (19%). Therefore, when both classifier results are combined, a 'possible OSA' category was added to differentiate these patients who may require additional screening. These patients may suffer from mild OSA, or might not suffer from OSA but from other sleep disturbances that resemble OSA. Even the non-OSA patients in DS1 were referred to the hospital with suspected OSA and many were diagnosed with other forms of SBD or sleep disturbances. Figure 8.2 presents the details of percentage of patients in each OSA severity categorised as OSA, no OSA, or possible OSA. As mentioned in the previous chapters, one of most important advantages of using video recordings for screening patients is the possibility of long-term monitoring. A more definite OSA category may be possible for the patients if they are video-screened over a few days as opposed to just one night.



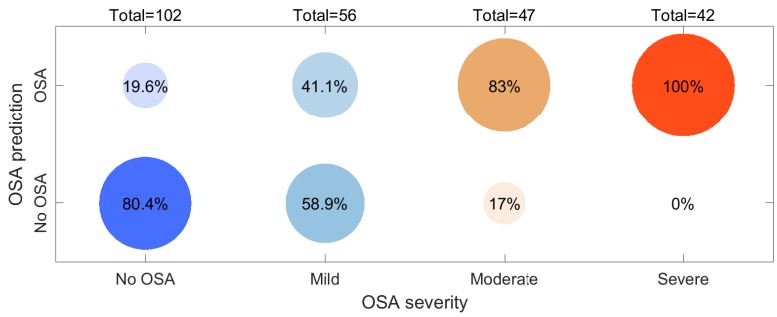
(a) Features: MSE_{Vid}, BMI, and Neck size



(b) Features: MSE_{Aud}, BMI, and Neck size



(c) Features: MSE_{Vid}, MSE_{Aud}, BMI, and Neck size



(d) Features: VRDI

Figure 8.1: Percentage of patients in each OSA severity correctly classified using the features derived in Chapters 5, 6, and 7.

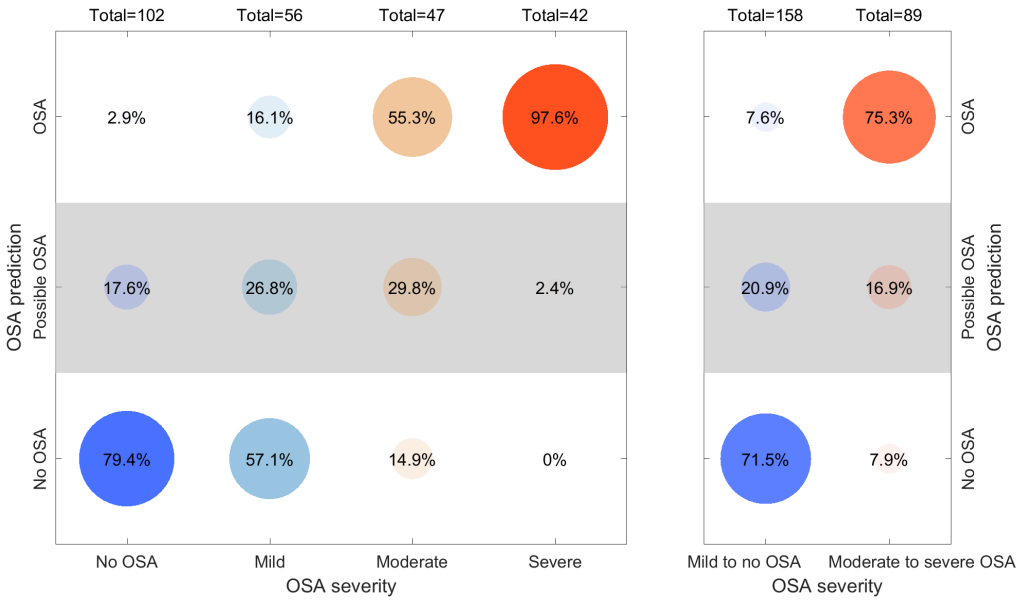


Figure 8.2: Percentage of patients in each OSA severity categorised as OSA, no OSA, or possible OSA using the combination of results from Chapters 5 and 7.

8.3 Limitations of thesis

Although this thesis has made several contributions to the development of non-contact OSA diagnosis, there are some limitations to the work. First, the majority of the recordings did not have reference respiratory signals and therefore no AHI values were calculated for the patients. To address this problem, the oxygen saturation signal and ODI were used in place of AHI. ODI has a strong correlation with AHI, however in some patients pulse oximetry may underestimate or over-estimate the actual number of OSA events. Therefore, in future work data collected from more patients with full PSG recordings (including respiratory signals) would help to ensure that the standard signals are used as ground truth for assessing the methods proposed in this thesis.

Secondly, no expert scoring of the OSA events was available. Despite the use of the AASM guidelines to automatically identify the events, artefacts in the data might have caused incorrectly identified events. Expert scoring would be valuable to

compare the automatically-identified OSA events using PSG signals, VDR signals, and audio recordings with the expert scoring separately.

Thirdly, there was a slight delay between the video recordings and the rest of the PSG recordings used for this thesis. Better synchronisation of channels would improve the identification of events and comparison of the identified events with the reference events.

8.4 Future work

8.4.1 Inclusion of arousals

As video recordings capture arousal-related movements as well as drops in the breath amplitude, a natural extension of the work would be to collect patient data with EEG recordings and compare the VDR-derived OSA events with reference OSA events that take patient arousals into account.

8.4.2 Identifying events in audio recordings

As discussed previously audio recordings of OSA patients include a few patterns that are specific to OSA, namely loud snoring, pauses in snoring followed by choking/snorting sounds. Therefore, a possible extension of the research presented in this thesis would be to analyse audio recordings to detect individual apnoeic events and combine these events with those identified in VDR signals.

