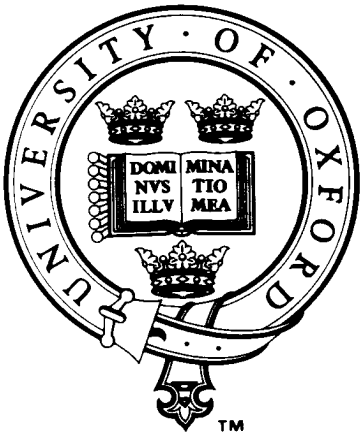
Investigation of Breathing-Disordered Sleep Quantification Using the Oxygen Saturation Signal



Lisa Lazareck
St John's College

Supervised by
Professor Lionel Tarassenko
Submitted: Trinity Term 2008

This DPhil Thesis is submitted to the
Department of Engineering Science, University of Oxford.

Lisa Lazareck
St John's College

Doctor of Philosophy
Trinity Term, 2008

Investigation of Breathing-Disordered Sleep Quantification Using the Oxygen Saturation Signal

Abstract

This thesis investigates the feasibility of using the non-invasive biomedical signal of oxygen saturation, or SpO_2, to diagnose a sleep disorder known as Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS). Epidemiologically, OSAHS is the most common condition investigated by sleep clinics. In a patient suspected of having the disorder, the upper airway is obstructed during sleep and a cessation in respiration results. An apnoea is defined as a temporary cessation of breathing. Similarly, a hypopnoea is defined as any reduction in breathing (i.e., less severe than an apnoea).

The work has three main objectives; the first being to establish automated evaluation procedures for methods of quantifying apnoeic activity from the SpO_2 signal, the second being to accurately identify apnoeic and normal activity on a minute-by-minute basis, the third being to create a Respiratory Disturbance Index (RDI) based on the analysis which is comparable to the gold-standard Apnoea Hypopnoea Index (AHI) derived by experts.

The detection of apnoeic activity is determined using three separate analyses: time domain, frequency domain, and autoregressive modelling with an incorporated amplitude criterion. A training dataset is utilised for algorithm development, and an independent dataset is employed for testing . All three methods result in comparable overall classification accuracies of: 81.2% (time domain), 82.1% (frequency domain), and 80.0%(autoregressive modelling with amplitude). In addition, particular attention is given to the resultant sensitivity, specificity, and accuracy values partitioned according to patient category; i.e., patients with OSAHS may be divided into normal, mild, moderate and severe.

Lastly, a simple RDI is computed based on the automated analyses; i.e., the number of apnoeic segments detected divided by the total number of segments used. A comparison between computed RDI and AHI values for the test database show correlation values above 0.8. In conclusion, this thesis shows that through the automated analysis of the SpO_2 signal, OSAHS severity in patients suspected of having the disorder can be quantified. The AR-modelling with an incorporated amplitude criterion, in particular, shows the most promise for further work in this area.

Acknowledgments

This thesis is entirely dedicated to my parents, Arnold and Libby, who have been there for me - always. From the bottom of my heart, I thank you and I love you.

I would like to thank my supervisor, Professor Lionel Tarassenko, for his encouragement, patience, tenacity, and advice throughout the course of my D.Phil. I am grateful to Ms Fiona Buchanan and Dr. Adrian Kendrick of the Bristol General Hospital for their clinical input and time.

I would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada and the University of Manitoba for their financial support.

Thankyou to all those who I worked with in the SPANNER and BSP research groups during my time in Oxford. In particular, thanks to: Dr. Steve Collins, Dr. Dileepan Joseph, Ms Jan Minchington, Ms Val Mitchell, Dr. Stephen Payne, and countless others within the Department of Engineering Science for your support over the years.

Thankyou to the Institute of Electrical and Electronics Engineers, Inc. (IEEE). Thankyou to the Engineering in Medicine and Biology Society (EMBS) - especially the AdCom and other volunteers. I am eternally grateful for the opportunities you have given me - thankyou so much Laura Wolf, and Jodi Janiszewski. Also, thankyou to the volunteers of Graduates Of the Last Decade (GOLD).

Thankyou to all those in St John's College - both staff and friends - who have made my life in Oxford comfortable and enjoyable. Thankyou Sir Michael and Lady Scholar for your kindness. Thankyou to the MCR (2003-2008) for the memories. Thankyou to Dr. Timothy Crafter and family and cats for your steadfast support. Thankyou linguists, bad boys, and especially Emilie Holland for a final year of giggles (and a keen proofreading eye). Thankyou to the Buttery Staff - in particular Anne, Kay, Maureen - for your caring. Thank you Paul Ashman, Lynda Morgan, and Andy O'Donnell. Thankyou Debs for the laughs. Thankyou to the Porters of SJC for always being there (and in no particular order: John, Sid, David, Martin, Brian, Roger, Barry, Terry, Peter, David, Trevor, Janet, and Jane). Thankyou Jackie Couling, Eileen Marston. Thankyou Noel and Alicia for the smiles. Thankyou Andrew and Jenny Davis for your friendship.

Thankyou to my friends in London for the much-needed perspective and fun, of course; especially to Claire, Danielle, Karen, Pascale, and the Maurice Family. Thankyou to my friends world-wide that always keep in touch: Andi, Aurora, Helena, Joerg, and Ulrich. Thankyou Otto for making me smile during the last few months of the project.

Thankyou Professor James Peters for your guidance and inspiration to try Oxford. Thankyou Brenda Peterson for your optimism. Thankyou Harold Lugsdin for your love. Thinking of you Ann Ludgsdin.

Thankyou to my sister Jennifer.

Contents

1 Introduction **1**
- 1.1 Aim . . . 1
- 1.2 Motivation . . . 1
  - 1.2.1 Epidemiology . . . 2
  - 1.2.2 OSAHS Diagnosis . . . 3
  - 1.2.3 Home Studies . . . 3
- 1.3 Objectives . . . 4
- 1.4 Overview of thesis . . . 5

2 Sleep **7**
- 2.1 General sleep definition . . . 7
- 2.2 Quantification of sleep . . . 8
  - 2.2.1 Recording the EEG . . . 9
  - 2.2.2 Sleep staging . . . 10
  - 2.2.3 Sleep architecture . . . 14
- 2.3 Summary . . . 16

3 Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS) **18**
- 3.1 Physiology . . . 18
- 3.2 Cause and effect . . . 19
- 3.3 Treatment . . . 20
- 3.4 Definitions: apnoea and hypopnoea . . . 21
  - 3.4.1 Classification of sleep disorders . . . 23
- 3.5 Arousal . . . 23
- 3.6 Respiratory Disturbance Indices (RDIs) . . . 25
- 3.7 Polysomnography . . . 27
- 3.8 Summary . . . 28

4 Pulse Oximetry **31**
- 4.1 The need for oxygen monitoring . . . 31
- 4.2 Oxygen and the human body . . . 32
  - 4.2.1 Definition of oxygen saturation . . . 32
  - 4.2.2 Oxygen content in plasma . . . 33
  - 4.2.3 Oxygen-haemoglobin dissociation curve . . . 34
- 4.3 Absorbance at two wavelengths . . . 35
- 4.4 Beer-Lambert law for oximetry . . . 36
- 4.5 Pulsatile nature of arterial blood . . . 38
- 4.6 Pulse oximeter instrumentation . . . 40
  - 4.6.1 Calibration of pulse oximeters . . . 41
- 4.7 Signal patterns during OSAHS . . . 43

4.8 Review of previous work . . . 45
4.8.1 SpO_2 Signal: Time Domain . . . 46
4.8.2 SpO_2 Signal: Frequency Domain . . . 49
4.8.3 SpO_2 Signal: plus others . . . 50

5 Databases **58**
5.1 Introduction . . . 58
5.2 Training database . . . 59
5.2.1 MIT-BIH Polysomnography . . . 59
5.2.2 Processing sleep annotations . . . 60
5.3 Testing database . . . 61
5.3.1 Bristol . . . 61
5.3.2 Processing sleep annotations . . . 64

6 SpO_2 Time Domain **67**
6.1 Introduction . . . 67
6.2 MIT-BIH PSG Training Database . . . 68
6.2.1 Pre-processing . . . 68
6.2.2 Processing . . . 72
6.2.3 Selection of an appropriate threshold . . . 76
6.2.4 Classification accuracy and results . . . 76
6.3 Bristol Test Database . . . 80
6.3.1 Processing . . . 80
6.3.2 Receiver Operating Characteristic (ROC) method . . . 82
6.3.3 Classification accuracy and results . . . 83
6.4 Discussion and summary . . . 92

7 SpO_2 Frequency Domain **94**
7.1 Introduction . . . 94
7.2 Bristol Database training dataset . . . 97
7.2.1 Pre-processing . . . 97
7.2.2 Selection of an appropriate threshold . . . 101
7.2.3 Classification results on the training dataset . . . 103
7.3 Bristol Database test dataset . . . 106
7.4 Discussion and summary . . . 112

8 AR-modelling of the SpO_2 Signal **114**
8.1 Introduction . . . 114
8.2 Tracking resonant frequencies . . . 115
8.2.1 Pre-processing the SpO_2 signal . . . 115
8.2.2 Illustration of AR-modelling . . . 115
8.3 Identifying dominant frequencies . . . 118
8.4 Results of test dataset analysis . . . 130
8.5 Discussion and summary . . . 131

9 Analysing an Entire Night's Sleep **134**
9.1 Motivation . . . 134
9.2 Comparison of automated analyses . . . 135
9.3 Apnoeic segment identification . . . 138
9.4 An automated RDI . . . 141
9.5 Per-hour analysis of the SpO_2 signal . . . 141
9.6 Overnight analysis of the SpO_2 signal . . . 142

9.7 Overnight analysis using a histogram-based RDI . . . 146
9.8 Optimisation of the RDI threshold values . . . 147
9.9 RDI result comparison with previous work . . . 152
9.10 Summary . . . 153

10 Conclusions **154**
10.1 Overview and conclusions . . . 154
10.1.1 Contributions . . . 158
10.2 Further work . . . 160
10.2.1 Obstructive vs. central apnoeas . . . 160
10.2.2 Apnoeas and hypopnoeas . . . 160
10.2.3 Event duration . . . 161
10.2.4 Other work . . . 162
10.3 Home monitoring using the SpO_2 signal . . . 163

A Pulse Oximetry: Background **165**
A.1 Haemoglobin as oxygen carrier . . . 165
A.2 History of oximetry . . . 167

B Comparison with Expert Scores **169**

C AR Modelling & Pole-Tracking **172**
C.1 Autoregressive Modelling . . . 172
C.2 Pole-Tracking Technique . . . 175
C.3 Pole Assignment Algorithm . . . 180

Bibliography **181**

List of Figures

2.1 The 10-20 International System of Electrode Placement Montage. 12
2.2 Examples of basic waveforms within the EEG. 13
2.3 An illustration of typical EEG signals from each nREM stage. 13
2.4 An example of typical 7.5-hour sleep architecture. 17

3.1 Normal and obstructed view of a patient lying in the supine position. . . . 20
3.2 A 10-channel example of a polysomnogram for one patient. 30

4.1 Shift in oxyhaemoglobin dissociation curve with change in temperature levels. 35
4.2 Absorption spectra of HbO_2 and Hb, including the ranges of wavelength for CO-oximetry and pulse oximetry. 37
4.3 Absorption of light due to the pulsation of arterial blood, venous blood, and tissue 40
4.4 A typical pulsatile signal recorded with a probe when light is shone through the finger. 41
4.5 Components of transmission pulse oximeter probe. 42
4.6 Two relationships between the ratio R and the oxygen saturation of the patient. 42
4.7 A four-minute example of arterial oxygen saturation measured by finger pulse oximetry in a patient with OSAHS. 46

6.1 SpO_2 signal (%) in the time domain from a patient with OSAHS, with 10-minute exemplary highlights of apnoea and normal breathing. 70
6.2 MIT-BIH PSG Database: SpO_2 signals in the time domain from all four patients. 71
6.3 slp60: standard deviation for each 30-sec window (trinary code). 73
6.4 slp59: standard deviation for each 30-sec window (colour-coded). 74
6.5 slp60: standard deviation for each 30-sec window (colour-coded). 74
6.6 slp61: standard deviation for each 30-sec window (colour-coded). 75
6.7 slp67: standard deviation for each 30-sec window (colour-coded). 75
6.8 Time domain analysis: histograms for patients in the MIT-BIH PSG Database, with the 0.7, 1.0, and 1.4 unit s.d. threshold values highlighted. 78
6.9 Bristol Database: ROC curves for the training dataset using time domain analysis and four window sizes: 30-second, 60-second, 90-second, 120-second. 84
6.10 Bristol 11: NORMAL patient, time domain analysis (σ_{SpO_2}). 89
6.11 Bristol 8: MILD patient, time domain analysis (σ_{SpO_2}). 89
6.12 Bristol 21: MODERATE patient, time domain analysis (σ_{SpO_2}). 90
6.13 Bristol 57: SEVERE patient, time domain analysis (σ_{SpO_2}). 90
6.14 Highlighted portion of the SpO_2 signal and Expert Scores for Bristol 8: MILD patient example, time domain analysis. 91

7.1 Highlighted apnoeic portion of the SpO_2 signal in the time domain, with marked period T, for Bristol 21: MODERATE patient example. 96
7.2 Apnoeic and normal breathing highlights of the SpO_2 signal for Bristol 21: MODERATE patient example, and the corresponding PSD curves. 100
7.3 ROC curves for the Bristol Database training dataset using frequency domain analysis ($P_{ave,apnoea}$) and four window sizes: 30-second, 60-second, 90-second, 120-second. 105
7.4 ROC curves for the Bristol Database training dataset using frequency domain analysis ($PowerRatio$) and four window sizes: 30-second, 60-second, 90-second, 120-second. 105
7.5 Bristol 11: NORMAL patient, frequency domain analysis ($P_{ave,apnoea}$). . . . 110
7.6 Bristol 8: MILD patient, frequency domain analysis ($P_{ave,apnoea}$). 110
7.7 Bristol 21: MODERATE patient, frequency domain analysis ($P_{ave,apnoea}$). . 111
7.8 Bristol 57: SEVERE patient, frequency domain analysis ($P_{ave,apnoea}$). . . . 111

8.1 Spectral estimation with a 10^{th}-order AR-model of a noisy input signal. . . 117
8.2 Highlighted 120-second segment magnified of the SpO_2 signal and Expert Scores for Bristol 21: MODERATE patient example, using the 5^{th}-order AR-model. 122
8.3 Bristol 11: NORMAL patient, preliminary pole-zero analysis. 123
8.4 Bristol 8: MILD patient, preliminary pole-zero analysis. 123
8.5 Bristol 21: MODERATE patient, preliminary pole-zero analysis. 124
8.6 Bristol 57: SEVERE patient, preliminary pole-zero analysis. 124
8.7 Bristol 11: NORMAL patient, pole-zero analysis. 126
8.8 Bristol 8: MILD patient, pole-zero analysis. 127
8.9 Bristol 21: MODERATE patient, pole-zero analysis. 128
8.10 Bristol 57: SEVERE patient, pole-zero analysis. 129

9.1 Linear regression analysis using the indices on a per hour basis. 144
9.2 Linear regression analysis using the indices for the entire night. 145
9.3 Linear regression analysis using the indices for the entire night using histogram data from pole-zero analysis. 149

B.1 A Receiver Operating Characteristic (ROC) curve, where the area under the curve (AUC) is equal to classification accuracy. 171

C.1 Linear prediction using a p^{th} order autoregressive model. 174
C.2 Evaluating the magnitude of an AR transfer function at a resonant frequency. 179
C.3 Pole-zero assignment algorithm flowchart. 180

List of Tables

2.1 Standard EEG frequency bands: delta, theta, alpha, sigma, beta, gamma. . 11
2.2 Sleep stages and their most prominent features according to R & K scoring. 14
2.3 Division into sleep stages by EEG, EMG and EOG activities. 14
2.4 Changes in sleep length and content with age. 15

3.1 Classification of the degrees of OSAHS severity according to AHI values. . 26

4.1 Summary of SpO_2 time-domain studies, 1989-1995. 49

5.1 MIT-BIH Polysomnography Database: annotation coding scheme. 60
5.2 MIT-BIH Polysomnography Database: percentage of apnoeas and hypopnoeas per patient, with respect to the total time spent asleep. 61
5.3 Bristol Database: typical channel configuration for an all-night, full, polysomnography (PSG). 64
5.4 Bristol Database: annotation scheme. 65
5.5 Bristol Database: the total event durations divided by the total duration of the sleep recordings for each patient on a per-category basis. 66
5.6 Bristol Database: the total number of each event type divided by the total number of events scored, shown as a percentage, for each patient on a per-category basis. 66

6.1 MIT-BIH PSG Database: number of segments analysed, total number of segments, and percentage of artefact for the four patients. 69
6.2 Time domain analysis: classification sensitivities (Se), specificities (Sp), and accuracies (Acc) for each patient in the MIT-BIH PSG Database, using threshold values of 0.7, 1.0, 1.4 units s.d. 79
6.3 Time domain analysis resultant Se, Sp, and Acc percentages across the 47-patient Bristol Database training dataset for 30-second, 60-second, 90-second, 120-second window lengths, and corresponding optimal (ROC-based) threshold values. 83
6.4 Time domain analysis resultant Se, Sp, and Acc percentages across the 44-patient Bristol Database training dataset, on an overall and *per category* basis, for 120-second window length and 0.778 units s.d. threshold. 87
6.5 NORMAL patients' confusion matrix (time domain analysis). 88
6.6 MILD patients' confusion matrix (time domain analysis). 88
6.7 MODERATE patients' confusion matrix (time domain analysis). 88
6.8 SEVERE patients' confusion matrix (time domain analysis). 88

7.1 Expected period and frequency ranges for apnoeic and normal breathing activity. 95

7.2 Selected N-point DFT and corresponding frequency ranges (apnoea, normal breathing) for selected window sizes: 30-second, 60-second, 90-second, 120-second. 98
7.3 Frequency domain analysis resultant Se, Sp, and Acc percentages across the 47-patient Bristol Database training dataset for 30-second, 60-second, 90-second, 120-second window lengths, and corresponding optimal (ROC-based) threshold values. 104
7.4 Frequency domain analysis resultant Se, Sp, and Acc percentages across the 44-patient Bristol Database testing dataset, on an overall and *per category* basis, for 120-second window length and 0.51 units threshold. 108
7.5 NORMAL patients' confusion matrix (frequency domain analysis). 109
7.6 MILD patients' confusion matrix (frequency domain analysis). 109
7.7 MODERATE patients' confusion matrix (frequency domain analysis). . . . 109
7.8 SEVERE patients' confusion matrix (frequency domain analysis). 109

8.1 Expected period, frequency, and Θ ranges for apnoeic and breath events. . 118
8.2 Pole-Zero analysis resultant Se, Sp, and Acc percentages across the 44-patient Bristol Database testing dataset, on an overall and *per category* basis, for 120-second window length and 0.778 units s.d.threshold. 131
8.3 NORMAL patients' confusion matrix (pole-zero analysis). 133
8.4 MILD patients' confusion matrix (pole-zero analysis). 133
8.5 MODERATE patients' confusion matrix (pole-zero analysis). 133
8.6 SEVERE patients' confusion matrix (pole-zero analysis). 133

9.1 Se, Sp, and Acc percentages across the 44-patient test dataset, on an overall and *per category* basis, using a 120-second window for segmentation and the ROC-based threshold (from the training dataset) for all three methods. 137
9.2 Percentage of apnoeic segments found using manual, time domain, frequency domain, and AR-modelling with amplitude criterion analyses. . . . 139
9.3 Manual Analysis: the percentage of apnoeic segments as found by the experts are shown for the test example from each patient category. 140
9.4 Automated Analysis: Time Domain, with the percentage of apnoeic segments shown for the test example from each patient category. 140
9.5 Automated Analysis: Frequency Domain, with the percentage of apnoeic segments shown for the test example from each patient category. 140
9.6 Automated Analysis: AR-modelling with amplitude criterion, with the percentage of apnoeic segments shown for the text example from each patient category. 140
9.7 Time domain analysis confusion matrix for the 44-patient test dataset, for RDI-AHI comparison. 150
9.8 Frequency domain analysis confusion matrix for the 44-patient test dataset, for RDI-AHI comparison. 150
9.9 AR-modelling with amplitude criterion analysis confusion matrix for the 44-patient test dataset, for RDI-AHI comparison. 150
9.10 AR-modelling with amplitude criterion analysis ($\mathrm{RDI}_{hist-apnoea/breath}$) confusion matrix for the 44-patient test dataset, for RDI-AHI comparison. . . 151
9.11 AR-modelling with amplitude criterion analysis ($\mathrm{RDI}_{hist-apnoea}$) confusion matrix for the 44-patient test dataset, for RDI-AHI comparison. 151

B.1 Confusion matrix describing the metrics corresponding to the 'manual' and 'automated' decisions. 171

Glossary

a.u.	arbitrary units
AASM	American Academy of Sleep Medicine
Acc	Accuracy
AHI	Apnoea Hypopnoea Index
ASDA	American Sleep Disorders Association
CPAP	continuous positive airway pressure
ECG	electrocardiogram
EDS	Excessive Daytime Sleepiness
EEG	electroencephalogram
EMG	electromyogram
EOG	electrooculogram
FN	False Negative
FP	False Positive
Hb	reduced haemoglobin
HbCO	carboxyhaemoglobin
HbO_2	oxyhaemoglobin
HRV	Heart Rate Variability
metHb	methaemoglobin
nCPAP	(or nasal-CPAP), CPAP applied via nasal mask
nREM	(or non-REM), non-Rapid Eye Movement
OSAHS	Obstructive Sleep Apnoea Hypopnoea Syndrome
PCO_2	partial pressure of carbon dioxide
PO_2	partial pressure of oxygen
PSG	polysomnography
R&K	Rechtschaffen and Kales, standardised sleep stage scoring technique
RDI	Respiratory Disturbance Index
REM	Rapid Eye Movement
RIP	respiratory inductance plethysmography

ROC	Receiver Operator Characteristic
s.d.	standard deviation (units of)
SaO_2	arterial oxygen saturation
SDB	Sleep Disordered Breathing
Se	Sensitivity
Sp	Specificity
SpO_2	oxygen saturation signal, as measured by pulse oximetry
SPT	Sleep Period Time
TN	True Negative
TP	True Positive
TST	Total Sleep Time

Chapter 1

Introduction

1.1 Aim

The aim of this thesis is to investigate the automated quantification of sleep disorder severity from a non-invasive sensor. The signal of interest is the oxygen saturation, or SpO_2, signal. The SpO_2 signal is chosen for investigation because it can be obtained non-invasively, it is a direct marker of respiration, and is currently utilised in clinical environments. Hence a successful solution based on this signal would not require additional hardware or monitoring devices.

1.2 Motivation

Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS) is a condition whereby the upper airway is obstructed and a cessation in respiration results. An apnoea is defined as a temporary cessation of breathing. There are three main types of sleep apnoea: obstructive, central, and mixed. Similarly, a hypopnoea is defined as any reduction in breathing (i.e., less severe than an apnoea). The term OSAHS is not, strictly speaking, applicable to the entire spectrum of sleep-disordered breathing (SDB); however, it is best reserved for individuals with the most severe form of the disease - the mild end of

the spectrum being snoring[1] [58]. Currently, the most common metric used to describe and quantify SDB is the respiratory-disturbance index (RDI) referred to as the apnoea-hypopnoea index, or AHI [94]. The AHI, a manual calculation of the average number of apnoeas plus hypopnoeas observed per hour of sleep, is insufficient alone for diagnosing OSAHS, and must be accompanied by characteristic (subjective) complaints, most notably daytime somnolence or sleepiness [60, 94].

1.2.1 Epidemiology

Epidemiologically, OSAHS is the most common condition investigated by sleep clinics [87]. In a 2002 study, a random selection of middle-aged working adults revealed that 9% of women and 24% of men had some sleep-disordered breathing [58]. Within this same study, and with the addition of the characteristic complaint of hypersomnolence, 2% of women and 4% of men met the minimal criteria for OSAHS. OSAHS frequency has been reportedly greater for the elderly, with estimates of 20-54% in women and 28-67% in men [58]. Additional statistics reveal that men with OSAHS are mostly middle-aged (85%) and approximately half are at least 30% or more than their ideal body weight [126]. Also, 20-30% of patients with hypertension (high blood pressure, a notable OSAHS-related complication) in the USA have unrecognised sleep apnoea [60]. Lastly, Koenig and Suratt report that the incidence of sleep disorders is as high as 46% for commercial truck drivers. For those with OSAHS, there is a seven-fold rate increase, in comparison with the general population, in automobile (and other work-related) accidents, which is presumed to be due to daytime sleeping and visual-motor incoordination (at repetitive tasks requiring vigilance) [58]. Overall, it is estimated that 18 million Americans suffer from OSAHS, costing $42 million annually in hospitalisation alone. Furthermore, 95% of cases are unsuspected.

[1] Interestingly, snoring is present in more than 80% of individuals with OSAHS, as the most common site of upper airway abnormality and narrowing is the region around the soft palate [58].

1.2.2 OSAHS Diagnosis

The current gold-standard method for the diagnosis of OSAHS and other sleep-related disorders is the 'sleep test' or polysomnography (PSG). PSG is a multi-channel recording of sleep and breathing variables that is traditionally performed overnight in a sleep laboratory, with a trained technician in attendance. Some believe the entire PSG method to be flawed, as the sleep test only became standard practice (during classical sleep studies) when the associated pathophysiology and range of obstructive sleep apnoea hypopnoea syndromes were not adequately understood [87]. The main disadvantages of PSG are such that it is an expensive system to set up and run; the sleep laboratory is an artificial environment; the definitions of apnoea and hypopnoea are not uniform amongst sleep labs world-wide; and the interpretation time for the various PSG-acquired signals may be from 4 to 6 hours [6, 101]. Sleep laboratories are currently under-funded, and the pressure of increasing referrals of OSAHS is on the rise [6]. Furthermore, sleep scoring is currently being performed using the rules outlined in the 1968 manual by Rechtschaffen and Kales [93]. When this manual was originally produced, however, it was designed to be pro-forma for a standard rather than a gold-standard [74]. Thus, new techniques need to be developed to help assess the performance of the PSG in terms of diagnostic potential for OSAHS subjects, and more importantly, investigate the ability of methods to identify apnoeic episodes by computer and assess the degree of sleep disorder severity automatically.

1.2.3 Home Studies

The replacement of full PSG with more simple and convenient tests for home studies, is a current issue amongst sleep researchers. Stradling, in 1993, notes that 'if all night records of oxygen saturation, heart rate, snoring and movement fail to reveal any abnormality, it is unlikely that the subject has a breathing-related sleep disorder' [26, 106]; i.e., PSG is not necessarily always required to confirm SDB in patients. Cost constraints

typically dictate whether sleep studies will be continuously monitored by a technician from a centralised hub or unattended entirely, such as that for home studies [65]. The most cost effective pathway for the diagnosis and management of OSAHS has yet to be established [6]; however, many sleep laboratories have reportedly used a combination of diagnostic modalities, such as: limited channel PSG; a combination of video, microphone and oximetry monitoring; nasal pressure monitoring; and overnight oximetry both within and outside of the clinical setting [6]. The simplest test is the measurement of the SpO_2 signal using overnight pulse oximetry, which has three major advantages: it is readily available, relatively inexpensive, and can be performed at the patient's home for a more 'typical' night's sleep for subsequent analysis [6]. Furthermore, pulse oximetry is currently being used in several sleep labs as an initial screening test for most patients with suspected OSAHS. Several methods have been proposed which automatically quantify apnoeic behaviour from the SpO_2 signal, and a description of these is found in Chapter 4. However, the link between the automated analysis of the SpO_2 signal and the resultant scoring via PSG for diagnosing OSAHS has not been fully researched. This thesis concentrates on the comparison of the automated analysis of the SpO_2 signal acquired overnight with pulse oximetry with the resultant clinical scoring by experts using the gold-standard PSG method.

1.3 Objectives

The objective of this thesis is to use signal processing techniques to provide respiratory information from the non-invasive SpO_2 signal, which is routinely obtained in both sleep laboratory and home environments. The main objectives of the work are three-fold.

First, there are many published works which examine the quantification of apnoeic activity from the SpO_2 signal including [39, 41, 63, 80, 119]; however, none of these present entirely objective and reproducible evaluations of the performance of their methods, as discussed at length in Chapter 4, Section 4.8. Such evaluation is necessary to compare

the accuracy of one method with another. Thus, the first aim of this thesis is to establish objective, reproducible and automated evaluation procedures for methods of quantifying apnoeic activity from the SpO_2 signal using time and frequency domain analyses.

Methods developed in this work aim to quantify individual apnoeic activity at resolution of seconds and minutes. The methods described in the literature reportedly use a resolution of hours, or diagnose OSAHS in patients using the SpO_2 signal from the *entire* night. Therefore, the second objective of this thesis is to develop a method which finds an accurate estimate of apnoeic activity on at least a minute-by-minute basis and is capable of working across the entire population of OSAHS patients (categorised as normal, mild, moderate, and severe - see Chapter 3).

A robust description of the entire night's sleep fragmentation may then be achieved through the accurate identification of apnoeic segments in the SpO_2 signal. The third objective of this thesis is to create a simple RDI based on the above results, which is comparable to the AHI derived by experts. We seek to answer whether or not the degree of apnoeic severity in patients with suspected OSAHS may be automatically quantified using the non-invasive SpO_2 signal, and whether diagnosis of the OSAHS may be performed with accuracy comparable to that achieved by the experts.

1.4 Overview of thesis

This thesis is organised as follows:

Chapter 2 provides the general definition of sleep and introduces the relevant biomedical signals, namely the electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG). The recording of these signals and their intrinsic patterns are discussed, followed by the definitions of sleep staging and sleep architecture.

Chapter 3 introduces OSAHS, including its associated physiology, consequences and treatment. Monitoring and diagnosis of OSAHS are discussed, including extensive definitions of events: apnoea, hypopnoea, arousal; and various Respiratory Disturbance Indices

(RDIs).

Chapter 4 provides a review of pulse oximetry - or the method of measuring arterial oxygen saturation within the human body. The oxygen saturation signal patterns during OSAHS are described followed by the literature review of previous approaches to detecting OSAHS using this signal, and other biomedical signals in conjunction with the oxygen saturation signal.

Chapter 5 introduces the databases used in this thesis, including the MIT-BIH Polysomnography Database and the Bristol Database. The former database of 4 patients is used for preliminary algorithm development. The latter database of 91 patients is used for further algorithm training and testing.

Chapter 6 begins the investigation of algorithms for detecting apnoeic activity using the SpO_2 signal. A time domain approach using standard deviation values is implemented and provides a set of benchmark results against which a frequency domain technique is then compared.

Chapter 7 focuses on frequency domain analysis to determine the average power in the two areas of interest within the SpO_2 signal spectrum: apnoea and normal breathing. Metrics based on average power metric and on the ratio of average power (apnoea to normal breathing) are investigated.

Chapter 8 introduces a new method to identify the dominant frequencies within the SpO_2 signal using both autoregressive modelling of the segmented SpO_2 signals, with the aim of using these frequencies to identify apnoeic and normal breathing activity.

Chapter 9 compares the sensitivity, specificity and accuracy results of Chapters 6 through 8. The identification of apnoeic segments is explored in order to assess the entire night's sleep record in terms of OSAHS severity. A simple RDI is computed based on the results, from which a comparison with the expert-derived AHI values follows.

Chapter 10 summarises the findings and draws conclusions on the viability of automated methods for detecting apnoeic activity from the SpO_2 signal. Possible avenues of further work are also discussed.

Chapter 2

Sleep

2.1 General sleep definition

Sleep is defined as the unconscious state or condition regularly recurring and naturally assumed by man, during which the activity of the nervous system and consciousness are almost or entirely suspended [1]. Sleep may also be explained as a reversible behavioural state of perceptual disengagement from and unresponsiveness to the environment [15]. Approximately one third of our adult lives are spent asleep. Currently, there is agreement amongst sleep researchers that sleep is regulated by two factors: duration of wakefulness (homoeostatic drive to sleep) and time of day (circadian drive to sleep). The absolute drive to sleep at any point in time is therefore a combination of these two factors [60]. The exact function of sleep has yet to be determined; however, some theories believe the possible function of sleep to be for any one of the following: protective behaviour, energy conservation, brain restoration, homeostasis, improving immune function, and temperature regulation [60].

Sleep is a complex combination of physiological and behavioural processes [15]. Typically, but not necessarily, sleep is accompanied by postural recumbence, behavioural quiescence, and closed eyes [15]. Sleep tends to emerge in the late evening when the lights are dim and the core body temperature begins to decline [24]. Furthermore, during

sleep, most hormones show individual patterns of secretion that intervene importantly in the homeostasis of body systems [24]. As will be discussed shortly, two distinct states of being dominate brain activity during sleep: non-Rapid Eye Movement (non-REM, or nREM) and Rapid Eye Movement (REM). nREM sleep is associated with quiescence, repose, low body temperature, and hypometabolism. This sleep state is further divided into four different stages: stage 1 (light sleep), stage 2 (consolidated sleep), stages 3 and 4 (collectively known as deep or slow wave sleep) [60]. REM sleep is a state of neural instability and high cerebral activity [24]. Control of respiration differs significantly between sleep and wakefulness. The respiratory system is designed to maintain homeostasis of partial pressures of oxygen (PO_2) and carbon dioxide (PCO_2) in the blood, as well as pH[1] regardless of the state of consciousness - see Chapter 4 for further details on oxygen and the human body and pulse oximetry [103]. Sleep modifies the mechanical, chemical, and neural processes controlling ventilation (defined as the oxygenation of the blood during the act of respiration [1]). The two significant factors that influence respiration during wakefulness are removed during sleep, including: the effect of wakefulness on respiratory control and volitional control of ventilation [103]. When a normal subject is awake, ventilation is controlled by two pathways: one driven by the brain-stem respiratory control centre, and the other by the cortex [125]. During sleep, this respiratory centre remains active but the cortex drive disappears causing regular breathing, a fall in ventilation, and rise in the CO_2 concentration [125]. Sleep is not a homogeneous state, and the effects of sleep vary according to the characteristic sleep stages.

2.2 Quantification of sleep

For accurate and consistent division of sleep into its various stages, the 'R & K Scoring' standardised technique is used. In their 1968 publication, Rechtschaffen and Kales produced an atlas of tracings and detailed descriptions on how to score sleep for the sleep

[1]pH is a number expressing the acidity or alkalinity of a solution equal to the logarithm to the base 10 of the reciprocal of the effective concentration (activity) of hydrogen ions (in moles per litre) [1].

research community [93]. The standardisation, known as R & K, is based on physiological tracings of the electroencephalogram (EEG), electromyogram (EMG) and electrooculogram (EOG). First, a brief introduction to the EEG is provided, followed by a description of each sleep stage.

2.2.1 Recording the EEG

The electroencephalogram or EEG is a graphical record of the electrical activity of the brain [60], which was first measured non-invasively in 1929 by Hans Berger.[2] Today, the EEG is used in most medical areas for a variety of purposes; for example, the neurologist uses the EEG for assessment of epilepsy; the general physician for prognostic help with unconscious patients after cardiac arrest and the psychiatrist for evaluation of pre-senile dementias. The EEG is especially helpful with the investigation of various sleep disorders, which is the main focus of this work.

The EEG is recorded by amplifying the difference of electrical potential between two electrodes located on the scalp. First, the scalp EEG signal is amplified, because the signals are small - in a normal adult, a typical range is between 20 and 150 microvolts (μV). Second, the signal is filtered with the frequency band of interest lying between 0.5 and 50 Hz [31]. Third, the signal is recorded electronically via multi-channel digital recorders with storage capabilities.[3]

On the scalp, electrodes are arranged according to the standard '10-20 International System of Electrode Placement' arrangement, as recommended by the International Federation of Societies for Electroencephalography and Clinical Neurophysiology (SEEGCN). The arrangement is usually bipolar, such that the EEG is recorded from one electrode with respect to another. (A monopolar arrangement exists for which the EEG is recorded from multiple electrodes with respect to a common reference electrode.) The 10-20 mon-

[2] In 1875, however, Caton first recorded the electrical activity of the brain using rabbits, monkeys and other small animals [16, 125].

[3] Originally, and as late as 1976, EEG signals were recorded on paper; for example, a routine recording of 15-minute EEG required 90 traced pages [99].

tage is proportional to skull size and shape and entirely dependant upon distances (10% and 20% of the total distance) measured between four scalp landmarks (inion, nasion, left and right mastoid). The conventional configuration is shown in Figure 2.1.

The EEG record is subject to artefact divided into technical and biological source categories. Technical artefact may originate from the electrode-scalp interface (e.g., poor contact) and environmental noise (e.g., hospital apparatus) [99]. Biological artefact may originate from head and body movement (e.g., palate twinge and leg twitch), sweating or any other undesired signal from the body.

2.2.2 Sleep staging

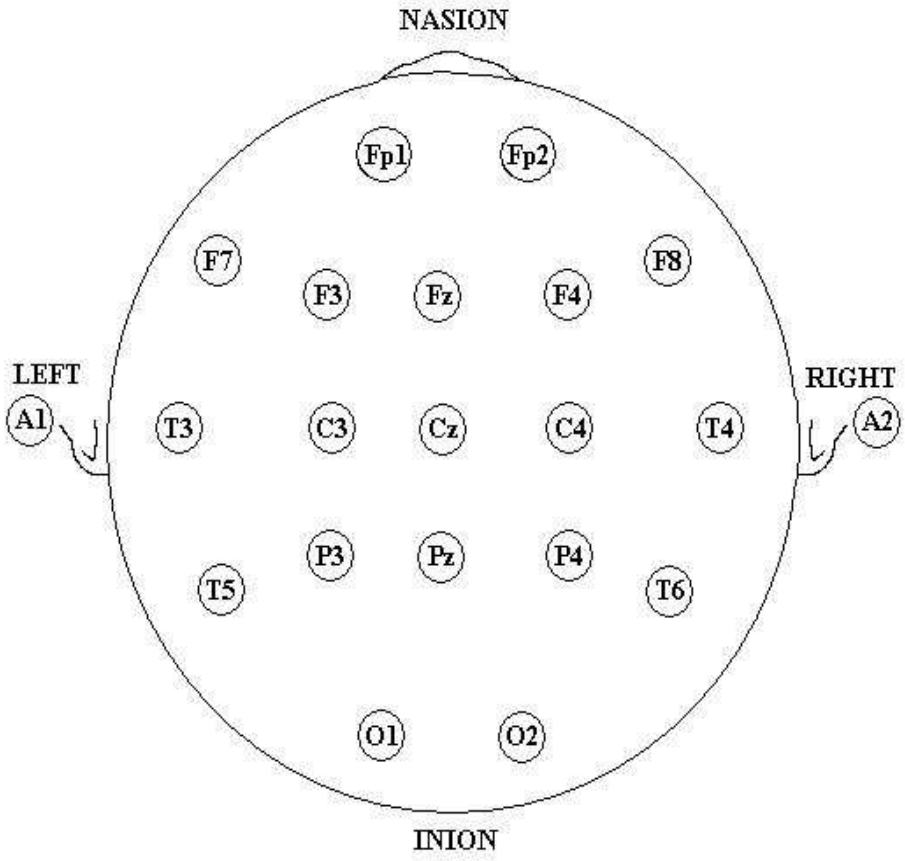
Electrical activity in the brain is divided into several frequency bands, chosen arbitrarily for ease of clinical interpretation. The six band labels are described in Table 2.1, with Figure 2.2 showing typical delta (δ), theta (θ), alpha (α), and beta (β) waveforms.

The R & K standardisation takes measurements from C4/A1 or C3/A2 for the EEG (see Figure 2.1), muscle areas on and beneath the chin (mental, submental) for the EMG, and areas laterally surrounding the eye for the EOG. Both EMG and EOG are acquired with the ear acting as the site for the reference electrode. The scoring procedure is based on an epoch-by-epoch approach, with intervals of 30 or 20 second durations. Each epoch is not considered in isolation, and the scoring system is based on a set of rules incorporating years of clinical experience. Initial epoch-assessment is based on the most prominent features of the EEG signals (Table 2.2). The R & K scoring rules require all three physiological measures to be used and a summary of the measures for each sleep stage is provided in Table 2.3. For example, in Stage 2 (nREM) sleep, the EEG signal exhibits mixed-frequency EEG activity, including θ, sleep spindles, and K-complexes on a background of relatively low voltage. Sleep spindles and K-complexes are transient phenomena. A sleep spindle is defined as EEG activity between 12 and 14 Hz, at least 0.5 seconds in duration. K-complexes are EEG waveforms with well-delineated negative sharp waves immediately followed by a positive component, with total duration exceeding

0.5 seconds [93]. In Stage 2 (nREM) sleep, there is some EMG activity, whereas EOG activity is rare. Figure 2.3 shows typical EEG signals from each nREM (non-Rapid Eye Movement) stage. In-depth and more specific scoring rules are found in [93].

Activity	Symbol	Range (Hz)
delta	δ	[0.5 - 3.5]
theta	θ	[4 - 8)
alpha	α	[8 - 13]
sigma	σ	[14 - 15)
beta	β	[15 - 25]
gamma	γ	[30 - 50]

Table 2.1: Standard EEG frequency bands; "["and "] " meaning inclusive, ")" meaning exclusive. Note that delta activity is limited to [0.5-2] Hz in sleep studies.



Fp1, Fp2	pre-frontal	P3, P4	parietal
F3, F4	frontal	O1, O2	occipital
C3, C4	central	F7, F8	anterior
T3, T4	mid-temporal	Fz	frontal mid-line
T5, T6	posterior temporal	Cz	central vertex
A1, A2	mastoid	Pz	parietal mid-line

Figure 2.1: The 10-20 International System of Electrode Placement Montage.

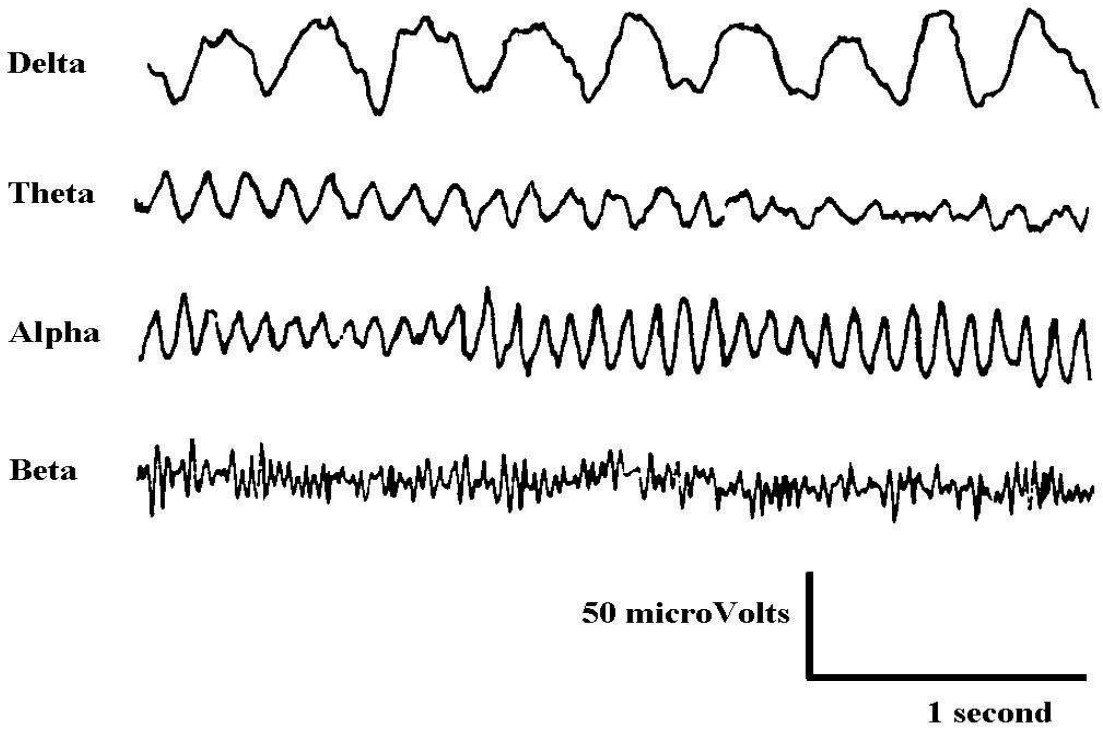


Figure 2.2: Examples of basic waveforms within the EEG, from top to bottom: delta (δ), theta (θ), alpha (α), and beta (β) rhythms [31].

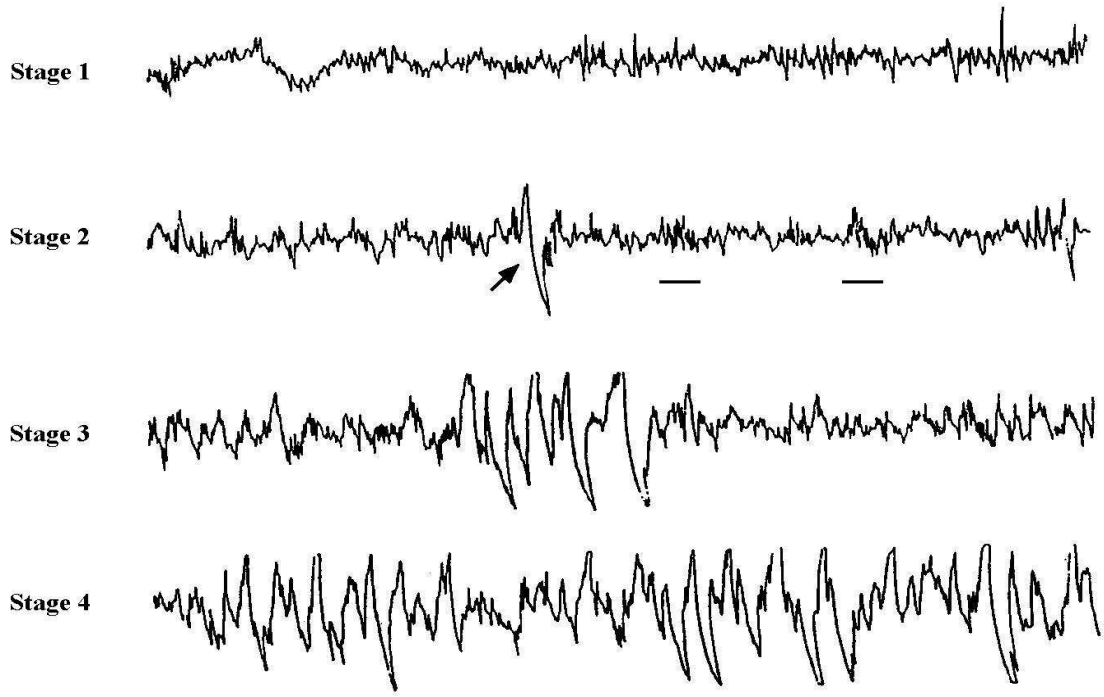


Figure 2.3: An illustration of typical EEG signals from each nREM stage (Stages 1 to 4). The arrow marks a K-complex, while the underlined areas indicate sleep spindles [60].

R & K	Most Prominent Features
Stage W	Wakefulness EEG contains alpha activity and/or low voltage, mixed-frequency activity.
MT	Movement Time Scoring epoch during which the record is obscured by subject movement.
Stage 1	Relatively low voltage, mixed-frequency EEG activity without rapid eye movements.
Stage 2	Sleep spindles and K-complexes on a background of relatively low voltage, mixed-frequency EEG activity.
Stage 3	Moderate amounts of high amplitude, slow wave activity.
Stage 4	Large amounts of high amplitude, slow wave activity.
Stage nREM	non-Rapid Eye Movement Stages 1, 2, 3, and 4 combined. Also known as Stage non-REM.
Stage REM	Rapid Eye Movement A relatively low voltage, mixed-frequency EEG activity in conjunction with episodic REMs and low amplitude EMG.

Table 2.2: Sleep stages and their most prominent features according to R & K scoring [93].

Stage	EEG	EMG	EOG
Stage W	α, β	very high	rapid
Stage 1	some θ	high	slow
Stage 2	θ, spindles, complexes	some	rare
Stage 3	moderate δ	low	rare
Stage 4	much δ	low	rare
Stage REM	α and θ	absent	rapid

Table 2.3: Division into sleep stages by EEG, EMG and EOG activities [60].

2.2.3 Sleep architecture

Sleep is organised into a series of ultradian periods, or cycles of physiological activity, that last approximately one hour in infants and small children and 1.5 hours in adults. The typical 'sleep cycle' is sequentially composed of the following: Stage W, Stage 1, Stage 2, Stage 3, Stage 4, and REM [60]. Thus, for a characteristic sleep period (of say 7.5 hours), the sleep cycle is repeated four or five times during the night [126]. This sleep construct is referred to as 'sleep architecture,' with a typical time plot shown in Figure 2.4.

In general, the relative percentage of slow wave sleep (Stages 3 and 4) is highest in the first sleep cycle and decreases as the night progresses. In Figure 2.4, slow wave sleep occurs approximately 45 minutes after sleep commencement, as expected [99]. Conversely, the relative percentage of REM sleep increases as the night progresses. In the same Figure, REM sleep first appears approximately 1.5 hours after sleep commencement. Periods of REM sleep are of at least 20 minutes in duration, and increase in length with the majority found towards the morning hours [60]. In general, sleep architecture reveals that lightening of sleep takes place in an abrupt fashion whilst deepening of sleep is more gradual [60].

Frequently within the sleep architecture, sleep cycles are incomplete; i.e., not all sleep stages are present for every sleep cycle. In addition, the length and content of sleep architecture varies according to age (infant to elderly status, described in Table 2.4), stress, exercise, mood changes, amount of previous sleep deprivation and various pathological conditions [60]. Lastly, it is common to have short awakenings during sleep cycles which interrupt sleep architecture (or cause 'sleep fragmentation'). These awakenings or *arousals* are discussed in greater detail in the following chapter.

Age Category	Sleep Time (hours)	Stages 1-2 (%)	Stages 3-4 (%)	REM (%)
Infants	13-16	10-30	30-40	40-50
Children	8-12	40-60	20-30	20-30
Adults	6-9	45-60	15-25	15-25
Elderly	5-8	50-80	5-15	15-25

Table 2.4: Changes in sleep length and content with age [60].

2.3 Summary

This Chapter has provided the necessary background on sleep including its physiology and architecture, and sleep staging via the R & K standardisation rules. A brief introduction to the EEG signal was provided in order to fully appreciate sleep scoring. Next, abnormal sleep is described - in particular, the Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS) sleep disorder is introduced with a description of the sleep analyses and subsequent sleep indices used for assessment in the clinical domain.

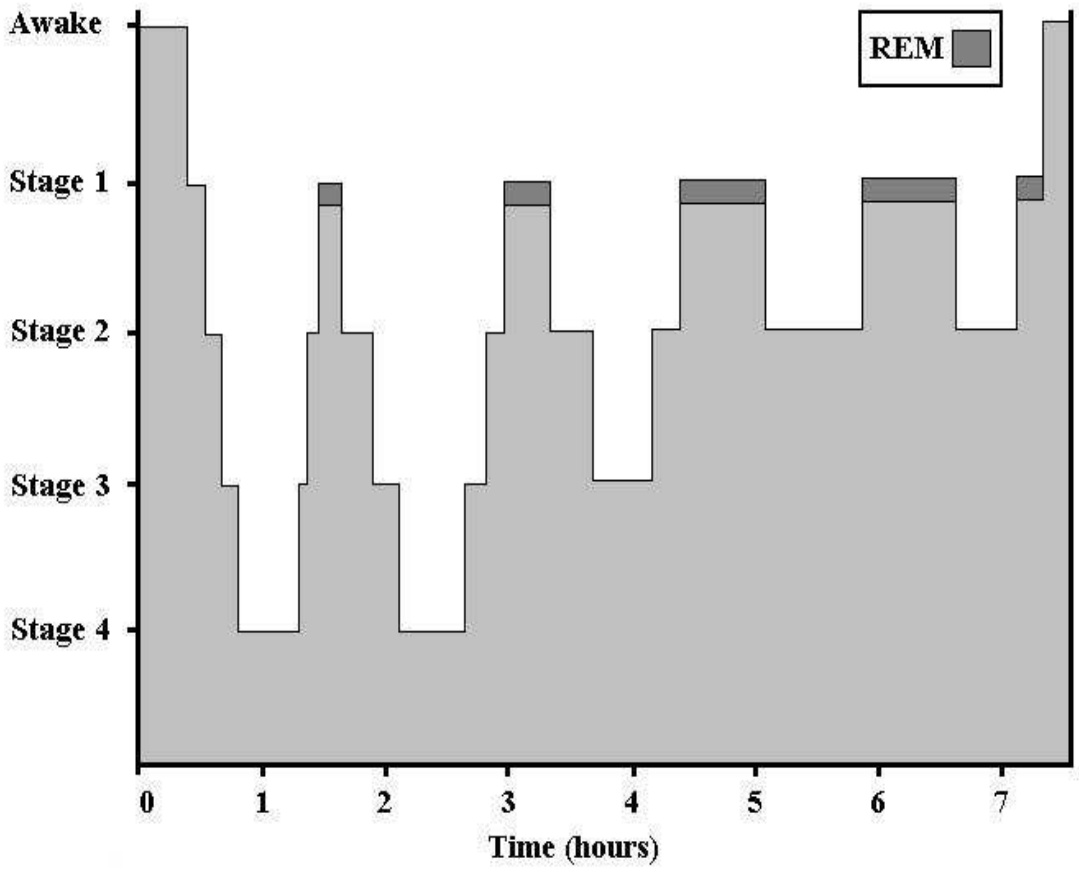


Figure 2.4: An example of typical 7.5-hour sleep architecture [60].

Chapter 3

Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS)

3.1 Physiology

Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS) is a relatively newly recognised disease that causes daytime sleepiness and generally impaired health status. It is also the most common condition investigated by sleep clinics [7,87]. In the most general sense, with the fall into sleep from wakefulness, there is a gradual increase in upper airway narrowing and resistance - even in normal individuals; however, normal compensatory mechanisms exist that counteract this situation, such as increases in upper airway tone [87]. In individuals susceptible to OSAHS, there is an imbalance of forces able to counteract this situation. Thus, in these individuals, narrowing of the upper airway causes an increase in respiratory effort. With this increase, the intrathoracic pressure drops and either an apnoea or hypopnoea transpire. Broadly speaking, apnoeas/hypopnoeas can respectively be complete and near-complete cessations of airflow. These events cause hypoxia (oxygen deficiency) and hypercapnia (carbon dioxide excess) whilst respiratory effort continues to counteract the airway narrowing. As this behaviour continues, an eventual critical point is reached whereby there is a maximum drop in oxygen level, also known as an **oxygen**

desaturation and/or an excess of carbon dioxide. The brain triggers an arousal, which essentially alarms the body and causes the individual to wake up. (Arousals are discussed in Section 3.5.) The 'critical' level varies for different subjects. Upon subsequent arousal, muscle tone returns, pressures become balanced, the airway becomes unblocked, the person resumes breathing, and the oxygen in the arterial blood increases, also known as **oxygen resaturation**. The oxygen is distributed throughout the body after several cardiac cycles, which may take up to several minutes. This vicious cycle may continue throughout the night, whether the patient is aware of the arousals or not. Further details of the signal patterns during OSAHS are provided in Chapter 4 (see Section 4.7).

3.2 Cause and effect

Obesity is the single most important predictor of apnoea [60]; however, several other features are also prevalent among those with OSAHS. The most common features include male gender, age greater than 40, habitual snoring, nocturnal gasping, choking or resuscitative snorting, witnessed apnoea, a body mass index (BMI) greater than 25 kg/m^2, or neck circumference greater than or equal to 17/16 inches in males/females, and systemic hypertension or high blood pressure. Additionally, there may be craniofacial and upper airway abnormalities to consider - relating to the general health status of the tongue, soft palate, uvula, nares and pharynx, for example. Finally, OSAHS has been shown to aggregate significantly with families, suggesting a possible inherited basis for the disorder [58].

Overall, these features lead to upper airway narrowing as illustrated in Figure 3.1 [52]. A variety of detrimental consequences are the result, with the most common effects including: sleep fragmentation, Excessive Daytime Sleepiness (EDS), personality changes, intellectual deterioration, visual-motor incoordination, impotence, insomnia, restlessness, choking, gagging, gasping, and resuscitative snorting. In addition, some long-term effects include pulmonary heart disease and hypertension, chronic hypercapnia, morning and nocturnal headache, left-sided congestive heart failure, cardiac dysrhythmias, nocturnal

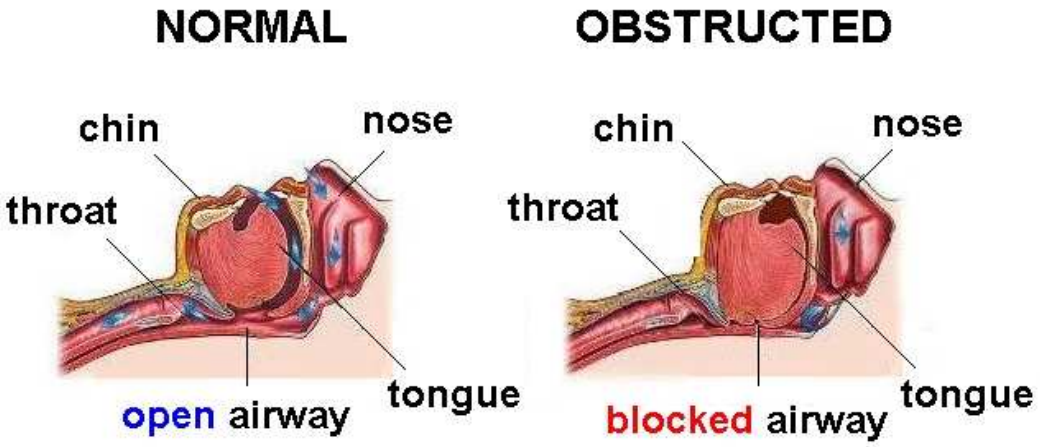


Figure 3.1: Normal and obstructed view of a patient lying in the supine position. The open and blocked airway have been labelled accordingly [67].

angina, and diurnal system hypertension. It should be noted that the mechanism whereby sleep-disordered breathing increases the risk of cardiovascular disease and other long-term consequences is unclear at this time [58]. Additionally, untreated severe OSAHS has been associated with excess mortality [58].

3.3 Treatment

The objectives of any disease management program are to contain costs without compromising quality of care [65]. The treatment options available for OSAHS may be divided into four options, including; mechanical, behavioural, surgical, and pharmacological.

Mechanical - Since its initial description in 1981, continuous positive airway pressure or CPAP (applied via a nasal mask) has been the primary therapy for patients with OSAHS [60]. CPAP or nCPAP (nasal-CPAP) effectively eliminates upper airway collapse in most individuals that use it. The major problem with nCPAP is user-compliance. Reportedly, two out of three patients who are offered nCPAP are willing to use it; however, they only do so for 4-5 hours of their full-night's sleep. Various strategies have been employed to improve patient adherence [60]. Further mechanical options include oral appliances (tongue retaining devices and mandible advancing oral appliances) [9, 81, 98].

Behavioural - As obesity is the single most important predictor of apnoea [60]; weight loss is an obvious solution. It has been shown that weight loss can lead to reduced blood pressure, improved sleep efficiency, decreased snoring and improved oxygenation (with a consequent reduction in number of apnoeas and hypopnoeas). In addition, other options include avoidance of alcohol and central nervous system depressants, and positional therapy including external nasal dilators [9, 104].

Surgical - A variety of surgical operations exist depending on the main source of obstruction. Examples include nasal surgery, combined nasal and pharyngeal surgery, hypopharyngeal airway surgery, somnoplasty with radiofrequency (RF) tissue ablation, laser-assisted uvulopalatoplasty (LAUP), uvulopalatopharynoplasty (UPPP) and skeletal facial corrections [48, 54, 55, 60, 64, 116].

Pharmacological -A variety of medications exist for a wide range of ailments; for example, nasal decongestants, oxygen, and drugs are available that increase ventilatory response and/or augment upper airway dilator muscle activity - each administered during sleep without actually disturbing sleep quality [10, 60].

3.4 Definitions: apnoea and hypopnoea

The American Academy of Sleep Medicine (AASM) Task Force defines an apnoea as a complete or near-complete cessation of airflow[1] that lasts for at least 10 seconds or longer [3], divided into one of three types [60]:

Obstructive Apnoea - ventilatory effort exists but there is no airflow, as there is an obstruction in the upper airway. This type of apnoea is usually associated with Excessive Daytime Sleepiness or EDS. It is the most prevalent sleep disorder seen in diagnostic sleep laboratories worldwide, accounting for 75 to 80% of diagnoses [60].

Central Apnoea - ventilatory effort is completely absent. This type of apnoea is usually associated with insomnia. It is considerably less prevalent with the exception of specific

[1]The standard measurement for airflow, or the lack thereof, is by pneumotachometer [3].

patient populations, such as patients with chronic heart failure or patients with neurological disorders.

Mixed Apnoea - this type of apnoea is initially due to absent ventilatory effort (a 'central' pattern) and subsequently persists despite resumption of ventilatory efforts (an 'obstructive' pattern).

A minimum of two signals, such as airflow and respiratory effort, are required to differentiate between apnoea types. A hypopnoea is less serious than an apnoea. Again, it can be obstructive or central and is defined by the AASM Task Force as an event that lasts for at least 10 seconds or longer and satisfies one of two criteria. First - a clear decrease from baseline (<50%) in the amplitude of a valid measure of breathing during sleep (respiratory effort or airflow measurement). Baseline is defined as the mean amplitude of stable breathing and oxygenation in the two minutes preceding onset of the event (in individuals who have a stable breathing pattern during sleep) or the mean amplitude of the three largest breaths in the two minutes preceding onset of the event (in individuals without a stable breathing pattern)[2] [3]. Second - a clear amplitude reduction of a valid measure of breathing during sleep (respiratory effort or airflow measurement) that does not reach the above criterion but is associated with either an oxygen desaturation of $> 3\%$ or an arousal [3]. In routine clinical practice, it is not always necessary to differentiate apnoeas from hypopnoeas because both type of events have similar pathophysiology which may be treated in the same manner [3]. Although not simultaneously, a patient can experience both types of events during sleep. Apnoeic and hypopnoeic events are more prominent when the patient is in the supine position. Other common causes of apnoeic/hypopopnoeic activity are provided in Section 3.2. Overall, little is known about the pathogenic mechanisms that differentiate between the two events, and the different short-term or long-term outcomes in patients with predominantly apnoeas as compared to hypopnoeas [3].

[2]Prior to the definition provided by the AASM, 'baseline' had been given several meanings in sleep literature. In agreement, however, is that the characteristics of the oxygen saturation signal change with virtually every apnoeic event over the course of the night [39].

3.4.1 Classification of sleep disorders

In this work, the term 'OSAHS' is used as a generic term to describe the syndrome whereby patients exhibit either apnoeic and/or hypopnoeic activity; i.e., a severe form of sleep disordered breathing (SDB). (The mild end of the SDB spectrum is snoring [58].) In the literature, other acronyms have been used interchangeably to mean the same thing. For example, the term 'OSAS' is typically used in population studies as a condition of recurrent episodes of obstructive apnoea and hypopnoea. The acronym 'OSAS' stands for Obstructive Sleep Apnoea Syndrome and indicates clinically diagnosed OSAHS (presumed symptomatic) [120]. Regardless of nomenclature, those patients who experience apnoeas and hypopnoeas throughout the night fall under the sleep disturbance category: *hypersomnia*, which describes a condition whereby patients remain sleepy despite adequate sleep time [60]. There are three other sleep disturbance categories, including: *insomnia*, *parasomnia*, and *sleep-wake schedule disorders*. Insomnia describes a condition whereby patients are tired, desire sleep, yet suffer from any combination of difficulties falling asleep, maintaining sleep, and waking up in the early morning [60]. Parasomnia describes any undesired event occurring during sleep. Sleep-wake schedule disorders, also known as circadian rhythm abnormalities, describe situations whereby patients sleep at undesired times [60].

3.5 Arousal

Apnoeas and hypopnoeas essentially cause an **arousal** in a sleeping individual; i.e., the sleeping person is woken up even though he/she might not be aware of it. These arousals can be recurrent throughout the sleep continuum and regardless of cause, may result in sleep fragmentation. An arousal typically occurs within three seconds of the termination of an event (apnoea/hypopnoea) [3]. Sleep monitoring often demonstrates this fragmentation of sleep as an increase in Stage 1 sleep and a reduction in Stage 3, Stage 4, and REM sleep. The Sleep Disorders Task Force of the American Sleep Disorders Association (ASDA)

defines an arousal in terms of its associated EEG signal where there is an abrupt shift in EEG frequency, which may include θ, α and/or frequencies greater than 16 Hz but not sleep spindles, subject to a set of 11 rules and conditions for arousal scoring purposes. The ASDA scoring was created to overcome the deficient scoring of arousals via the R & K standard, as the latter technique was not designed to score transient events shorter than 10 seconds. The ASDA rules stipulate that the minimum arousal duration is 3 seconds. In-depth and more specific scoring rules are found in [4].

There are several types of arousals. First, there are 'cortical' arousals that are due to those sleep disturbances that arise in the brain cortex. The above described ASDA definition of **arousal** is solely for cortical or EEG arousals, based on the EEG signal alone. The causes for cortical arousals are not all necessarily related to OSAHS and may be related to other factors such as external noise, changes in light, snoring, leg movement, bowel disturbance, bladder distension and gastroesophageal reflux [125]. The term 'micro-arousal' or 'μ-arousal' is reserved for cortical arousals only. Micro-arousals, recurrent throughout the sleep continuum, result in sleep fragmentation - without the person even being aware of their occurrence. It is believed that micro-arousals are the primary mechanism responsible for daytime hypersomnolence in patients with a variety of sleep disorders [52]. The change in level of arousal (i.e., when an arousal transpires) is probably the most important and difficult parameter to assess in EEG work, and it may cause confusion when the sleep technician is determining the amount of abnormality [99].

Second, there are 'sub-cortical' or 'autonomic' arousals that are due to those sleep disturbances that do not reach the brain cortex. These arousals may be detected by monitoring the heart rate and beat-to-beat blood pressure, looking for an increase in the pulse rate and blood pressure along with an increase in the respiratory effort following an apnoea/hypopnoea event [125].

3.6 Respiratory Disturbance Indices (RDIs)

Standardised indices of respiratory disturbance, based on events (apnoea/hypopnoea) described above, have been developed for patient comparison. The indices are dependent upon event definitions. The events may be defined according to respiratory effort or airflow measurements or according to oxygen desaturation measurements. Firstly, there is the apnoea index (AI), or average number of apnoeas per hour of sleep. AI is calculated by dividing the total number of apnoeas during a recording period by the Total Sleep Time or TST. TST is defined as time spent during REM and nREM, but also includes movement time which occurs during Sleep Period Time or SPT. SPT is defined as sleep onset to the cessation of sleep according to the R & K definition (see Section 2.2). Arousals are included in SPT; whereas, such periods of wakefulness are **not** included in TST. Similarly defined are the HI (hypopnoea index) and AHI (apnoea and hypopnoea index). The AHI, or number of apnoeas plus hypopnoeas per hour of sleep, is one of the most common RDIs in literature, even though the definition of what constitutes an apnoea and hypopnoea differs amongst sleep-clinics. Note that the AHI will **only** be used to refer to the index created by **experts in the clinical setting**. (In subsequent chapters, the term RDI will be used to refer to the index created by the automated analysis of the oxygen saturation signal - see Chapter 9.)

The AHI is employed to quantify OSAHS severity [94]. A patient with the most mild case of OSAHS would have an AHI value of five; i.e., five apnoeas/hypopnoeas for every hour of sleep [58]. This minimal requirement for the clinical assessment of OSAHS is based on epidemiological data that suggests that health effects (such as hypertension, sleepiness, and motor vehicle accidents) may be observed at an AHI threshold of five [3]. In addition, an increased risk of hypertension becomes substantial at an AHI of approximately 30, according to the Wisconsin Sleep Cohort [3]. Note that no adequate prospective studies have validated the severity criteria for sleepiness and thus the proposed classifications of OSAHS severity based on the AHI are 'operational definitions' [3]. Table 3.1 describes

the AHI corresponding to various degrees of OSAHS intensity from normal to severe. Clinically, patients with severe OSAHS may experience *both* a higher incidence of events and events of longer duration over the course of the night. The AHI has also been shown to increase in normal subjects with age, and poorly correlates with symptoms of excessive daytime somnolence [6].

For diagnosing OSAHS, however, the AHI is insufficient as stand-alone criterion; i.e., transient hypoxia is just one of the physiological consequences of OSAHS and the absence of significant hypoxia does not rule out OSAHS [26, 33, 42, 113]. Therefore, the AHI must be accompanied by a second severity criteria: daytime somnolence or sleepiness [3, 60]. Daytime somnolence may be associated with two or more of the following symptoms: loud snoring, choking or gasping during sleep, recurrent nocturnal awakening, unrefreshing sleep, daytime fatigue, and impaired concentration [6]. The three basic levels of sleepiness correspond to unwanted sleepiness or involuntary sleep episodes that occur during activities that require little attention (mild sleepiness), some attention (moderate sleepiness), and a great deal of attention (severe sleepiness). Examples of activities are, respectively: reading, watching television, travelling as a passenger (mild); attending meetings, presentations, concerts (moderate); driving, eating, walking, having a conversation (severe) [3].

Degree of OSAHS (categorisation)	Apnoea Hypopnoea Index, AHI (events per hour)
normal	0 - <5
mild	5 - 14
moderate	>14 - 30
severe	>30

Table 3.1: Classification of the degrees of OSAHS severity according to AHI values [3].

Defining the prevalence of the OSAHS using the AHI is problemmatic due to the disparate definitions of apnoeas/hypopnoeas used amongst sleep clinics. For example, for the definition of a hypopnoea, some sources of variability include: differences in sensors for detecting breathing amplitude change, the absolute amplitude criteria considered sufficient

to identify a breathing pattern as 'reduced,' and the use of corroborative data [3]. Further methodological research recommendations, from the AASM Task Force, include the need to establish standard criteria for quantifying and interpreting data collected from sensors used to detect hypopnoeas, with the adequacy of these criteria judged by the accuracy of any measurement or set of measurements compared with a gold-standard (or by the ability of these criteria to predict clinically relevant short and long term outcomes) [3]. Prior to any such investigation, it is essential to first accurately identify suitable patient studies through a retrospective analysis of clinical records.

3.7 Polysomnography

The current gold-standard method for the accurate diagnosis of OSAHS and other sleep-related disorders is the 'sleep test' or polysomnography (PSG) - a multi-channel recording of sleep and breathing variables that is traditionally performed overnight, with a trained technician in attendance. The standard full PSG test records the electrical brain activity or electroencephalogram 'EEG,' eye activity or electrooculogram 'EOG,' muscle tone or muscle activity or electromyogram 'EMG,' oro-nasal airflow and respiratory effort. Typical additional recorded measures include body position and limb movement, oxygen saturation, and electrocardiogram (ECG).

Figure 3.2 illustrates a typical and general polysomnogram (no units of measurement provided) with 10 signals recorded. (Signals recorded for standard PSG vary according to sleep clinic.) The top four traces in Figure 3.2 show plots of the essential signals for scoring sleep stages via Rechtschaffen and Kales (R & K) rules. For accurate and consistent division of sleep into its various stages, the 'R & K Scoring' standardised technique is used (as published by Rechtschaffen and Kales in their 1968 atlas of tracings, with detailed descriptions on how to score sleep for the sleep research community [93] - see Section 2.2 in Chapter 2). Leg movement can also be measured with electrodes placed on the anterior tibialis muscle. The lower five traces show airflow, respiratory effort and

arterial oxygen saturation measurements. Airflow is measured by both 'thermistor' and 'nasal cannulae.' The former senses alteration in heat exchange during inspiration and expiration indirectly reflecting changes in airflow [52]. The latter uses a sensitive pressure transducer that records pressure fluctuation at the nose during inspiration and expiration, indirectly reflecting inspiratory and expiratory airflow [52]. Respiratory effort through 'rib' and/or 'abdomen' movement is measured by respiratory inductance plethysmography (RIP) or piezoelectric strain gauges [60]. RIP detects changes in the circumference of the rib cage and abdomen during breathing. When properly calibrated, the sum of these two signals provides a measure of tidal volume or airflow [52]. Lastly, arterial oxygen saturation is measured non-invasively and continuously by pulse oximetry. The recorded signal is commonly known as SpO_2 (reference name for the remainder of this work), SaO_2 (Figure 3.2), O2Sat or PulseOx. A description of pulse oximetry, and the literature review of previous approaches to detecting OSAHS using the oxygen saturation signal (as measured by pulse oximetry) follows in Chapter 4.

Overall, the analysis of signals acquired during PSG is time and labour-intensive, leading to a consequent cost and resource burden in many sleep centres. Some believe the entire PSG method to be flawed, as the sleep test only became standard practice (during classical sleep studies) when the associated pathophysiology and range of obstructive sleep apnoea syndromes were not adequately understood [87]. In addition, when the R & K manual [93] was originally produced, it was designed to be a pro-forma for a standard rather than a gold-standard [74].

3.8 Summary

This Chapter has provided an overview of OSAHS, including its physiology, cause and effect, and applicable treatment. Definitions have been given for apnoea, hypopnoea, arousal, AHI, and RDI. As our primary signal of interest, the principles underlying the oxygen saturation, or SpO_2, signal are described in the following Chapter. Also, a review

of previous work is provided detailing the automated diagnosis of OSAHS using the SpO_2 signal - both on its own and in conjunction with other biomedical signals.

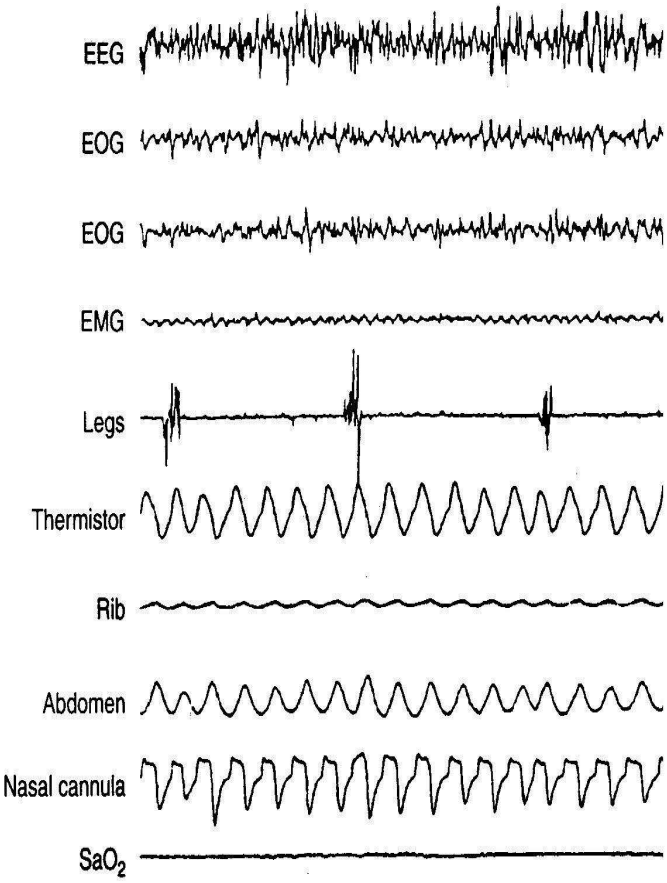


Figure 3.2: A 10-channel example of a polysomnogram for one patient [31].

Chapter 4

Pulse Oximetry

4.1 The need for oxygen monitoring

Oxygen is the most acutely necessary substrate for aerobic life [28], and hence adequate oxygen delivery to tissues is of vital importance to patients who suffer from sleep apnoea and other sleep-related disorders that trigger recurrent fluctuations in oxygen level overnight. Hypoxic events (lack of oxygen in the blood) can occur repeatedly during sleep and may lead to detrimental pulmonary and heart complications and, in the most severe case, death. For patients at risk of respiratory failure, it is important to monitor the efficiency of gas exchange in the lungs - with such information preferably available to clinicians on a continuous basis (i.e., rather than every few hours) [110]. As introduced in Section 3.7, the current gold-standard method for accurate diagnosis of the OSAHS is polysomnography, or PSG. PSG is a multi-channel recording of sleep and breathing variables, and includes the non-invasive and continuous monitoring of blood oxygen levels overnight. This data is an invaluable aid to the physician for the management of the patients' sleep therapy.

The non-invasive monitoring of blood oxygen content using pulse oximetry has been developed over the last 20 years. Severinghaus and Astrup describe the pulse oximeter as 'arguably the most significant technological advance ever made in monitoring the

well-being and safety of patients during anaesthesia, recovery and critical care' [5, 28]. A brief history of oximetry is provided in Appendix A. In this Chapter, the basic principles of measuring oxygen saturation within the human body are provided followed by a description of signal patterns during OSAHS. A literature review of previous approaches to detecting OSAHS using the oxygen saturation signal is detailed thereafter.

4.2 Oxygen and the human body

Oxygen delivery to tissues depends on the oxygen entering the lungs, the adequacy of the pulmonary gas exchange, blood flow to the tissues, and the oxygen-carrying capability of the blood [20, 28]. Blood flow depends on cardiac output and the degree of constriction of the vascular tissue bed. The amount of O_2 (oxygen) in the blood is determined by the amount of haemoglobin in the blood. (A more complete discussion on haemoglobin as an oxygen carrier is found in Appendix A.) A haemoglobin molecule rich in oxygen is referred to as oxygenated haemoglobin, or **oxyhaemoglobin, HbO_2**. Conversely, a haemoglobin molecule completely devoid of oxygen is referred to as deoxygenated haemoglobin, or **reduced haemoglobin, Hb**.

4.2.1 Definition of oxygen saturation

In general, the degree of blood oxygenation is known as *oxygen saturation*, which is defined as the ratio of HbO_2 concentration to the sum of the HbO_2 and Hb concentrations in a blood sample, expressed as a percentage:

$$oxygen\ saturation = \frac{HbO_2}{HbO_2 + Hb} \times 100\% \tag{4.1}$$

The degree of oxygen saturation in arterial (arteries, arterioles) blood is called arterial oxygen saturation or SaO_2. In a normal healthy adult, the normal operating range for SaO_2 is greater than 90%, usually between 96 to 100%. Arterial saturations between 70

and 80% result in drowsiness, dizziness and visual impairment, and any further decrease in saturation results in deterioration of mental function and coordination [28].

The degree of oxygen saturation in venous blood is called venous oxygen saturation or SvO_2. Under normal physiological conditions, SvO_2 has a much lower oxygen saturation than SaO_2. Specifically, it is usually between 70 to 75%, and much lower in value during vigorous exercise [28, 110]. Overall, oxygen stores exist in the following order: lungs, arteries, tissue, and veins; however, oxygen depletion occurs first in tissue, then the veins, lungs, and arteries. Thus, desaturation of arterial blood occurs only after desaturation in other areas [100].

4.2.2 Oxygen content in plasma

The majority of oxygen transported by the body is bound to haemoglobin; however, a very small amount of oxygen is also dissolved in plasma. (For the same amount of plasma, 65 times more oxygen can be transported with haemoglobin than would be possible without haemoglobin [96]). The volume of oxygen carried by haemoglobin per 100mL of blood, or C_{HbO_2}, is defined as follows:

$$C_{HbO_2} = 1.37 \times Hb \times SaO_2 \tag{4.2}$$

where 1.37 is the number of mL of oxygen bound to one gram of fully saturated haemoglobin. Typically, C_{HbO_2} is around 19 mL O_2 /100 mL blood, and the weight of Hb is 14 g Hb /100 mL blood [20], [84]. The volume of oxygen dissolved in the plasma per 100 mL of blood, or C_DO_2, is defined as follows:

$$C_DO_2 = 0.003 \times PaO_2 \tag{4.3}$$

where 0.003 is the solubility of oxygen in blood as the percent by volume per mmHg, and PaO_2 is the partial pressure of oxygen in the arteries measured in mmHg. Typically,

C_DO_2 is around 0.3 mL O_2/100 mL blood [20]. The overall oxygen content of the blood, or C_AO_2, is therefore a combination of oxygen bound to haemoglobin and oxygen dissolved in plasma, or:

$$C_AO_2 = C_{HbO_2} + C_DO_2 \tag{4.4}$$

The amount of oxygen available to tissue is defined as the oxygen delivery index, and is typically 550 to 650 mL/(min m^2). However, normal oxygen consumption, a measure of the oxygen actually diffused into the tissue, is typically 115 to 165 mL/(min m^2); i.e., not all of the available oxygen diffuses into the tissue [84], [20]. The residual oxygen is distributed to the remaining oxygen stores in the body.

4.2.3 Oxygen-haemoglobin dissociation curve

The relationship between arterial oxygen saturation (SaO_2) as defined above, and the partial pressure of oxygen (PO_2) in the blood is described by the oxygen-haemoglobin dissociation curve - see Figure 4.1. The curve describes the percentage of oxygen bound to haemoglobin (normalised by the maximum percentage) and its relation to PO_2. The curve's sigmoid shape is crucial for oxygen transport. A normal healthy adult should have an SaO_2 above 0.9 on the curve (90%), corresponding to a PO_2 of approximately 60-100 mmHg [2, 20, 100]. The flat top section of the curve describes oxygen uptake, where the haemoglobin is almost completely saturated over a wide range of PO_2 [28]. The steep section of the curve (less than 80% SaO_2) describes the oxygen release to tissues. For a small drop in PO_2, a large amount of oxygen is released to the blood flowing through the capillaries; thus, the driving force for diffusion is maintained [28].

The oxygen-haemoglobin dissociation curve varies (left-shift, right-shift) according to such factors as carbon dioxide concentration, pH and temperature (as illustrated in Figure 4.1) [20]. For example, for any specific oxygen partial pressure, a decrease in temperature from 40 to 28°C yields a higher amount of oxygen saturation; i.e., the dissociation curve

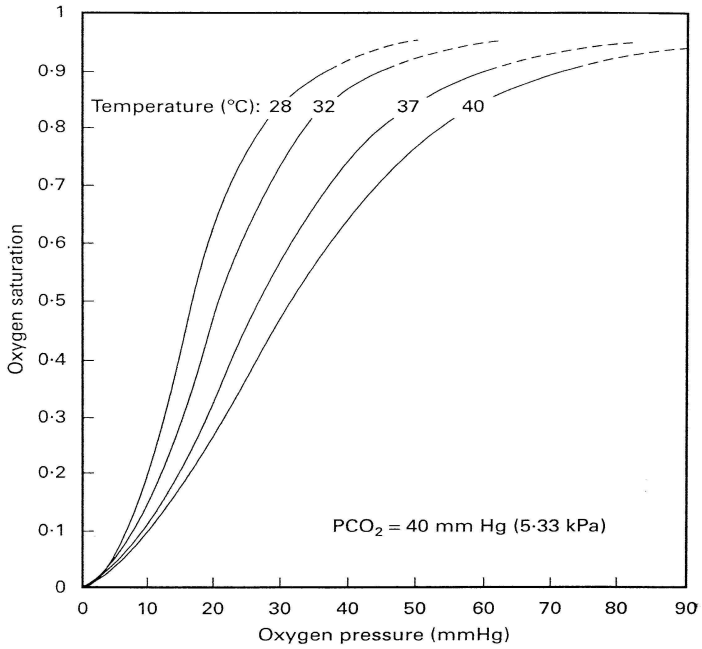


Figure 4.1: Shift in oxyhaemoglobin dissociation curve with change in temperature levels [20].

of Figure 4.1 shifts to the left. Left displacements to the curve are also caused by decreases in ATP[1] in red blood cells, and increases in pH, carboxyhaemoglobin (HbCO), and methaemoglobin (metHb) (see Appendix A for definitions). The converse (in all cases) results in right displacements to the curve.