8.4.3 Long-term monitoring of patients

One of the limitations of current in-laboratory or at-home PSG screening systems is that at least one sensor (usually at least a pulse oximeter probe) is attached to the patients. Pulse oximetry is very informative in terms of identifying episodes of SBD. However, pulse oximeters can become uncomfortable if worn over a long period of time. Moreover, one of the main issues with on-body sensors is that they can become detached from the body or move, causing artefacts in the signal and invalid recordings. Since the positions of camera and bed are fixed in a bedroom, the long-term monitoring of patients becomes more feasible. Therefore, a future direction for this thesis would be to collect home PSG recordings of patients over a few days along with video recordings. This data could then be used to compare the accuracy of the results from a single night's recordings to the results from the combination of several nights' recordings.

Appendix A

Sleep questionnaires

A.1 Epworth Sleepiness Scale (ESS)

How likely are you to doze off or fall asleep in the following situations, in contrast to feeling just tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently, try to work out how they would have affected you. Use the following scale to choose the most appropriate number for each situation:

0 = Would never doze

1 = Slight chance of dozing

2 = Moderate chance of dozing

3 = High chance of dozing

Situation	Chance of Dozing
Sitting and Reading	
Watching TV	
Sitting, inactive in a public place (e.g. a theatre or a meeting)	
As a passenger in a car for an hour without a break	
Lying down to rest in the afternoon when circumstances permit	
Sitting and talking to someone	
Sitting quietly after lunch without alcohol	
In a car, while stopped for a few minutes in traffic	

ESS scores from 0-8 are classified as normal, 9-12 are classified as mild, 13-16 are classified as moderate, and greater than 16 are classified as severe.

A.2 STOP-BANG questionnaire

Answer the following questions to find out if you are at risk of Obstructive Sleep Apnoea.

STOP		
Do you SNORE loudly (louder than talking or loud enough to be heard through closed doors)?	YES	NO
Do you often feel TIRED, fatigued, or sleepy during daytime?	YES	NO
Has anyone OBSERVED you stop breathing during your sleep?	YES	NO
Do you have or are you being treated for high blood PRESSURE?	YES	NO
BANG		
BMI more than 35 $\frac{kg}{m^2}$?	YES	NO
AGE over 50 years old?	YES	NO
NECK circumference >16 inches (40cm)?	YES	NO
GENDER: Male?	YES	NO
TOTAL SCORE:		

Number of questions answered yes:

0-2: Low risk of OSA

3-4: Intermediate risk of OSA

5-8: High risk of OSA

A.3 Berlin Questionnaire

The Berlin Questionnaire includes three categories of questions related to the risk of having sleep apnoea. The person's response to individual questions and the total score in each category will determine whether the he/she is in the low OSA risk or high OSA risk group.

Complete the following:

- Height:
- Weight:
- Age:
- Gender:

Please choose the correct response to each question.

Category 1

1. Do you snore?

   a. Yes
   b. No
   c. Don't know

If you answered 'yes':

2. Do you snore?

   a. Slightly louder than breathing
   b. As loud as talking
   c. Louder than talking

3. How often do you snore?

   a. Almost every day
   b. 3-4 times per week
   c. 1-2 times per week
   d. 1-2 times per month
   e. Rarely or never

4. Has your snoring ever bothered other people?

   a. Yes
   b. No

c. Don't know

5. Has anyone noticed that you stop breathing during your sleep?

   a. Almost every day
   b. 3-4 times per week
   c. 1-2 times per week
   d. 1-2 times per month
   e. Rarely or never

Category 2

6. How often do you feel tired or fatigued after your sleep?

   a. Almost every day
   b. 3-4 times per week
   c. 1-2 times per week
   d. 1-2 times per month
   e. Rarely or never

7. During your waking time, do you feel tired, fatigued or not up to par?

   a. Almost every day
   b. 3-4 times per week
   c. 1-2 times per week
   d. 1-2 times per month
   e. Rarely or never

8. Have you ever nodded off or fallen asleep while driving a vehicle?

   a. Yes
   b. No

If you answered 'yes':

9. How often does this occur?

   a. Almost every day
   b. 3-4 times per week
   c. 1-2 times per week
   d. 1-2 times per month
   e. Rarely or never

Category 3

10. Do you have high blood pressure?

   a. Yes
   b. No
   c. Don't know

A.3.1 Scoring Berlin Questionnaire

The questionnaire consists of 3 categories related to the risk of having sleep apnoea. Patients can be classified into High Risk or Low Risk based on their responses to the individual items and their overall scores in the symptom categories (adapted from [44]).

Categories and Scoring:

Category 1: Questions 1, 2, 3, 4, and 5.

Q 1: if 'Yes', assign 1 point

Q 2: if 'c' or 'd' is the response, assign 1 point

Q 3: if 'a' or 'b' is the response, assign 1 point

Q 4: if 'a' is the response, assign 1 point

Q 5: if ‘a’ or ‘b’ is the response, assign 2 points

Add the points. Category 1 is positive if the total score is 2 or more points.

Category 2: Questions 6, 7, 8 (item 9 should be noted separately).

Q 6: if ‘a’ or ‘b’ is the response, assign 1 point

Q 7: if ‘a’ or ‘b’ is the response, assign 1 point

Q 8: if ‘a’ is the response, assign 1 point

Add the points. Category 2 is positive if the total score is 2 or more points.

Category 3: is positive if the answer to item 10 is ‘Yes’ or if the BMI of the patient is greater than 30 $\frac{kg}{m^2}$ (BMI is defined as weight (kg) divided by height (m) squared).

High Risk: if there are 2 or more categories where the score is positive. **Low Risk**: if there is only 1 or no categories where the score is positive.

Additional Question: item 9 should be noted separately.

Appendix B

Classification performance metrics

- True positive (TP): When the model correctly predicts a positive outcome.
- True negative (TP): When the model correctly predicts a negative outcome.
- False positive (FP): When the model falsely predicts a positive outcome (the actual outcome is negative).
- False negative (FN): When the model falsely predicts a negative outcome (the actual outcome is positive).

Based of these definitions, the performance of a classification model can be reported in a combination of the following parameters [175].