4.3 Absorbance at two wavelengths

The absorption of visible light by a haemoglobin solution varies with oxygenation, such that an HbO_2 molecule is bright red and a Hb molecule is dark red [110]. These two states have significantly different optical spectra in the wavelengths: 660nm (red) and 940 nm (near infrared, or NIR) [20, 110]. Figure 4.2 shows the absorption spectra for the two forms of the molecule. For haemoglobin, the isobestic point (wavelength at which the

[1]ATP is adenosine triphosphate, or the nucleotide that stores and transports chemical energy.

absorption by the two forms of the molecule is the same) is at 805 nm [77,110]. The NIR wavelength is used because the absorption spectra is reasonably flat; i.e., slight variations in the peak wavelength makes very little difference to calibration. The red wavelength is used because there is a comparatively large difference in the absorption between HbO_2 and Hb at this point on the spectrum; i.e., detectable absorption changes may be achieved with small changes in oxygen saturation. At wavelengths shorter than 600 nm, tissue and pigmentation (e.g., red skin pigment or melanin) absorb blue, green and yellow light; whereas, at wavelengths longer than 1300 nm, there is strong absorption due to water in the tissues [77,110]. Overall, it is possible to use the differences in absorption at different wavelengths *in vivo* because - for the desired range of 660 to 940 nm - there is the least attenuation of light by body tissue.

4.4 Beer-Lambert law for oximetry

The Beer-Lambert law describes the relationship between the absorption of light and the amount of absorbent [77]. The law relates the fraction of radiant energy absorbed by the substance to the concentration and amount of the substance:

$$A = log\left(\frac{I_o}{I}\right) = \varepsilon lc \tag{4.5}$$

where A is the absorbance of the sample under investigation, I_o is incident light (or the intensity of the energy without the sample), I is transmitted light (or the intensity with the sample), l is the optical path length (or depth of the sample), c is the concentration of the sample, and ε is the extinction coefficient (wavelength-dependent constant characterising the sample) [77].

If it is assumed that the transmission of light through the arterial bed is influenced only by the relative concentrations of HbO_2, Hb, and their absorption coefficients at the two measurement wavelengths, the light intensity will decrease logarithmically with path length according to the Beer-Lambert law. Additionally, the absorption coefficients must

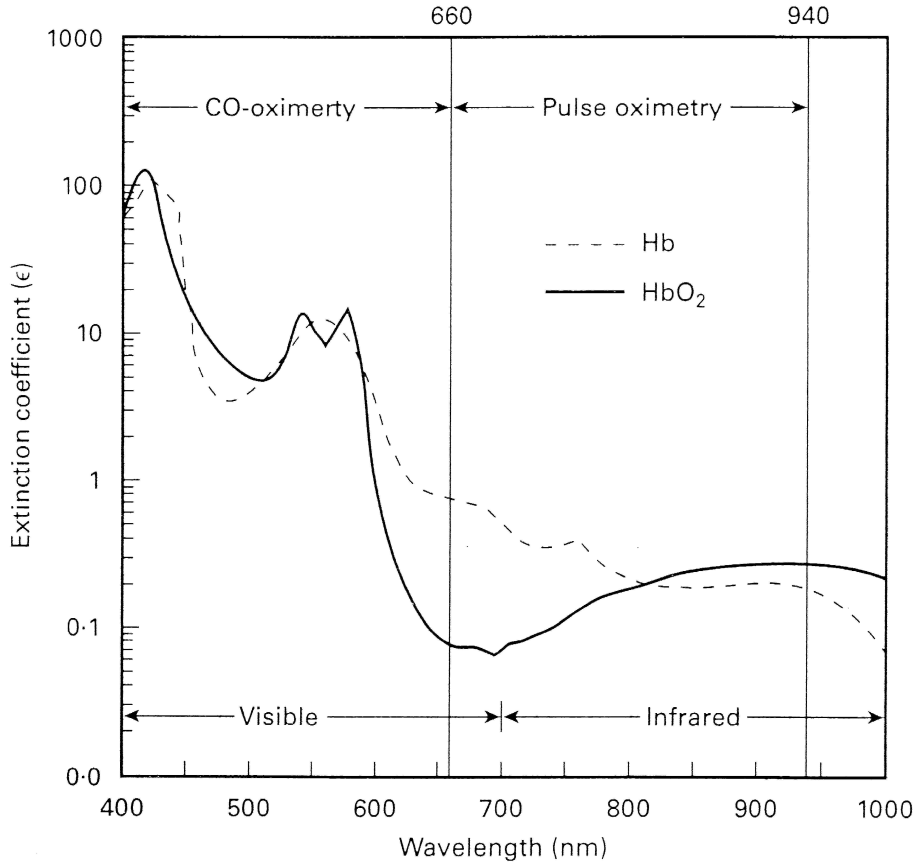


Figure 4.2: Absorption spectra of HbO_2 and Hb, with the isobestic wavelength (equal absorption by both forms of the molecule) to be 805 nm. The ranges of wavelength for both CO-oximetry and pulse oximetry are provided [77].

be sensitive enough to provide adequate variation (for comparison) whilst robust enough so that the blood does not appear completely opaque [100].

If we consider an artery, for example, of length l through which light (initial intensity I_{in}) is shone, the Beer-Lambert law states that the intensity I is given by (at wavelengths λ_1 and λ_2):

$$I_{\lambda_1} = I_1 = I_{in1} 10^{-(\alpha_{o1} C_o + \alpha_{r1} C_r) l} \tag{4.6}$$

$$I_{\lambda_2} = I_2 = I_{in2} 10^{-(\alpha_{o2} C_o + \alpha_{r2} C_r) l} \tag{4.7}$$

where Co and Cr are the concentrations of HbO_2 and Hb, and α_{on} and α_{rn} are the absorption coefficients of HbO_2 and Hb at wavelength λ_n (with $n = 1, 2$). The ratio R may be written as:

$$R = \frac{log_{10}(I_1/I_{in1})}{log_{10}(I_2/I_{in2})} \tag{4.8}$$

Oxygen saturation may then be simplified as:

$$SaO_2 = \frac{C_o}{C_o + C_r} = \frac{\alpha_{r2}R - \alpha_{r1}}{(\alpha_{r2} - \alpha_{o2})R - (\alpha_{r1} - \alpha_{o1})} \tag{4.9}$$

However, the Beer-Lambert law does not take into account the multiple scattering of light by the red blood cells. More specifically, the Beer-Lambert law holds only for monochromatic radiation through a homogeneous (blood is a non-homogeneous liquid) and isotropic medium (refractive index is the same in all directions) with negligible scattering, and no reaction between absorbent and solvent [77]. Additionally, scattering is a function of both wavelength and particle size. Although oximetry is a differential technique, the effect of scattering is only partially compensated for [110]. In 1951, Kramer *et al.* found considerable deviations (i.e., conflict between theory and practice) of the Beer-Lambert law for high concentrations of red blood cells [59].

4.5 Pulsatile nature of arterial blood

Pulse oximetry determines the oxygen saturation of the arterial blood by measuring the light transmitted through an extremity (typically a finger, or toe, earlobe, bridge of the nose) at two different wavelengths [77,110]. The chosen area must have a reasonably short optical path length for sufficient translucence at the chosen wavelengths. Pulse oximetry assumes that the attenuation of light can be split into three independent components: arterial blood, venous blood and tissues - see Figure 4.3 [110]. The incident light that is not absorbed is transmitted.

Arterial pulsation is used to differentiate between absorbance of arterial blood and other absorbers - such as venous blood and other tissue - that are more constant or slow varying. Specifically, the transmitted light intensity is a pulsatile signal superimposed on a direct current (d.c.) level, and time-varying with the heart beat. (The amplitude of this cardiac-synchronous pulsatile signal is approximately 1% of the direct current level [110].)

If it is assumed that the increase in light attenuation is caused only by the inflow of arterial blood into the (say) fingertip, oxygen saturation of the arterial blood may be calculated by subtracting the direct current component (of the attenuation) from the total attenuation [110]. What is left is the cardiac-synchronous pulsatile component for the dual-wavelength determination of oxygen saturation [110]. Figure 4.4 shows a typical cardiac-synchronous pulsatile waveform recorded when red or near infra-red light is shone through the finger. The secondary peak shown is called the dicrotic notch, and is relatively common [110]. (Dicrotic denotes a pulse in which a double beat is detectable for each beat of the heart [1].) Owing to the cardiac rhythm, the absorption of light due to the haemoglobin in the arteries and arterioles increases during systole and decreases during diastole, as shown in Figure 4.4 [77].

The general oximetry ratio R (see Equation 4.8) is equally valid for pulse oximetry, if R is given by:

$$R = \frac{log_{10}((I_{dc+ac})/(I_{dc}))_{\lambda_1}}{log_{10}((I_{dc+ac})/(I_{dc}))_{\lambda_2}} \tag{4.10}$$

Furthermore, Equation 4.10 may be rewritten with the d.c component eliminated:

$$R \approx R' = \frac{log_{10}(I_{ac})_{\lambda_1}}{log_{10}(I_{ac})_{\lambda_2}} \tag{4.11}$$

Limitations in pulse oximetry have more to do with the actual limitation in monitoring arterial oxygen saturation as opposed to the technique of oximetry itself [100]. For example, pulse oximetry is a poor measure of hypoventilation when inspired oxygen concentration is high [50, 69, 100].

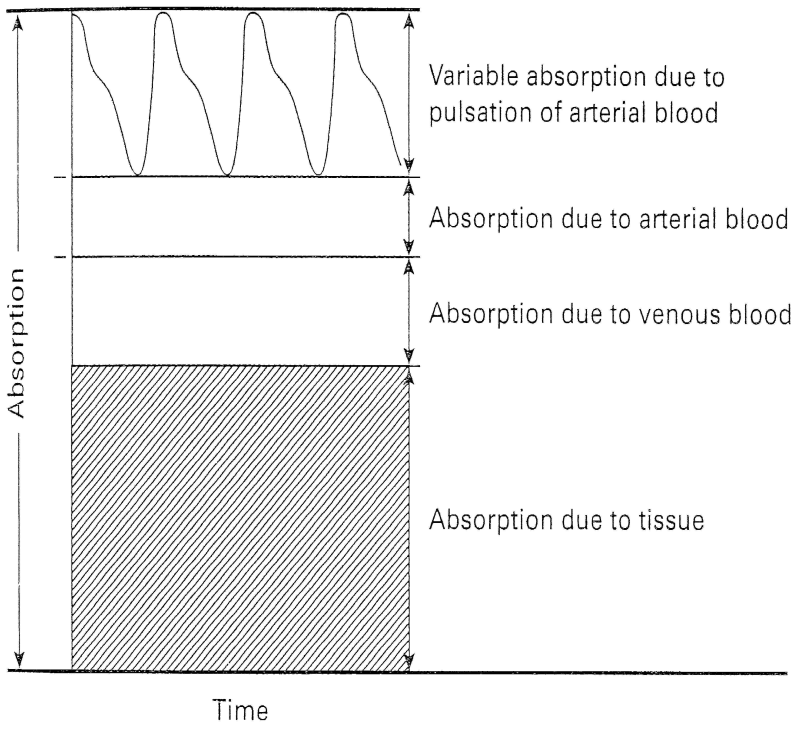


Figure 4.3: Absorption of light due to the pulsation of arterial blood, venous blood, and tissue [77].

4.6 Pulse oximeter instrumentation

Pulse oximeters use two light emitting diodes (LEDs) that work in the red and NIR part of the spectrum (a separate energy source is needed for each wavelength), and a photodiode placed on opposite sides of a vascular bed [28, 77]. The LEDs are built with an internal lensing system that produces high intensity output detectable by a simple, compact, solid-state photodetector, such as the silicon photodiode. LEDs can be switched on and off very rapidly, and therefore only a single photodetector is required. In addition, the special-purpose LEDs utilised do not overheat tissue although they are placed in close contact with the skin surface; i.e., although the peak output power is very high, the average output power (which produces the heating effect) is very low.

The components of a typical transmission pulse oximeter probe are shown in Figure 4.5. (*Transmission pulse oximetry* is based on light diffraction [77].) The extremity (finger) is held snugly in place between the LEDs and photodetector, so that energy cannot bypass

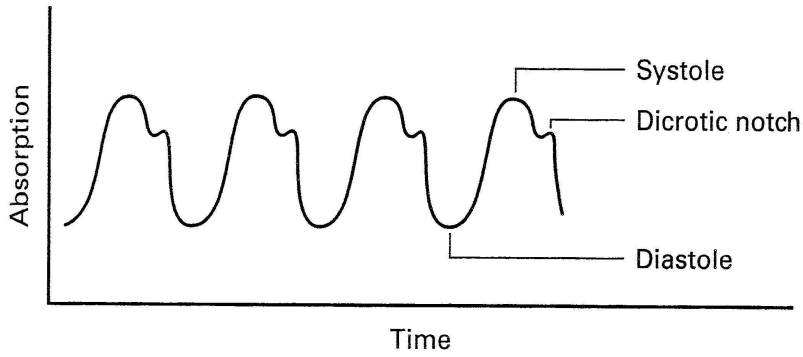


Figure 4.4: A typical pulsatile signal recorded with a probe when light is shone through the finger [77].

the tissue. The entire probe assembly is protected from extraneous light, over the range of wavelengths to which the detector is sensitive, by a light-excluding enclosure [77].

4.6.1 Calibration of pulse oximeters

Instruments based on the Beer-Lambert law, which ignores scattering, tend to give erroneous estimates of the true value of oxygen saturation. Some attempts have been made to develop theoretical models which take scattering into account [28]; however, most current pulse oximeters require the use of empirical look-up tables derived from calibration studies on large numbers of healthy volunteers whose oxygen saturation is measured invasively using the CO-oximeter. Figure 4.6 shows the relationship between oxygen saturation and the ratio R (Equation 4.8) for the two cases: Beer-Lambert law, empirical data [110].

The gold-standard for the calibration of pulse oximetry is the CO-oximeter. The CO-oximeter requires small samples of heparinised[2] arterial blood to be taken and provides oxygen saturation readings only for the instant at which blood is withdrawn. Most CO-oximeters operate in the visible range (see Figure 4.2) as they work only with haemoglobin solution in plasma, and not with skin, muscle, bone, etc. [77].

[2]The arterial blood is treated with heparin, an anti-coagulant, to reduce blood clotting.

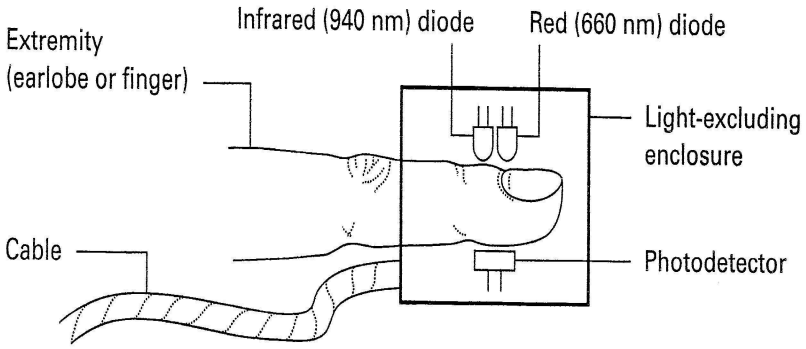


Figure 4.5: Components of transmission pulse oximeter probe [77].

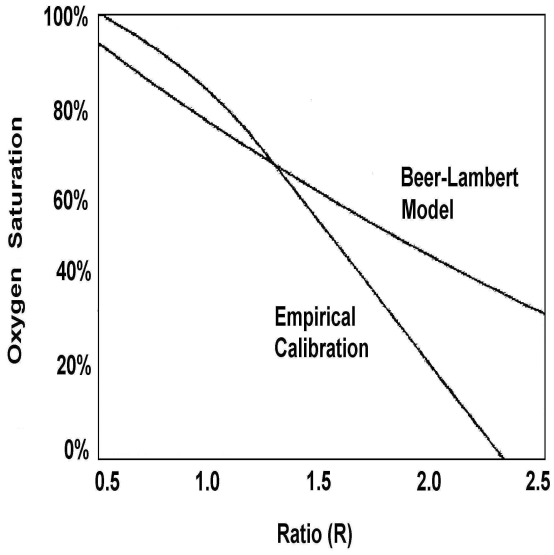


Figure 4.6: Two relationships between the ratio R and the oxygen saturation of the patient [110].

4.7 Signal patterns during OSAHS

Apnoeas (and hypopnoeas) produce various nocturnal physiological markers - including SpO_2 desaturation. There are only two possible types of signals found within the SpO_2 waveform with different characteristics in the time domain: normal breathing, or low amplitude data and apnoea, or high amplitude data. The advantage of using the SpO_2 signal in the time domain is the large variations in signal amplitude typically observed during the sleep of patients with OSAHS. Specifically, the SpO_2 signal exhibits a distinct cyclical pattern, also referred to as a saw-tooth waveform. (Single central apnoeas, not in conjunction with Cheyne-Stokes respiration,[3] can also exhibit a saw-tooth waveform in the SpO_2 signal [78].) During OSAHS, airflow ceases because of occlusion of the upper airway (nasopharynx and oropharynx), by the movement of the tongue and palate into apposition with the posterior pharyngeal wall. The occlusion is complete during apnoeas and partial during hypopnoeas. A cessation (apnoea) or reduction (hypopnoea) of airflow results, despite continued respiratory efforts and activation of inspiratory muscles [12]. Subsequently, asphyxia develops, but only until there is a brief arousal from sleep, restoration of upper airway patency, and resumption of airflow. During asphyxia,[4] the oxygen level drops and continues to decrease even past the resumption of breathing, or oxygen desaturation. The oxygen saturation nadir level is reached within (typically) 30 seconds[5] of the resumption of breathing (or equivalently, the termination of the obstructed breath) [3]. The subsequent delay is due to the time needed for oxygen to resaturate throughout the body, depending on heart rate [114], plus the time needed for oxygen to flow from the lungs to the periphery; i.e., ear/finger where the pulse oximeter is located. With ear probes, the average delay is 9 to 10 seconds; whereas a 23.6 second delay has been shown with a finger probe (with somewhat longer delays reported for toe probes) [34,78,113,119]. The depth of each oxygen desaturation is dependent upon the duration of the apnoea,

[3] Cheyne-Stokes respiration is characterised by a cyclical fluctuation in breathing with periods of central apnoeas or hypopnoeas alternating with periods hyperpnoea in a gradual waxing and waning fashion [3].

[4] An asphyxia is defined as a lack of oxygen or excess of carbon dioxide in the body.

[5] Netzer *et al.* note that the typical cyclical drop in SpO_2 in patients with OSAHS may lag between 45 and 60 seconds behind a respiratory event.

the baseline level of the SpO_2 signal, the expiratory reserve volume, and the apnoea type (see Section 3.4) [37, 101]. Upon resumption of airflow, the oxygen level increases to an SpO_2 percentage slightly above the pre-apnoeic level for a short time; i.e., the majority of patients compensate for the cessation of breathing with hyperventilation [90]. Reoxygenation is, however, somewhat impaired during the cyclic recurrence of apnoeas and it is more typical that the next apnoea would start from a lower baseline than the previous apnoea [90]. For recurrent apnoeic events, the SpO_2 signal will typically exhibit oxygen desaturation-resaturation cycles, hence the saw-tooth pattern observed in the waveform, as previously mentioned. Note that it is possible for an event (apnoea/hypopnoea) to occur without a visible dip in the SpO_2 signal, such as for patients with high levels of stored oxygen in their blood [101].

Figure 4.7 shows three oxygen desaturation-resaturation cycles for a patient with OSAHS. (Further examples are provided in Chapters 6 through 9 of this work.) The SpO_2 signal falls in the range of 84 to 96%. In a normal healthy adult, the normal operating range for SpO_2 is greater than 90%, usually between 96 and 100%. Arterial saturations between 70 and 80% result in drowsiness, dizziness and visual impairment, as mentioned in Section 4.2.1 [28]. Furthermore, saturation below 50% leads to depressed respiration and, in a matter of minutes, unconsciousness, permanent brain damage and ultimately death [28]. Beneath the SpO_2 signal is the corresponding airflow with the envelope (FIR-filtered with a 10-point moving average window) superimposed in black. For one event, 'Breath Stop,' 'Breath Start,' and 'Nadir' have been highlighted - respectively corresponding to the cessation of airflow, resumption of airflow and the lowest point of SpO_2 achieved. The highlighted event lasts for 23.9 seconds (or the time between 'Breath Stop' and 'Breath Start.'). The 'Nadir' is reached 32.6 seconds after breathing has stopped (or 8.7 seconds after breathing has resumed).

The AASM Task Force stipulates that an apnoea must be at least 10 seconds in duration; with $T/2 = 10$ sec, $T = 20$ sec. (The apnoea corresponds to the desaturation portion of the SpO_2 signal; i.e., $T/2$.) Conversely, the maximum cycle length of apnoea

observed in OSAHS is 120 seconds during REM sleep [102]; with $T/2 = 60$ sec, $T = 120$ sec. Lastly, the AASM Task Force stipulates that a hypopnoea also must be at least 10 seconds in duration. The maximum length of a hypopnoea has been noted as several minutes in duration during REM sleep [3]. According to Findley *et al.*, the factors which determine the duration of an apnoea probably relate to neural reflexes, central chemosensitivity, and sleep stage [35]. For example, in 1999, Hilton *et al.* used the first continuous (no-wake epochs) 15 minutes sections of non-REM (stage 1 or 2 sleep) and REM sleep for their study of the Heart Rate Variability (HRV) and SpO_2 signals from 40 patients with suspected OSAHS. They discuss the fact that due to the overall reduction in muscular tone during REM sleep, there are associated increases in HRV and more pronounced oxygen desaturations [47,85] - see Section 4.8 for further details. (Also, there are very few studies investigating the effect of OSAHS on sleep stages - see Chapter 10.) In an animal model of obstructive apnoea, Fletcher *et al.* showed that the rate of fall of SpO_2 was independent of apnoeic duration but had a consistent negative correlation with pre-apnoeic mixed venous oxyhaemoglobin saturation (see Section 4.2.1). They also noted that mixed venous oxyhaemoglobin saturation may be affected by pre-existing lung disease and by the close repetitiveness of apnoeas (not allowing sufficient time for oxyhaemoglobin saturation recovery in the veins) [37].

4.8 Review of previous work

Periodicity of ventilation produces periodicity in the oxygen saturation signal, SpO_2. The non-invasive monitoring of blood oxygen saturation using pulse oximetry (ear or finger pulse oximeter) has been developed over the last 20 years and is the simplest PSG signal to acquire [3]. Subsequently, the analysis of SpO_2 changes has been thoroughly investigated as a potential replacement for PSG analysis for apnoeic/hypopnoeic event detection.

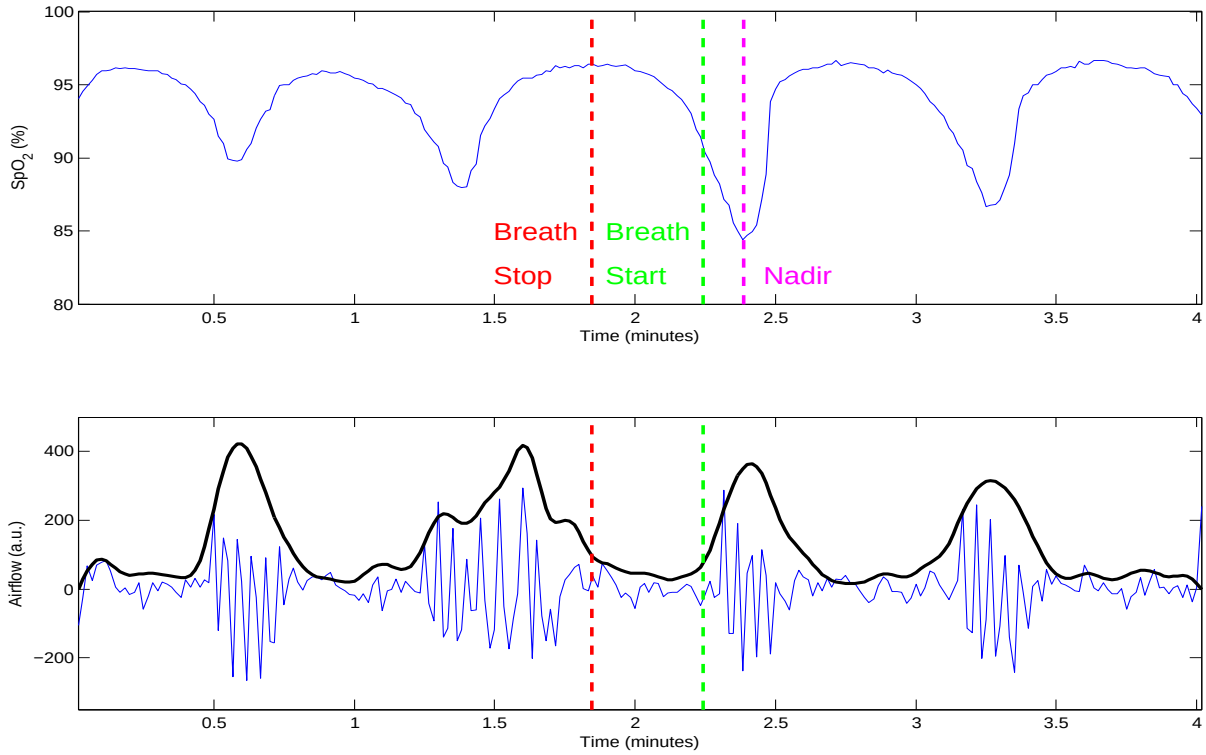


Figure 4.7: A four-minute example of arterial oxygen saturation measured by finger pulse oximetry in a patient with OSAHS, showing recurrent episodes of desaturation, with its corresponding airflow plotted in arbitrary units (a.u.) whereby the signal's envelope is superimposed in black (filtered and smoothed with a 10-point moving average window). 'Breath Stop' = cessation of airflow, 'Breath Start' = resumption of airflow; 'Nadir' = lowest level of the SpO_2 signal.

4.8.1 SpO_2 Signal: Time Domain

In 1988, George *et al.* produced a brief report on the identification and quantification of apnoeas by computer-based analysis of oxygen saturation [39]. Using the cyclical nature of the SpO_2 signal in the time-domain, for 9 patients with severe OSAHS (the average $\pm$ standard deviation of the patient AHI indices was 78 $\pm$ 20 events per hour). Apnoeic episodes were identified using peak/trough detection and a minimum episode period of 10 seconds. A window was moved along the SpO_2 signal in the time domain, which could detect a drop in saturation from baseline. Their definition of 'baseline' was not provided (and the study precedes the 1999 definition of 'baseline' as provided by the AASM [3]). Both the window size and drop in saturation were user definable; however,

the minimum drop in SpO_2 signal had to have been greater than 3%. A high sensitivity was reported when automated results were compared with manual scoring. The method was limited to oxygen desaturation events only; i.e., apnoeas could not be distinguished from hypopnoeas, nor could central/mixed/obstructive apnoeas be differentiated.

In a 1995 study, Yamashiro and Kryger calculated the number of desaturations >3% from baseline on a per hour basis for 269 patients with Excessive Daytime Sleepiness (EDS) and a suspicion of Sleep Disordered Breathing (SDB) [119]. Six scorers each computed a desaturation index (DI) (number of desaturations per hour) and an averaged diagnosis was made for each patient. The definition of 'baseline' was not provided. Using a DI-classification scheme (normal: DI<5, mild OSAHS: 5<DI<20, moderate OSAHS: 20<DI<40, severe OSAHS: DI>40) all moderate and severe OSAHS patients were correctly detected; however, among those 66 patients classified as normal, 1 had mild OSAHS, 20 had UARS (Upper Airway Resistance Syndrome), 9 had PLMs (periodic limb movements), 4 had narcolepsy and 2 had parasomnia.[6]

In 1996, Lévy *et al.* used an oxygen saturation variability based index to detect apnoeic events in subjects with sleep apnoea syndrome [63]. The 'Delta Index' (Δ-index) is defined as the variability of oxygen saturation quantified by the mean of the absolute differences in successive oxygen saturation values sampled at 12-second intervals [11]. For example, if oxygen saturation was nearly constant during the night, the computed index was low. With a variability index below 0.6, they were able to rule out the diagnosis of OSAHS in 301 screened patients, with only three negative test results, giving a sensitivity of 98% and specificity of 46%. In 1999, however, Olson *et al.* re-examined this variability index for 793 patients suspected of having OSAHS [80]. (They also looked at the cumulative percentage of time when the SpO_2 signal was less than 90%.) After an unattended overnight home oximetry test, and next day follow-up laboratory PSG, they concluded that the diagnostic specificity was too low to warrant further use. Conversely, Golpe *et al.* showed that oximetry had high specificity when analysing 116 patients with

[6]OSAHS, UARS, PLMs, narcolepsy and parasomnia all fall under the SDB category.

moderate to severe OSAHS, and concluded that oximetry analysis was more useful to confirm rather than to exclude OSAHS diagnoses [41]. Three numerical oximetry indices were evaluated, including: the average number of dips (desaturations) in the SpO_2 signal $\geq$4% per hour, the average number of resaturations of the SpO_2 signal $\geq$3% per hour, and the cumulative percentage of time spent at saturations below 90%. Chiner *et al.* noted the obvious contradictions reported up to date (1999) between the potential successes and failures of overnight oximetry analyses for the diagnosis of OSAHS. In their study of 275 patients with a clinical suspicion of OSAHS, a desaturation was defined as a fall in SpO_2 signal lower than 4% from baseline, computed as the average value of the SpO_2 signal in the previous minute. Patients were classified as OSAHS or non-OSAHS based on their computed index, or number of oxygen desaturations per hour. The automated and manual assessment of each patient were compared, whereby Chiner *et al.* concluded that should oximetry analysis had been utilised, over 100 PSG tests would not have had to have been performed [19]. The automated assessment of oximetry data, therefore, has potential for clinical diagnoses. Cooper *et al.* also concluded, in their 41-patient study calculating the number of SpO_2 dips >5%, that pulse oximetry is an effective tool for screening patients with moderate-to-severe sleep apnea [22]. Deegan and McNicholas reported in their 250-patient study, that patient history and pulse oximetry would have been sufficient to make a diagnosis of OSAHS in one third of the patients; whereas, in the other two thirds, a final diagnosis would have to be established using PSG [29, 78].

Several research groups have investigated the correlation between RDI from oximetry and AHI from polysomnography [87]. For example, in 1991, Rauscher *et al.* observed a linear relationship between the RDI derived from oximetry alone (using either desaturation or resaturation events) and the AHI derived manually from PSG, in 53 patients under investigation for possible OSAHS [90]. In 2000, Vázquez *et al.* produced an oximeter derived RDI that was highly correlated ($\rho = 0.97$) with the PSG derived AHI for 246 patients with suspected OSAHS. The AHI was created by a single trained observer using the R & K criteria for staging sleep, and the AASM definitions for identifying apnoeas

and hypopnoeas. (However, a $\geq$4% oxygen desaturation was utilised instead of the >3% criteria specified by the AASM Task Force [3,111]). For each patient, the RDI was created by the retrospective analysis of the SpO_2 signal whereby an event was marked when the percentage of oxygen saturation value dropped below baseline, defined as the average of the SpO_2 signal of the five minutes preceding the marked event. Patients were classified as OSAHS or non-OSAHS based on their new RDI values. A summary of previous studies was included in Yamashiro *et al.*, and is reproduced here for convenience - see Table 4.1 [119]. (Flemons *et al.* also provide a systemic review of the literature in their 2003 work [36].)

Year	Authors	No. Patients	SpO_2 Drop or Other Criteria
1989	Farney *et al.* [33]	54	>4%
1991	Cooper *et al.* [23]	41	>5%
1991	Pepin *et al.* [86]	26	Defined 'Delta Threshold'
1991	Williams *et al.* [115]	40	>4% and <90%
1992	Douglas *et al.* [30]	200	>4%
1992	Stoohs *et al.* [105]	56	>3%
1993	Gyulay *et al.* [44]	98	>4% (>2%, >3%)
1993	Rauscher *et al.* [91]	116	Cyclic Oscillations > 30 minutes
1993	Series *et al.* [101]	240	Not Defined
1995	Yamashiro *et al.* [119]	300	>3%

Table 4.1: Summary of SpO_2 time-domain studies, 1989-1995.

4.8.2 SpO_2 Signal: Frequency Domain

As discussed above, the desaturation-resaturation cycle illustrated in the SpO_2 signal for subjects with OSAHS has been well documented in the time domain. In 1999, Zamarrón *et al.* investigated the power density of SpO_2 signals using the Fast-Fourier Transform (FFT), whereby OSAHS was suspected if a peak in the spectrum was observed (of a subject's overnight SpO_2 signal, in its entirety) between 0.014 and 0.033 Hz (corresponding to a period range between 30 and 70 seconds) [124]. Their database included 233 subjects clinically suspected of having OSAHS, 27 of whom had underlying lung diseases. In

two later papers, Zamarrón *et al.* showed that the frequency of cyclical apnoeic activity corresponds to a spectral peak in both the periodograms[7] of the SpO_2 and heart rate signals, again between 30 and 70 seconds, using databases consisting of 197 and 300 patients with suspected OSAHS [121, 123]. In 2005, Zamarrón *et al.* broadened their period range from 30 to 120 seconds, in their 201-patient study investigating the relationship between SpO_2 and heart rate signals using coherence analysis (a measure of the linear covariance between oxygen saturation and heart rate spectra) [122]. In all of their studies, Zamarrón *et al.* consider the SpO_2 (and heart rate signals) per patient in their entirety; i.e., the all-night recordings as a whole. Most recently, Marcos *et al.* extracted four spectral features from the estimated power spectral density (PSD) of the SpO_2 signal for input into a neural network for patient classification in its entirety (OSAHS or non-OSAHS) [70]. The features include peak amplitude within the band of interest - as defined in [121], total average power, average power within the band of interest, and the ratio of the latter two values [70].

4.8.3 SpO_2 Signal: plus others

Signal patterns of the heart during OSAHS

OSAHS also elicits a unique heart rate rhythm of brady-tachycardia [43, 57].[8] Inspiratory efforts are made against an occluded upper airway during obstructive events. This is similar to performing the Valsalva Manoeuvre (inspiration against a closed glottis), thus producing vagal or parasympathetic stimulation and bradycardia, a reduction in heart rate [102]. An obstructive event terminates with an arousal, thus producing sympathetic stimulation and tachycardia, an increase in heart rate [46, 47]. The brady-tachycardia rhythm disrupts the body's sympathovagal balance - a complicated interaction between

[7] A periodogram is the graphical representation of the power spectral density (PSD) estimate of a given signal, calculated in units of power per radians per sample. A vector of corresponding frequencies is computed in radians per sample, with the same length as the PSD estimate [73].

[8] The normal (resting) heart rate is 70 beats per minute (bpm). The heart rate is lower during sleep, and an abnormally low heart rate $\leq$ 60 bpm could indicate the disorder *bradycardia*. The heart rate is much higher during vigorous activity (200 bpm), but a high resting heart rate $\geq$ 100 bpm could indicate the disorder *tachycardia* [89].

the sympathetic and parasympathetic branches of the central nervous system (CNS) which affects the autoregulation of the cardiovascular system and autonomic tone. As the focus of this work is on the oxygen saturation signal, further discussion of OSAHS and heart rate is restricted to the use of the Heart Rate Variability (HRV) signal with the SpO_2 signal for apnoeic activity identification. (More detailed information on the HRV signal and OSAHS may be found in the following references: [14, 21, 38, 46, 57, 71, 72, 97].)

Heart Rate Variability (HRV)

Some research, however, has been published that claims superior diagnostic accuracy by the combination of SpO_2 and Heart Rate Variability (HRV) parameters. Stoohs and Guilleminault noted that a combination of indices - derived from oxygen saturation, heart rate, and snoring - would have eliminated the two false-positive cases observed in their 56-subject detection of OSAHS study [105]. They monitored oxygen saturation, heart rate, snoring and body position in order to screen subjects for OSAHS. In 1996, Shiomi *et al.* reported a significant increase in very low frequency activity (0.008-0.04 Hz, or 25-120 seconds) within the HRV signal, which was synchronised with a decrease in SpO_2 signal, for 12 patients with OSAHS and a subsequent decrease in the very low frequency activity upon treatment with prosthetic mandibular advancement (PMA) [102]. An apnoea was defined as a respiratory arrest lasting for 10 seconds or more, and with an additional five second 'awakening response,' the minimum length of an apnoea according to their definition was set as 25 seconds ($T/2 = 10$ sec, $T = 20$ sec + 5 sec) [102]. However, their results were based on 30-minute epochs of the HRV and SpO_2 signals. Raymond *et al.* utilised a 'cardiac-oximetry disturbance index' (CODI) in their retrospective study of 100 patients with suspected OSAHS. The number of respiratory event-related arousals were estimated by counting the number of autonomic arousals that were coincident with a rise in oximetry [92]. Autonomic arousals were assessed on the basis of changes in the RR-interval tachogram.[9] Specifically, as arousal from sleep has been found to cause a characteristic

[9]The interval between two consecutive QRS markers of the ECG signal gives the RR-interval, which can be averaged over a number of beats to obtain a good estimate of the inter-beat interval [89]. Instantaneous

pattern of rise and fall in heart rate [92]; therefore, autonomic arousals were identified according to the following template: a decrease of at least 60 ms over a 10-second window in the filtered RR-interval followed within 20 seconds by a rise of at least 85 ms over a 10-second window. In their study, they found that oximetry alone (based on 4% dips) was a less effective screening test than CODI. This index was significantly correlated with AHI and was also better correlated with the EEG arousal index than was oximetry. Lastly, Hilton *et al.* examined spectral analysis of HRV and SpO_2 signals in their study of 40 patients with suspects OSAHS, as introduced in Section 4.7. They used two methods of analysis: Discrete-Fourier Transform and Discrete Harmonic Wavelet Transform. Receiver operator characteristic (ROC) curves were used to assess their methods' diagnostic potential, and a comparison was made between HRV and oximetry analysis, whereby the AHI was computed using the SpO_2 dips >2%, >3%, and >4%. The study concludes that spectral analysis of HRV 'appears to be a significantly' better indicator of OSAHS than oximetry for nREM sleep. However, for REM sleep, spectral analysis of HRV is a comparable indicator of OSAHS to oximetry, owing to the fact that there is an overall reduction in muscular tone during REM sleep; i.e., desaturation is more pronounced and the variability in heart rate increases [47, 85].

Airflow

Conflicting reports have been made regarding the combination of airflow and oxygen saturation parameters for OSAHS event detection. Taha *et al.* analysed 10 PSG records containing 1,938 Sleep Disordered Breathing (SDB) events (apnoea, hypopnoea) [109]. By combining pulse oximetry and sum respiratory inductive plethysmography (RIP) data, 93.1% of the manually detected events were detected by their automated algorithm. For every detected desaturation event, the program checked for the presence of an apnoea first and hypopnoea second. If neither was found, the desaturation event was flagged for a subsequent manual review. The comparison of automated and manual event detection

heart rate values are defined as the inverse of the RR-interval of each beat [13, 89].

was not 1:1; i.e., a variable delay exists between the onset of each SDB event and the subsequent desaturation [114] and therefore apnoeas/hypopnoeas were restricted to occur within a window beginning prior to the start of the desaturation and ending at the time of minimum saturation. Baltzan *et al.* combined oximetry with thermistor airflow information to evaluate the diagnostic accuracy for OSAHS. In their OxiFlow device, they used decrements in airflow (between 20 and 50%) and decrements in oxygen saturation (between 2 and 4%) to identify apnoeic episodes with marginal success. Visually, they found airflow information helpful in providing additional information for recording interpretation. With automatic analysis, however, they concluded that it was 'unclear' whether their device was superior to oximetry alone.

In 2000, Redline *et al.* reviewed the various methods of identifying respiratory disturbances for sleep apnoea assessment [94]. Presently, inter-laboratory result variations may be attributed to the use of different sensors for characterising breathing patterns, different amplitude criteria for identifying any reduction in breathing signal, and different use of corroborative data to discriminate normal from hypopnoeic breaths [94]. Redline *et al.* defined 11 discrete apnoea-hypopnoea indices (AHI) calculated according to various combinations of airflow, oxygen desaturation, and micro-arousal measures [94]. Four examples, with "hypopnoeas and apnoeas identified on the basis of amplitude changes in airflow," include: AHI-T: regardless of associated desaturation and micro-arousal; AHI-2D: associated with a $\geq$2% desaturation;[10] AHI-2D, A: associated with a $\geq$2% desaturation or micro-arousal; and AHI-A: associated with a micro-arousal regardless of desaturation. From over 5000 subjects (who underwent overnight PSG), several conclusions were drawn [94]. First, the magnitude of the median AHI values varied 10-fold (or from 2.9 to 29.3) depending on the definition used by the experts. Second, the correlation between AHIs ranged from 0.99 to 0.68, with highest correlations being found among AHIs requiring apnoeas and hypopnoeas to be associated with some level of desaturation [94]. Overall, it was suggested that different approaches for measuring AHI contribute to sub-

[10] Additional definitions replaced the 2% oxygen desaturation limit with 3, 4 and 5%; and also included various other combinations not discussed above.

stantial variability in the identification and classification of OSAHS. The extent of the variability between AHI values was also exemplified in a reported survey of 44 ASDA-accredited sleep laboratories, where no two labs used the same equipment and definition of hypopnoea [76, 94]. Redline *et al.* emphasised the need to understand which markers of physiological stress during sleep-breathing are most closely linked to adverse outcomes, as well as the need to derive standard AHI definitions that best take into account the underlying population distribution.

Respiration and EEG

In 1998, Pitson and Stradling reviewed new measures for estimating the degree of sleep fragmentation in patients with OSAHS [88]. Three conventional indices of sleep fragmentation (micro-arousals, AHI (based on airflow only), SpO_2 dip rate) and two autonomic indices (heart rate and blood pressure rises) were compared. They also looked at the relationship of these markers during sleep with micro-arousals, respiratory events and subjective sleepiness. Their study utilised one-night PSG recordings from 40 sleep clinic referrals undergoing investigation for possible OSAHS. Correlations (r) between the five indices ranged from 0.38 to 0.73; correlations between blood pressure rises with SpO_2 dips and micro-arousals were stronger ($r = 0.71$ and $r = 0.65$) than for heart rate rises with SpO_2 dips and micro-arousals ($r = 0.55$ and $r = 0.51$); and all indices (except heart rate increases) were similar in their correlation with subjective sleepiness. Overall, they concluded that autonomic arousals obtained from blood pressure rises could be measured easily, objectively, and appeared 'no worse' at predicting subjective sleepiness than either micro-arousals or AHI [88].

In 1999, Bennett *et al.* reviewed the health status in OSAHS patients and the relationship between this status and sleep fragmentation, daytime sleepiness and the effects of nCPAP treatment [7]. First, they confirmed that in comparison with the general population, OSAHS patients have impaired health. Second, they confirmed that the health status of OSAHS patients is improved with nCPAP treatment. Confirmations

were achieved through the use of a subjective but quantitative questionnaire. In the 51-patient study, with subjects ranging from non-snoring individuals to those with severe OSAHS, PSG tests were administered with autonomic (blood pressure related) and EEG indices of sleep fragmentation calculated. The EEG-based indices included the following variables scored manually: sleep stages (scored via R & K rules); micro-arousals (scored via ASDA rules); and respiratory arousals (scored as micro-arousals preceded by an apnoea or hypopnoea, defined by airflow measurements only). Daytime sleepiness was also measured both subjectively (with the Epworth Sleepiness Scale, ESS) and objectively (with the Oxford Sleep Resistance Test, OSLER). Although respiration was also recorded by pulse oximetry, in addition to airflow and rib/abdomen movement, no related sleep fragmentation index was described or utilised in subsequent analysis. Overall, significant associations were found between all sleep fragmentation indices and health status improvement with nCPAP; however, these correlations were not close. Thus, subsequent health status improvement may be more directly related to symptoms not directly related to OSAHS. Bennet *et al.* do not reach any firm conclusions regarding the optimal choice of sleep index for OSAHS analysis - thus supporting the notion that more research is required.

In 1999, Stradling *et al.* continued their work with OSAHS and patterns, first noting that the main reason for treating OSAHS is to control Excessive Daytime Sleepiness (EDS); however, correlation between OSAHS severity measures and EDS assessment results are not close [108]. Some potential reasons for this poor correlation include night-to-night sleep apnoea severity variation, lack of familiarity with single-night hospital stay, possibility that symptoms are cumulative results of variable sleep fragmentation (patients may vary nocturnal sleep time or take extra daytime sleep for compensation), variation in the extent to which a patient is forced to wake up per respiratory event to re-establish adequate airflow across pharynx, and invisible events varying between individuals (i.e., not all apnoeas end in visible EEG arousals). Thus, Stradling *et al.* investigate the magnitude of the micro-arousal that terminates respiratory events in sleep apnoea, especially

the inter-individual variations, the differences caused by posture, and whether events are apnoeas or hypopnoeas (airflow-based). Over 10 patients with severe OSA (AHI>30) were investigated. From the studies, it was found that apnoeas were more disturbing to sleep than hypopnoeas; i.e., the arousals accompanying apnoeic events caused greater variability in quantified sleep depth, and also lasted longer based on visual scoring, than the hypopnoeic events. Also, OSAHS did not worsen in the supine position as initially hypothesised. Overall, highly significant inter-individual variations were found between patients, in the degree to which each event (at night) disturbs the EEG. The development of methods to identify reliably which events are most disruptive to sleep may improve the ability of sleep studies to identify those OSAHS subjects most likely to respond to nCPAP [108].

In 2002, Pepperell *et al.* extensively reviewed sleep apnoea studies [87]. Their report described the pathophysiology of OSAHS and more importantly, the historical development of the standard laboratory-based sleep studies and their associated limitations. In terms of analyses based on EEG data, [87] explain how classical sleep laboratories had a long history of sleep investigation centred on the EEG and its varying patterns during sleep. In the 1960s, routine sleep studies expanded beyond stand-alone EEG and utilised EEG, EMG and EOG to stage sleep and define total sleep times. Although Pepperell *et al.* believe that classical PSG is a flawed gold-standard; they suggest two EEG-based methods for micro-arousal detection. More specifically, they note that it is the ability to score micro-arousals by EEG data that supports the argument for continued inclusion of the EEG channel during sleep studies. First, spectral analysis of the raw EEG signal is suggested for display alongside a conventional sleep stage histogram. Thus, EEG periods may be identified (via frequency bands) and subsequently used to detect arousals and calculate sleep fragmentation. Second, neural network analysis is suggested to automate the process of identifying micro-arousals. They also note, however, that the availability of EEG-scored micro-arousals does not increase the ability to diagnose OSAHS in comparison with simple markers of respiratory disturbance or body movement [8, 87]. Pepperell

et al. also discuss autonomic markers such as heart rate and blood pressure; oesophageal pressure and mechanical forces such as rib and abdomen movement; and oxygen desaturation. Overall, they conclude that classical PSG cannot and should not be regarded as the gold-standard for sleep analyses, and argue that PSG has never been properly validated as a tool to measure the actual pathological processes resulting primarily from reduced airway tone at sleep onset, or to answer some of the usual clinical questions about patients with sleep apnoea [87].

Chapter 5

Databases

5.1 Introduction

Accurate detection of apnoeic activity is key to the assessment of Obstructive Sleep Apnoea Hypopnoea Syndrome (OSAHS). As introduced in Chapter 3, clinicians use the Apnoea Hypopnoea Index (AHI), number of apnoeas and hypopnoeas per hour of sleep, to determine the severity of OSAHS in patients. The definition of apnoea and hypopnoea vary widely amongst sleep clinics, and many studies have been conducted to diagnose OSAHS using the automated analysis of the SpO_2 (and other) signals, as discussed in Section 4.8.

This Chapter describes the preliminary processing of SpO_2 signals to be used for the automated detection of *individual* events for comparison with those events hand-scored by experts. Two sets of databases are used for developing and testing the automated algorithms. The four-patient **MIT-BIH Polysomnography Database** is used for preliminary algorithm development, and the 91-patient **Bristol Database** is used for further algorithm training and subsequent testing. Each database is described in terms of its acquisition, subsequent scoring by experts, and processing of the expert sleep annotations in preparation for analyses in both time and frequency domains.

5.2 Training database

5.2.1 MIT-BIH Polysomnography

The database utilised for algorithm development is the 'MIT-BIH Polysomnography' database, which was originally developed for inclusion in the PhysioNet archival system developed by Moody *et al.* in 2000 [40]. PhysioNet is a large research resource, complete with an archive of well-characterised digital recordings of physiological signals and related data for use by the biomedical research community. Of the 16 continuous polysomnography recordings, four include the SpO_2 signal (all anonymised male patients who suffer from OSAHS), and airflow signal. The four SpO_2 and airflow recordings are between one and seven hours in duration and are digitised at a sampling rate of 250 Hz with a resolution of 12 bits/sample.

Each record was annotated by two or more experts; discrepancies were resolved by consensus, or by another expert if the original annotators were unavailable. The 'Expert Scores' include sleep stages, movement time, and the following events - with and without an associated arousal: obstructive apnoeas, central apnoeas, hypopnoeas, leg movements (plus unspecified arousals). Unfortunately, the exact definitions of apnoea and hypopnoea are not available for this database; it is not known whether or not the AASM definitions (see Chapter 3) were utilised for scoring events.

Each sleep stage annotation, and event annotation if applicable, is assigned to 30-second segments of non-overlapping data. The annotation coding scheme is such that one or more events may be assigned to any one 30-second segment. For example, the annotation 'H OA' means that both a hypopnoea and obstructive apnoea have been found to occur within the same 30-second segment. The order of the annotation is successive; i.e., the hypopnoea occurs before the obstructive apnoea. Furthermore, an event may be annotated for a 30-second (n^{th}) segment as long as it starts within the segment; i.e., the event may finish entirely within the n^{th} segment or finish in the $(n+1)^{th}$ segment. For example, for the annotation 'H OA,' the obstructive apnoea starts within the same

segment as the hypopnoea but may not end until the following segment. Table 5.1 details the annotation coding scheme.

Annotation	Meaning
W	patient is awake
1	sleep stage 1
2	sleep stage 2
3	sleep stage 3
4	sleep stage 4
R	REM sleep
H	hypopnoea
HA	hypopnoea with arousal
OA	obstructive apnoea
X	obstructive apnoea with arousal
CA	central apnoea
CAA	central apnoea with arousal
L	leg movements
LA	leg movements with arousal
A	unspecified arousal
MT	movement time

Table 5.1: MIT-BIH Polysomnography Database: annotation coding scheme, such that one or more events may be assigned to any one 30-second segment.

5.2.2 Processing sleep annotations

As discussed in Section 3.4, a minimum of two signals (such as airflow and respiratory effort) are required according to the AASM definition, to differentiate between apnoea types. However, in routine clinical practice, it is not always necessary to differentiate apnoea (and hypopnoea) types because both events have similar pathophysiology which may be treated in the same manner [3]. Furthermore, this work focuses on the exclusive use of the SpO_2 signal for the automated detection of events. Therefore, obstructive apnoeas, central apnoeas, and hypopnoeas are grouped into an overall 'apnoea' category. For each patient, the Expert Score annotations are reviewed for obstructive apnoeic, central apnoeic, and hypopnoeic events - with or without arousal. In total, there are six possible Expert Score annotations corresponding to apnoea (OA, X, CA, CAA, H, HA

- see Table 5.1). A label of '1' is assigned to every 30-second window with an apnoeic event, and a label of '0' otherwise. The duration of each scored event is unknown for each patient within the database; however, the number of events and the total time spent asleep is known. Therefore, both the AHI (see Section 3.6) and the percentage of events with respect to the total time spent asleep may be computed.

The data for all four patients were analysed, whereby the experts assessed the records as being apnoeic (obstructive and central) for 30% to 40% of the total time asleep (see Table 5.2). The computed AHI values for patients (slp59, slp60, slp61, and slp67) are respectively (54.3, 58.3, 41.2, and 56.9). According to the AASM definition, all four patients suffer from *severe* OSAHS.

Patient	Obstructive Apnoea (%)	Central Apnoea (%)	Total Apnoea (%)	Hypopnoea (%)
slp59	17.5	11.7	29.2	16.0
slp60	30.6	10.4	41.0	7.6
slp61	28.8	0.9	29.7	4.6
slp67	9.7	33.1	42.8	4.5

Table 5.2: MIT-BIH Polysomnography Database: percentage of apnoeas and hypopnoeas per patient, with respect to the total time spent asleep.

5.3 Testing database

5.3.1 Bristol

The database utilised for algorithm testing is the 'Bristol' database, which was originally developed for the clinical assessment of patients at the Bristol General Hospital (Bristol, UK) from 2005 to 2008. Each patient who visited the sleep clinic with a suspected sleep condition (OSAHS),[1] was first sent home to undergo two nights of pulse oximetry recording. The recorded SpO_2 signals were then analysed using the 'Download 2001'

[1]Some OSAHS patients included in this database underwent full-night PSG for research purposes instead of diagnostic purposes.

software from Stowood Scientific Instruments, whereby a sleep index is created for experts/clinicians. Specifically, 'Download 2001' computes the number of dips >4% of the SpO_2 signal per hour[2] [107]. From the clinical perspective, nCPAP (see Section 3.3) treatment was prescribed for patients with a sleep index of 10 to 15 dips per hour; whereas, one full-night polysomnography (PSG) was recommended for patients with a sleep index of <10 dips per hour. All 91 patients in the Bristol Database underwent PSG. Note that the Download 2001 results for the 91-patient database are not available.

At the Bristol General Hospital, full-night PSG includes the acquisition of up to 21 channels of information. A typical channel configuration (which varies slightly across the patient database) is found in Table 5.3. Each record was annotated by one expert. The 'Expert Scores' include sleep stages and the events listed in Table 5.4. Note that each event is marked as it occurs and is not assigned to any fixed-size window (such as in the MIT-BIH PSG Database). The software employed for scoring the PSG signals is the Alice ®Sleepware ™version 2.4.8.a ©1999, 2006 from Respironics, Inc. The 'Alice Sleepware' software is able to to analyse the SpO_2 signal; however, its results are **not** compared with those of the 'Download 2001' software whatsoever.

All 91 recordings include the SpO_2 signal. Of the patients in the Bristol Database, 25 are female and 66 are male patients with a suspected sleep condition. The SpO_2 recordings are between 3.5 and 11 hours in duration and are digitised at a sampling rate of 1 Hz with a resolution of 10 bits/sample. Each patient's data was anonymised and given a randomly-generated patient identification number.

The exact definitions of apnoea and hypopnoea used by the Bristol experts are the same as those provided by the American Academy of Sleep Medicine (AASM) Task Force (see Chapter 3), and repeated here for convenience. An apnoea is defined as the complete or near-complete cessation of airflow that lasts for at least 10 seconds or longer. A hypopnoea is defined as an event that lasts for at least 10 seconds or longer and must

[2]The 'Download 2001' software provides, in addition to artefact rejection, three sets of analyses, including SpO_2, Pulse Rate, and PCO_2. In addition to the calculation of the dip rate, the SpO_2 analysis also includes (for example) the computation of the mean, median, mode, mean nadir, desaturation index, and time spent below user-specified percentages of the SpO_2 signal [27].

satisfy one of the following criteria: (i) a clear decrease (>50%) from baseline in the amplitude of a valid measure of breathing during sleep; (ii) a clear amplitude reduction of a valid measure of breathing during sleep that does not reach the above criterion but is associated with either an oxygen desaturation of >3% or an arousal. As shown in Table 5.3, channels 7 through 10 provide airflow information. According to the experts, the preferred channel is airflow via nasal cannulae (channel 8), followed by airflow via thermistor (channel 7). The experts only use the cleanest respiratory signals. The thoracic and abdominal channels (9 and 10) are difficult to rely upon as movement shifts the acquired signals dramatically. In the event that all four respiratory signals are not suitable for scoring, the experts use a combination of the other PSG signals, including the SpO_2 signal. (Note that a plethysmograph (channel 13) is an instrument for recording and measuring variation in the volume of a part of the body, or of the whole body, especially as caused by changes in blood pressure [1], and the quality of the electrode-scalp connection - for acquiring EEG signals - is checked via electrical impedance (channel 21).)

The experts define an AHI for use in categorising the OSAHS severity (normal, mild, moderate, severe) - as discussed in Section 3.6. The experts' AHI is defined as the number of 'events' divided by the Total Sleep Time or TST. An 'event' includes obstructive apnoeas, mixed apnoeas, central apnoeas, and hypopnoeas. TST is defined as time spent during REM and nREM, but also includes movement time which occurs during Sleep Period Time or SPT. SPT is defined as sleep onset to the cessation of sleep according to the R & K definition (see Section 2.2). Arousals are included in SPT; whereas, such periods of wakefulness are **not** included in TST. According to these rules, the experts categorised 33 patients in the database as 'normal' (AHI: 0-<5 events per hour); 23 patients as 'mild' (AHI: 5-14 events per hour); 23 patients as 'moderate' (AHI: >14-30 events per hour); and 12 patients as 'severe' (AHI: >30 events per hour).

Channel	Signal Acquired
1	EEG - C3 A2
2	EEG - O2 A1
3	EEG - A1 A2
4	EOG - Left
5	EOG - Right
6	EMG
7	Airflow - Thermistor
8	Airflow - Nasal Cannulae
9	Thoracic
10	Abdominal
11	Microphone (Snoring)
12	SpO_2
13	Plethysmograph
14	Pulse Rate
15	Body Movement
16	ECG - Lead 1
17	ECG - RR Detector
18	ECG - Lead 2
19	Leg Movement - 1
20	Leg Movement - 2
21	Electrode Impedance (EEG)

Table 5.3: Bristol Database: typical channel configuration for an all-night, full, polysomnography (PSG).

5.3.2 Processing sleep annotations

As discussed in Section 5.2.2, a minimum of two signals (such as airflow and respiratory effort) are required according to the AASM definition, to differentiate between apnoea types. Therefore, obstructive, central, and mixed apnoeas are grouped into an overall 'apnoea' class. It is also impossible to distinguish between an 'apnoea' and hypopnoea using the SpO_2 signal only; therefore, the hypopnoeic events are also included in the 'apnoea' class. For each patient, the Expert Score annotations are reviewed for 'apnoeic' events. For each patient, the number of apnoeic events with duration greater than 10, 20, 30, 60, 90 and 120 seconds were computed. Next, the percentage of patients who at least *one* event greater than (10, 20, 30, 60, 90 and 120) seconds was calculated and were found to be (100%, 89%, 71%, 21%, 5% and 3%). For example, 71% of the 91

Annotation
Central Apnoea
Obstructive Apnoea
Mixed Apnoea
Hypopnoea
Bradcardia
Tachycardia
Leg Movement
Snore
Desaturation (4%)
Arousal
Long-RR
Micro-Arousal
Heart Rate Drop
Heart Rate Rise

Table 5.4: Bristol Database: annotation scheme.

patients had at least one event greater than 30 seconds in duration. Only 21% had at least one event greater than 60 seconds in duration. 100% of the patients had at least one event greater than 10 seconds in duration - as 10 seconds is, by definition, the minimum duration required in order to score an event. The value of the maximum event duration for each patient was on average 45 seconds. The values of the maximum event duration for each patient, averaged according to categorisation status, were found to be 30 seconds (normal); 42 seconds (mild); 63 seconds (moderate); and 63 seconds (severe).

For the database of 91 Bristol patients, the experts assessed the records as being apnoeic (obstructive, mixed, central) for 0.2 to 19.6% of the total time and hypopnoeic for 0.7 to 12.0% of the total time - see Table 5.5 - depending on OSAHS severity (normal, mild, moderate, severe) as defined in Section 3.6. For example, the patients categorised as 'moderate' had apnoeic and hypopnoeic events for 3.8% and 7.1% of their total time sleeping. As the length of each patient's record is not the same, the total apnoeic/hypopnoeic events were divided by the total SPT for that category of patients. Similarly, the breakdown of events with respect to the total number of events scored is shown in Table 5.6 - also according to OSAHS severity. For example, the patients categorised as 'normal' had 371 events scored overall, of which 46 (or 12.4%) were obstructive apnoeas. In con-

trast, the patients categorised as 'severe' had 4142 events scored overall, of which 2364 (or 57.0%) were obstructive apnoeas.

To assess the diagnostic accuracy of the signal processing methods to be described in Chapters 6 to 8, a comparison must be made between the output for these methods and the corresponding 'Expert Scores.' Comparison of the automated methods to the Expert Scores requires the calculation of the following parameters: True Positives, True Negatives, False Positives, False Negatives, Sensitivity, Specificity, and Accuracy - as described in Appendix B.

Category	Central Apnoea (%)	Obstructive Apnoea (%)	Mixed Apnoea (%)	Total No. of Apnoeas (%)	Hypopnoea (%)
Normal	0.1	0.1	0.0	0.2	0.7
Mild	0.2	0.3	0.0	0.5	3.2
Moderate	0.6	3.1	0.1	3.8	7.1
Severe	0.6	18.3	0.7	19.6	12.0

Table 5.5: Bristol Database: the total event durations divided by the total duration of the sleep recordings for each patient on a per-category basis.

Class	Central Apnoea	Obstructive Apnoea	Mixed Apnoea	Total No. of Apnoeas	Hypopnoea	Total No. of Events
Normal	11.9 % (44)	12.4 % (46)	0.3 % (1)	24.5 % (91)	75.5 % (280)	100 % (371)
Mild	9.1 % (101)	10.0 % (110)	1.2 % (13)	20.3 % (224)	79.7 % (880)	100 % (1104)
Moderate	7.4 % (194)	26.4 % (696)	1.2 % (31)	34.9 % (921)	65.1 % (1717)	100 % (2638)
Severe	2.3 % (97)	57.0 % (2364)	2.1 % (89)	61.6 % (2550)	38.4 % (1592)	100 % (4142)

Table 5.6: Bristol Database: the total number of each event type (in brackets) divided by the total number of events scored, shown as a percentage, for each patient on a per-category basis.

Chapter 6

SpO_2 Time Domain

6.1 Introduction

The SpO_2 waveform for an apnoeic patient is characterised either by small-amplitude, rapid oscillations (normal breathing) or high-amplitude, slow oscillations (apnoea). Figure 6.1 illustrates typical normal breathing (green highlight) and apnoeic (red highlight) behaviour in two ten-minute sections of the SpO_2 signal from a patient with OSAHS (patient slp59 from the MIT-BIH PSG training database). Large oscillations in signal amplitude corresponding to dips in the value of SpO_2 (oxygen desaturation and resaturation) are typically observed during the sleep of OSAHS patients. The depth of each oxygen desaturation is dependent upon the duration of the apnoea, the baseline level of the SpO_2 signal, the expiratory reserve volume, and the apnoea type (see Section 3.4) [37,101]. It is believed that SpO_2 'changes with virtually every apnoea event' [39]. For the SpO_2 signal illustrated in Figure 6.1, the shift in baseline is obvious; i.e., for the highlighted areas of interest, the baseline values are approximately 96% and 90% for apnoeic and normal breathing activity, respectively.

In Section 4.7, it was noted that the normal operating range for SpO_2 is greater than 90%, usually between 96 and 100% [28]. Thus, the 96% baseline shown for the apnoea highlight is as expected. However, reoxygenation is somewhat impaired during the cyclic

recurrence of apnoeas; i.e., it is more typical that the next apnoea would start from a lower baseline than the previous apnoea [90]. Thus, the 90% baseline shown for the breath highlight is justified. As introduced in Section 4.7, it should be possible to detect apnoeic activity automatically from the SpO_2 signal by tracking the amplitude of the dips in the time domain. In this Chapter, we describe the signal processing of the SpO_2 signal using the standard deviation of the signal amplitude to identify apnoeic episodes.

6.2 MIT-BIH PSG Training Database

6.2.1 Pre-processing

The patient data from the MIT-BIH Polysomnography Database is divided into two classes: normal breathing and apnoea (whereby obstructive apnoeas, central apnoeas, and hypopnoeas are all grouped together, as explained in Section 5.2.2). Each SpO_2 signal sampled at a frequency f_s of 250 Hz is first segmented into non-overlapping windows 30 seconds in duration. The 30-second window size is chosen to correspond to the Expert Score analysis which, with this database, is applied on a 30-second timescale.

Next, the windowed SpO_2 signal is detrended and visually and automatically assessed for artefact. An artefact is defined as a region of SpO_2 signal with sudden transitions, such as a sharp change in baseline value. For example, for a patient with an SpO_2 level of around 90%, a sharp dip below 60% is indicative of an artefact rather than a brief physiological deterioration, as SpO_2 values of less than 60% are associated with mental function and coordination impairment [28]. For example, if the pulse oximeter probe's position briefly shifts on the finger during recording, an artefact will occur. Figure 6.2 shows the SpO_2 signals in the time domain for all four patients in the MIT-BIH PSG Database. In patient slp67, bottom plot of Figure 6.2, there are two noticeable artefact between 0.2 and 0.4 hours, during which the SpO_2 signal suddenly drops from 95% to 75% and below. (The first artefact is at 0.3 hours and the second artefact is just before 0.4 hours.) An artefact is also automatically detected when the SpO_2 signal drops to 0% or

the mean value of the SpO_2 signal in the time domain (average value of the SpO_2 signal within the 30-second window) is lower than 60%; i.e., the signal falls abruptly due to artefact but does not necessarily reach 0%. Note that over 30% peak-to-peak oscillations in SpO_2 values *are* physiologically possible, but only for extreme cases of OSAHS in patients (see Chapter 3). Segments with artefact are not included in any further analysis; i.e., for each window under analysis, if 50% or more of the window consists of artefact, that window is discarded. Only non-artefactual segments of the processed signals are then compared with the Expert Scores. The total number of 30-second segments for patients (slp59, slp60, slp61, slp67) are respectively (480, 710, 740, 154), of which (88, 46, 110, 22) segments are further removed from analysis due to artefact - see Table 6.1 where the percentage of artefact varies from 6.6% to 18.5%.

Subject	No. of Segments Analysed	Total No. of Segments	Artefact (%)
slp59	391	480	18.5
slp60	663	710	6.6
slp61	622	740	15.9
slp67	131	154	14.9

Table 6.1: MIT-BIH PSG Database: number of segments analysed for each of the four patients compared with the total number of segments available per patient, with the percentage of artefact found for that record shown accordingly.

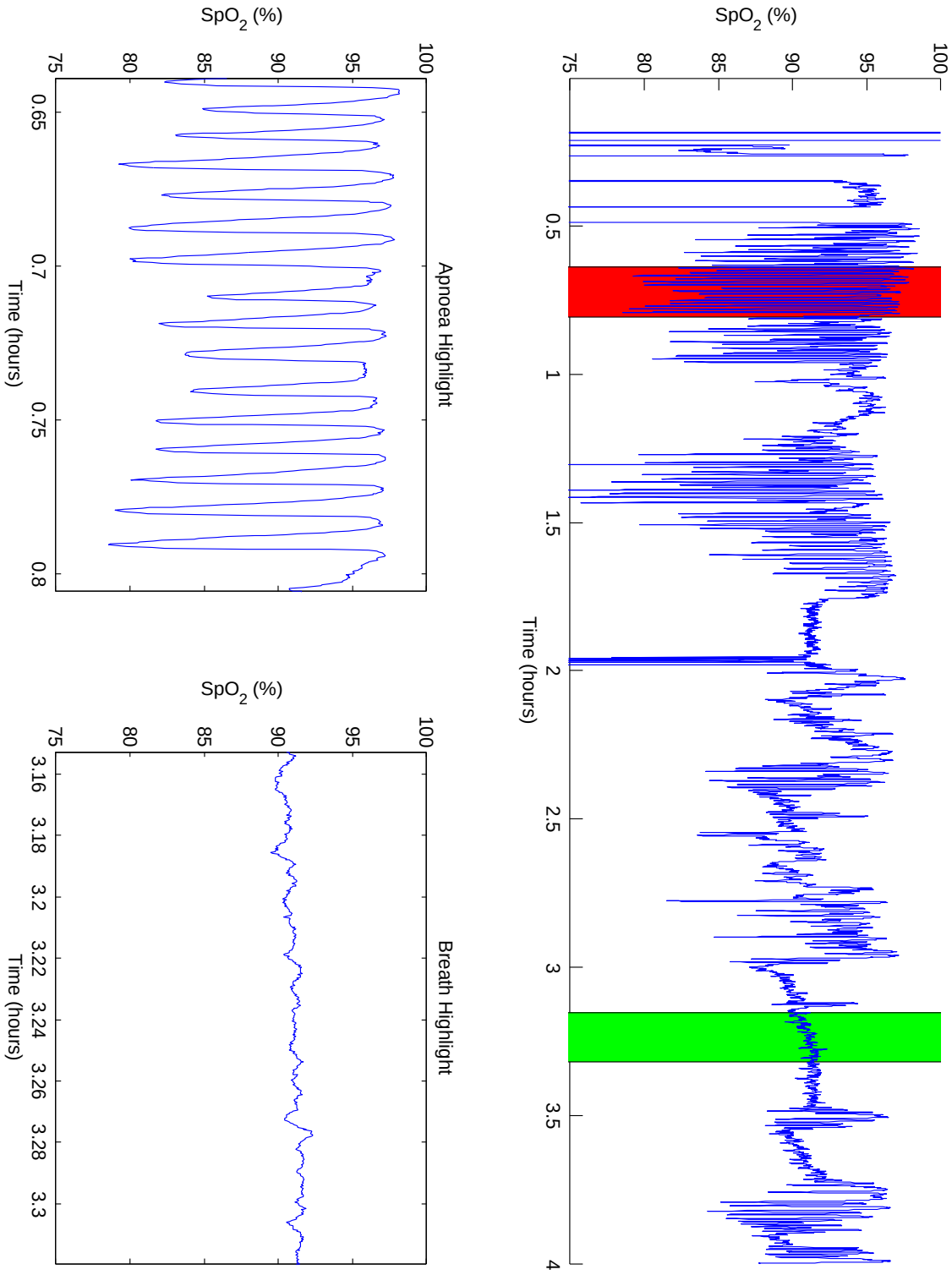


Figure 6.1: From top to bottom: SpO_2 signal (%) in the time domain from a patient with OSAHS; left and right highlights of the SpO_2 signal where the 10-minute red highlight corresponds to apnoea, and 10-minute green highlight corresponds to normal breathing.

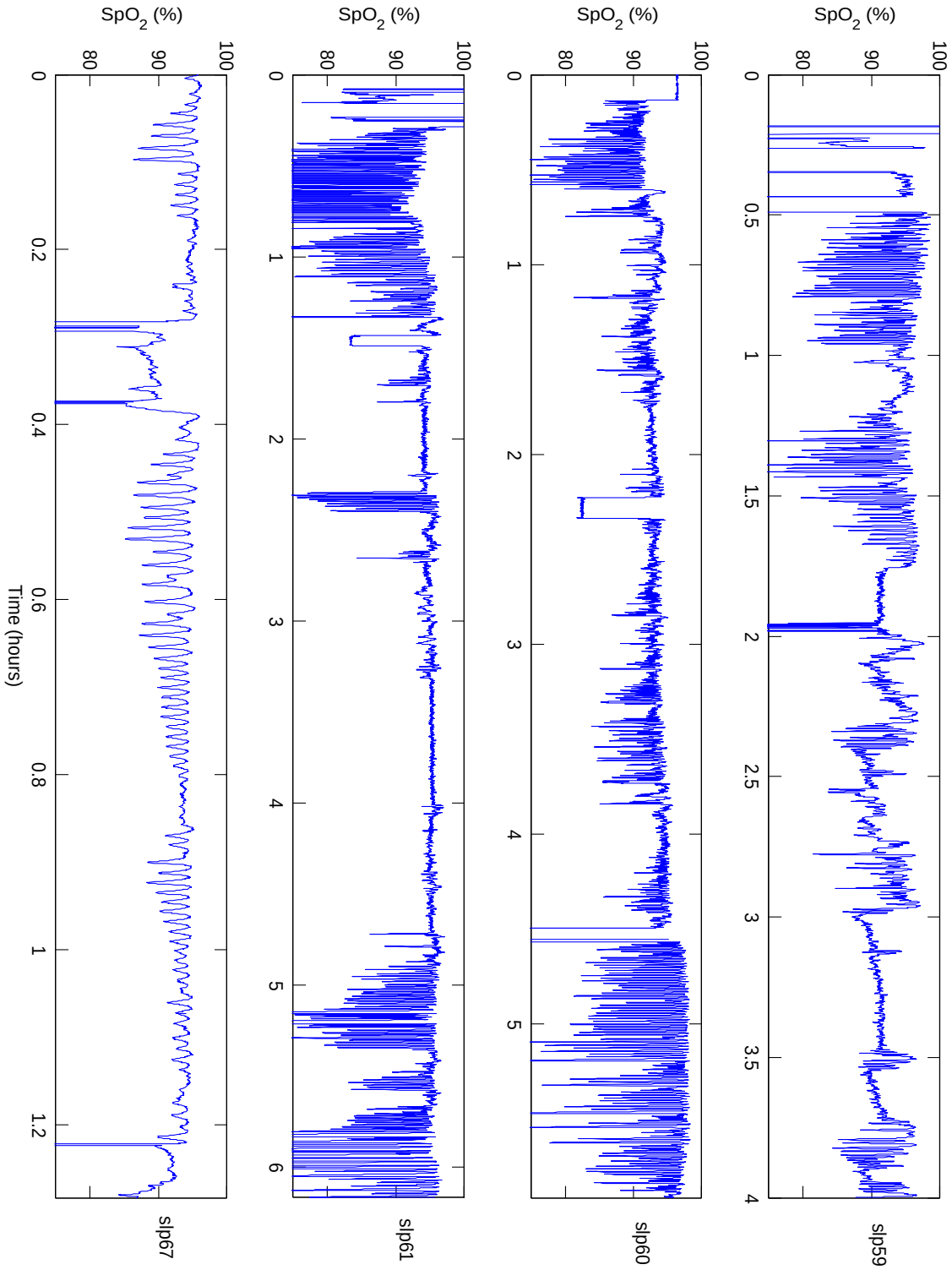


Figure 6.2: MIT-BIH PSG Database: SpO_2 signals in the time domain from all four patients (slp59 through slp67). Values are shown from 75% to 100% oxygen saturation, with the x-axis giving the time from the start of the recording in hours.

6.2.2 Processing

Following segmentation and detrending, the standard deviation of each 30-second window is computed. The more pronounced the apnoeic activity (desaturation and resaturation of the SpO_2 signal), the larger the standard deviation value. Conversely, the less pronounced the apnoeic activity, the smaller the standard deviation value. Figure 6.3 shows the SpO_2 signal from patient slp60 and the corresponding standard deviation values and Expert Scores. The Expert Scores are labelled such that an 'apnoea' is given a score of '1,' a 'normal breathing' a score of '0,' and any artefact windows a score of '-1.'

An alternative method of visualisation uses colour-coding of the Expert Scores, superimposed on the standard deviation values. Figures 6.4 through 6.7 show the SpO_2 signals and their corresponding standard deviation values colour-coded, such that segments labelled as 'apnoea' are shown in red, segments labelled as 'normal breathing' are shown in green, and 'artefact' segments are shown in black. Also included in these figures are standard deviation values above 10 units, although they are cut off from view (y-axis). Following this section, segments which are labelled artefact will be omitted from view in subsequent figures.

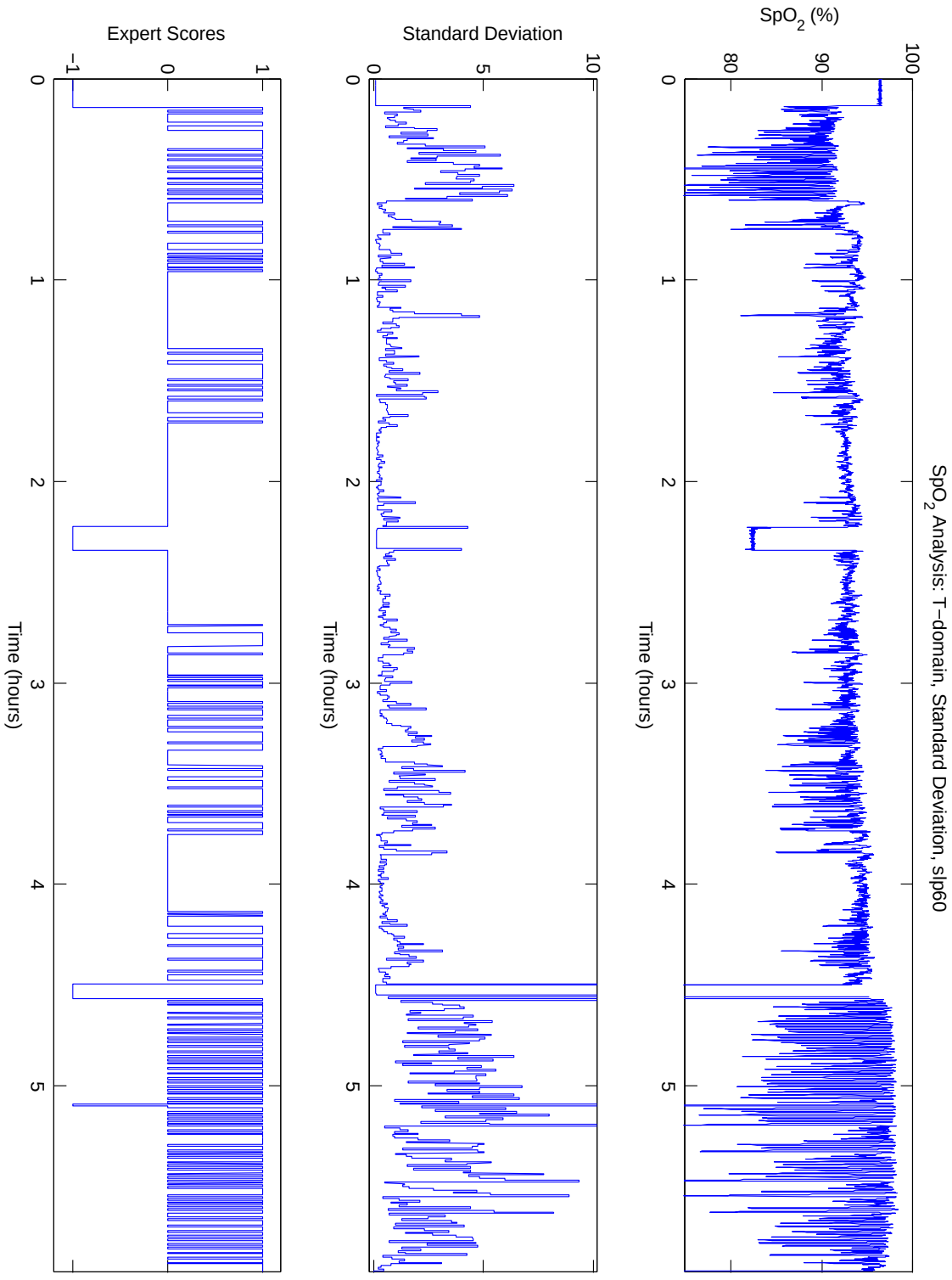


Figure 6.3: Subject slp60. From top to bottom: SpO_2 signal (%); standard deviation for each 30-second window; Expert Scores where '1' corresponds to apnoea, '0' corresponds to breath, and '-1' corresponds to artefact.

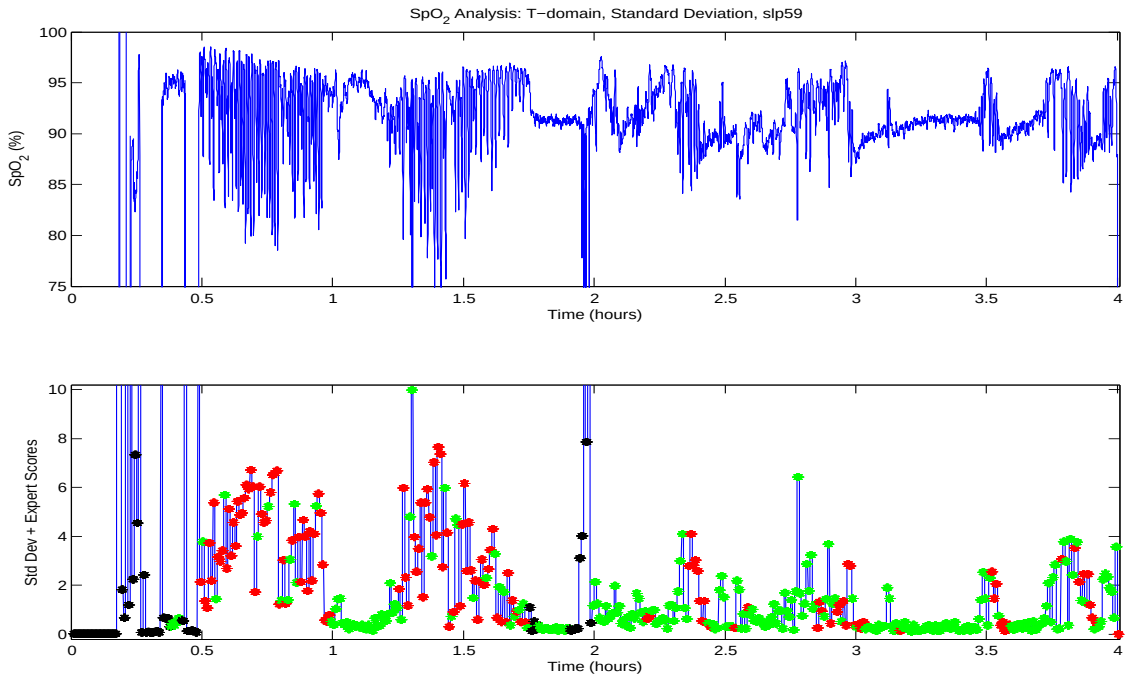


Figure 6.4: Subject slp59. From top to bottom: SpO_2 signal (%); standard deviation for each 30-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea, green = normal breathing, black = artefact.