- Accuracy (Ac) = $\frac{TP+TN}{TP+TN+FP+FN}$
- Sensitivity (Se) or recall or true positive rate = $\frac{TP}{TP+FN}$
- Specificity (Sp) or true negative rate = $\frac{TN}{TN+FP}$
- False positive rate, 1-Sp: $\frac{FP}{TN+FP}$

- Positive predictive value (PPV) or precision = $\frac{TP}{TP+FP}$
- Negative predictive value (NPV) = $\frac{TN}{TN+FN}$
- $F_1 score$: $2 \times \frac{Se \times PPV}{Se+PPV} = \frac{2 \times TP}{2 \times TP+FP+FN}$
- AUC: Area under the receiver operating characteristics (ROC) curve. ROC graphs plot the TP rate on the Y axis and FP rate on the X axis. For detailed explanation of ROC curves and AUC calculation see [175].

Appendix C

List of Publications

Journal and conference articles

- E. Gederi and G.D. Clifford. Fusion of image and signal processing for the detection of obstructive sleep apnea. In *Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on*, pages 890-893. IEEE, 2012.

- E. Gederi, S. Nemati, B. A. Edwards, G. D. Clifford, A. Malhotra, and A. Wellman. Model-based estimation of loop gain using spontaneous breathing: a validation study. *Respiratory physiology & neurobiology*, 201:84-92, 2014.

- A. Roebuck, V. Monasterio, E. Gederi, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G.D. Clifford. A review of signals used in sleep analysis. *Physiological measurement*, 35(1):R1, 2014.

- J. Behar, A. Roebuck, J. S. Domingos, E. Gederi, and G. D. Clifford. A review of current sleep screening applications for smartphones. *Physiological measurement*, 34(7):R29, 2013.

- G. D. Clifford and E. Gederi. Out of touch: From audio recordings to phone apps to mattress sensors, noncontact systems offer a less cumbersome way to monitor sleep. *IEEE pulse*, 5(5):19-21, 2014.

Journal articles in preparation

- E. Gederi, L. Tarassenko, G. D. Clifford, Multiscale analysis of audio and video recordings of sleep for the detection of OSA, *Physiological measurement* (associated with chapters 4 and 5)
- E. Gederi, G. D. Clifford, L. Tarassenko, Respiratory waveform estimation from video recordings of sleep for the screening of OSA, *Physiological measurement* (associated with chapters 6 and 7)

Bibliography

[1] L. A. Irish, C. E. Kline, H. E. Gunn, D. J. Buysse, and M. H. Hall. The role of sleep hygiene in promoting public health: A review of empirical evidence. *Sleep Medicine Reviews*, 22:23–36, 2015.

[2] M. H. Kryger, T. Roth, and W. C. Dement. *Principles and Practice of Sleep Medicine*. Elsevier LTD, Oxford, 2010.

[3] T. Young, P. E. Peppard, and D. J. Gottlieb. Epidemiology of obstructive sleep apnea: a population health perspective. *American Journal of Respiratory and Critical Care Medicine*, 165(9):1217–1239, 2002.

[4] J. E. Ferrie, M. Kumari, P. Salo, A. Singh-Manoux, and M. Kivimäki. Sleep epidemiology–a rapidly growing field. *International Journal of Epidemiology*, 40(6):1431–1437, 2011.

[5] E. Aserinsky and N. Kleitman. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 118(3062):273–274, 1953.

[6] M. A. Carskadon and W. C. Dement. Normal human sleep: an overview. In M. H. Kryger, T. Roth, and W. C. Dement, editors, *Principles and Practice of Sleep Medicine*, pages 17–26. Elsevier, Philadelphia, 2010.

[7] A. L. Loomis, E. N. Harvey, and G. A. Hobart. Cerebral states during sleep, as studied by human brain potentials. *Journal of Experimental Psychology*, 21(2):127, 1937.

[8] A. Rechtschaffen, A. Kales, Los Angeles. Brain Information Service University of California, and NINDB Neurological Information Network. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. US Dept. of Health, Education, and Welfare Bethesda, Md., 1968.

[9] C. Iber, S. Ancoli-Israel, A. Chesson, S. F. Quan, et al. *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, volume 1. American Academy of Sleep Medicine Westchester, IL, 2007.

[10] B. M. Altevogt, H. R. Colten, et al. *Sleep disorders and sleep deprivation: an unmet public health problem*. National Academies Press, 2006.

[11] S. Keenan and M. Hirshkowitz. Monitoring and staging human sleep. In William C. Dement Meir H. Kryger, Thomas Roth, editor, *Principles and Practice of Sleep Medicine*, pages 1602–1609. Elsevier, Philadelphia, 2010.

[12] N. J. Douglas, D. P. White, C. K. Pickett, J. V. Weil, and C. W. Zwillich. Respiration during sleep in normal man. *Thorax*, 37(11):840–844, 1982.

[13] American Academy of Sleep Medicine et al. International classification of sleep disorders-third edition (icsd-3). *AASM Resource Library*, 2014.

[14] A. Elizabeth, M. A. Martin, R. Aggrewal, J. C. Jobling, and E. Taylor. Oxford concise medical dictionary, 2003.

[15] N. Collop. The effect of obstructive sleep apnea on chronic medical disorders. *Cleveland Clinic Journal of Medicine*, 74(1):72–78, 2007.

[16] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley. Access to diagnosis and treatment of patients with suspected sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 169(6):668–672, 2004.

[17] H. Gastaut. Etude polygraphique des manifestations episodiques,(hypniques et respiratoires), diurnes et nocturnes, du syndrome de pickwick. *Rev Neurol*, 112:568–579, 1965.

[18] R. Jung and W. Kuhlo. Neurophysiological studies of abnormal night sleep and the pickwickian syndrome. *Progress in Brain Research*, 18:140–159, 1965.

[19] C. Guilleminault, A. Tilkian, and W. C. Dement. The sleep apnea syndromes. *Annual Review of Medicine*, 27(1):465–484, 1976.

[20] D. Alvarez-Estevez and V. Moret-Bonillo. Computer-assisted diagnosis of the sleep apnea-hypopnea syndrome: A review. *Sleep Disorders*, 2015, 2015.

[21] G. J. Gibson. Obstructive sleep apnoea syndrome: underestimated and undertreated. *British Medical Bulletin*, 72(1):49–64, 2004.

[22] M. T. Cao, C. Guilleminault, and C. A. Kushida. Clinical features and evaluation of obstructive sleep apnea and upper airway resistance syndrome. In M. H. Kryger, T. Roth, and W. C. Dement, editors, *Principles and Practice of Sleep Medicine*, pages 1206–18. Elsevier, Philadelphia, 2010.

[23] M. Hirshkowitz and M. H. Kryger. Monitoring techniques for evaluating suspected sleep-disordered breathing. In M. H. Kryger, T. Roth, and W. C. Dement, editors, *Principles and Practice of Sleep Medicine*, pages 1610–1623. Elsevier, Philadelphia, 2010.

[24] N. A. Antic, C. Buchan, A. Esterman, M. Hensley, M. T. Naughton, S. Rowland, B. Williamson, S. Windler, S. Eckermann, and R. D. McEvoy. A randomized controlled trial of nurse-led care for symptomatic moderate–severe obstructive sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 179(6):501–508, 2009.

[25] AASM. Obstructive sleep apnea (OSA). Technical report, American Academy of Sleep Medicine, 2008.

[26] P. Lévy, M. Kohler, W. T. McNicholas, F. Barbé, R. D. McEvoy, V. K. Somers, L. Lavie, and J. Pepin. Obstructive sleep apnoea syndrome. *Nature Reviews. Disease Primers*, 1:15015–15015, 2014.

[27] T. Young, L. Evans, L. Finn, M. Palta, et al. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep*, 20(9):705, 1997.

[28] S. P. Patil, H. Schneider, A. R. Schwartz, and P. L. Smith. Adult obstructive sleep apnea. *Chest*, 132(1):325, 2007.

[29] WHO. Obesity and overweight. Fact sheet N311, 03 2011.

[30] European markets for sleep diagnostics and its data management systems. Technical report, Frost & Sullivan, 2006.

[31] C. Guilleminault and V. C. Abad. Obstructive sleep apnea. *Current Treatment Options in Neurology*, 6(4):309–317, 2004.

[32] K. A. Ferguson, R. Cartwright, R. Rogers, W. Schmidt-Nowara, et al. Oral appliances for snoring and obstructive sleep apnea: a review. *Sleep*, 29(2):244, 2006.

[33] D. A. Hanzel, N. G. Proia, and D. W. Hudgel. Response of obstructive sleep apnea to fluoxetine and protriptyline. *Chest*, 100(2):416–421, 1991.

[34] I. E. Smith and T. G. Quinnell. Pharmacotherapies for obstructive sleep apnoea. *Drugs*, 64(13):1385–1399, 2004.

[35] R. C. Heinzer, D. P. White, A. S. Jordan, Y. L. Lo, L. Dover, K. Stevenson, and A. Malhotra. Trazodone increases arousal threshold in obstructive sleep apnoea. *European Respiratory Journal*, 31(6):1308–1312, 2008.

[36] C. Guilleminault, J. H. Lee, and A. Chan. Pediatric obstructive sleep apnea syndrome. *Archives of Pediatrics & Adolescent Medicine*, 159(8):775–785, 2005.

[37] A. R. Schwartz, D. W. Eisele, A. Hari, R. Testerman, D. Erickson, and P. L. Smith. Electrical stimulation of the lingual musculature in obstructive sleep apnea. *Journal of Applied Physiology*, 81(2):643–652, 1996.

[38] A. Oliven, R. P. Schnall, G. Pillar, N. Gavriely, and M. Odeh. Sublingual electrical stimulation of the tongue during wakefulness and sleep. *Respiration Physiology*, 127(2):217–226, 2001.

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

[40] M. W. Johns. Reliability and factor analysis of the epworth sleepiness scale. *Sleep*, 15(4):376–381, 1992.

[41] M. Hirshkowitz, N. Gokcebay, S. Iqbal, et al. Epworth sleepiness scale and sleep disordered breathing: replication and extension. *Sleep Research*, 24:249, 1995.

[42] A. Sharafkhaneh and M. Hirshkowitz. Contextual factors and perceived self-reported sleepiness: a preliminary report. *Sleep Medicine*, 4(4):327–331, 2003.

[43] F. Chung, B. Yegneswaran, P. Liao, S. A. Chung, S. Vairavanathan, S. Islam, A. Khajehdehi, and C. M. Shapiro. Stop questionnairea tool to screen patients for obstructive sleep apnea. *The Journal of the American Society of Anesthesiologists*, 108(5):812–821, 2008.

[44] N. C. Netzer, R. A. Stoohs, C. M. Netzer, K. Clark, and K. P. Strohl. Using the berlin questionnaire to identify patients at risk for the sleep apnea syndrome. *Annals of Internal Medicine*, 131(7):485–491, 1999.

[45] K. E. Bloch. Polysomnography: a systematic review. *Technol Health Care*, 5(4):285–305, 1997.

[46] A. S. Jordan, D. G. McSharry, and A. Malhotra. Adult obstructive sleep apnoea. *The Lancet*, 383(9918):736–747, 2014.

[47] S. A. Sands, R. L. Owens, and A. Malhotra. New approaches to diagnosing sleep-disordered breathing. *Sleep Medicine Clinics*, 11(2):143–152, 2016.

[48] R. Ferber, R. Millman, M. Coppola, J. Fleetham, C. F. Murray, C. Iber, V. McCall, G. Nino-Murcia, M. Pressman, and M. Sanders. Portable recording in the assessment of obstructive sleep apnea. asda standards of practice. *Sleep*, 17(4):378–392, 1994.