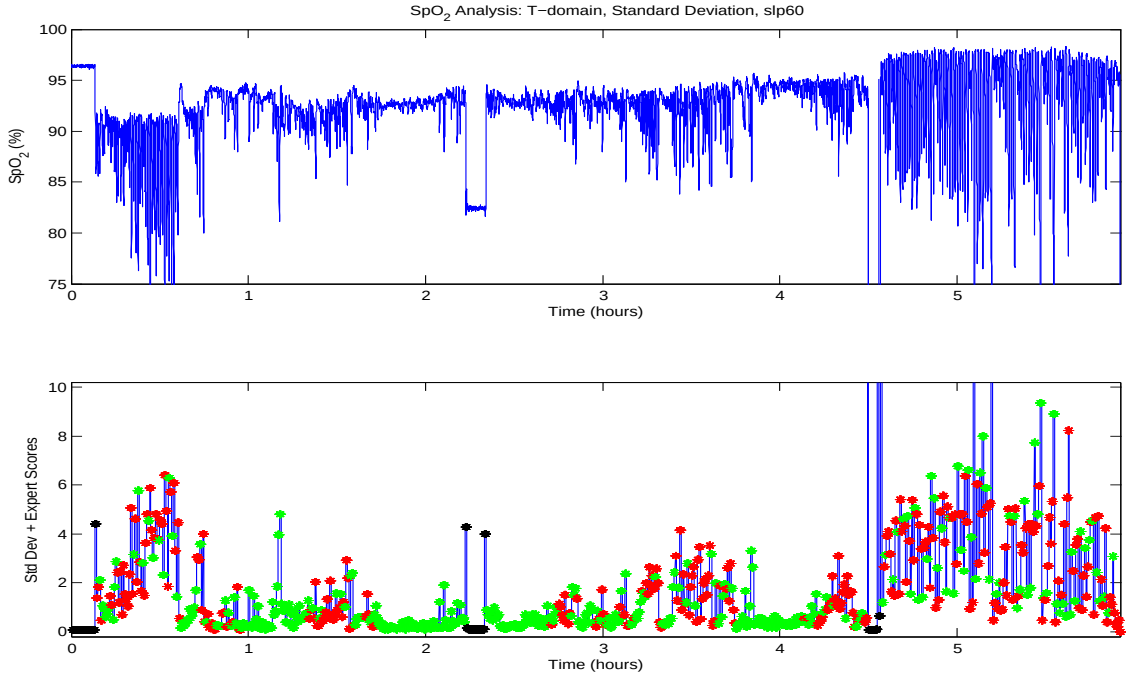


Figure 6.5: Subject slp60. From top to bottom: SpO_2 signal (%); standard deviation for each 30-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea, green = normal breathing, black = artefact.

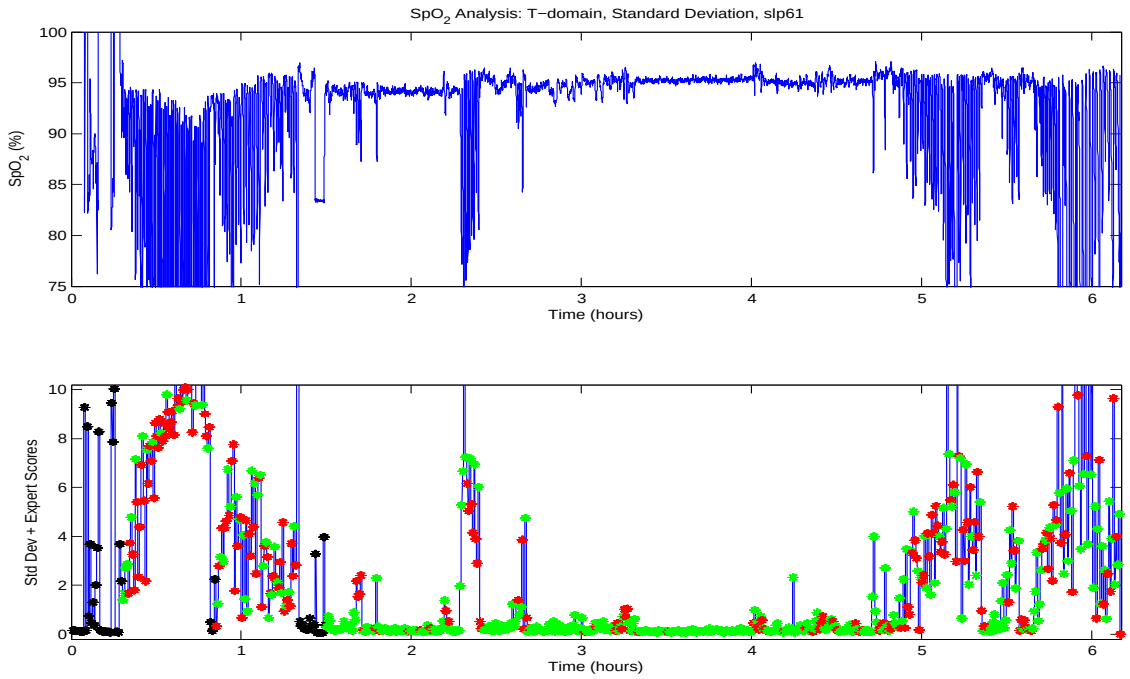


Figure 6.6: Subject slp61. From top to bottom: SpO_2 signal (%); standard deviation for each 30-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea, green = normal breathing, black = artefact.

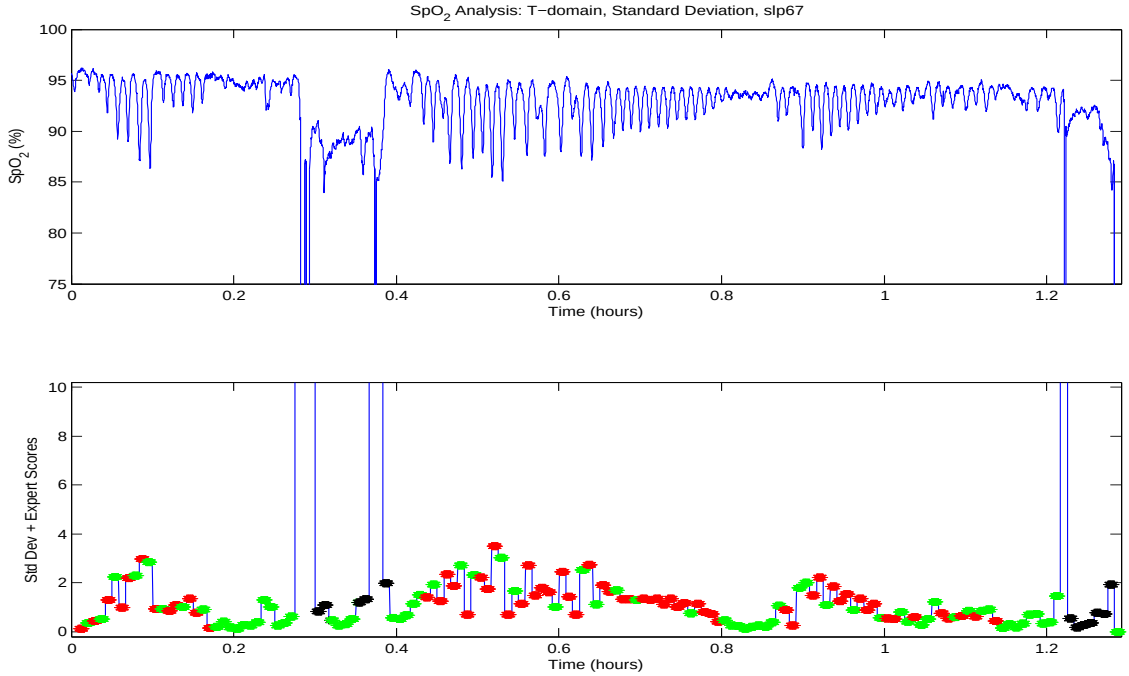


Figure 6.7: Subject slp67. From top to bottom: SpO_2 signal (%); standard deviation for each 30-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea, green = normal breathing, black = artefact.

6.2.3 Selection of an appropriate threshold

For each SpO_2 record, there can only be two types of signals: small-amplitude oscillations at breathing frequencies corresponding to normal breathing, OR high-amplitude oscillations at a much lower frequency corresponding to apnoeas. From Figures 6.4 to 6.7, it is clear that apnoea can be identified by applying an appropriate threshold to the values of the SpO_2 standard deviation, such that values above the threshold correspond to regions of apnoea (and are given a score of 1) and values below correspond to regions of normal breathing (and are are given a score of 0). The standard deviation values computed for each window of the segmented signals will be referred to as σ_{SpO_2} from now on, given in standard deviation units or 'units s.d.'

Based on previous research, see Table 4.1 in Section 4.8, three typical threshold values used as markers of the presence of apnoeic event activity include: 2%, 3%, and 4% dips in the SpO_2 signal, which correspond to σ_{SpO_2} values of 0.7, 1.0, and 1.4 units s.d. respectively. The classifications determined with the use of these three threshold values can be compared with the Expert Scores in terms of Sensitivity (Se), Specificity (Sp) and Accuracy (Acc).

6.2.4 Classification accuracy and results

The distribution of σ_{SpO_2} values computed per segment for each patient are shown in the histograms in Figure 6.8 with threshold values superimposed as a dotted line (red = 0.7 units s.d., green = 1.0 units s.d., blue = 1.4 units s.d.). Notice that the maximum values of each distribution, i.e., the y-axis, differ from each other, and artefactual segments are not included in the Figures. For each patient, a peak in the distribution of σ_{SpO_2} values is visible to the left of the thresholds; i.e., the breathing segments. The data distribution to the right of the thresholds is more uniform; i.e., a larger range of σ_{SpO_2} values is visible for apnoeic segments.

The classifications (Se, Sp, Acc) found using the three selected thresholds for each

patient are given in Table 6.2. For the four patients, as the threshold increases from 0.7 to 1.4 units s.d., the sensitivity decreases; for all four, a threshold of 2% SpO_2 desaturation is optimal for the detection of apnoeic activity. Conversely, as the threshold increases from 0.7 to 1.4 units s.d., both the specificity and accuracy increase, with the exception of patient slp67 for whom the accuracy decreases. For all three patients (slp67 excluded), as the threshold increases to 1.4 units s.d., fewer False Positives (i.e., breath identified as apnoea) are scored, but at the cost of the True Positives (i.e., apnoea identified as apnoea) being missed. The highest accuracy, for each threshold value, is found for patient slp61, with: 71.4% (0.7 units s.d.), 72.7% (1.0 units s.d.), and 73.0% (1.4 units s.d.). The SpO_2 signal of this patient exhibits the most extreme behaviour: for long periods of time, normal breathing is present; at other times, apnoeic activity is severe. It is therefore not surprising that the patient's record gives the highest accuracy, when using standard deviation as the discriminating parameter. The corresponding histogram of σ_{SpO_2} values (Figure 6.8) illustrates the extreme behaviour well, with peaks to the left and right of the thresholds; i.e., a bimodal distribution is clearly observed. Conversely, the lowest accuracy is found for patient slp67, with a value of 57.3% for a threshold of 1.4 units s.d. This is not surprising as the SpO_2 signal of patient slp67 does not exhibit stable periods of normal breathing, but includes apnoeic activity throughout. Based on the maximum number of correctly classified segments in Table 6.2, the optimal threshold is 1.4 units s.d.

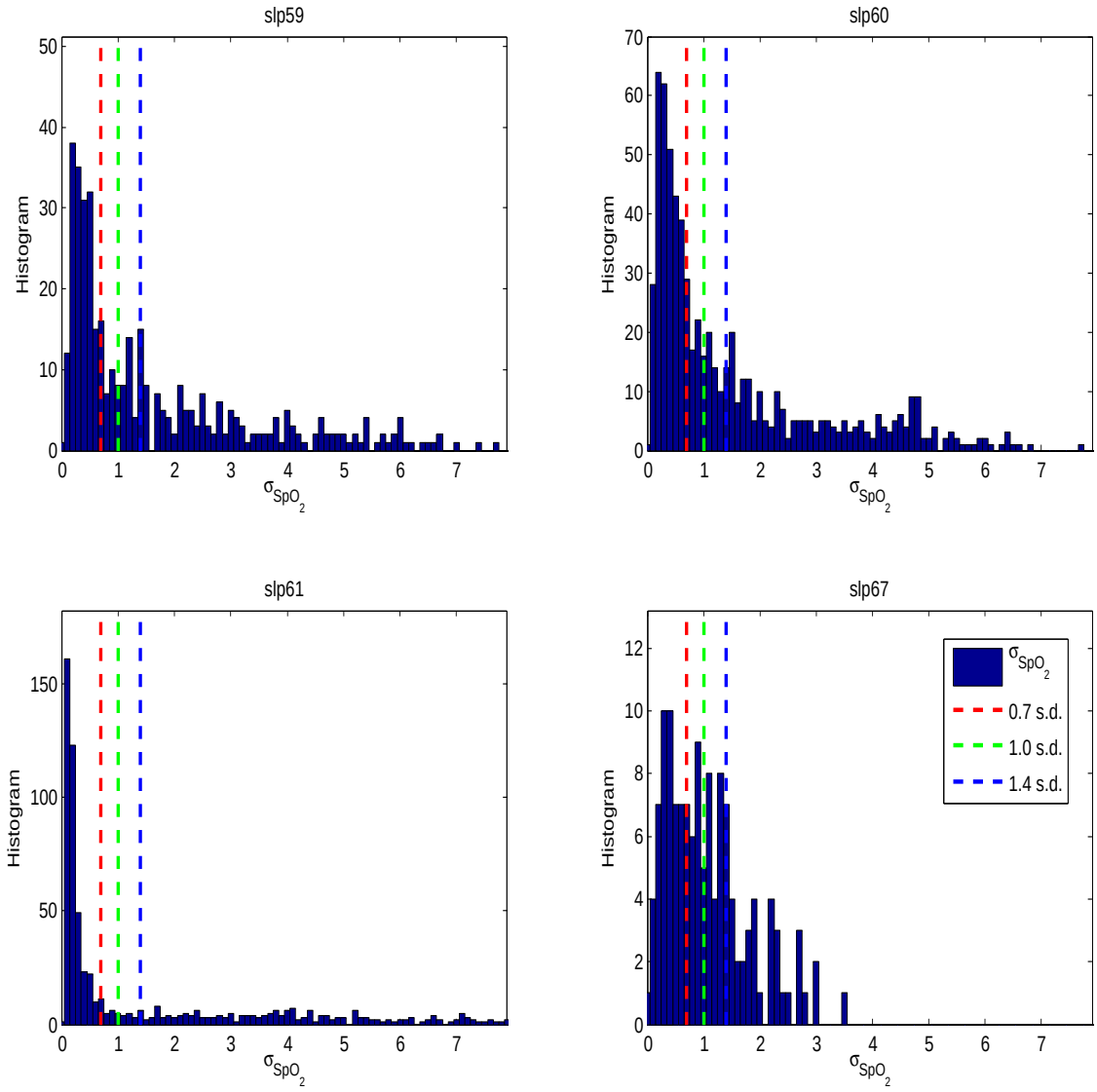


Figure 6.8: Time domain analysis (σ_{SpO_2}) histograms for all four patients in the MIT-BIH PSG Database, with the threshold values under consideration shown by the red (0.7 units s.d.), green (1.0 units s.d.) and blue (1.4 units s.d.) dotted lines.

Subject	Threshold (units s.d.)	Se (%)	Sp (%)	Acc (%)	No. of Segments Correctly Classified	Total No. of Segments
slp59	0.7	78.7	55.7	63.7	249	391
	1.0	74.3	64.7	68.0	266	391
	1.4	64.0	74.5	70.8	277	391
slp60	0.7	75.6	59.4	66.1	438	663
	1.0	65.7	68.9	67.6	448	663
	1.4	55.7	76.8	68.2	452	663
slp61	0.7	64.2	74.2	71.4	444	622
	1.0	61.3	77.1	72.7	452	622
	1.4	57.2	79.1	73.0	454	622
slp67	0.7	77.8	51.5	64.1	84	131
	1.0	61.9	64.7	63.4	83	131
	1.4	33.3	79.4	57.3	75	131

Table 6.2: SpO_2 time domain analysis: classification sensitivities (Se), specificities (Sp), and accuracies (Acc) for each patient in the MIT-BIH PSG Database, using threshold values of 0.7, 1.0, 1.4 units s.d. The number of correctly classified segments out of the total number of segments analysed is also provided.

6.3 Bristol Test Database

The MIT-BIH PSG training database consists of four patients who suffer from *severe* OSAHS; whereas, the test database consists of 91 patients with varying degrees of apnoeic severity. Therefore, the threshold determined using a training database (1.4 units s.d.) of just four patients may be sub-optimal for the general population of OSAHS patients. The Bristol Database is therefore used both for training purposes (∼50% of the patients) and subsequent evaluation on an independent test set of patients (the other ∼50%). The assignment of patients to either the training or test dataset is performed randomly: out of the 33 normal patients, 17 and 16 patients are used for training and testing; out of the 23 mild patients, 12 and 11 patients are used for training and testing; out of the 23 moderate patients, 12 and 11 patients are used for training and testing; and out of the 12 severe patients, 6 patients are assigned in each case to the training and testing datasets. In total, the new training and test datasets, based on a much more representative sample of OSAHS patients, consist of 47 and 44 patients respectively. The training dataset will be used to optimise the standard deviation threshold value, which will then be applied to the test dataset for classifications (Se, Sp, Acc). In addition, restricting the choice of threshold values to correspond to 2%, 3%, and 4% dips in the SpO_2 signal is too coarse a choice of possible threshold values. Also, as discussed in Section 4.8, there is no consensus amongst sleep researchers as to the choice of oxygen desaturation percentage which optimises event detection. Therefore, a range of threshold values from 1% to 5%, in 0.1% increments will be investigated.

6.3.1 Processing

As with the MIT-BIH Polysomnography data, the patient data from the Bristol Database is divided by the experts into two classes: normal breathing and apnoea (whereby obstructive, central, and mixed apnoeas, and hypopnoeas are all grouped together, as explained in Section 5.3.2). Each SpO_2 signal sampled at a frequency f_s of 1 Hz is first segmented

into overlapping windows. In Chapter 5, the average maximum duration of events for the Bristol (testing) database was shown to be 30 seconds (normal category); 42 seconds (mild category); 63 seconds (moderate category); and 63 seconds (severe category). Therefore, four window sizes are explored: 30-second, 60-second, 90-second, and 120-second. (In Section 5.3.2, it was noted that 21% and 5% of the 91-patient database had at least one event greater than 90 and 120 seconds in duration respectively.) For patients categorised as 'normal,' a 30-second window may be long enough to capture the average 30-second desaturation of the SpO_2 signal. For patients categorised as 'mild,' a 60-second window would be long enough to capture the average 42-second desaturation of the SpO_2 signal. However, for patients categorised as 'moderate' and 'severe,' a 90-second or 120-second window would be required to capture the average 63-second desaturation of the SpO_2 signal. Each window starts 30 seconds after the previous one to maximise the number of windows for analysis.

As each SpO_2 signal must be segmented for analysis, so must the Expert Score for comparative purposes. As with the MIT-BIH PSG Database, if an apnoeic event has been scored within the window of choice a label of '1' is assigned, and a label of '0' otherwise. The minimum duration of an apnoeic event is 10 seconds (see Section 3.4); therefore, for a 30-second window, at least 10 seconds must be apnoeic. The minimum apnoeic requirement per window must change accordingly; i.e., doubling/tripling/quadrupling the window size requires a doubling/tripling/quadrupling of the minimum apnoeic requirement. For the 60-second, 90-second, and 120-second window sizes, the minimum apnoea requirements are therefore 20 seconds, 30 seconds, and 40 seconds.

Next, the windowed SpO_2 signal is detrended and automatically assessed for artefact - as described in Section 6.2.1 for the MIT-BIH PSG Database. Segments with artefact are not included in any further analysis; i.e., for each window under analysis, if 50% or more of the window consists of artefact, that window is discarded. Only non-artefactual segments of the processed signals are compared with the Expert Scores.

Following segmentation and detrending, the standard deviation (σ_{SpO_2}) of each window

is computed. An event is identified by applying an appropriate threshold to the values of the σ_{SpO_2}, such that values above the threshold correspond to regions of apnoea (and are given a score of 1) and values below correspond to regions of normal breathing (and are are given a score of 0).

6.3.2 Receiver Operating Characteristic (ROC) method

The Receiver Operating Characteristic (ROC) curve is used for identifying the *optimal* or ROC-threshold for the training dataset for each window size - see Appendix B for an illustration of the generic ROC curve [127]. (Note that the term 'optimal threshold' is reserved for the threshold identified from the training dataset which is subsequently applied to the testing dataset.) For each subject, a range of threshold values is investigated, from 1% to 5% dips in the SpO_2 signal, with increments of 0.1%. Alternatively, the range of threshold values is from 0.35 units s.d. to 1.77 units s.d., in increments of 0.035 units s.d. For each threshold, the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values are computed. The length of SpO_2 signal recordings varies between patients; therefore, computing (Se, Sp, Acc) for each patient in the (Bristol) training dataset and then taking the average across all 47 patients in the training dataset would not be correct. Instead, the TP, TN, FP, and FN are summed across the entire database and used accordingly. From this point onwards, this method will be referred to as *averaging* the results across all the segments from the available patient data. Each Se-Sp pair is determined by comparing a non-artefact segment of the SpO_2 signal with the Expert Score for the same segment, for the given threshold value. The threshold value which yields the highest classification accuracy corresponds to the point on the ROC curve which is closest to the upper left corner (0,1) - where both Se and Sp are 100% [127]. In this manner, the ROC-threshold value is found for all the segments in the entire dataset for each window size. Next, the average (Se, Sp, Acc) values are computed. The optimal window size and corresponding threshold are then chosen, for subsequent application to the test dataset - for which both overall and per-category results are calculated.

6.3.3 Classification accuracy and results

Bristol Database: training dataset

The classification sensitivity, specificity and accuracy found using the four segmentation windows (30-second, 60-second, 90-second, 120-second) and the optimal threshold values determined from the ROC curves for the 47-patient Bristol training dataset are provided in Table 6.3. The optimal threshold values range from 1.4 to 2.2% dips in the SpO_2 signal, or 0.495 to 0.778 units s.d. Figure 6.9 shows the ROC curve computed for each window size, for this dataset. As the window size gets larger from 30 to 120 seconds, both the Se (number of detected True Positives, or apnoeic segments) and Sp (number of detected True Negatives, or normal breathing segments) values increase. The lowest (Se, Sp) pairing is for the 30-second window at (74.1%, 78.5%). The highest (Se, Sp) pairing is for the 120-second window at (83.3%, 82.0%). Subsequently, Acc values increase as the window size gets larger from 78.0% (30-second window) to 82.1% (120-second window). For the 120-second window, the (Se, Sp, Acc) values are (83.3%, 82.0%, 82.1%); i.e., the Se value is slightly lower than that reported for the 90-second window, but the Sp and Acc are the highest overall at 82.0% and 82.1% respectively.

Window Length	Optimal Threshold (units) [Percentage Dip SpO_2]	Sensitivity (%)	Specificity (%)	Accuracy (%)
30 sec	0.495 s.d. [1.4%]	74.1	78.5	78.0
60 sec	0.636 s.d. [1.8%]	80.0	80.1	80.1
90 sec	0.707 s.d. [2.0%]	83.5	80.8	81.0
120 sec	0.778 s.d. [2.2%]	83.3	82.0	82.1

Table 6.3: Time domain analysis resultant Se, Sp, and Acc percentages across the 47-patient Bristol Database training dataset for 30-second, 60-second, 90-second, 120-second window lengths, and corresponding optimal (ROC-based) threshold values.

Bristol Database: test dataset

Table 6.4 shows the overall (Se, Sp, Acc) values computed for the 44-patient test dataset using a 120-second window and 0.778 units s.d. threshold value: 85.3%, 80.5%, and 81.2%.

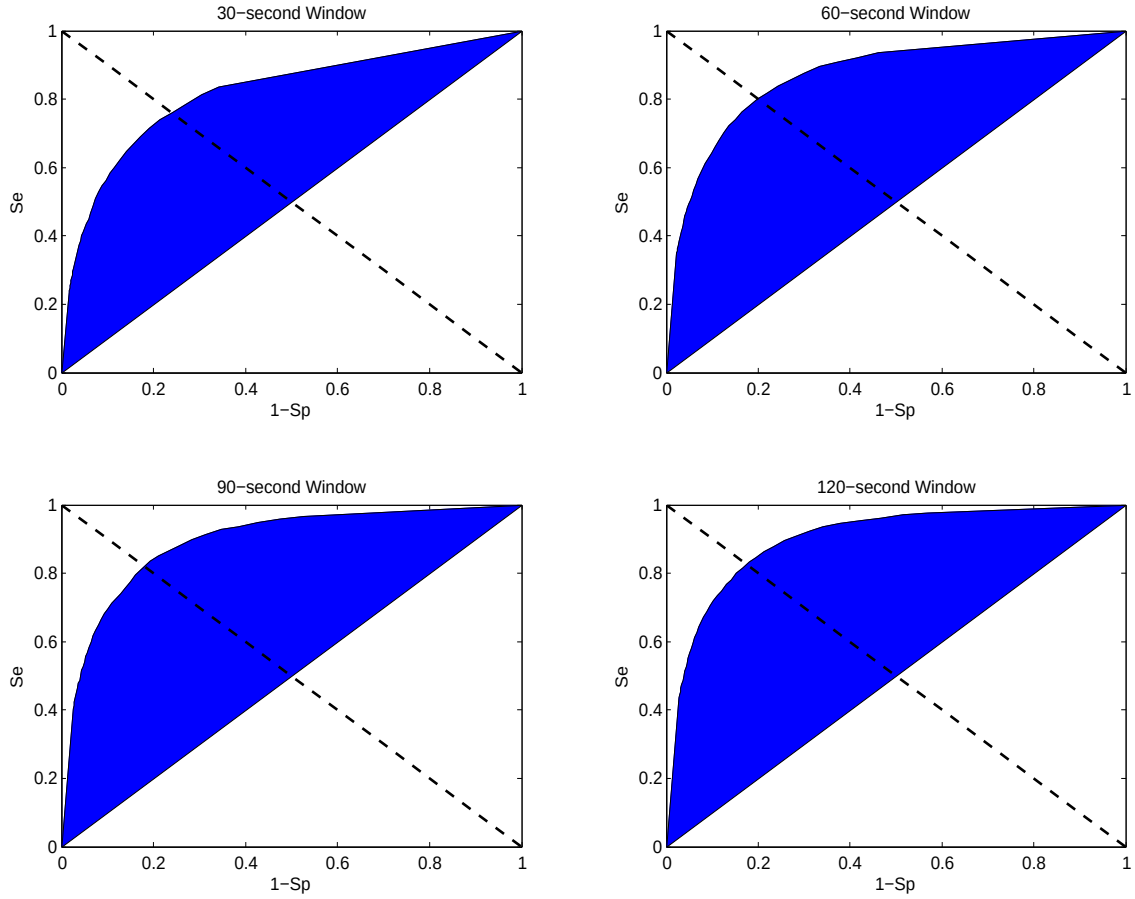


Figure 6.9: Bristol Database: ROC curves for the training dataset using time domain analysis (standard deviation) and four window sizes: 30-second, 60-second, 90-second, 120-second. Maximum classification accuracy occurs at the point on the ROC curve closest to (0,1).

In addition, the (Se, Sp, Acc) values are presented according to normal, mild, moderate, and severe patients within the dataset. The highest Se value reported is for the severe category missing very few apnoeas (96.6%), followed by the normal category (91.4%), moderate category (72.3%), and finally the mild category (44.3%). The severe category also exhibits the lowest Sp value (48.0%); however, for such patients, consistent periods of normal breathing are rare and therefore a low Sp value is not significant. Conversely, Sp values are more significant for normal and mild categories, whereby stable breathing dominates the SpO_2 signal and there are few apnoeas to detect. The highest Sp value reported is for the normal category (87.9%), followed by the mild category (83.4%), and lastly the moderate category (77.2%).

For the automated analysis of patients for whom apnoeic activity dominates (moderate, severe), a high sensitivity is preferable (as apnoea is the "event" to be detected). Conversely, for the automated analysis of patients where normal breathing activity dominates (normal, mild), a high specificity is required. It is therefore worthwhile to consider sensitivity, specificity and accuracy with respect to the severity of the patient condition. The confusion matrices, Tables 6.5 through 6.8, illustrate the results for the 120-second window analysis with the threshold of 0.778 units s.d. The percentage of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) with respect to the total number of segments under analysis is also given.

For the normal category (Table 6.5) there are 35 out of 11348 segments identified as apnoeic, or 0.3%. The number of correctly identified segments is 32 out of 35, or 91.4% (i.e., sensitivity - see Table 6.4). For the mild category (Table 6.6) there are 305 out of 8891 segments identified as apnoeic, or 3.4%. The number of correctly identified segments is 135 out of 305, or 44.3%. For the moderate category (Table 6.7) there are 1525 out of 8375 segments identified as apnoeic, or 18.2%. The number of correctly identified segments is 1103 out of 1525, or 72.3%. Lastly, for the severe category (Table 6.8) there are 2849 out of 5498 segments identified as apnoeic, or 51.8%. The number of correctly identified segments is 2752 out of 2849, or 96.6%. The percentage of apnoeic activity increases with the severity of OSAHS: 0.3% (normal), 3.4% (mild), 18.2% (moderate), 51.8% (severe). However, the percentage of normal breathing segments identified as apnoea (FP) are 12.1%, 16.0%, 18.6%, and 25.1% for the normal, mild, moderate and severe categories, respectively. There is therefore an increasing percentage of False Positives (the system is over-sensitive) but because the percentage of False Positives is also proportional to the severity of the condition, it should not affect the categorisation of the patient by the automated system (see Chapter 9). Also, similarly computed, the percentage of normal breathing activity decreases with the severity of OSAHS: 99.7% (normal), 96.6% (mild), 81.8% (moderate), 48.2% (severe).

Figures 6.10 through 6.13 illustrate the time domain analysis results for the SpO_2

signals of four typical patients, one from each category as assigned by the experts (see Section 5.3.2). Patients 11, 8, 21, and 57 from the Bristol Database are examples of patients with normal, mild, moderate, and severe cases of OSAHS respectively. In all four Figures, a 120-second window is utilised along with the optimal threshold of 0.778 units s.d. For example, the top panel in Figure 6.10 shows the SpO_2 signal for the entire night's sleep of 'normal' patient 37. The bottom panel shows the corresponding σ_{SpO_2} (computed per window) in blue, and colour-coded according to the Expert Score attached to each window (red corresponds to apnoea, green corresponds to normal breathing). The superimposed black dotted line shown here is the optimal threshold. The most straightforward example of σ_{SpO_2} exhibiting large values when there is apnoeic activity, and small values otherwise, is in Figure 6.12 - for the moderate category. However, notice the large σ_{SpO_2} values around the second and sixth hour of the SpO_2 signal; whereby, the experts have scored the majority of segments as normal breathing, but the automated analysis have scored the majority of segments as apnoeic (i.e., the segments fall above the threshold). Similarly, the mild and severe category examples illustrate large σ_{SpO_2} values, but an Expert Score that corresponds to normal breathing. Figure 6.14 shows a highlighted portion of the mild category example (Bristol 8) whereby desaturation-resaturation cycles of the SpO_2 signal are visible (especially between 5.3 and 5.4 hours), the σ_{SpO_2} values fall above the 0.778 units s.d. threshold. However, normal breathing is scored for the majority of these segments.

Lastly, as discussed in Section 3.6, patients with severe OSAHS experience *both* a higher incidence of apnoeic events and apnoeic events of longer duration over the course of the night. In Figure 6.13, dips in the SpO_2 signal as large as 20% are visible; therefore, the resultant 96.6% Se value is not surprising with the optimal threshold of 0.778 units s.d. From the four examples, it is clear that the dips in the value of the SpO_2 signals increase as the severity of patient OSAHS increases from mild to severe.

Categorisation	Sensitivity (%)	Specificity (%)	Accuracy (%)
OVERALL	85.3	80.5	81.2
Normal	91.4	87.9	87.9
Mild	44.3	83.4	82.1
Moderate	72.3	77.2	76.3
Severe	96.6	48.0	73.2

Table 6.4: Time domain analysis resultant Se, Sp, and Acc percentages across the 44-patient Bristol Database training dataset, on an overall and *per category* basis, for 120-second window length and 0.778 units s.d. threshold.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	32 (0.3%)	3 (0.0%)
	Normal Breathing	1370 (12.1%)	9943 (87.6%)

Table 6.5: NORMAL patients' confusion matrix (time domain analysis): 120-second window, 0.778 units s.d. threshold. A total of 11348 segments are used, with the relative percentages shown in brackets accordingly.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	135 (1.5%)	170 (1.9%)
	Normal Breathing	1424 (16.0%)	7162 (80.6%)

Table 6.6: MILD patients' confusion matrix (time domain analysis): 120-second window, 0.778 units s.d. threshold. A total of 8891 segments are used, with the relative percentages shown in brackets accordingly.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	1103 (13.2%)	422 (5.0%)
	Normal Breathing	1559 (18.6%)	5291 (63.2%)

Table 6.7: MODERATE patients' confusion matrix (time domain analysis): 120-second window, 0.778 units s.d. threshold. A total of 8375 segments are used, with the relative percentages shown in brackets accordingly.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	2752 (50.1%)	97 (1.8%)
	Normal Breathing	1378 (25.1%)	1271 (23.1%)

Table 6.8: SEVERE patients' confusion matrix (time domain analysis): 120-second window, 0.778 units s.d. threshold. A total of 5498 segments are used, with the relative percentages shown in brackets accordingly.

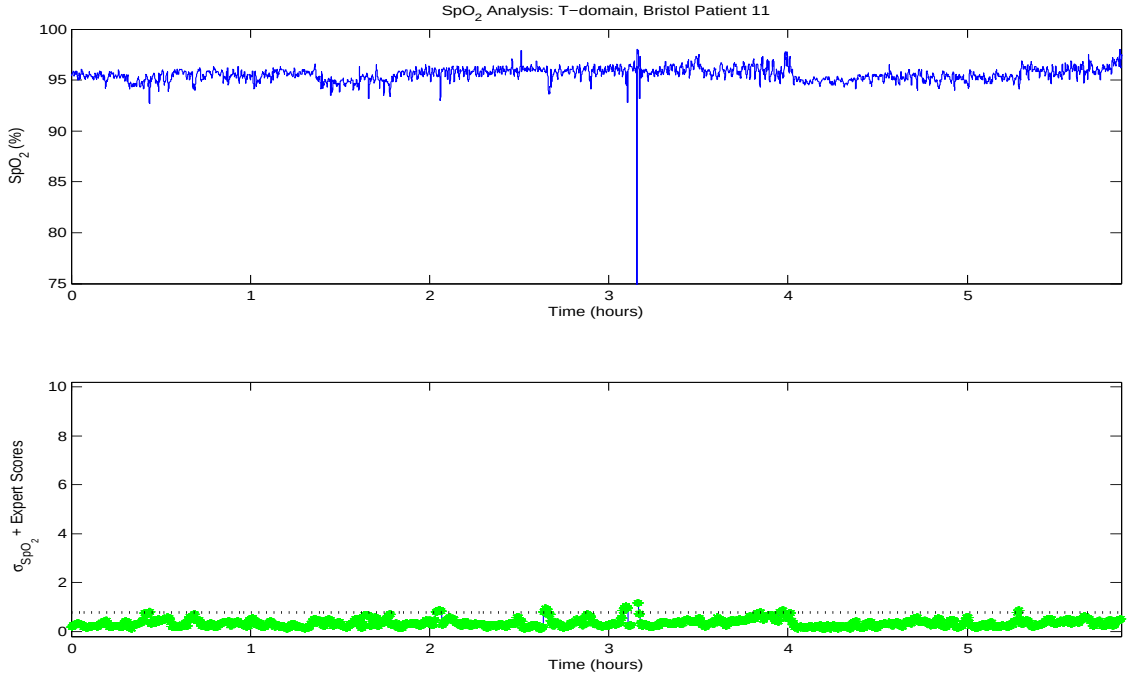


Figure 6.10: Bristol 11: NORMAL patient example. From top to bottom: SpO_2 signal (%); standard deviation per 120-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with 0.778 units s.d. threshold shown as dotted line.

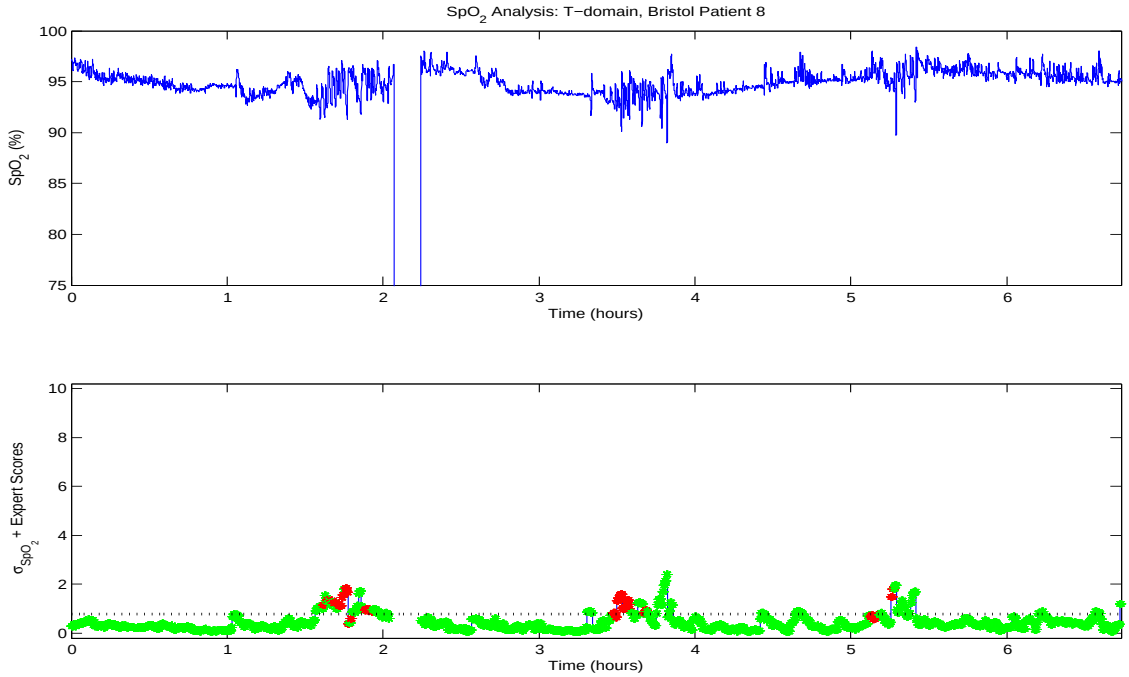


Figure 6.11: Bristol 8: MILD patient example. From top to bottom: SpO_2 signal (%); standard deviation per 120-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with 0.778 units s.d. threshold shown as dotted line.

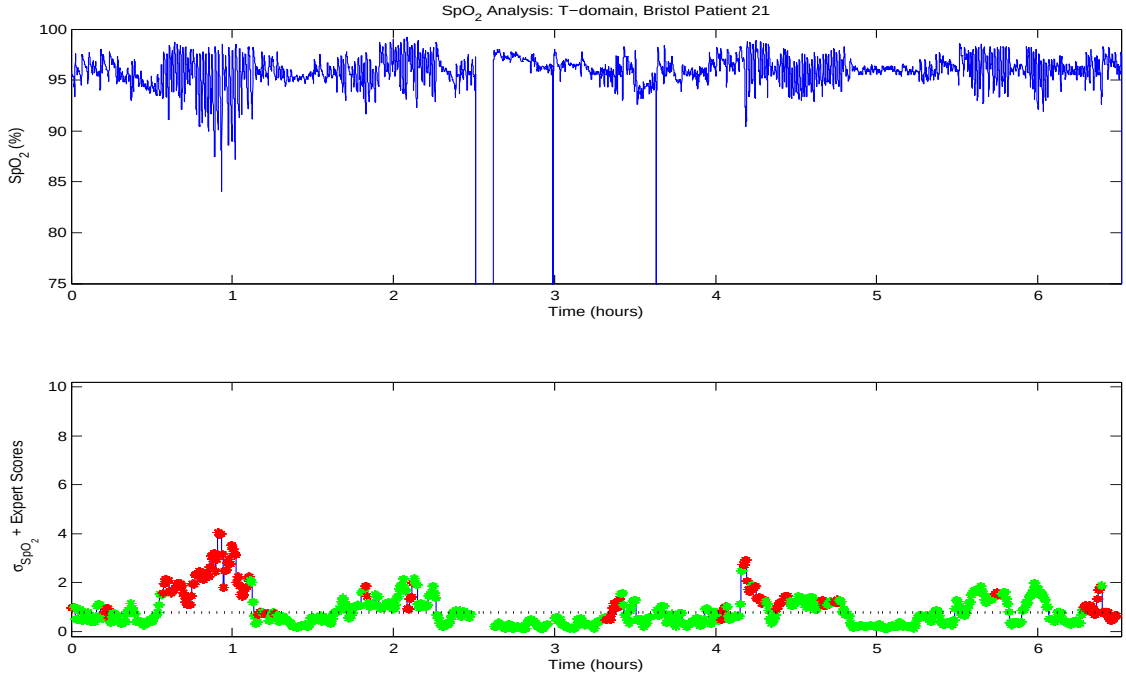


Figure 6.12: Bristol 21: MODERATE patient example. From top to bottom: SpO_2 signal (%); standard deviation per 120-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with 0.778 units s.d. threshold shown as dotted line.

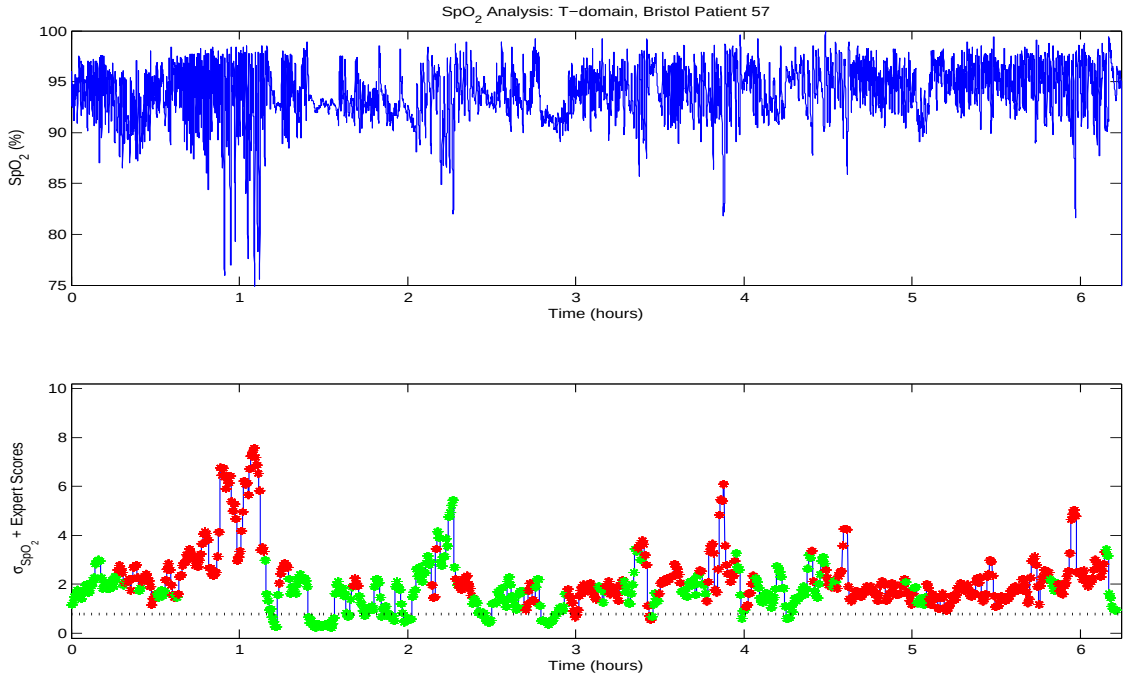


Figure 6.13: Bristol 57: SEVERE patient example. From top to bottom: SpO_2 signal (%); standard deviation per 120-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with 0.778 units s.d. threshold shown as dotted line.

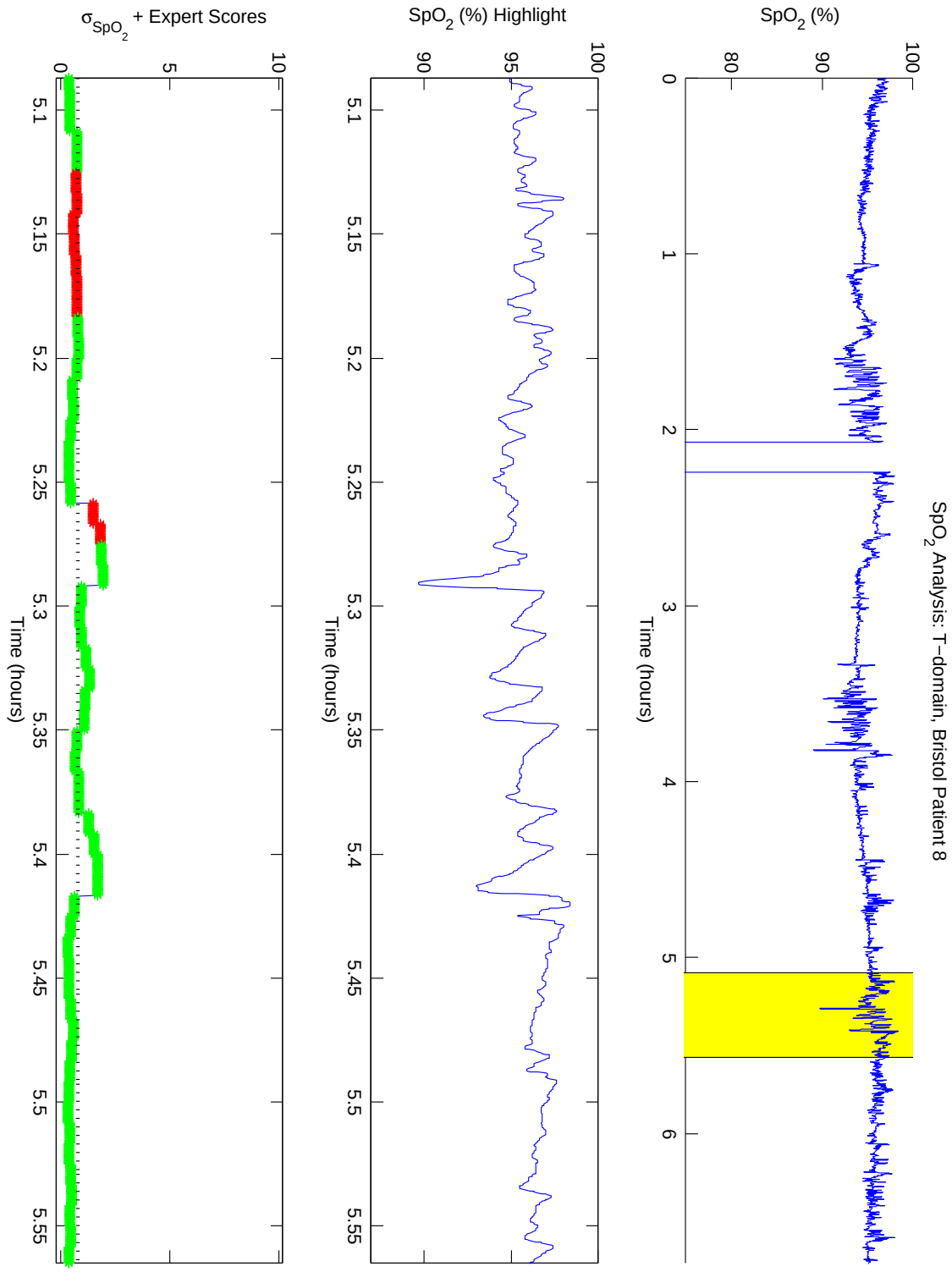


Figure 6.14: Bristol 8: MILD patient example highlighted. From top to bottom: SpO_2 signal (%); standard deviation per 120-second window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with 0.778 units s.d. threshold shown as dotted line.

6.4 Discussion and summary

In this Chapter, the analysis of the SpO_2 signal has been based on the computed σ_{SpO_2} for segmented windows per patient. Classification accuracy is dependent upon the choice of threshold used for the division of data into apnoea or normal breathing. The optimal choice of threshold has been based on ROC analysis on a training dataset with 47 patients, balanced in the same way as the overall dataset across the four categories of apnoea. Four window sizes were investigated: 30-second, 60-second, 90-second, and 120-second - with the threshold values in the range of 0.35 to 1.77 units s.d., in increments of 0.035 units s.d. (Equivalently, the threshold values were in the range of 1% to 5% dips in the SpO_2 signal, with increments of 0.1%.) Said range of threshold values included the 2%, 3%, and 4% dips of the SpO_2 signal used in various sleep studies as reported in the literature. For example, Hilton *et al.* sought to classify 15-minute epochs of SpO_2 signal as OSAHS or non-OSAHS using amplitude criteria, or the number of 2%, 3% and 4% dips found in the time domain [47]. The optimal threshold and window size was found using the Bristol Database training dataset, whereby the 120-second window with 0.778 units s.d. threshold yielded the highest Sp and Acc, with only a very marginally reduced value of Se when compared with the optimal threshold for a 90-second window. Specifically, the (Se, Sp, Acc) values were (83.3%, 82.0%, 82.1%) for the training dataset (47 patients). (Note that there is almost a 10% difference in Se values between 30-second and 120-second analyses, but only a 4% difference in Acc values.) The results of the application of the 120-second window and 0.778 units s.d. threshold to the test dataset (44 patients) were similar; with the overall (Se, Sp, Acc) of (85.3%, 80.5%, 81.2%); i.e., virtually the same accuracy but higher sensitivity and lower specificity.

Improvement of apnoeic event detection may be possible through further analysis of the SpO_2 signals in the frequency domain. In the SpO_2 waveform, each of the two possible types of signals have different characteristics in the frequency domain: normal breathing occurs at "high-frequency," and apnoea occurs at "very low-frequency." The advantage of

using the SpO_2 signal in the frequency domain is the presence of a strong spectral peak in very low frequencies corresponding to apnoeas or OSAHS-type behaviour [102, 121]. The differentiation of frequency content for apnoeic and normal breathing events may be less variable than that achieved using apnoea-related amplitude dips in the SpO_2 signal across a wide range of patients. Frequency domain analysis of the SpO_2 signal is presented in the next Chapter.

Chapter 7

SpO_2 Frequency Domain

7.1 Introduction

As introduced in Chapter 3, oxygen saturation levels drop when a person experiences an apnoea (desaturation). Upon resumption of normal breathing, the oxygen saturation level rises back to its baseline value (resaturation). The desaturation-resaturation cycle is typical for patients with OSAHS and is indicative of apnoeic behaviour. Figure 7.1 shows the SpO_2 signal from Bristol patient 21 (moderate example), with five desaturation-resaturation cycles (of the magnified section of the SpO_2 signal) shown in the bottom panel. The period, T, is marked accordingly for one cycle. The AASM Task Force stipulates that an apnoea must be at least 10 seconds in duration; with $T/2 = 10$ sec, $T = 20$ sec or $f = 0.05$ Hz. Note that the minimum cycle length of apnoea usually occurs during nREM sleep [102]. Conversely, the maximum cycle length of apnoea observed in OSAHS is 120 seconds during REM sleep [102]; with $T/2 = 60$ sec, $T = 120$ sec or $f = 0.008$ Hz. Thus, the expected range of frequency for an apnoea is from 0.008 to 0.05 Hz (see Table 7.1). In Figure 7.1, $T = 65$ sec, and this value therefore falls within the range for apnoea according to Table 7.1.

The breathing rate of adults during sleep varies from 6 breaths per minute ($f = 6/60$ Hz) to 24 breaths per minute ($f = 24/60$ Hz) - or 0.1 to 0.4 Hz (see Table 7.1). It should

therefore be possible to differentiate apnoeic from normal breathing activity automatically from the SpO_2 signal in the frequency domain by finding the frequency corresponding to the peak power of each segment. In this Chapter, we describe the signal processing of the SpO_2 signal including the use of the Discrete-Fourier Transform (DFT) to estimate the power spectral density (PSD) of each segment, and identify the frequencies corresponding to apnoeic and normal breathing activity.

	Apnoea	
T/2:	10 sec	60 sec
T:	20 sec	120 sec
f:	0.05 Hz	0.008 Hz
	Normal Breathing	
f:	6 breaths/min	24 breaths/min
f:	6/60 Hz	24/60 Hz
f:	0.1 Hz	0.4 Hz

Table 7.1: Expected period and frequency ranges for apnoeic and normal breathing activity.

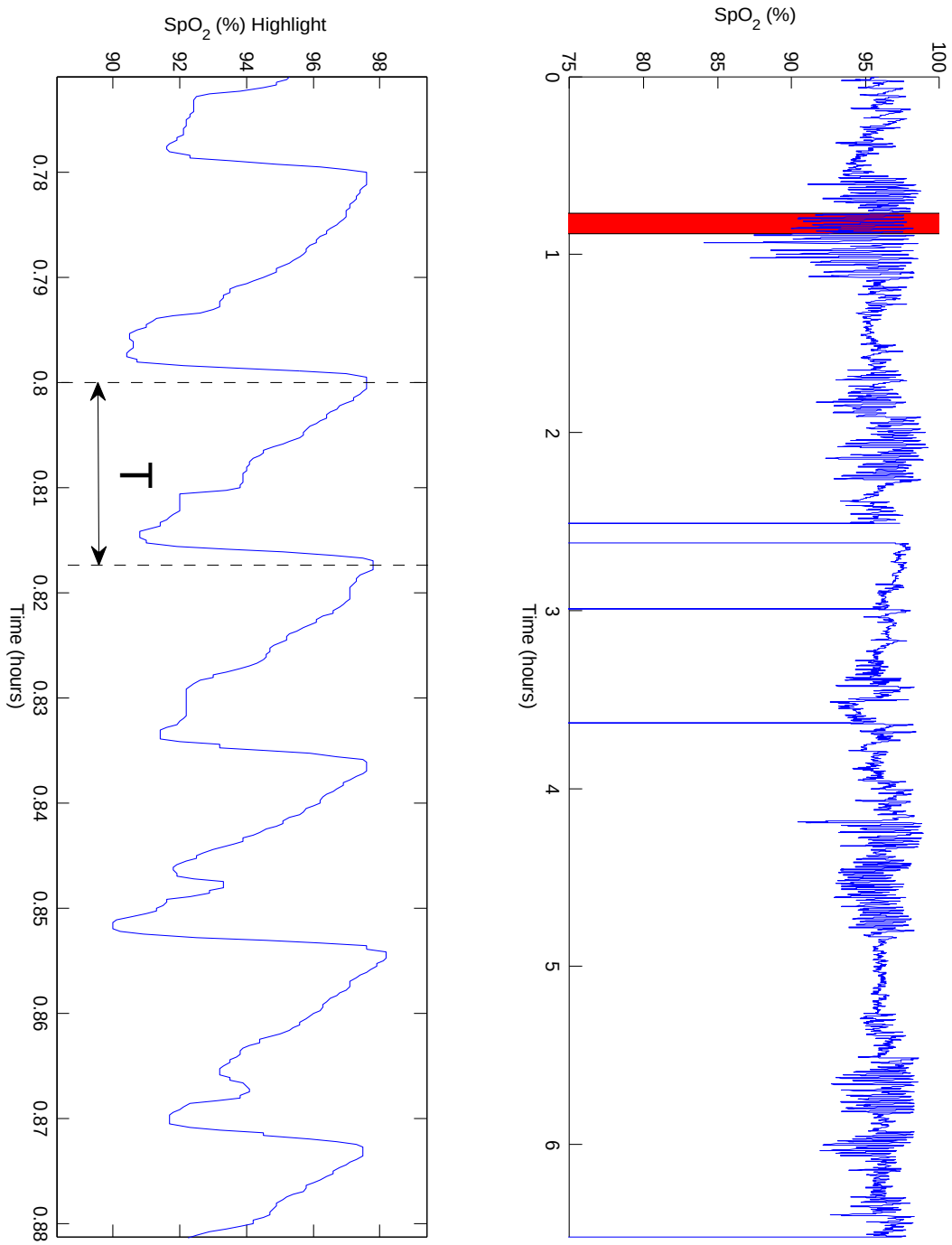


Figure 7.1: Bristol 21: MODERATE patient example. From top to bottom: SpO_2 signal in the time domain; highlighted apnoeic portion with period, T, indicated between peaks marked by the dotted lines.

7.2 Bristol Database training dataset

The MIT BIH PSG training database initially investigated at the beginning of Chapter 6 consists of four patients who suffer from *severe* OSAHS; whereas, the Bristol Database later investigated in that Chapter includes 91 patients with varying degrees of apnoeic severity. The computation of a threshold, which best differentiates between apnoea and normal breathing, determined using a database of just four patients with an extreme version of the condition is likely to be sub-optimal for the general population of OSAHS patients. For this reason, analysis of the SpO_2 signal shall concentrate exclusively on the data from the Bristol Database. This was the approach adopted for time-domain analysis in Chapter 6, and it will be the approach for all subsequent analyses, including analysis in the frequency domain in this Chapter. In the following section, pre-processing of the SpO_2 signals is described for the Bristol Database training dataset only.

7.2.1 Pre-processing

The expected range of period for an apnoea is from 20 to 120 seconds (see Table 7.1); therefore, four window sizes are explored as in the previous Chapter: 30-second, 60-second, 90-second, and 120-second. In each case, the interval between consecutive windows is 30 seconds to maximise the number of windows for analysis. Each windowed SpO_2 signal is detrended, and a Hamming window is applied to reduce spectral leakage. The power spectral density (PSD) of each window is then computed. (The goal of spectral estimation is to describe the distribution (over frequency) of the power contained in a signal based on a finite set of data [73].) For a windowed time series, such as the SpO_2 signal used in this Chapter, the power spectral density (PSD) is typically estimated using the Discrete Time-Fourier Transform (DTFT) of the samples within the segmented window. The periodogram computes the squared magnitude of the Discrete-Fourier Transform (or DFT) computed per window and divides it by the square of the norm (or Euclidean length) of the data window applied to the signal to ensure that the estimate is asymptotically unbiased;

i.e., as the number of samples increases, the expected value of the periodogram approaches the true PSD [73]. The frequency resolution Δf achieved using the DTFT is based on the sampling frequency f_s and the number of points N used to compute the DFT, or:

$$\Delta f = \frac{f_s}{N} \tag{7.1}$$

For example, for a 30-second window and a sampling frequency of $f_s = 1$ Hz (frequency at which the SpO_2 signal is stored in the Bristol Database), a 30-point DFT will give a frequency resolution of 0.0333 Hz.

The frequency bands for power calculation using the DFT are described in Table 7.2, for all four window sizes. For example, for a 30-second window (and 30-point DFT) the first and second coefficients are utilised to define the frequency range for apnoea: 0.033 to 0.067 Hz. The fourth to thirteenth coefficients are used to define the frequency range for normal breathing: 0.1 to 0.4 Hz. The 120-second window uses coefficients 2 through 7 to define the frequency range for apnoea: 0.008 to 0.05 Hz - which is the closest match to the physiologically based range as described in Table 7.1. For all four window sizes, the frequency range for normal breathing (0.1 to 0.4 Hz) matches the physiologically based range - see Table 7.1. Note the trade-off between window size and spectral resolution; i.e., the longer window provides higher resolution (defined as the ability to discriminate spectral features [73]).

Window Size	Coefficients for Apnoea	f Range for Apnoea (Hz)	Coefficient for Normal Breathing	f Range for Normal Breathing (Hz)
30-second	1,2	0.033-0.667	4-13	0.1-0.4
60-second	1,2,3,4	0.017-0.05	7-25	0.1-0.4
90-second	1,2,3,4,5,6	0.011-0.056	10-37	0.1-0.4
120-second	2,3,4,5,6,7	0.008-0.05	13-49	0.1-0.4

Table 7.2: Selected N-point DFT, coefficients utilised for the definition of apnoeic and normal breathing frequency ranges, and corresponding frequency ranges in Hz for the selected window sizes: 30-second, 60-second, 90-second, 120-second.

Figure 7.2 illustrates the PSD values for two typical 120-second segments with ap-

noeic and normal breathing activity respectively, from the SpO_2 signal of Bristol patient 21 (moderate example). The apnoeic segment of the SpO_2 signal is highlighted in red. Almost two complete desaturation-resaturation cycles are visible as well as an 8% difference between maximum and minimum SpO_2 percentage values (in the time domain). The corresponding PSD plot shows a peak at 26.5 dB corresponding to a frequency of $f = 0.017$ Hz, which is within the expected apnoeic frequency range (see Table 7.1). The normal breathing segment of the SpO_2 signal is highlighted in green and maintains a consistent value of 96% SpO_2 throughout the 120 seconds. Also, the peak within the expected apnoeic frequency range is at 3.6 dB; i.e., 22.9 dB below the peak for the apnoeic segments indicating that the amplitude of oscillations in the SpO_2 signal is 200 times less than for the apnoeic segment (as can be verified by inspection of Figure 7.2).

Expert scores

As introduced in Section 6.3.1, for a 30-second window, at least 10 seconds must be apnoeic. The minimum apnoeic requirement per window should change accordingly. For the 60-second, 90-second, and 120-second window sizes, the minimum apnoea requirement is therefore taken to be 20 seconds, 30 seconds, and 40 seconds of total apnoea time scored by the expert for that window to be classified as "apnoeic."

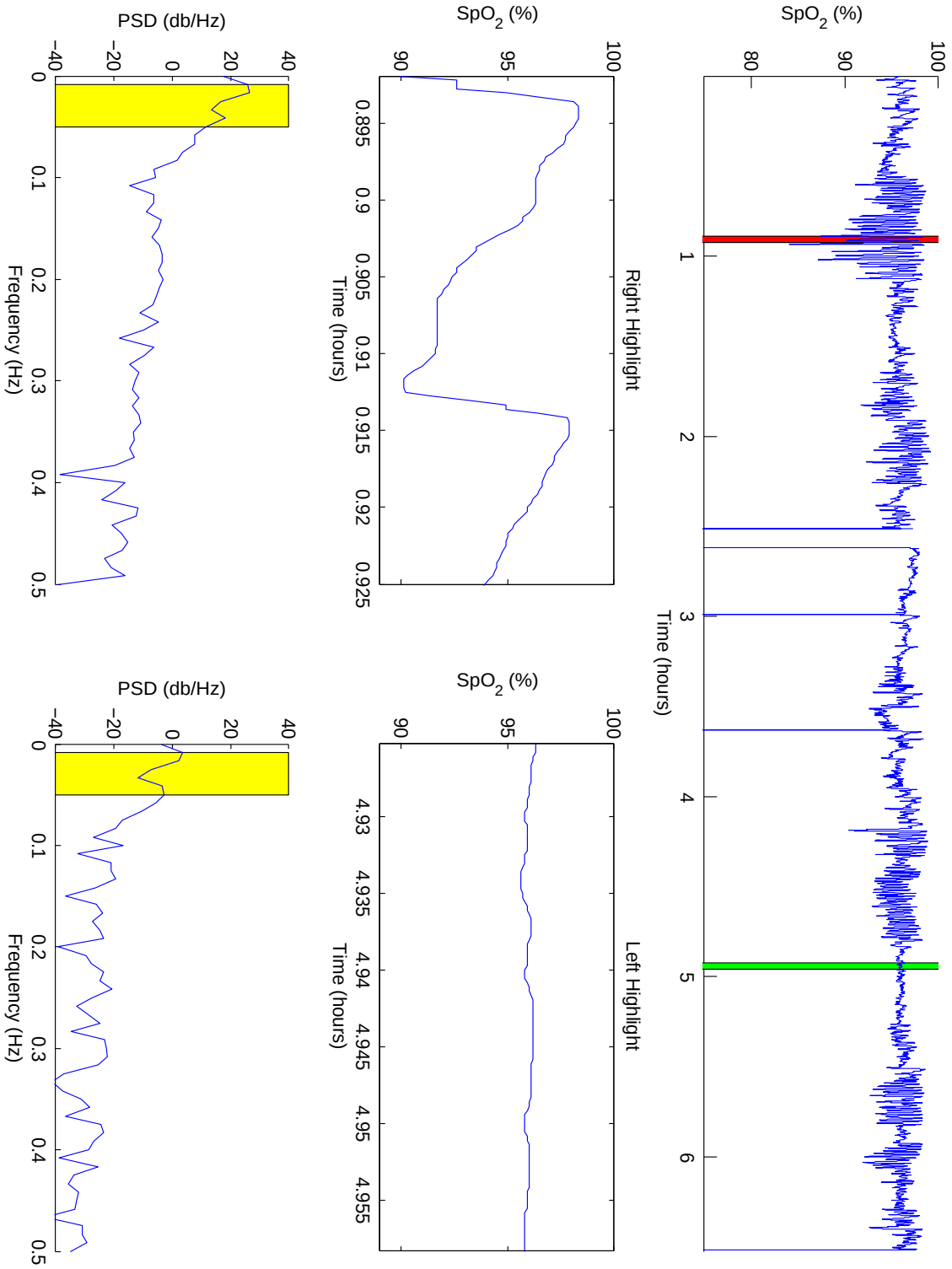


Figure 7.2: Bristol 21: MODERATE patient example. From top to bottom: SpO_2 signal (%) in the time-domain, with highlighted examples of apnoea (red) and normal breathing (green); (left) 120-second apnoeic segment of the SpO_2 signal magnified and (right) 120-second normal breathing segment of the SpO_2 signal magnified; corresponding PSD curves for each magnified segment.

7.2.2 Selection of an appropriate threshold

Following segmentation, detrending, windowing (Hamming) and computation of the PSD for each window, the average power is computed within the two regions of interest: apnoea ($P_{ave,apnoea}$) and normal breathing ($P_{ave,breath}$). As shown in Table 7.2, there are different frequency ranges defined for apnoeic activity according to the DFT used. (The frequency range defined for normal breathing activity is 0.1 to 0.4 Hz for all four DFT computations utilised.) For example, the frequency range for apnoeic activity using a 90-point DFT is 0.011 to 0.056 Hz, using coefficients one through six. The average power within this frequency range is therefore computed by summing the power at each coefficient (bins one through six) and multiplying the result by the bin width. Average power is equivalent to the area under the PSD curve. In addition to the computation of total power within each frequency range of interest (for each chosen N-point DFT), a ratio of power values is computed as follows:

$$PowerRatio = \frac{P_{ave,apnoea}}{P_{ave,breath}} \tag{7.2}$$

The advantage of using the ratio of Equation 7.2, is such that the metric is *independent* of any gain present in the equipment used to acquire the SpO_2 signal. On the other hand, $P_{ave,apnoea}$ and $P_{ave,breath}$ are dependent upon the gain of the equipment; i.e., only the same equipment - as in this study - may be used if these metrics are to be used further.

Two metrics will be investigated: $P_{ave,apnoea}$ and the $PowerRatio$ of Equation 7.2. An event (apnoea, normal breathing) is identified by applying an appropriate threshold to the power values, such that values above the threshold correspond to regions of apnoea (and are given a score of 1) and values below correspond to regions of normal breathing (and are are given a score of 0). As in Section 6.3.1, the Receiver Operating Characteristic (ROC) curve is used for identifying the *optimal* threshold for the training dataset for each window size. For each subject, a range of threshold values is investigated from 0.01 to 1.0, with increments of 0.01 ($P_{ave,apnoea}$) and from 1 to 100, with increments of 1.0 ($PowerRatio$).

(These ranges are selected empirically, based on the observed values of $P_{ave,apnoea}$ and $PowerRatio$ for all four categories of patients: normal, mild, moderate, severe.) As in Chapter 6, the threshold value which yields the highest classification accuracy is optimal for that training set. The window size (with corresponding optimal threshold value) which yields the highest overall accuracy in that training set is then selected for application to the test dataset.

7.2.3 Classification results on the training dataset

The classification sensitivity, specificity and accuracy found using the four segmentation windows (30-second, 60-second, 90-second, 120-second) and the optimal-threshold values determined from the ROC curves for the 47-patient training dataset are provided in Table 7.3. The optimal threshold values range from 0.12 to 0.51 units ($P_{ave,apnoea}$) and from 18 to 91 units ($PowerRatio$). Figures 7.3 and 7.4 show the ROC curve computed for each window size for the $P_{ave,apnoea}$ and $PowerRatio$ metrics respectively, computed for this training dataset.

For $P_{ave,apnoea}$, as the window size gets larger from 30 to 120 seconds, both the Se (number of detected True Positives, or apnoeic segments) and Sp (number of detected True Negatives, or normal breathing segments) values increase. The lowest (Se, Sp) pairing is for the 30-second window at (71.4%, 77.7%), followed by (79.0%, 79.2%) for the 60-second window. The 30-second window utilises a 30-point DFT; whereby the first and second coefficients are used to define the frequency range for apnoeic activity: 0.033 to 0.667 Hz. This range does not include the lowest expected frequency for apnoeic activity (0.008 Hz), and exceeds the highest expected frequency for apnoeic activity (0.05 Hz). The 60-second window utilises a 60-point DFT, with an apnoeic frequency range of interest: 0.017 to 0.05 Hz. Again, this range does not include the lowest expected frequency (0.008 Hz) but does match the upper range of expected activity (0.05 Hz). These resultant sensitivity and specificity pairs reported reflect this accordingly. The highest (Se, Sp) pairing is for the 120-second window at (82.0%, 83.1%), and followed closely by the 90-second window at (83.6%, 80.0%). The 90-second window uses a 90-point DFT, and a corresponding frequency range for apnoeic activity more close to that expected physiologically: 0.011 to 0.056 Hz; whereas, the 120-second window and 120-point DFT yield a frequency range for apnoeic activity *exactly* as desired: 0.008 to 0.05 Hz. Overall, it may be concluded that the larger window sizes are optimal as the frequency resolution is high enough to capture the lower frequencies corresponding to apnoeic activity. Specifically, the use of the 120-second window, and 0.51 unit threshold, yields the highest classification accuracy

of 83.0% - see Table 7.3.

For *PowerRatio*, also as described in Table 7.3, there is no particular pattern in sensitivity, specificity and accuracy as the window size gets larger from 30 to 120 seconds. The results for each window/DFT size are comparable to each other, all falling within the range of 62.0 to 68.5%. For example, for the 60-second and 90-second window, the (Se, Sp) pairs are respectively (68.5%, 64.5%) and (65.6%, 67.6%). The lowest and highest accuracy values are 62.1% (30-second window) and 67.4% (90-second window). The results using the *PowerRatio* metric are worse than those results found using the $P_{ave,apnoea}$ metric. We find that although there is a large difference between $P_{ave,apnoea}$ in apnoeic and normal breathing segments, as previously shown by Figure 7.2, the $P_{ave,breath}$ is variable; i.e., $P_{ave,breath}$ is not always higher in normal breathing than apnoeic segments. Due to this variability, the *PowerRatio* metric does not perform as well as $P_{ave,apnoea}$; thus, any subsequent frequency domain analysis will be restricted to the use of the $P_{ave,apnoea}$ metric only.

Window Length	Optimal-Threshold (units)	Sensitivity (%)	Specificity (%)	Accuracy (%)
$P_{ave,apnoea}$				
30 sec	0.12	71.4	77.7	76.9
60 sec	0.30	79.0	79.2	79.1
90 sec	0.40	83.6	80.0	80.4
120 sec	0.51	82.0	83.1	83.0
PowerRatio				
30 sec	18	63.1	62.0	62.1
60 sec	60	68.5	64.5	64.9
90 sec	88	65.6	67.6	67.4
120 sec	91	66.7	66.2	66.2

Table 7.3: Frequency domain analysis resultant Se, Sp, and Acc percentages across the 47-patient Bristol Database training dataset for 30-second, 60-second, 90-second, 120-second window lengths, and corresponding optimal (ROC-based) threshold values.

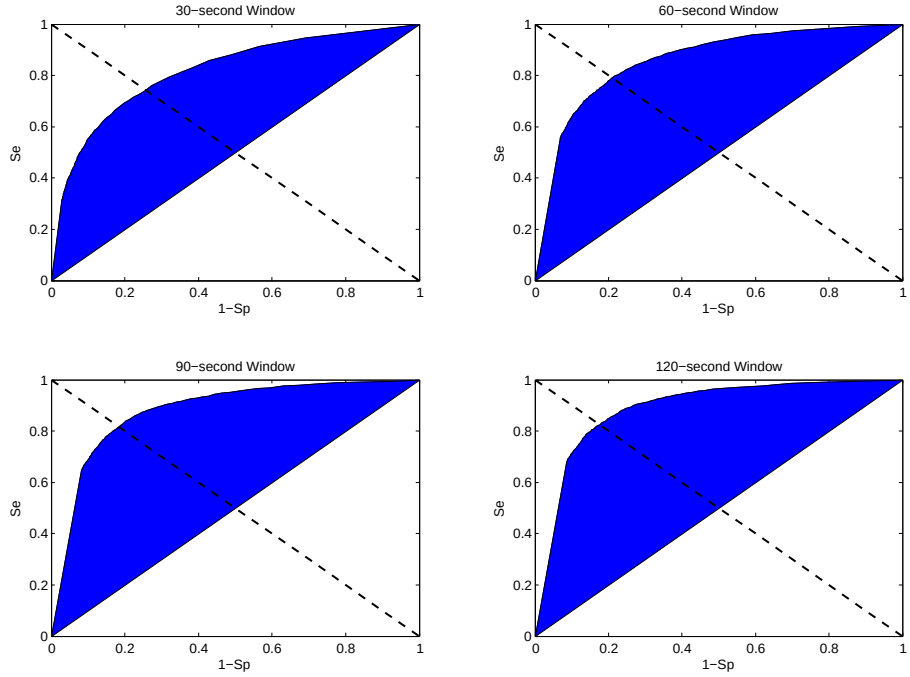


Figure 7.3: ROC curves for the Bristol Database training dataset using frequency domain analysis ($P_{ave,apnoea}$) and four window sizes: 30-second, 60-second, 90-second, 120-second. Maximum classification accuracy occurs at the point on the ROC curve closest to (0,1).

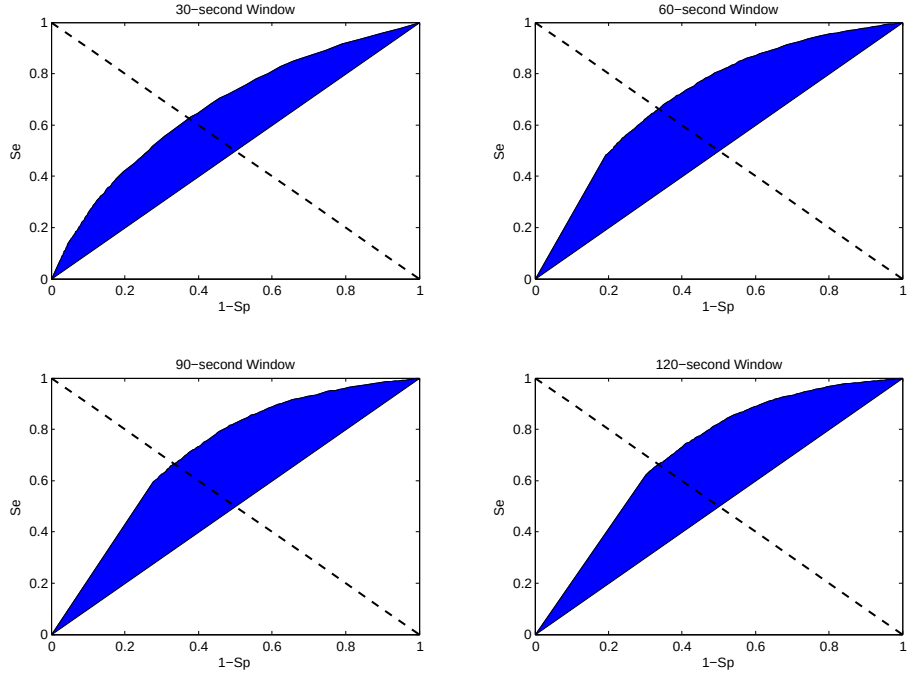


Figure 7.4: ROC curves for the Bristol Database training dataset using frequency domain analysis ($PowerRatio$) and four window sizes: 30-second, 60-second, 90-second, 120-second. Maximum classification accuracy occurs at the point on the ROC curve closest to (0,1).

7.3 Bristol Database test dataset

Table 7.4 shows the overall (Se, Sp, Acc) values computed for the 44-patient test dataset using a 120-second window and optimal threshold of 0.51 units using the $P_{ave,apnoea}$ metric: 84.9%, 81.7%, and 82.1%. In addition, the (Se, Sp, Acc) values are presented according to normal, mild, moderate, and severe patients within the dataset. The highest Se value reported is for the severe category missing very few apnoeas (96.3%), followed by the normal category (82.9%), moderate category (71.6%), and finally the mild category (45.2%). The severe category also exhibits the lowest Sp value (51.3%); however, for such patients, consistent periods of normal breathing are rare and therefore a low Sp value is not significant. Conversely, specificity values are more significant for normal and mild categories, whereby stable breathing dominates the SpO_2 signal and there are few apnoeas to detect. The highest Sp value reported is for the normal category (88.7%), followed by the mild category (84.4%), and lastly the moderate category (78.4%).

As presented in Chapter 6 (see Section 6.3.3) for the automated analysis of patients for whom apnoeic activity dominates (moderate, severe), a high sensitivity is preferable (as apnoea is the "event" to be detected). Conversely, for the automated analysis of patients where normal breathing activity dominates (normal, mild), a high specificity is required. It is therefore worthwhile to consider sensitivity, specificity and accuracy with respect to the severity of the patient condition. The confusion matrices, Tables 7.5 through 7.8, illustrate the results for the 120-second window analysis with the threshold of 0.51 units for the $P_{ave,apnoea}$ metric. The percentage of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) with respect to the total number of segments under analysis is also given.

For the normal category (Table 7.5) there are 35 out of 11348 segments identified as apnoeic, or 0.3%. The number of correctly identified segments is 29 out of 35, or 82.9% (i.e., sensitivity - see Table 7.4). For the mild category (Table 7.6) there are 305 out of 8891 segments identified as apnoeic, or 3.4%. The number of correctly identified segments

is 138 out of 305, or 45.2%. For the moderate category (Table 7.7) there are 1525 out of 8375 segments identified as apnoeic, or 18.2%. The number of correctly identified segments is 1092 out of 1525, or 71.6%. Lastly, for the severe category (Table 7.8) there are 2849 out of 5498 segments identified as apnoeic, or 51.8%. The number of correctly identified segments is 2743 out of 2849, or 96.3%. The percentage of apnoeic activity increases with the severity of OSAHS: 0.3% (normal), 3.4% (mild), 18.2% (moderate), 51.8% (severe). Similarly to the results described in Section 6.3.3, the percentage of normal breathing segments identified as apnoea (FP) are 11.2%, 15.0%, 17.7%, and 23.5% for the normal, mild, moderate and severe categories, respectively; i.e., the automated system of analysis is over-sensitive. However, it should not affect the categorisation of the patient by the automated system as will be discussed in Chapter 9.

Figures 7.5 through 7.8 illustrate the frequency domain analysis results for the SpO_2 signals of four typical patients, the same as those Bristol patients utilised in Chapter 6: 11 (normal), 8 (mild), 21 (moderate), 57 (severe). In all four Figures, a 120-second window is utilised along with the optimal threshold of 0.51 units. For example, the top panel in Figure 7.5 shows the SpO_2 signal for the entire night's sleep of 'normal' patient 11. The bottom panel shows the corresponding $P_{ave,apnoea}$ (computed per window) in blue, and colour-coded according to the Expert Score attached to each window (red corresponds to apnoea, green corresponds to normal breathing). The superimposed black dotted line shown here is the optimal threshold. The most straightforward example of the $P_{ave,apnoea}$ exhibiting large values when there is apnoeic activity, and small values otherwise, is in Figure 7.7 - for the moderate class. However, notice the large $P_{ave,apnoea}$ values around the second and sixth hour of the SpO_2 signal, for which the experts have scored the majority of segments as normal breathing, whereas the automated analysis has assessed most of the segments being as apnoeic (i.e., the segments fall above the threshold). Similarly, the severe class example illustrates very large $P_{ave,apnoea}$ values throughout, even when the experts have scored the segments as normal breathing. (These results are similar to those of Chapter 6 (see Section 6.3.3); however, further comparison and discussion are reserved

for Chapter 8.) From the four examples, it is clear that the $P_{ave,apnoea}$ of the SpO_2 signal increases as the severity of patient OSAHS varies from normal to severe.

Categorisation	Sensitivity (%)	Specificity (%)	Accuracy (%)
OVERALL	84.9	81.7	82.1
Normal	82.9	88.7	88.7
Mild	45.2	84.4	83.1
Moderate	71.6	78.4	77.2
Severe	96.3	51.3	74.6

Table 7.4: Frequency domain analysis resultant Se, Sp, and Acc percentages across the 44-patient Bristol Database testing dataset, on an overall and *per category* basis, for 120-second window length and 0.51 units threshold.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	29 (0.3%)	6 (0.1%)
	Normal Breathing	1276 (11.2%)	10037 (88.4%)

Table 7.5: NORMAL patients' confusion matrix (frequency domain analysis): 120-second window, 0.51 units threshold. A total of 11348 segments are analysed, with the relative percentages shown in brackets.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	138 (1.6%)	167 (1.9%)
	Normal Breathing	1338 (15.0%)	7248 (81.5%)

Table 7.6: MILD patients' confusion matrix (frequency domain analysis): 120-second window, 0.51 units threshold. A total of 8891 segments are analysed, with the relative percentages shown in brackets.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	1092 (13.0%)	433 (5.2%)
	Normal Breathing	1479 (17.7%)	5371 (64.1%)

Table 7.7: MODERATE patients' confusion matrix (frequency domain analysis): 120-second window, 0.51 units threshold. A total of 8375 segments are analysed, with the relative percentages shown in brackets.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	2743 (49.9%)	106 (1.9%)
	Normal Breathing	1291 (23.5%)	1358 (24.7%)

Table 7.8: SEVERE patients' confusion matrix (frequency domain analysis): 120-second window, 0.51 units threshold. A total of 5498 segments are analysed, with the relative percentages shown in brackets.

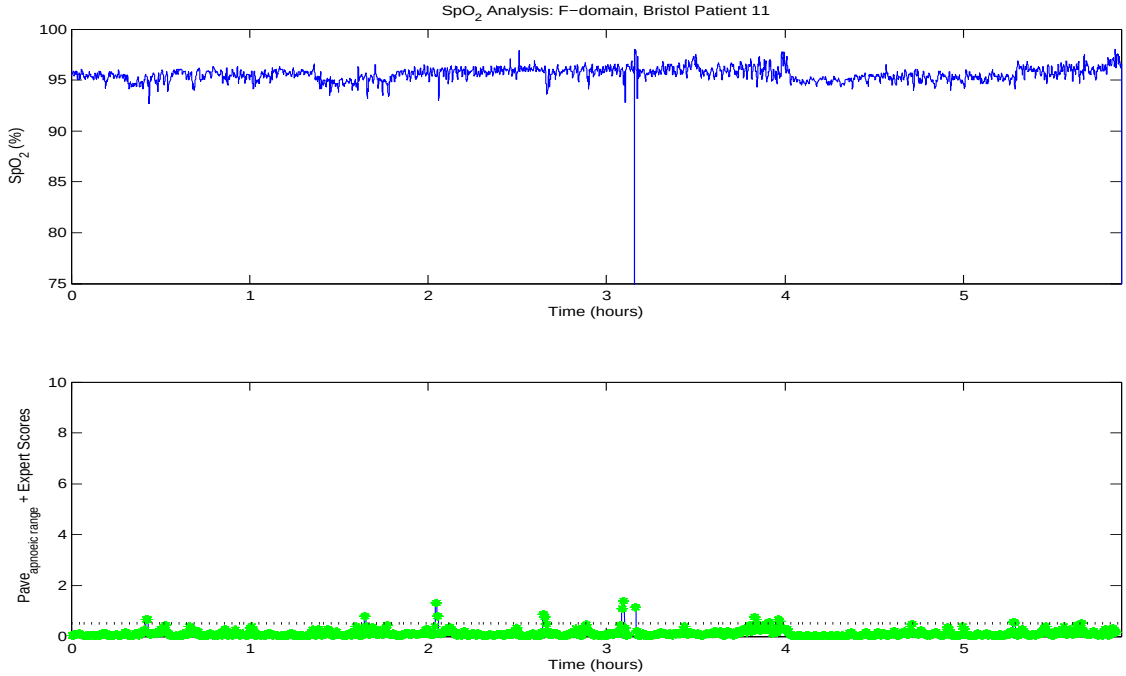


Figure 7.5: Bristol 11: NORMAL patient example. From top to bottom: SpO_2 signal (%); $P_{ave,apnoea}$ per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with threshold shown as dotted line. A 120-second window is utilised with 0.51 units threshold.