[49] M. J. Murray, S. H. Rose, D. J. Wedel, C. T. Wass, B. A. Harrison, and J. T. Mueller. *Faust's Anesthesiology Review E-Book: Expert Consult*. Elsevier Health Sciences, 2014.

[50] L. Loerup. *Physiological Patterns in Pregnancy*. DPhil thesis, University of Oxford, 2016.

[51] A. V. J. Challoner. Photoelectric plethysmography for estimating cutaneous blood flow. *Non-invasive Physiological Measurements*, 1:125–151, 1979.

[52] The abcs of pulse oximetry. `http://www.homecaremag.com/senior-care-products/abcs-pulse-oximetry`. Accessed: 15-06-2017.

[53] J. Law and R. Rennie. *A Dictionary of Physics*. OUP Oxford, 2015.

[54] N. Finer and T. Leone. Oxygen saturation monitoring for the preterm infant: the evidence basis for current practice. *Pediatric Research*, 65(4):375–380, 2009.

[55] W. W. Flemons, M. R. Littner, J. A. Rowley, P. Gay, W. M. D. Anderson, D. W. Hudgel, R. D. McEvoy, and D. I. Loube. Home diagnosis of sleep apnea: A systematic review of the literature. *Chest*, 124(4):1543, 2003.

[56] M. Chambrin. Alarms in the intensive care unit: how can the number of false alarms be reduced? *Critical Care*, 5(4):184, 2001.

[57] M. T. Petterson, V. L. Begnoche, and J. M. Graybeal. The effect of motion on pulse oximetry and its clinical significance. *Anesthesia & Analgesia*, 105(6):S78–S84, 2007.

[58] V. Monasterio, F. Burgess, and G. D. Clifford. Robust classification of neonatal apnoea-related desaturations. *Physiological Measurement*, 33(9):1503, 2012.

[59] S. J. Barker. "motion-resistant" pulse oximetry: a comparison of new and old models. *Anesthesia & Analgesia*, 95(4):967–972, 2002.

[60] N. Netzer, A. H. Eliasson, C. Netzer, and D. A Kristo. Overnight pulse oximetry for sleep-disordered breathing in adults: a review. *Chest Journal*, 120(2):625–633, 2001.

[61] D. J. Pitson, A. Sandell, R. Van den Hout, et al. Use of pulse transit time as a measure of inspiratory effort in patients with obstructive sleep apnoea. *European Respiratory Journal*, 8(10):1669, 1995.

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

[63] S. Zhang, A. V. Rowlands, P. Murray, T. L Hurst, et al. *Physical activity classification using the GENEA wrist-worn accelerometer*. PhD thesis, Lippincott Williams and Wilkins, 2012.

[64] A. Sadeh, P. J. Hauri, D. F. Kripke, and P. Lavie. The role of actigraphy in the evaluation of sleep disorders. *Sleep*, 18(4):288, 1995.

[65] A. M. Berger, K. K. Wielgus, S. Young-McCaughan, P. Fischer, L. Farr, and K. A. Lee. Methodological challenges when using actigraphy in research. *Journal of Pain and Symptom Management*, 36(2):191–199, 2008.

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

[67] E. J. Paavonen, M. Fjällberg, M. R. Steenari, and E. T. Aronen. Actigraph placement and sleep estimation in children. *Sleep*, 25(2):235–237, 2002.

[68] A. Oksenberg, I. Khamaysi, D. S. Silverberg, and A. Tarasiuk. Association of body position with severity of apneic events in patients with severe nonpositional obstructive sleep apnea. *Chest*, 118(4):1018–1024, 2000.

[69] S. R. Lloyd and R. D. Cartwright. Physiologic basis of therapy for sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, 136(2):525–526, 1987.

[70] N. B. Kavey, A. Blitzer, S. Gidro-Frank, and K. Korstanje. Sleeping position and sleep apnea syndrome. *American Journal of Otolaryngology*, 6(5):373–377, 1985.

[71] R. D. Cartwright. Effect of sleep position on sleep apnea severity. *Journal of Sleep Research & Sleep Medicine*, 1984.

[72] K. Yoshida. Influence of sleep posture on response to oral appliance therapy for sleep apnea syndrome. *Sleep*, 24(5):538–544, 2001.

[73] E. R. Van Kesteren, J. P. van Maanen, A. A. Hilgevoord, D. M. Laman, and N. de Vries. Quantitative effects of trunk and head position on the apnea hypopnea index in obstructive sleep apnea. *Sleep*, 34(8):1075, 2011.

[74] N. Ben-Israel, A. Tarasiuk, and Y. Zigel. Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults. *Sleep*, 35(9):1299–1305, 2012.

[75] H. Nakano, K. Hirayama, Y. Sadamitsu, A. Toshimitsu, H. Fujita, S. Shin, and T. Tanigawa. Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept. *Journal of Clinical Sleep Medicine*, 10(1):73, 2014.

[76] Y. Sivan, A. Kornecki, and T. Schonfeld. Screening obstructive sleep apnoea syndrome by home videotape recording in children. *European Respiratory Journal*, 9(10):2127, 1996.

[77] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A Keenan, M. H. Kryger, T. Penzel, et al. The visual scoring of sleep in adults. *Jounal of Clinical Sleep Medicine*, 3(2):121–131, 2007.

[78] W. W. Flemons, D. Buysse, S. Redline, A. Oack, K. Strohl, J. Wheatley, T. Young, N. Douglas, P. Levy, W. McNicolas, et al. Sleep-related breathing disorders in adults. *Sleep*, 22(5):667–689, 1999.

[79] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. L. Marcus, and B. V. Vaughn. The AASM manual for the scoring of sleep and associated

events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 2012.

[80] M. M. Grigg-Damberger. The aasm scoring manual four years later. *Journal of Clinical Sleep Medicine*, 8(3):323–332, 2012.