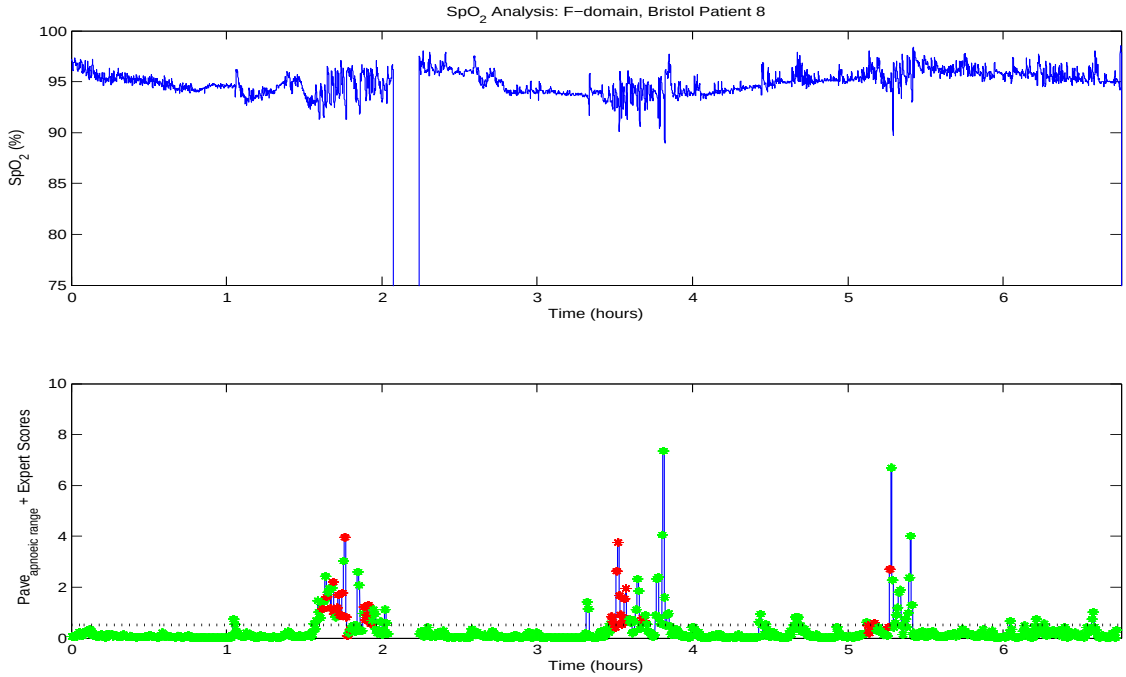


Figure 7.6: Bristol 8: MILD patient example. From top to bottom: SpO_2 signal (%); $P_{ave,apnoea}$ per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with threshold shown as dotted line. A 120-second window is utilised with 0.51 units threshold.

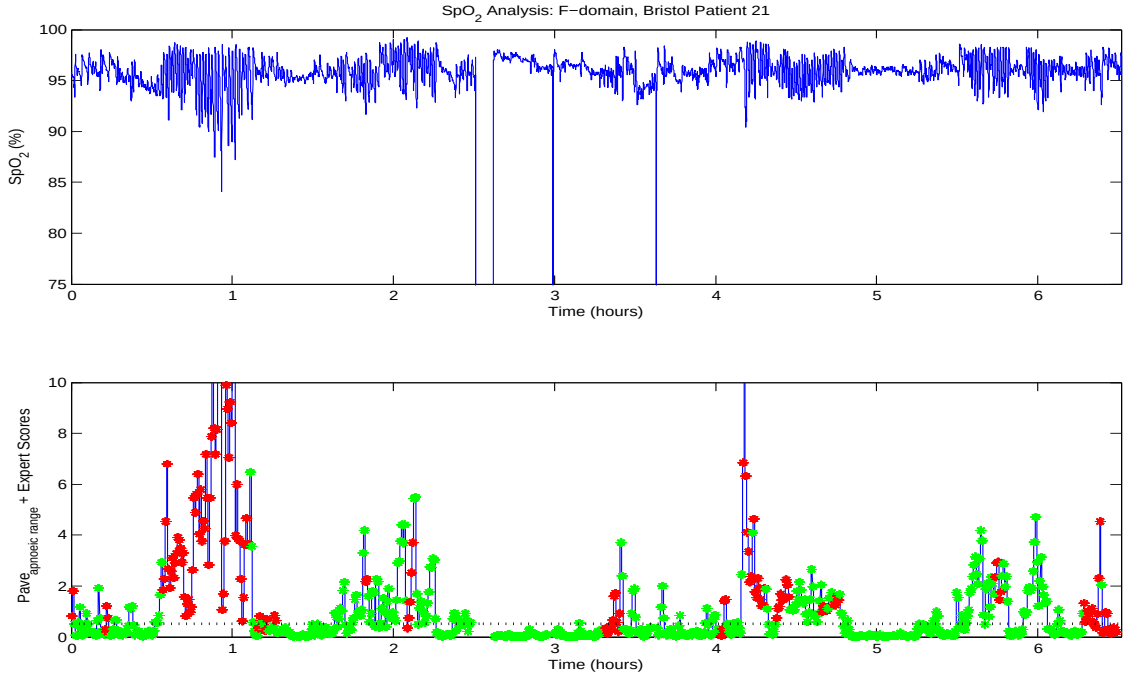


Figure 7.7: Bristol 21: MODERATE patient example. From top to bottom: SpO_2 signal (%); $P_{ave,apnoea}$ per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with threshold shown as dotted line. A 120-second window is utilised with 0.51 units threshold.

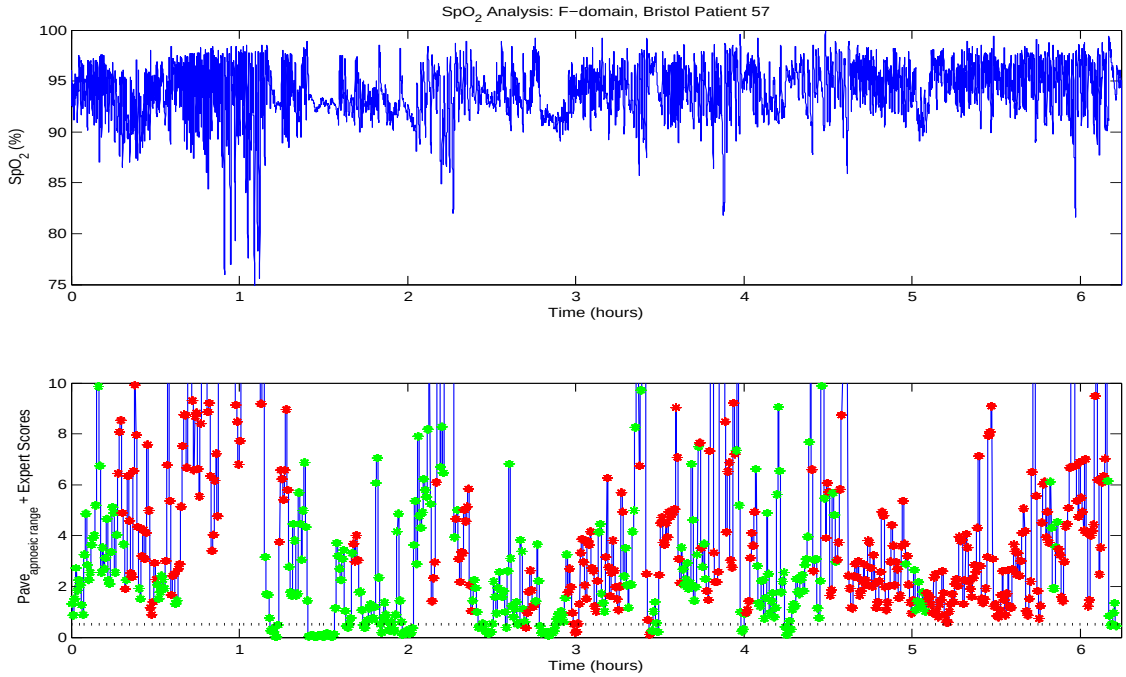


Figure 7.8: Bristol 57: SEVERE patient example. From top to bottom: SpO_2 signal (%); $P_{ave,apnoea}$ per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing, with threshold shown as dotted line. A 120-second window is utilised with 0.51 units threshold.

7.4 Discussion and summary

In this Chapter, the automated analysis of the SpO_2 signal has been based on the computation of the average power for the apnoeic and normal breathing frequency ranges of interest. Two power metrics were computed: $P_{ave,apnoea}$ and $PowerRatio$, or the average power within the apnoeic frequency range and the ratio of the average power within the apnoeic frequency range with respect to the average power within the normal breathing frequency range. The choice of the optimal classification threshold was based on ROC analysis of a training dataset with 47 patients, balanced in the same way as the overall dataset across the four categories of apnoea. Four window sizes were investigated: 30-second, 60-second, 90-second, and 120-second.

The analysis of the training dataset provided sensitivity, specificity and accuracy values for all four window sizes using the two computed power metrics. For all four window sizes, the computed (Se, Sp, Acc) values were higher for the $P_{ave,apnoea}$ than for the $PowerRatio$ metric; i.e., the difference in (Se, Sp, Acc) percentages were as follows: (8.3%, 15.7%, 14.8%) for the normal category, (10.5%, 14.7%, 14.2%) for the mild category, (18.0%, 12.4%, 13.0%) for the moderate category, and (15.3%, 16.9%, 16.8%) for the severe category. Hence, the detailed analysis was carried out with the former metric only. The $P_{ave,apnoea}$ optimal window size and threshold were found to be 120 seconds and 0.51 units, yielding an (Se, Sp, Acc) of (82.0%, 83.1%, 83.0%) on the training dataset (47 patients). The results of the application of this window and threshold to the test dataset (44 patients) were similar; with the overall (Se, Sp, Acc) of (84.9%, 81.7%, 82.1%); i.e., virtually the same accuracy but higher sensitivity and lower specificity. Thus, as with the time-domain analysis of Chapter 6, the characteristics of the test dataset in the frequency domain must be very similar to those of the training dataset, providing a post hoc validation of the division into the two datasets. The per-category accuracies were computed as follows: 88.7% (normal), 83.1% (mild), 77.2% (moderate), 74.6% (severe). It is not surprising that the 120-second point DFT yielded the highest (Se, Sp, Acc) as

the resultant spectral resolution matched the desired (physiologically-based) frequency resolution exactly.

In recent sleep literature, emphasis has been placed on the strong presence of a spectral peak within the apnoeic frequency range of interest, and its potential for OSAHS diagnosis in patients with the suspected disorder [70, 102, 121]. However, no attempts have been made to identify *individual* apnoeic episodes during the night and to compare the results of automated analysis with expert scores on a time resolution of minutes, as opposed to 30 minutes [102] or an entire night [70, 121]. The results of frequency domain analysis are best discussed with respect to the prior analysis of the SpO_2 signal in the time domain in the previous Chapter. Both time and frequency domain analyses yield comparable results, with overall (Se, Sp, Acc) of (85.3%, 80.5%, 81.2%) and (84.9%, 81.7%, 82.1%), respectively. (Time domain analysis, however, is not as dependent upon window size as frequency domain analysis.) It is not possible to combine these two methods such that new information would be extracted; i.e., the peak in the frequency spectrum is related to the amplitude in the time domain. The question arises: is there any *independent* information in the frequency domain which could be combined to advantage with amplitude information in the time domain? Indeed, a solution exists which is described at length in the following Chapter.

Chapter 8

AR-modelling of the SpO_2 Signal

8.1 Introduction

As introduced in Chapter 7, the question of whether or not independent information in the frequency domain can be combined to advantage with amplitude information in the time domain is now addressed. The location of the main frequencies in the SpO_2 signal **can indeed** be used to advantage; i.e., normal breathing activity occurs at much higher frequencies than apnoeic activity. The extrication of the location of dominant frequencies in the SpO_2 signal may be achieved through the use of the pole-zero plot instead of the Discrete-Fourier Transform (DFT) for three reasons. First, the DFT coefficients are harmonically related, whereas the pole-zero method has no such constraint. Second, the DFT is not designed to look for peaks specifically. Every frequency within the pass-band has equal weighting. For example, in Figure 7.2, the spectral peaks are not well defined, especially in the normal breathing frequency range of interest (0.1 to 0.4 Hz). Third, there are no frequency resolution issues in the pole-zero plot; i.e., poles can be placed at any angle within the unit circle. For these three reasons, the DFT is less useful in the identification of dominant frequencies and the pole-zero method is employed. In this Chapter, the pole-zero method is described which includes autoregressive modelling and dominant pole tracking.

8.2 Tracking resonant frequencies

For non-adaptive modelling, a best-fit model may only be computed if the signal is stationary; i.e., signal characteristics - such as mean and frequency content - must not vary with time (see Appendix C). In our study, the SpO_2 signals are non-stationary, but may be segmented into stationary epochs. The non-adaptive AR-model is used to analyse these *piecewise stationary signals* using the Burg algorithm, which is relatively simple and reliable (see Appendix C) [82].

8.2.1 Pre-processing the SpO_2 signal

To start, each Bristol Database SpO_2 signal sampled at a frequency $f_s = 1$ Hz is segmented using a 120-second window - chosen as it was found to be optimal in both Chapters 6 and 7 (see Sections 6.3.3 and 7.2.3). Each segmented window starts 30 seconds after the previous one to maximise the number of windows for analysis. Next, the trend of each windowed signal is removed. The frequency bands of interest - for both apnoeic and normal breathing activity - are defined as in Section 7.1, and provided again in Section 8.3 for convenience.

Following the segmentation, detrending, and windowing of the SpO_2 signal, the autoregressive (AR) model of each consecutive window is computed. From this model, the dominant frequencies within the window are found, by identifying the largest poles in the AR model. (Appendix C explains how the power at the dominant frequencies within a signal can be estimated from the magnitudes of the corresponding poles of its AR-model.) The use of AR-modelling to identify resonant frequencies is best illustrated by example.

8.2.2 Illustration of AR-modelling

A linear combination of two sinusoidal waves corrupted by Gaussian white noise, as defined in Equation 8.1, and shown in Figure 8.1, is used as a synthetic signal. The top panel shows the waveform in the time domain where the sampling interval, Δt, is chosen to

be 1.0 sec (or $f_s = 1$ Hz). The signal $x(t)$ includes two resonant frequencies (one at $\frac{1}{30\,\text{sec}} = 0.033$ Hz and the other at $\frac{1}{6\,\text{sec}} = 0.167$ Hz - the latter with half the amplitude of the first one) and a Gaussian noise term, $e(t)$. Thus:

$$x(t) = 2\sin\left(\frac{2\pi}{30\,\text{sec}}t\right) + \sin\left(\frac{2\pi}{6\,\text{sec}}t\right) + e(t) \tag{8.1}$$

The lower left panel of Figure 8.1 shows a 10^{th}-order AR model of the power spectral density (PSD) from zero to half the sampling frequency (see Appendix C for details of PSD derivation). The PSD (dB/Hz) is estimated via the Burg method with $f_s = 1$ Hz. Two spectral peaks occur at approximately the estimated frequencies 0.033 Hz and 0.165 Hz, marked A and B. The lower right panel of Figure 8.1 shows the poles of the AR model plotted in the complex plane (poles are marked with an 'x' and zeros are marked with an 'o'). As the pole-zero plot is symmetric about the real axis, all complex-conjugate pairs are marked accordingly (e.g., A-A*). The poles marked A and B each have a magnitude close to unity (0.98, 0.95) and phase angles equal to 12.5° and 59.1°. The relationship between pole-zero plot angle and the corresponding resonant frequency is as follows:

$$\Theta = 360^\circ * \frac{f_{res}}{f_s} \tag{8.2}$$

where Θ is the angle sub-tended by the pole, f_{res} is the pole frequency, and f_s is the sampling frequency. Poles A and B, therefore, have resonant frequencies equal to 0.035 Hz and 0.164 Hz which approximate the desired frequencies of 0.033 Hz and 0.167 Hz (or 12° and 60°) from Equation 8.1. As $x(t)$ is known to contain only two resonant frequencies, the three remaining estimated resonant frequencies (marked C, D, E) model the additional Gaussian noise. These poles have magnitude less than unity and with their phase angle evenly distributed around the unit circle (in an attempt to generate an equiripple estimate of a flat PSD [45, 56]). A 10^{th}-order model is selected such that there are enough poles to model the two resonant frequencies *and* additional noise in $x(t)$. Overall, both the PSD and pole-zero plots convey the same information, but in alternative forms.

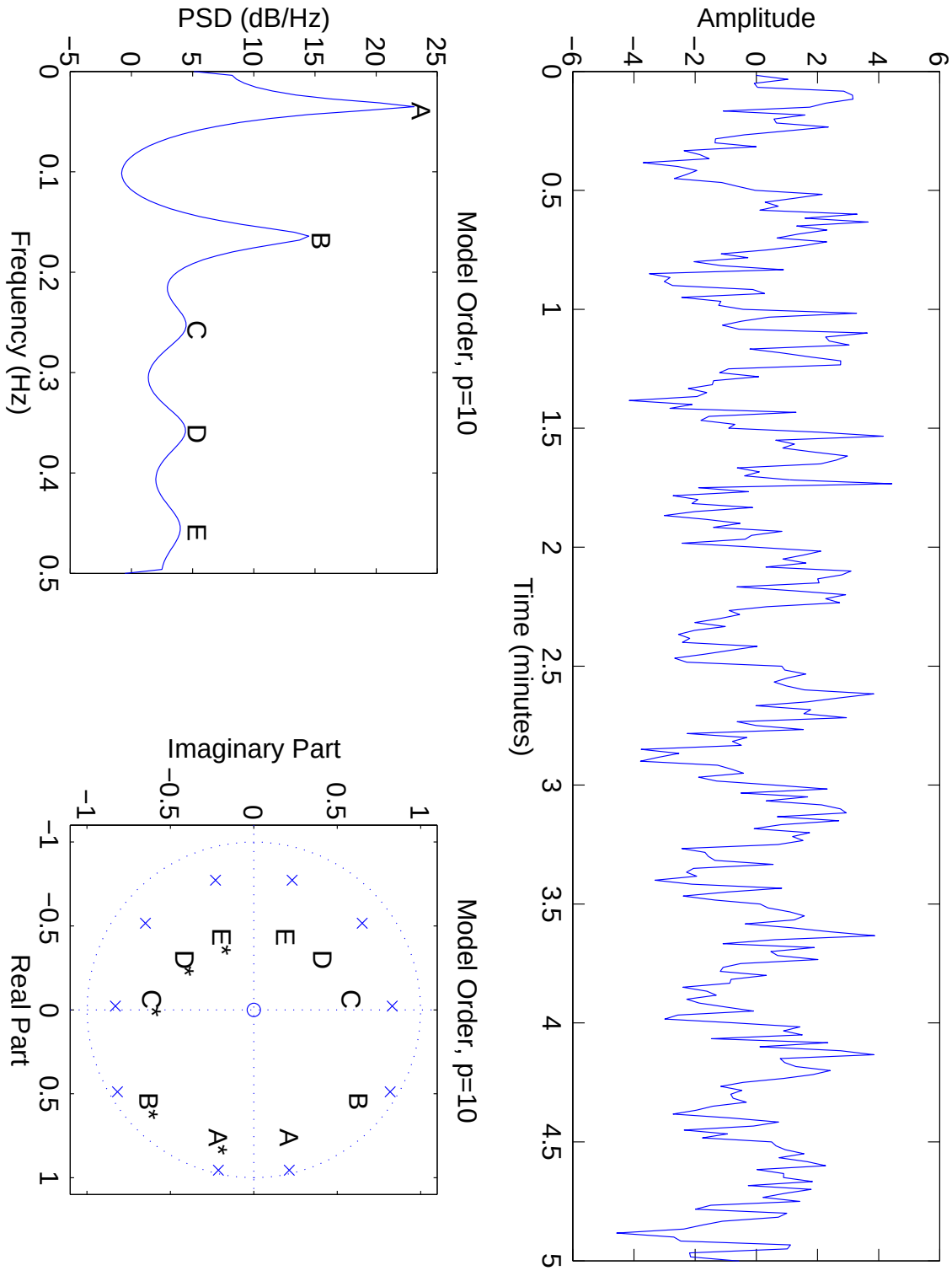


Figure 8.1: From top to bottom: Spectral estimation with an AR-model of noisy input signal - a linear combination of two sine waves corrupted by Gaussian white noise, $1 * e(t)$; 10^{th}-order AR model: (left) The PSD showing two large spectral peaks at resonant frequencies, marked A and B; (right) The pole-zero plot with two poles with magnitude near unity and phase angles corresponding to resonant frequencies, marked A and B.

8.3 Identifying dominant frequencies

We assume that the apnoeic and normal breathing frequencies in the SpO_2 signal may be identifiable from the pole-zero plot for the AR model of each two-minute window. As previously discussed in Chapters 6 and 7, the AASM Task Force stipulates that an apnoea must be at least 10 seconds in duration. Using Equation 8.2, and $f_s = 1$ Hz and $T_{res}/2 = 10$ sec (or $T_{res} = 20$ sec), the minimum computed value is $\Theta = 18°$. Conversely, the maximum cycle length of an apnoea is $T_{res} = 120$ sec, and therefore $\Theta = 3°$ (see Equation 8.2). Thus, the acceptable range of angular position on the pole-zero plot for an apnoea is from 3 to 18°. The breathing rate of adults during sleep ranges from 6 breaths per minute ($f_{res} = 6/60$ Hz) to 24 breaths per minute ($f_{res} = 24/60$ Hz) [68]. The corresponding range of angular position on the pole-zero plot for normal breathing is therefore from 36 to 144°. Table 8.1 illustrates the range of frequency, period, and Θ values for apnoeic and normal breathing activity. (It should be noted that a higher sampling rate would have resulted in poor resolution of Θ values on the unit circle; i.e., the angles would not be discernible from DC.)

	Apnoea	
T/2:	10 sec	60 sec
T:	20 sec	120 sec
f:	0.05 Hz	0.008 Hz
Θ:	18°	3°
	Breath	
f:	6 breaths/min	24 breaths/min
f:	6/60 Hz	24/60 Hz
f:	0.1 Hz	0.4 Hz
Θ:	36°	144°

Table 8.1: Expected period, frequency, and Θ ranges for apnoeic and breath events.

To estimate the value of the two resonant frequencies in $x(t)$, at least four poles are required or two complex-conjugate pairs ($r = 2$, see Appendix C). As detrending is not always perfect, an extra pole is used to model the dc-value. Using the model order $p = 5$ ensures that not more than one pole will normally fall within the acceptable range of

angular position around the origin for apnoea and breath [61]. No more than one pole is needed to model the dc-value of the clean SpO_2 signals; i.e., the SpO_2 signals are noise-free in comparison to that signal described in Section 8.2.2. For example, in Figure 8.2, the magnified 120-second segment of SpO_2 signal shows a typical apnoea or desaturation-resaturation cycle. This signal is much less noisy than that signal of Equation 8.1, shown in Figure 8.1 where three pairs of complex-conjugate poles were used to model the additional noise ($e(t)$). In addition, with a 5^{th}-order model, there are five different sectors around the unit circle where the poles could fall. The decision as to whether our two poles lie within apnoeic and/or normal breathing sectors would be made even more difficult had we used a 10^{th}-order model instead.

Using a 5^{th}-order model, each pole is assigned to one of five sectors around the unit circle: 0 to <3° (not of interest), 3 to 18° (apnoeic region of interest), >18 to <36° (not of interest), 36 to 144° (normal breathing region of interest), and >144 to 180° (not of interest). DC corresponds to 0° or 180° on the unit circle. The algorithm for pole assignment is described in the flowchart in Appendix C, see Figure C.3. However, as an example, we start with the pole with the largest angle - in the >144 to 180° sector (not of interest). (Logically, we search for poles from 180 to 0°, although a search in the converse direction would also be sufficient.) The second pole is either also in this sector, or in any of the other four sectors. For each combination of pole assignment, there are ultimately four cases assigned: 1. apnoea , 2. normal breathing, 3. mixed, and 4. cannot be analysed. There are two subsets to both the apnoea and normal breathing cases, as described below (1.(a),(b) and 2.(a),(b)). Returning to our example, briefly, when the first pole falls within the >144 to 180° sector (not of interest), and so does the second pole, Case 4 is assigned. When the second pole falls within the 36 to 144° sector (normal breathing region of interest), Case 2.(a) is assigned, and so on.

1. **(a)** One pole falls in the region not of interest, and the other pole falls in the apnoeic region of interest.

   (b) Both poles fall in the apnoeic region of interest.

2. **(a)** One pole falls in the region not of interest, and the other pole falls in the normal breathing region of interest.

   (b) Both poles fall in the normal breathing region of interest.

3. One pole falls in the normal breathing region of interest, and the other pole falls in the apnoeic region of interest.

4. No poles fall in either the apnoeic or normal breathing regions of interest; i.e., both poles fall in regions not of interest.

Figure 8.2 shows the poles of the AR-model in the complex plane for one highlighted 120-second (for example) segment of the SpO_2 signal for Bristol patient 21 (the moderate example). A complete desaturation-resaturation cycle (apnoea) is visible in the highlighted segment. Note how Pole 'A' lies within the acceptable angular region for an apnoea (shaded red) with magnitude $r = 0.94$; Pole 'B' lies within the acceptable angular region for normal breathing (shaded green) with magnitude $r = 0.68$; and Pole 'D' lies on $\Theta = 180°$, or DC, with magnitude $r = 0.39$. For this segment, the mixed case is assigned as one pole falls within each region of interest (apnoea, normal breathing). This is by far the most common case. Not surprisingly, in a 120-second window, there is evidence of both apnoea and normal breathing. In fact, the expert scores for this 120-second segment are 40-seconds apnoea (shaded red), 20-seconds normal breathing (shaded green), and 60-seconds apnoea (shaded red) as highlighted on the SpO_2 signal. There will therefore be poles in both sectors, but one will dominate with respect to the other, as indicated by their magnitudes.

The prevalence of the mixed case is clear from the four examples from the test database, as used in Chapters 6 and 7; i.e., Bristol patients 11 (normal), 8 (mild), 21 (moderate), and 57 (severe). As can be seen from Figures 8.3 through 8.6, the mixed case, which is coded in blue, is by far the most common. The green asterisks (at +1) correspond to the normal breathing case; the blue asterisks (at 0) correspond to the mixed case; and the red asterisks (at -1) correspond to the apnoeic case. For the 44-patient training dataset,

4.5%, 16.4%, 75.4%, and 3.7% of the data segments are apnoea (Case 1), breath (Case 2), mixed (Case 3), and not of interest/rare (Case 4).

Using the 44-patient training dataset, ROC analysis will be employed to find an optimal threshold on the value of the ratios that bests divides the mixed cases into the two classes (apnoea, normal breathing). However, the radius r of each pole depends on the relative amplitude of apnoeic and normal breathing components (the frequencies of which are given by the pole-angles). In Chapter 6, time domain analysis of the SpO_2 signal, ROC analysis was already used on the 44-patient training dataset to determine that a 0.778 units s.d. threshold gave the best discrimination between the two classes - see Section 6.3.3. This amplitude criterion may therefore by applied here to discriminate between apnoeic and normal breathing activity for the mixed class on the basis of amplitude. Thus, for the mixed segments, the standard deviation is computed and compared to the 0.778 units s.d. threshold. If the standard deviation falls above the threshold, the segment is identified as **dominant apnoea**; and otherwise, the segment is identified as **dominant normal breathing**. Hence, the 'blue segments' are resolved into 'red' and 'green' segments. For example, in the mixed case segment illustrated in Figure 8.2, the computed standard deviation is 2.44 units. Therefore, in comparison with the 0.778 units s.d. threshold, this segment would be resolved as apnoeic. With a window size of 120 seconds, the amplitude threshold of 0.778 units s.d., the pole-zero method is applied to the 47-patient test dataset and compared with the Expert Scores, which are also segmented into 120-second windows. As before, a 40-second minimum of total apnoea time scored by the experts is required for that window to be classified as apnoeic - see Section 6.3.1. Also, a 30-second offset is used for Expert Score segmentation, and only non-artefact regions of the processed signals are displayed.

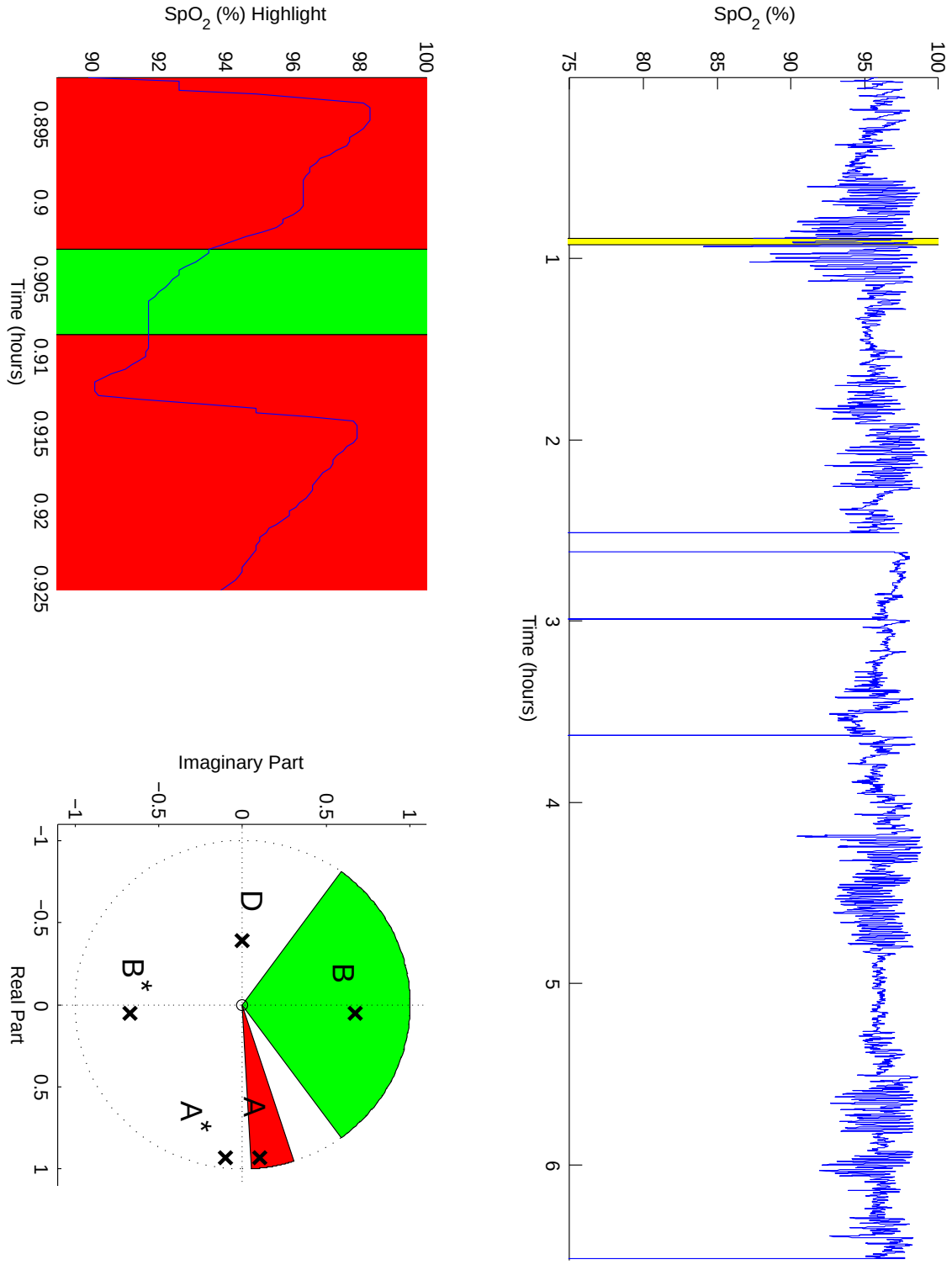


Figure 8.2: Bristol 21: MODERATE patient example highlighted. From top to bottom: SpO_2 signal in its entirety, with one 120-second segment highlighted in yellow; (left) Highlighted 120-second segment magnified with Expert Scores assigned (apnoea shaded red, normal breathing shaded green), (right) A pole-zero plot of the 5^{th}-order AR-model for one segment of the SpO_2 signal. Pole 'A' lies within the acceptable angular region for an apnoea, shaded red. Pole 'B' lies within the acceptable angular region for normal breathing, shaded green. Pole 'D' lies on DC. A* and B* are the other poles in the complex-conjugate pairs.

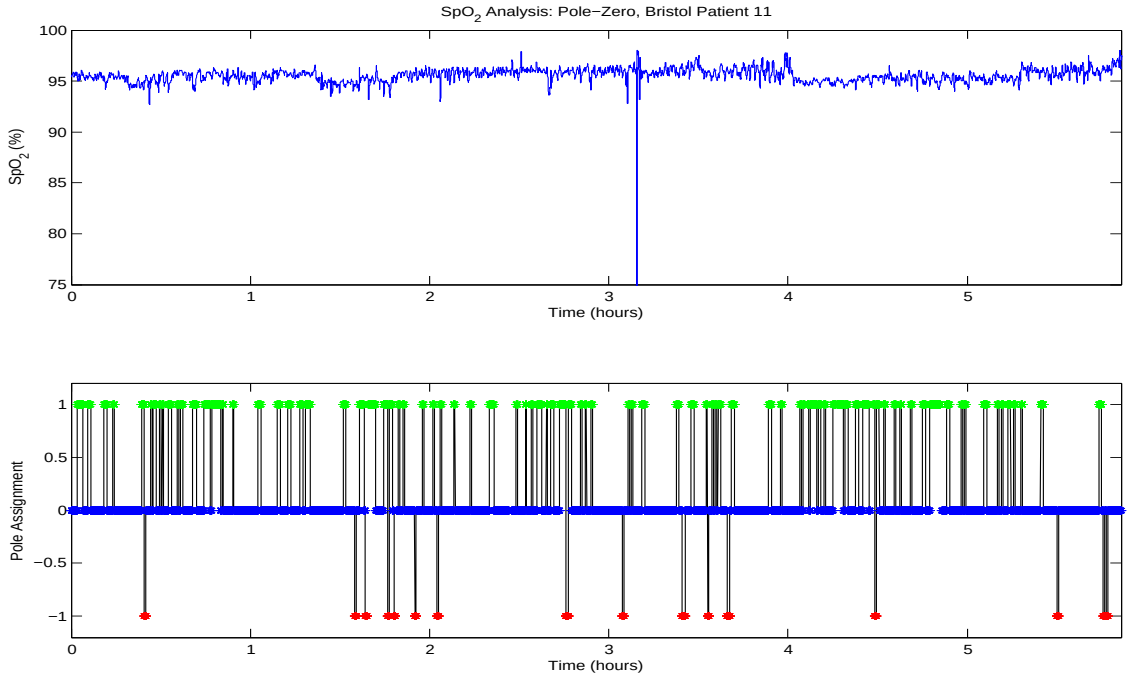


Figure 8.3: Bristol 11: NORMAL patient example. From top to bottom: SpO_2 signal (%); Automated Scores using pole-zero analysis, with a 120-second window, whereby the dominant frequencies correspond to apnoeic (red '*'), normal breathing (green '*'), and both apnoeic and normal breathing (blue '*') range(s) of interest.

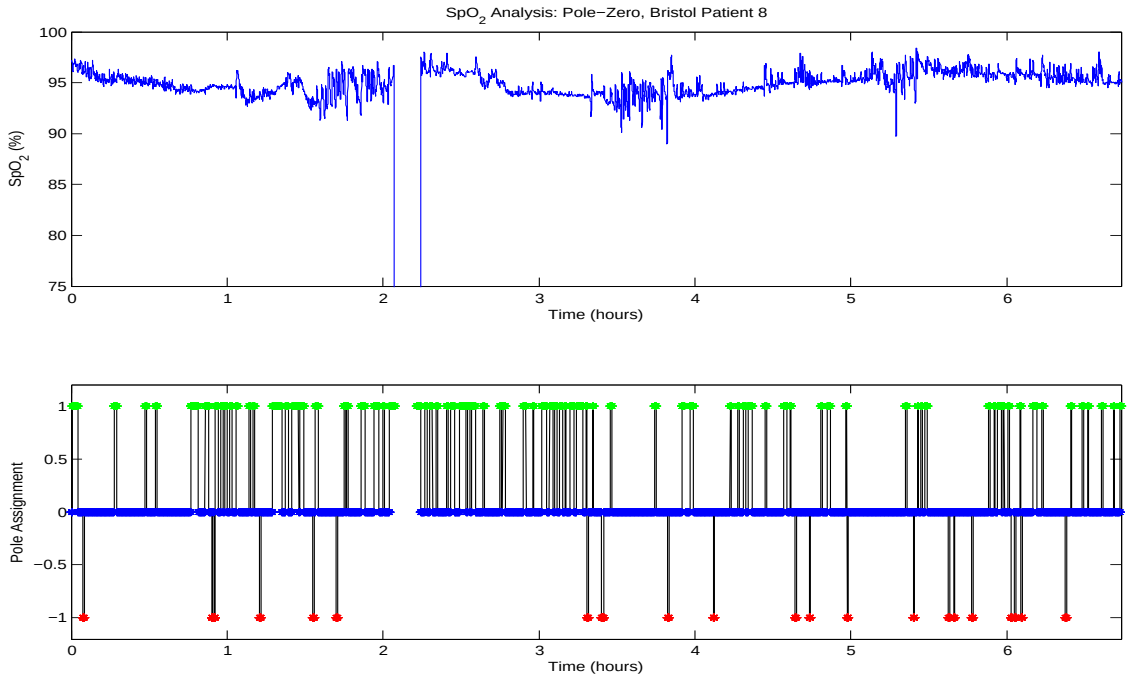


Figure 8.4: Bristol 8: MILD patient example. From top to bottom: SpO_2 signal (%); Automated Scores using pole-zero analysis, with a 120-second window, whereby the dominant frequencies correspond to apnoeic (red '*'), normal breathing (green '*'), and both apnoeic and normal breathing (blue '*') range(s) of interest.

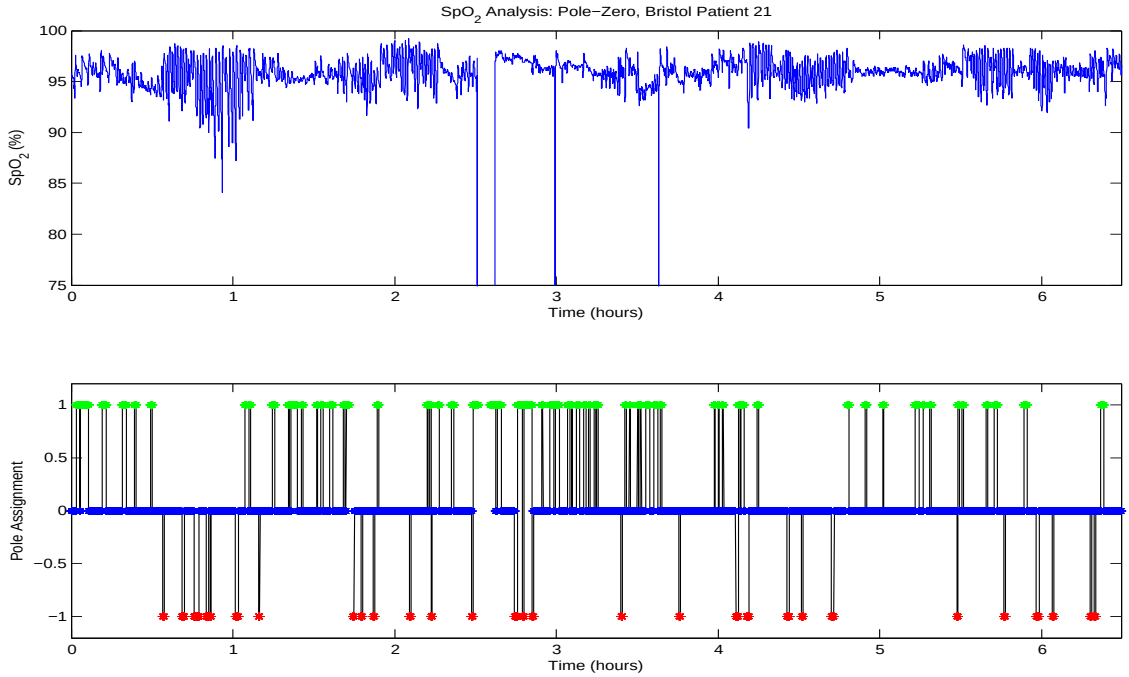


Figure 8.5: Bristol 21: MODERATE patient example. From top to bottom: SpO_2 signal (%); Automated Scores using pole-zero analysis, with a 120-second window, whereby the dominant frequencies correspond to apnoeic (red '*'), normal breathing (green '*'), and both apnoeic and normal breathing (blue '*') range(s) of interest.