[81] R. Heinzer, S. Vat, P. Marques-Vidal, H. Marti-Soler, D. Andries, N. Tobback, V. Mooser, M. Preisig, A. Malhotra, G. Waeber, et al. Prevalence of sleep-disordered breathing in the general population: the hypnolaus study. *The Lancet Respiratory Medicine*, 3(4):310–318, 2015.

[82] B. Duce, J. Milosavljevic, and C. Hukins. The 2012 aasm respiratory event criteria increase the incidence of hypopneas in an adult sleep center population. *Journal of Clinical Sleep Medicine*, 11(12):1425, 2015.

[83] A. S. BaHammam, A. Obeidat, K. Barataman, S. A Bahammam, A. H. Olaish, and M. M. Sharif. A comparison between the aasm 2012 and 2007 definitions for detecting hypopnea. *Sleep and Breathing*, 18(4):767–773, 2014.

[84] W. R. Ruehland, P. D. Rochford, F. J. O'Donoghue, R. J. Pierce, P. Singh, and A. T. Thornton. The new aasm criteria for scoring hypopneas: impact on the apnea hypopnea index. *Sleep*, 32(2):150, 2009.

[85] O. Ozeke, O. Erturk, M. Gungor, S. B. Hızel, D. Aydın, M. K. Celenk, H. Dıncer, G. Ilıcın, F. Ozgen, and C. Ozer. Influence of the right-versus left-sided sleeping position on the apnea-hypopnea index in patients with sleep apnea. *Sleep and Breathing*, pages 1–4, 2011.

[86] A. Wirz-Justice, C. M. Schröder, P. F. Gasio, C. Cajochen, and E. Savaskan. The circadian rest-activity cycle in korsakoff psychosis. *The American Journal of Geriatric Psychiatry*, 18(1):33–41, 2010.

[87] K. Wulff, S. Gatti, J. G. Wettstein, and R. G. Foster. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nature Reviews Neuroscience*, 11(8):589–599, 2010.

[88] E. Sforza, M. Johannes, and B. Claudio. The pam-rl ambulatory device for detection of periodic leg movements: a validation study. *Sleep Medicine*, 6(5):407–413, 2005.

[89] M. A. King, M. Jaffre, E. Morrish, J. M. Shneerson, and I. E. Smith. The validation of a new actigraphy system for the measurement of periodic leg movements in sleep. *Sleep Medicine*, 6(6):507–513, 2005.

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

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

[92] M. Elbaz, G. M. Roue, F. Lofaso, and M. A. SALVA. Utility of actigraphy in the diagnosis of obstructive sleep apnea. *Sleep*, 25(5):527–531, 2002.

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

[94] J. Tilmanne, J. Urbain, M. V Kothare, A. V. Wouwer, and S. V. Kothare. Algorithms for sleep–wake identification using actigraphy: a comparative study and new results. *Journal of Sleep Research*, 18(1):85–98, 2009.

[95] T. F. Anders and A. M. Sostek. The use of time lapse video recording of sleep-wake behavior in human infants. *Psychophysiology*, 13(2):155–158, 1976.

[96] K. Banno and M. H. Kryger. Use of polysomnography with synchronized digital video recording to diagnose pediatric sleep breathing disorders. *Canadian Medical Association Journal*, 173(1):28, 2005.

[97] R. Silvestri, A. Gagliano, I. Aricņ, T. Calarese, C. Cedro, O. Bruni, R. Condurso, E. Germanņ, G. Gervasi, R. Siracusano, et al. Sleep disorders in children with attention-deficit/hyperactivity disorder (adhd) recorded overnight by video-polysomnography. *Sleep Medicine*, 10(10):1132–1138, 2009.

[98] C. W. Wang, A. Hunter, N. Gravill, and S. Matusiewicz. Real time pose recognition of covered human for diagnosis of sleep apnoea. *Computerized Medical Imaging and Graphics*, 34(6):523–533, 2010.

[99] D. Ramanan, D. A. Forsyth, and A. Zisserman. Tracking people by learning their appearance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 65–81, 2007.

[100] W. H. Liao and C. M. Yang. Video-based activity and movement pattern analysis in overnight sleep studies. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.

[101] G. R. Bradski and J. W. Davis. Motion segmentation and pose recognition with motion history gradients. *Machine Vision and ApplicationsMotion segmentation and pose recognition with motion history gradients*, 13(3):174–184, 2002.

[102] W. H. Liao and J. H. Kuo. Sleep monitoring system in real bedroom environment using texture-based background modeling approaches. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–10, 2011.

[103] A. Heinrich, X. Aubert, and G. de Haan. Body movement analysis during sleep based on video motion estimation. In *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pages 539–543. IEEE, 2013.

[104] M. Costa, A. L. Goldberger, and C. K. Peng. Multiscale entropy analysis of biological signals. *Physical Review E*, 71(2):021906, 2005.

[105] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039, 2000.

[106] J. Ye, R. Janardan, and Q Li. Two-dimensional linear discriminant analysis. In *Advances in neural information processing systems*, pages 1569–1576, 2004.

[107] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 6. Springer, 2013.

[108] T. Li, S. Zhu, and M. Ogihara. Using discriminant analysis for multi-class classification: an experimental investigation. *Knowledge and Information Systems*, 10(4):453–472, 2006.

[109] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2 edition, November 2001.

[110] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[111] A. Roebuck. *Comparative analysis of polysomnographic signals for classifying Obstructive Sleep Apnoea*. DPhil thesis, University of Oxford, 2014.

[112] What is flac? `https://xiph.org/flac/index.html`. Accessed: 15-06-2017.

[113] G. J. Tortora and B. Derrickson. *Principles of Anatomy and Physiology*. John Wiley & Sons, 2008.

[114] M. Folke, L. Cernerud, M. Ekström, and B. Hök. Critical review of non-invasive respiratory monitoring in medical care. *Medical and Biological Engineering and Computing*, 41(4):377–383, 2003.

[115] P. B. Lovett, J. M. Buchwald, K. Stürmann, and P. Bijur. The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. *Annals of Emergency Medicine*, 45(1):68–76, 2005.

[116] W. Bianchi, A. F. Dugas, Y. H. Hsieh, M. Saheed, P. Hill, C. Lindauer, A. Terzis, and R. E. Rothman. Revitalizing a vital sign: improving detection of tachypnea at primary triage. *Annals of Emergency Medicine*, 61(1):37–43, 2013.

[117] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35(5):807, 2014.

[118] M. Z. Poh, D. J. McDuff, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762–10774, 2010.

[119] G. F. Lewis, R. G. Gatto, and S. W. Porges. A novel method for extracting respiration rate and relative tidal volume from infrared thermography. *Psychophysiology*, 48(7):877–887, 2011.

[120] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis. Touchless monitoring of breathing function. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 1, pages 1196–1199. IEEE, 2004.

[121] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt. Neonatal non-contact respiratory monitoring based on real-time infrared thermography. *BioMedical Engineering OnLine*, 10:93, 2011.

[122] D. Alinovi, L. Cattani, G. Ferrari, F. Pisani, and R. Raheli. Spatio-temporal video processing for respiratory rate estimation. In *Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium on*, pages 12–17. IEEE, 2015.

[123] C. B. Pereira, X. Yu, M. Czaplik, V. Blazek, B. Venema, and S. Leonhardt. Estimation of breathing rate in thermal imaging videos: a pilot study on healthy

human subjects. *Journal of Clinical Monitoring and Computing*, pages 1–14, 2016.

[124] A. H. Alkali, R. Saatchi, H. Elphick, and D. Burke. Thermal image processing for real-time non-contact respiration rate monitoring. *IET Circuits, Devices & Systems*, 2016.

[125] M. Frigola, J. Amat, and J. Pagès. Vision based respiratory monitoring system. In *Proceedings of the 10th Mediterranean Conference on Control and Automation MED2002 Lisbon, Portugal*, 2002.

[126] V. Makkapati, P. Raman, and G. Pai. Camera based respiration rate of neonates by modeling movement of chest and abdomen region. In *Signal Processing and Communications (SPCOM), 2016 International Conference on*, pages 1–5. IEEE, 2016.

[127] M. H. Li, A. Yadollahi, and B. Taati. Non-contact vision-based cardiopulmonary monitoring in different sleeping positions. *IEEE Journal of Biomedical and Health Informatics*, 2016.

[128] K. S. Tan, R. Saatchi, H. Elphick, and D. Burke. Real-time vision based respiration monitoring system. In *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*, pages 770–774. IEEE, 2010.

[129] M. Z. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *Biomedical Engineering, IEEE Transactions on*, 58(1):7–11, 2011.

[130] N. Koolen, O. Decroupet, A. Dereymaeker, K. Jansen, J. Vervisch, V. Matic, B. Vanrumste, G. Naulaers, S. V. Huffel, and M. D. Vos. Automated respiration detection from neonatal video data. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, pages 164–169, 2015.

[131] H. U. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman. Eulerian video magnification for revealing subtle changes in the world. 2012.

[132] D. M. Tveit, K. Engan, I. Austvoll, and Ø Meinich-Bache. Motion based detection of respiration rate in infants using video. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 1225–1229. IEEE, 2016.

[133] H. Aoki, K. Koshiji, H. Nakamura, Y. Takemura, and M. Nakajima. Study on respiration monitoring method using near-infrared multiple slit-lights projection. In *Micro-NanoMechatronics and Human Science, 2005 IEEE International Symposium on*, pages 291–296. IEEE, 2005.

[134] M. Martinez and R. Stiefelhagen. Breath rate monitoring during sleep using near-ir imagery and pca. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3472–3475. IEEE, 2012.

[135] M. Bartula, T. Tigges, and J. Muehlsteff. Camera-based system for contactless monitoring of respiration. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 2672–2675. IEEE, 2013.

[136] A. Heinrich, F. van Heesch, B. Puvvula, and M. Rocque. Video based actigraphy and breathing monitoring from the bedside table of shared beds. *Journal of Ambient Intelligence and Humanized Computing*, 6(1):107–120, 2015.

[137] J. Pardey, D. Roberts, and L. Tarassenko. A review of parametric modelling techniques for eeg analysis. *Medical Engineering & Physics*, 18(1):2–11, 1996.

[138] C. L. Mason and L. Tarassenko. Quantitative assessment of respiratory derivation algorithms. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 2, pages 1998–2001. IEEE, 2001.

[139] S. Nemati, S. Malhotra, and G. D. Clifford. Data fusion for improved respiration rate estimation. *Journal on Advances in Signal Processing*, 2010:10, 2010.

[140] C. Orphanidou, O. Brain, J. Feldmar, S. Khan, J. Price, and L. Tarassenko. Spectral fusion for estimating respiratory rate from the ecg. In *Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on*, pages 1–4. IEEE, 2009.

[141] J. F. Thayer, J. J. Sollers III, E. Ruiz-Padial, and J. Vila. Estimating respiratory frequency from autoregressive spectral analysis of heart period. *IEEE Engineering in Medicine and Biology*, 21(4):41–45, 2002.

[142] C. Orphanidou, S. Fleming, S. A. Shah, and L. Tarassenko. Data fusion for estimating respiratory rate from a single-lead ecg. *Biomedical Signal Processing and Control*, 8(1):98–105, 2013.

[143] S. G. Fleming and L. Tarassenko. A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram. *International Journal of Biological and Medical Sciences*, 2(4):232–6, 2007.

[144] J. Lee and K. H. Chon. An autoregressive model-based particle filtering algorithms for extraction of respiratory rates as high as 90 breaths per minute from pulse oximeter. *Biomedical Engineering, IEEE Transactions on*, 57(9):2158–2167, 2010.

[145] G. Mazeika and R. Swanson. Respiratory inductance plethysmography an introduction. `https://go.aastweb.org/Resources/focusgroups/rip_intro.pdf`.

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

[147] B. Krieger, D. Feinerman, A. Zaron, and F. Bizousky. Continuous noninvasive monitoring of respiratory rate in critically iii patients. *Chest*, 90(5):632–634, 1986.

[148] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A Clifton, R. Beale, and P. J. Watkinson. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. *Physiological measurement*, 37(4):610, 2016.

[149] A. Burgos, A. Goñi, A. Illarramendi, and J. Bermúdez. Real-time detection of apneas on a pda. *IEEE Transactions on Information Technology in Biomedicine*, 14(4):995–1002, 2010.

[150] E. Gil, J. M. Vergara, and P. Laguna. Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children. *Biomedical Signal Processing and Control*, 3(3):267–277, 2008.

[151] P. M. Macey, J. S. J. Li, and R. P. K. Ford. Expert system for the detection of apnoea. *Engineering Applications of Artificial Intelligence*, 11(3):425–438, 1998.

[152] P. Várady, T. Micsik, S. Benedek, and Z. Benyó. A novel method for the detection of apnea and hypopnea events in respiration signals. *IEEE Transactions on Biomedical Engineering*, 49(9):936–942, 2002.

[153] J. Han, H. B. Shin, D. U. Jeong, and K. S. Park. Detection of apneic events from single channel nasal airflow using 2nd derivative method. *Computer Methods and Programs in Biomedicine*, 91(3):199–207, 2008.

[154] H. Nakano, T. Tanigawa, T. Furukawa, and S. Nishima. Automatic detection of sleep-disordered breathing from a single-channel airflow record. *European Respiratory Journal*, 29(4):728–736, 2007.

[155] H. Nazeran, A. Almas, K. Behbehani, J. Burk, and E. Lucas. A fuzzy inference system for detection of obstructive sleep apnea. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 2, pages 1645–1648. IEEE, 2001.

[156] P. J. van Houdt, P. P. W. Ossenblok, M.G. Van Erp, K. E. Schreuder, R. J. J. Krijn, P. A. J. M. Boon, and P. J. M. Cluitmans. Automatic breath-to-breath analysis of nocturnal polysomnographic recordings. *Medical & Biological Engineering & Computing*, 49(7):819–830, 2011.

[157] D. Pitson and J. Stradling. Autonomic markers of arousal during sleep in patients undergoing investigation for obstructive sleep apnoea, their relationship to eeg arousals, respiratory events and subjective sleepiness. *Journal of Sleep Research*, 7(1):53–59, 1998.

[158] B. H. Taha, J. A. Dempsey, S. M. Weber, M. S. Badr, J. B. Skatrud, T. B. Young, A. J. Jacques, and K. C. Seow. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep*, 20(11):991, 1997.

[159] J. Y. Tian and J. Q. Liu. Apnea detection based on time delay neural network. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 2571–2574. IEEE, 2005.

[160] D. Sommermeyer, D. Zou, L. Grote, and J. Hedner. Detection of sleep disordered breathing and its central/obstructive character using nasal cannula and finger pulse oximeter. *Journal of Clinical Sleep Medicine*, 8(5):527–533, 2012.

[161] Y. Maali and A. Al-Jumaily. Genetic fuzzy approach based sleep apnea/hypopnea detection. *International Journal of Machine Learning and Computing*, 2(5):685, 2012.

[162] A. Otero, P. Félix, A. Barro, and C. Zamarrón. A structural knowledge-based proposal for the identification and characterization of apnoea episodes. *Applied Soft Computing*, 12(1):516–526, 2012.

[163] D. Álvarez-Estévez and V. Moret-Bonillo. Fuzzy reasoning used to detect apneic events in the sleep apnea-hypopnea syndrome. *Expert Systems with Applications*, 36(4):7778–7785, 2009.

[164] H. M. Al-Angari and A. V. Sahakian. Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier. *Information Technology in Biomedicine, IEEE Transactions on*, 16(3):463–468, 2012.

[165] Y. Takemura, J. Sato, and M. Nakajima. A respiratory movement monitoring system using fiber-grating vision sensor for diagnosing sleep apnea syndrome. *Optical Review*, 12:46–53, 2005. 10.1007/s10043-005-0046-6.

[166] C. W. Wang, A. Hunter, N. Gravill, and S. Matusiewicz. Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea. *Biomedical Engineering, IEEE Transactions on*, 61(2):396–404, 2014.

[167] W. H. Tsai, W. Flemons, W. A. Whitelaw, and J. E. Remmers. A comparison of apnea-hypopnea indices derived from different definitions of hypopnea. *American Journal of Respiratory and Critical Care Medicine*, 159(1):43, 1999.

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

[169] A. Roebuck and G. D. Clifford. Comparison of standard and novel signal analysis approaches to obstructive sleep apnoea classification. *Frontiers in Bioengineering and Biotechnology*, 3:114, 2015.

[170] A. Kulkas, P. Tiihonen, P. Julkunen, E. Mervaala, and J. Töyräs. Desaturation delay, parameter for evaluating severity of sleep disordered breathing. In *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China*, volume 39, page 336. Springer Science & Business Media, 2013.

[171] A. Malhotra, M. Younes, S. T. Kuna, R. Benca, C. A. Kushida, J. Walsh, A. Hanlon, B. Staley, A. I Pack, and G. W. Pien. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep*, 36(4):573–582, 2013.

[172] U. J. Magalang, N. H. Chen, P. A. Cistulli, A. C. Fedson, T. Gíslason, D. Hillman, T. Penzel, R. Tamisier, S. Tufik, G. Phillips, et al. Agreement in the scoring of respiratory events and sleep among international sleep centers. *Sleep*, 36(4):591–596, 2013.

[173] S. D. Pittman, M. M. MacDonald, R. B. Fogel, A. Malhotra, K. Todros, B. Levy, A. B. Geva, and D. P. White. Assessment of automated scoring of polysomnographic recordings in a population with suspected sleep-disordered breathing. *Sleep*, 27(7):1394–1403, 2004.

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

[175] T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.