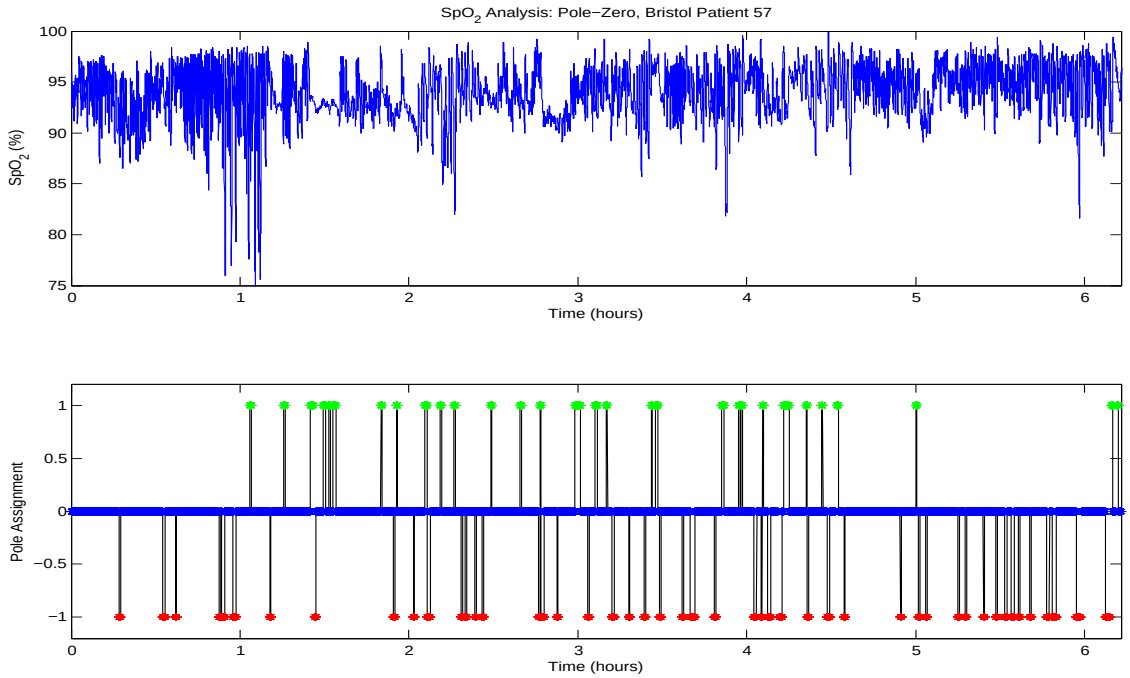


Figure 8.6: Bristol 57: SEVERE patient example. From top to bottom: SpO_2 signal (%); Automated Scores using pole-zero analysis, with a 120-second window, whereby the dominant frequencies correspond to apnoeic (red '*'), normal breathing (green '*'), and both apnoeic and normal breathing (blue '*') range(s) of interest.

Figures 8.7 through 8.10 illustrate the pole zero analysis results for the four SpO_2 records. For example, in Figure 8.7, the top panel shows the SpO_2 signal for the entire night's sleep of 'normal' patient 11. The middle panel shows the dominant frequency for each window, shown as angle 'theta' in degrees. The dominant frequency for the mixed segments has been selected according to whether or not the standard deviation for that segment is greater or smaller than 0.778 units s.d. The angles have been colour-coded according to the Expert Score attached to each window whereby red corresponds to apnoea, and green corresponds to normal breathing. Notice the dashed lines, also colour-coded, showing the angular regions of interest for apnoea and normal breathing. Also, for the cases where two poles lie in the apnoeic region of interest (Case 1.(b)) or normal breathing region of interest (Case 2.(b)), the pole with the larger amplitude (computed σ_{SpO_2}) is utilised. The bottom panel shows the distribution of dominant angles in a histogram, for which the fixed regions of apnoeic and normal breathing activity are highlighted by the red and green dashed lines, respectively. The histogram is included in these figures to illustrate the distribution of apnoeic and normal breathing activity for each patient example. For example, in Figure 8.7, there are three identifiable peaks in the histogram for mild patient 11: an apnoeic peak at 10°, and two normal breathing peaks at 95° and 120°. The peak within the apnoeic region of interest corresponds to a period of 36 seconds (see Equation 8.2); i.e., this patient experiences this length of apnoeas the most over the course of the night (whilst also experiencing shorter and longer apnoeas - with duration specified according to dominant frequency). The peaks within the normal breathing region of interest correspond to breathing frequencies of 16 and 20 breaths per minute. As previously discussed in Chapter 3, breathing rates do change throughout the night; therefore, it is not surprising to find a distribution of breathing frequencies for this patient. Notice the clear separation between peak frequencies for apnoeic and normal breathing activity, as shown in the histogram. This separation (i.e., two peak distribution of apnoeic and normal breathing frequencies) is obvious for the other patient examples as well.

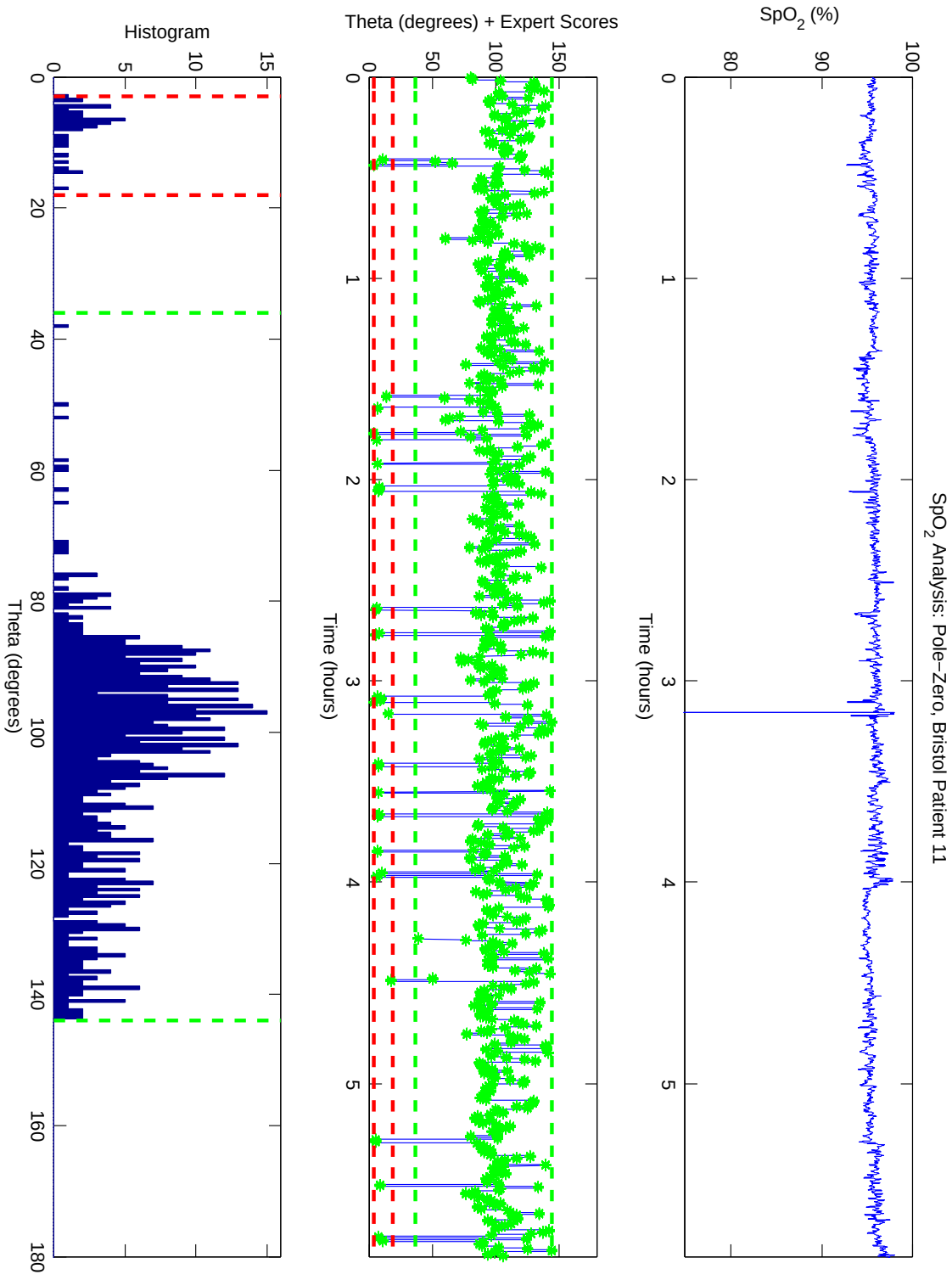


Figure 8.7: Bristol 11: NORMAL patient example. From top to bottom: SpO_2 signal (%); dominant frequency per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing; histogram of dominant angles with apnoeic and normal breathing ranges of interest indicated by the red and green dotted lines, respectively. A 120-second window is utilised for SpO_2 signal segmentation.

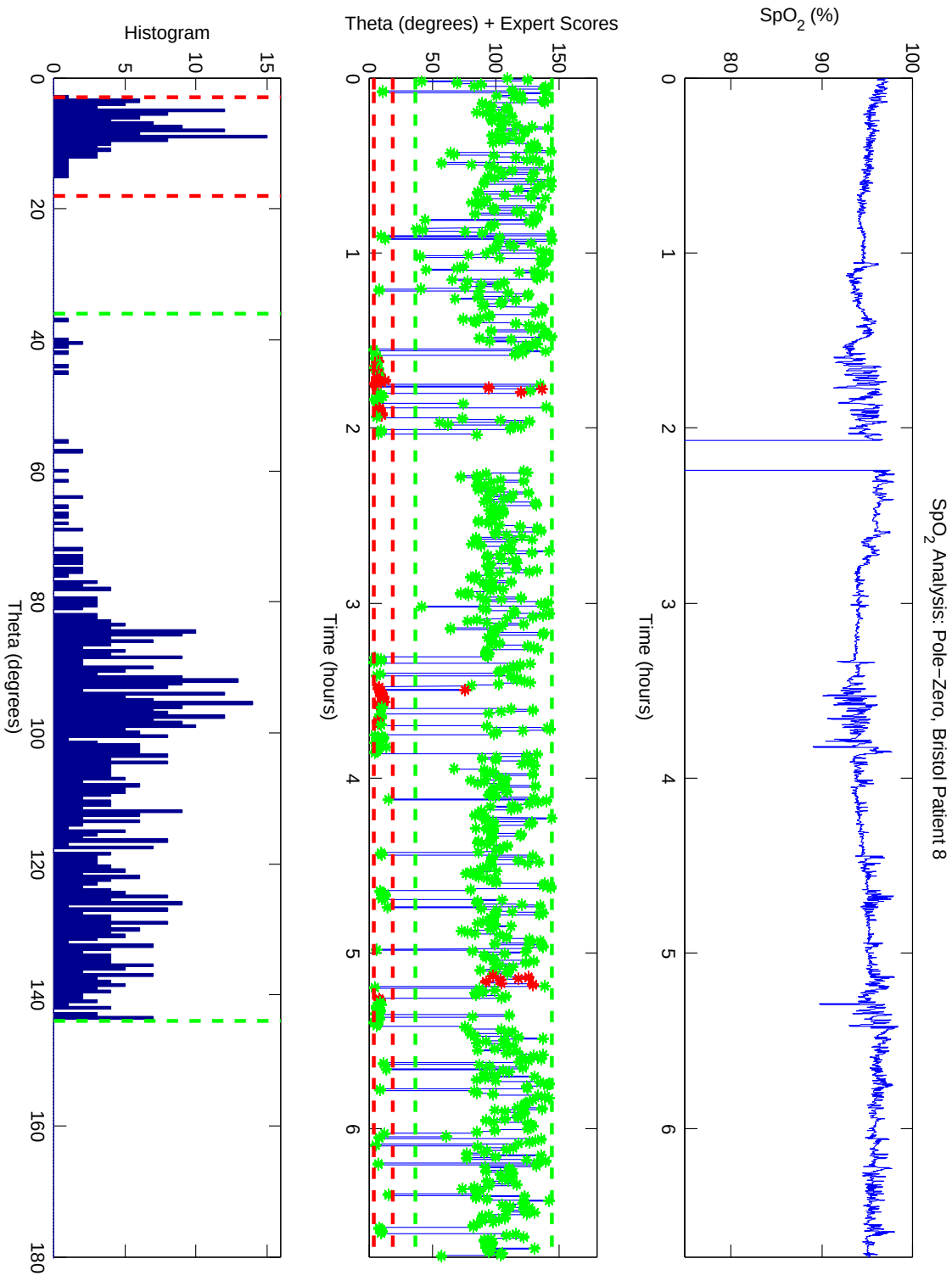


Figure 8.8: Bristol 8: MILD patient example. From top to bottom: SpO_2 signal (%); dominant frequency per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing; histogram of dominant angles with apnoeic and normal breathing ranges of interest indicated by the red and green dotted lines, respectively. A 120-second window is utilised for SpO_2 signal segmentation.

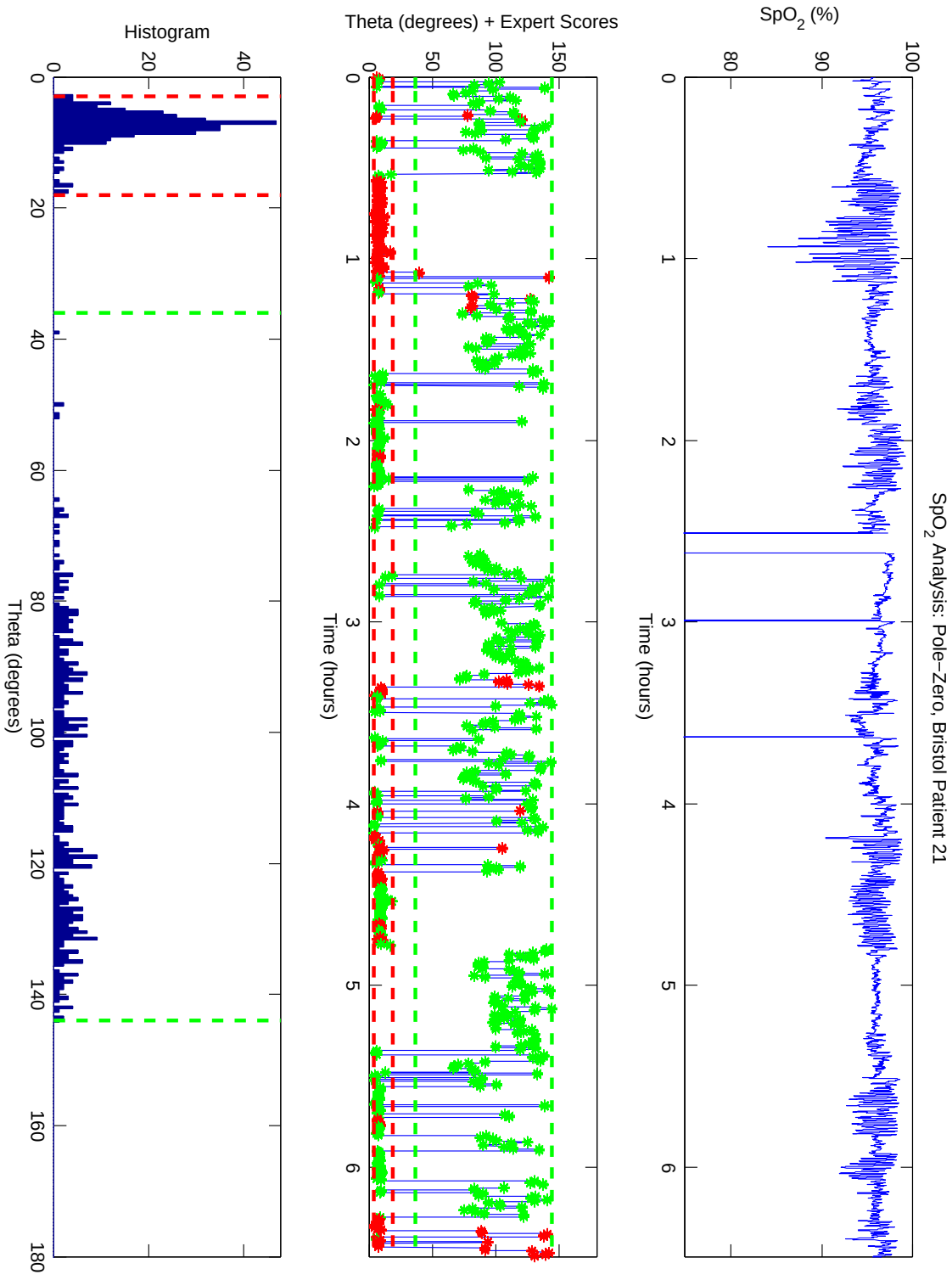


Figure 8.9: Bristol 21: MODERATE patient example. From top to bottom: SpO_2 signal (%); dominant frequency per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing; histogram of dominant angles with apnoeic and normal breathing ranges of interest indicated by the red and green dotted lines, respectively. A 120-second window is utilised for SpO_2 signal segmentation.

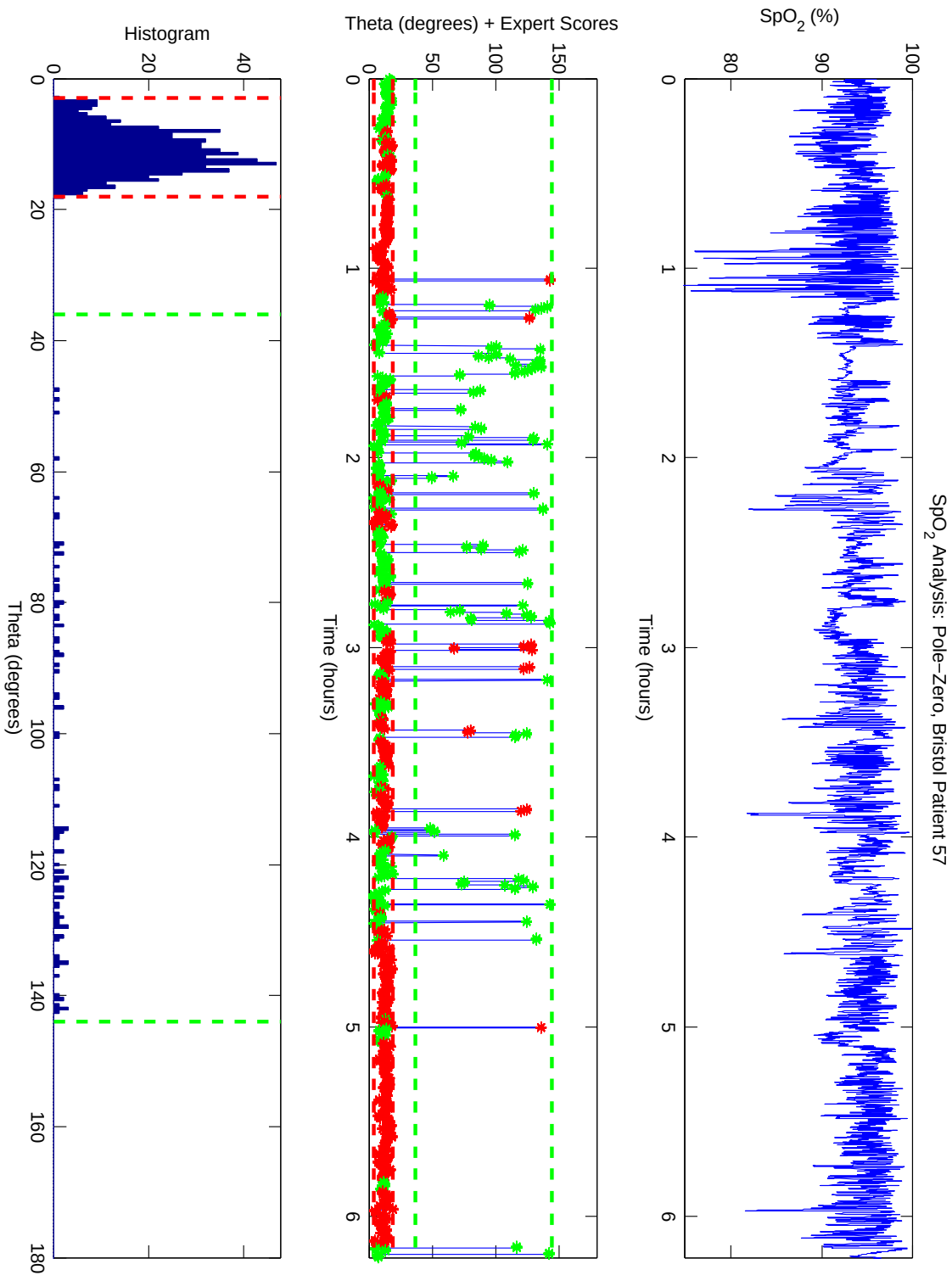


Figure 8.10: Bristol 57: SEVERE patient example. From top to bottom: SpO_2 signal (%); dominant frequency per window with Expert Scores superimposed and colour-coded where red corresponds to apnoea and green corresponds to normal breathing; histogram of dominant angles with apnoeic and normal breathing ranges of interest indicated by the red and green dotted lines, respectively. A 120-second window is utilised for SpO_2 signal segmentation.

8.4 Results of test dataset analysis

Table 8.2 shows the overall (Se, Sp, Acc) values computed for the 44-patient test dataset using a 120-second window using the pole-zero method combined with the amplitude criterion (0.778 units s.d.) to disambiguate the mixed segments. The highest Se value reported is for the normal category (91.4%), followed by the severe category missing very few apnoeas (88.2%), the moderate category (69.5%), and finally the mild category (44.0%). The severe category exhibits the lowest Sp value (51.0%); however, for such patients, consistent periods of normal breathing are rare and therefore a low Sp value is not significant. Conversely, specificity values are more significant for normal and mild categories, whereby stable breathing dominates the SpO_2 signal and there are few apnoeas to detect. The highest Sp value reported is for the normal category (86.4%), followed by the mild category (82.3%), and lastly the moderate category (78.0%).

As in Chapter 7, it is therefore worthwhile to consider sensitivity, specificity and accuracy with respect to the severity of the patient condition. The confusion matrices, Tables 8.3 through 8.6, illustrate the results for the 120-second window analysis using the pole-zero technique with the threshold of 0.778 units s.d. The percentage of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) with respect to the total number of segments under analysis is also given. For the normal category (Table 8.3) there are 35 out of 11165 segments identified as apnoeic, or 0.3%. The number of correctly identified segments is 32 out of 35, or 91.4% (i.e., sensitivity - see Table 8.2). For the mild category (Table 8.4) there are 300 out of 8784 segments identified as apnoeic, or 3.4%. The number of correctly identified segments is 132 out of 300, or 44.0%. For the moderate category (Table 8.5) there are 1505 out of 8247 segments identified as apnoeic, or 18.2%. The number of correctly identified segments is 1046 out of 1505, or 69.5%. Lastly, for the severe category (Table 8.6) there are 2823 out of 5444 segments identified as apnoeic, or 51.9%. The number of correctly identified segments is 2489 out of 2823, or 88.2%. The percentage of apnoeic activity increases with the

severity of OSAHS: 0.3% (normal), 3.4% (mild), 18.2% (moderate), 51.9% (severe). The over-sensitivity of this method (as with the methods presented in Chapters 6 and 7) is discussed further in Chapter 9.

Categorisation	Sensitivity (%)	Specificity (%)	Accuracy (%)
OVERALL	79.3	80.1	80.0
Normal	91.4	86.4	86.5
Mild	44.0	82.3	81.0
Moderate	69.5	78.0	76.5
Severe	88.2	51.0	70.3

Table 8.2: Pole-Zero analysis resultant Se, Sp, and Acc percentages across the 44-patient Bristol Database testing dataset, on an overall and *per category* basis, for 120-second window length and 0.778 units s.d.threshold.

8.5 Discussion and summary

In this Chapter, autoregressive modelling is used to target apnoeic and normal breathing activity from the frequency content of the segment. There two fixed angular regions of interest on the pole-zero plot, corresponding to apnoeic activity with period boundaries of 20 to 120 seconds (3 to 18°), and normal breathing activity of 6 to 24 breaths per minute (36 to 144°). First, AR-modelling is used to compute the location of 5 poles which are assigned accordingly (see Figure C.3 in Appendix C). In 75.4% of the total segments analysed in the training set, poles were found in *both* apnoeic and normal breathing regions of interest (i.e., the mixed case); and thus, amplitude criterion was incorporated to differentiate between classes (apnoea, normal breathing). A reduction in window size may help reduce the number of 'mixed' segments; i.e., higher frequency resolution may be possible with a shorter signal (window) length. In comparison with the nonparametric DFT method used in Chapter 7, AR-modelling is a parametric method which uses a different approach to spectral estimation [73]. Instead of estimating the PSD directly, the data is modelled by a system of linear equations that are relatively simple to solve. In particular, the main advantages of using the Burg method for AR spectral estimation

include resolving closely spaced sinusoids in signals with low noise levels, and estimating short data records, whereby the AR power spectral density estimates are very close to the true values [73].

In this Chapter, the combination of time criterion and frequency domain information allows the identification of apnoeic and normal breathing activity with an overall accuracy of 80.0%. In the following chapter, a comparison is made between the results presented here, and those of time and frequency domain analyses in Chapters 6 and 7. Furthermore, an automated method is investigated which summarises the entire night's sleep record on an hourly and overnight basis.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	32 (0.3%)	3 (0.0%)
	Normal Breathing	1509 (13.5%)	9621 (86.2%)

Table 8.3: NORMAL patients' confusion matrix (pole-zero analysis): 120-sec window, 0.778 units s.d. threshold. A total of 11165 segments are analysed, with the relative percentages shown in brackets.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	132 (1.5%)	168 (1.9%)
	Normal Breathing	1449 (17.1%)	6985 (79.5%)

Table 8.4: MILD patients' confusion matrix (pole-zero analysis): 120-second window, 0.778 units s.d. threshold. A total of 8784 segments are analysed, with the relative percentages shown in brackets.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	1046 (12.7%)	459 (5.6%)
	Normal Breathing	1480 (17.9%)	5262 (63.8%)

Table 8.5: MODERATE patients' confusion matrix (pole-zero analysis): 120-second window, 0.778 units s.d. threshold. A total of 8247 segments are analysed, with the relative percentages shown in brackets.

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	2489 (45.7%)	334 (6.1%)
	Normal Breathing	1284 (23.6%)	1337 (24.6%)

Table 8.6: SEVERE patients' confusion matrix (pole-zero analysis): 120-second window, 0.778 units s.d. threshold. A total of 5444 segments are analysed, with the relative percentages shown in brackets.

Chapter 9

Analysing an Entire Night's Sleep

9.1 Motivation

The utility of unattended oximetry for home diagnosis of OSAHS is currently dependent upon conventional AHI values, as introduced in Section 4.8. The AHI is defined as the number of apnoeas plus hypopnoeas identified by experts per hour of sleep (although the definition of what constitutes an apnoea and hypopnoea differs amongst sleep-clinics). The cut-off AH-indices between normal and pathological are based on two general criteria: first, the level at which there are no adverse effects of the disorder; second, the level at which treatment offers no net benefit [80]. Table 3.1 describes the AHI corresponding to various degrees of OSAHS intensity from normal to severe, repeated here for convenience: 0 to <5 events per hour (normal), 5 to 14 events per hour (mild), >14 to 30 events per hour (moderate), and >30 events per hour (severe). The relationship between AHI and the outcome of OSAHS treatment is not entirely well defined. There is no evidence, for example, that treating mild forms of OSAHS reduces blood pressure when it is elevated due to the disorder [80, 95]. On the other hand, benefits from treatment for patients with severe forms of OSAHS have been reported in [80, 117]. Despite the limitations and ambiguities associated with the AHI, it is the gold-standard assessment of an entire night's sleep, with which any automated RDI must be compared.

In Chapters 6 through 8, three methods of analysis were introduced to identify apnoeic activity within the SpO_2 signal of patients, with a time resolution between 30 and 120 seconds. This Chapter first compares the three methods: time, frequency, and AR-modelling with amplitude criterion. An automated method to summarise the entire night's sleep record is then proposed and compared to AHI values.

9.2 Comparison of automated analyses

Table 9.1 shows both the overall and per-category results from Chapters 6 through 8 using a 120-second window for SpO_2 signal segmentation. In the time domain, the overall results are (85.3%, 80.5%, 81.2%). In the frequency domain, the overall results are comparable, at (84.9%, 81.7%, 82.1%). With the pole-zero AR modelling and amplitude criterion, the overall results are slightly lower, at (79.3%, 80.1%, 80.0%). In addition, the sensitivity, specificity and accuracy values are presented according to normal, mild, moderate, and severe patients within the 44-patient test dataset. Classification accuracies, for time/frequency/AR-modelling and amplitude analyses, are as follows: 87.9%/88.7%/86.5% (normal), 82.1%/83.1%/81.0% (mild), 76.3%/77.2%/76.5% (moderate), 73.2%/74.6%/70.3% (severe). The sensitivity and specificity pairs are also comparable for each patient category using all three methods, with the exception of the normal category. For example, the (Se, Sp) pairs for the moderate category are as follows for the three methods: (72.3%, 77.2%), (71.6%, 78.4%), and (69.5%, 78.0%). On the other hand, for the normal category, the (Se, Sp) pairs are as follows: (91.4%, 87.9%), (82.9%, 88.7%), and (91.4%, 86.4%). The time domain and pole-zero results are such that apnoeic activity is correctly identified in three more segments (TP of 32 instead of 29) and normal breathing activity is incorrectly identified in three less segments (FN of 3 instead of 6). (These results are available in the confusion matrices of Tables 6.5, 7.5, and 8.3.) The 8.5% increase in Se values found using the time domain and AR-modelling with amplitude criterion analyses is actually misleading, as the misdiagnosed segments are so few. The

search for a global classification method is *not* dependent upon a missed apnoea or two.

In Chapters 6 through 8, it was concluded that each automated method is over-sensitive in the identification of apnoeic activity; however, because the percentage of False Positives is also proportional to the severity of the condition, it should not unduly affect the categorisation of the patient by the automated system - as discussed below. The over-estimation of apnoeic activity explains the high sensitivity value reported for the normal category; i.e., the few existing apnoeic events are mostly identified correctly. Unfortunately, this does not explain the low sensitivity values reported for the mild category of patients using time domain (44.3%), frequency domain (45.2%), and AR-modelling with amplitude criterion (44.0%) analyses. Visible desaturation-resaturation cycles of SpO_2 signal are not always scored apnoeic by the experts, as illustrated in the highlighted SpO_2 signal of mild patient 11 in Figure 6.14. The experts have additional signals to consult during scoring - an advantage that the automated methods presented here do not. In the cases where the experts seem to be incorrect, such as in Figure 6.14, airflow and other PSG-related signals would have been consulted. Lastly, the high sensitivity values reported for the moderate and severe patient categories are expected, due to the high number of apnoeic events present in each SpO_2 signal of these categories.

Any diagnostic test typically performs less well on patients with borderline disease [80]. Conversely, the higher the proportion of patients with severe disease, the better the diagnostic test [80]. In our four patient examples, we see poor Se values reported for mild cases of OSAHS (44.3%, 45.2%, and 44.0%) and high Se values reported for severe cases of OSAHS (96.6%, 96.3%, and 88.2%). The selection of the classification threshold was optimised on the training data using ROC analysis. It would be possible to re-tune the threshold such that the detection of apnoeic episodes in patients with mild OSAHS was optimal. However, the goal of this Chapter is to identify the best solution on average; i.e., find for the distribution of categories found in the dataset the method which automatically identifies apnoeic activity with the highest possible accuracy for all categories.

The optimal window size found in both Chapters 6 and 7 was 120-seconds, most likely

because the window is long enough to capture complete desaturation-resaturation cycles of the SpO_2 signal. In the 91-patient Bristol Database, as discussed in Chapter 5, the value of the maximum duration of an event (apnoea, hypopnoea) for each patient was on average 45 seconds. The values of the maximum event duration for each patient, averaged according to categorisation status, were found to be 30 seconds (normal); 42 seconds (mild); 63 seconds (moderate); and 63 seconds (severe).

Categorisation	Sensitivity (%)	Specificity (%)	Accuracy (%)
OVERALL			
Time Domain	85.3	80.5	81.2
Frequency Domain	84.9	81.7	82.1
AR-modelling with amplitude criterion	79.3	80.1	80.0
NORMAL			
Time Domain	91.4	87.9	87.9
Frequency Domain	82.9	88.7	88.7
AR-modelling with amplitude criterion	91.4	86.4	86.5
MILD			
Time Domain	44.3	83.4	82.1
Frequency Domain	45.2	84.4	83.1
AR-modelling with amplitude criterion	44.0	82.3	81.0
MODERATE			
Time Domain	72.3	77.2	76.3
Frequency Domain	71.6	78.4	77.2
AR-modelling with amplitude criterion	69.5	78.0	76.5
SEVERE			
Time Domain	96.6	48.0	73.2
Frequency Domain	96.3	51.3	74.6
AR-modelling with amplitude criterion	88.2	51.0	70.3

Table 9.1: Se, Sp, and Acc percentages across the 44-patient test dataset, on an overall and *per category* basis, using a 120-second window for segmentation and the ROC-based threshold (from the training dataset) - in both time domain (0.778 units s.d.); frequency domain (0.51 units); and with AR-modelling combined with the 0.778 units s.d. amplitude criterion.

9.3 Apnoeic segment identification

The number of segments identified as apnoeic according to manual and automated analyses, may be compared to the total number of segments in the respective SpO_2 signals. First, the percentage of apnoeic segments is computed on a *per-category* basis for the 44-patient test dataset; whereby the number of TP, TN, FP, and FN values are *averaged* per-category. Table 9.2 shows the percentage of apnoeic segments found using manual, time domain, frequency domain, and AR-modelling with amplitude criterion analyses. In all three cases, the percentage of identified apnoeic segments increases as the severity of OSAHS increases. Manual analysis yields the following results: 0.3% (normal), 3.4% (mild), 17.7% (moderate), and 51.3% (severe). In the time domain analysis, the results are as follows: 12.4% (normal), 17.5% (mild), 31.8% (moderate), and 75.1% (severe). In the frequency domain analysis, the results are not very different: 11.5% (normal), 16.6% (mild), 30.7% (moderate), and 73.4% (severe). With the AR-modelling with amplitude criterion, again there is not much difference: 13.8% (normal), 18.6% (mild), 30.6% (moderate), and 69.3% (severe). Note that the total number of segments used in the AR-modelling analysis is slightly lower than that for the time and frequency domain analyses, due to the classification corresponding to case 4 (no pole in either the apnoeic or the normal breathing region of interest) - see Section 8.3.

The four patient examples - used throughout Chapters 6 through 8 - are next examined in more detail. Table 9.3 shows the percentage of segments identified as apnoeic in comparison to the total number of segments available for analysis, according to the experts. The percentage of segments which they identified as apnoeic increases with the severity of OSAHS: 0% (normal 11), 5.6% (mild 8), 21.1% (moderate 21), and 53.5% (severe 57). Tables 9.4, 9.5, and 9.6 show the percentage of segments identified as apnoeic in comparison to the total number of segments used for analysis, noting that artefactual segments are excluded from the analysis. Using time domain analysis, the percentage of identified apnoeic segments increases with the severity of OSAHS: 2.7% (normal 11), 15.2% (mild

8), 46.5% (moderate 21), and 91.6% (severe 57) - see Table 9.4. The results obtained using frequency domain analysis are again similar (if slightly lower) than those reported using time domain analysis; i.e., the percentage of identified apnoeic segments increases with the severity of OSAHS: 2.7% (normal 11), 13.6% (mild 8), 45.6% (moderate 21), 89.5% (severe 57) - see Table 9.5. Finally, the results tabulated using AR-modelling with amplitude criterion are also comparable; i.e., the percentage of identified apnoeic segments increases with the severity of OSAHS: 4.8% (normal 11), 16.0% (mild 8), 44.8% (moderate 21), 87.8% (severe 57). For all four examples, the percentage of apnoeic segments is over-estimated in comparison to the experts' decision, and the percentage of apnoeic activity increases as OSAHS severity increases (from normal to severe), as expected.

Categorisation	Expert Scores (%)	Time Domain (%)	Frequency Domain (%)	AR-Modelling with Amplitude Criterion (%)
Normal	0.3	12.4	11.5	13.8
Mild	3.4	17.5	16.6	18.6
Moderate	17.7	31.8	30.7	30.6
Severe	51.3	75.1	73.4	69.3

Table 9.2: Percentage of apnoeic segments found using manual, time domain, frequency domain, and AR-modelling with amplitude criterion analyses, with the percentage of apnoeic segments averaged on a *per-category* basis.

Bristol Patient Example (category)	No. of Apnoeic Segments	No. of Segments Available	Percentage of Apnoeic Segments
Bristol 11 (normal)	0	783	0%
Bristol 8 (mild)	45	808	5.6%
Bristol 21 (moderate)	164	779	21.1%
Bristol 57 (severe)	399	746	53.5%

Table 9.3: Manual Analysis: the percentage of apnoeic segments as found by the experts are shown for the test example from each patient category.

Bristol Patient Example (category)	No. of Apnoeic Segments	No. of Segments Analysed	Percentage of Apnoeic Segments
Bristol 11 (normal)	19	699	2.7%
Bristol 8 (mild)	119	784	15.2%
Bristol 21 (moderate)	350	753	46.5%
Bristol 57 (severe)	683	746	91.6%

Table 9.4: Automated Analysis: Time Domain, with the percentage of apnoeic segments shown for the test example from each patient category.

Bristol Patient Example (category)	No. of Apnoeic Segments	No. of Segments Analysed	Percentage of Apnoeic Segments
Bristol 11 (normal)	19	699	2.7%
Bristol 8 (mild)	107	784	13.6%
Bristol 21 (moderate)	343	753	45.6%
Bristol 57 (severe)	668	746	89.5%

Table 9.5: Automated Analysis: Frequency Domain, with the percentage of apnoeic segments shown for the test example from each patient category.

Bristol Patient Example (category)	No. of Apnoeic Segments	No. of Segments Analysed	Percentage of Apnoeic Segments
Bristol 11 (normal)	33	689	4.8%
Bristol 8 (mild)	125	779	16.0%
Bristol 21 (moderate)	332	741	44.8%
Bristol 57 (severe)	650	740	87.8%

Table 9.6: Automated Analysis: AR-modelling with amplitude criterion, with the percentage of apnoeic segments shown for the text example from each patient category.

9.4 An automated RDI

From the automated analyses, it is possible to compute a simple RDI to summarise the entire night's sleep record, defined as follows:

$$RDI_{derived} = \left(\frac{\text{No. of apnoeic segments detected by automated system}}{\text{Total number of segments used in analysis}} \right) \times 100\% \quad (9.1)$$

The identification of apnoeic and normal breathing activity in this thesis has so far been on a minute-by-minute basis for the reasons explained in Section 1.3. The computation of the above RDI, therefore, enables the minute-by-minute analysis to be converted to a larger timescale. As the expert-derived, and gold-standard, AHI is a count of the number of events *per hour*, the initial analysis is performed, with linear regression, on a per-hour basis.

9.5 Per-hour analysis of the SpO_2 signal

Using Equation 9.1, the RDI is computed for the 44-patient test database. Linear regression analysis is used to model the relationship between the computed RDI and standard AHI values. A linear equation is fitted to the data using the method of least-squares. The strength of the relationship between the variables of interest is the correlation coefficient; whereby a value of one indicates perfect correlation (i.e., exact linear dependency) and a value of zero indicates no correlation whatsoever [112]. The lengths of each patient record do not (necessarily) divide evenly into hourly intervals; thus, some truncation is required. For example, the record length for Bristol severe patient 57 is 6.25 hours; the computed RDI and AHI (from the expert) are obtained for each of the first six hours, the partial hour not being analysed further. Six index pairs (expert AHI score, automated RDI) are therefore obtained for this patient and this is repeated for each patient in the test database, with each of the three automated methods. The top panel of Figure 9.1 shows the regression line between the indices derived by manual and time domain analyses, with

a resultant correlation coefficient of $\rho = 0.79$. The middle panel shows the corresponding regression line for the frequency domain analysis, with $\rho = 0.82$. Finally, the bottom panel gives the results for the AR-modelling with amplitude criterion, giving $\rho = 0.78$. All three methods yield comparable results.

There are many cases where the per-hour indices do not match whatsoever - this is especially clear when the RDI exceeds 40 events per hour, and the AHI shows no events whatsoever. We refer back to Figure 6.14 where desaturation-resaturation cycles were clear in the SpO_2 signal; however, the Expert Scores classified the majority of segments as normal breathing (green asterisks). For such a case, that hour would yield a comparable RDI-AHI comparison as described above: the RDI would be large as dips in the SpO_2 signal occurred, whereas the AHI would be zero as the experts did not score anything accordingly.

9.6 Overnight analysis of the SpO_2 signal

With a high correlation between manual and automated indices on a per-hour basis, overnight analysis of the SpO_2 is next computed. Figure 9.2 shows the regression lines fitted to the sleep index variables computed using all three methods. The top panel of Figure 9.2 shows the regression line between the indices derived by manual and time domain analysis, with a resultant correlation coefficient of $\rho = 0.84$. The middle and bottom panels show the result for the frequency-domain analysis ($\rho = 0.86$) and AR-modelling with amplitude criterion ($\rho = 0.82$). The three values of correlation coefficients are higher than those obtained for the per-hour analysis. Returning to the above discussion, the lack of expert scoring of some activity throughout the night (see Figure 6.14) does not necessarily correspond to a lack of scoring of apnoeic activity for the entire night. For example, in Figure 6.11 (time domain analysis of the mild patient, Bristol 8), there is some apnoeic detection around 1.5, 3.5, and 5 hours - see the red asterisks for apnoea. However, not all of the segments with σ_{SpO_2} values falling above the optimised threshold

are scored apnoeic - see the green asterisks for normal breathing. In these such areas, the 'lack' of scoring by the experts may be attributed to their use of multiple signals, such as airflow and EEG, to identify events.

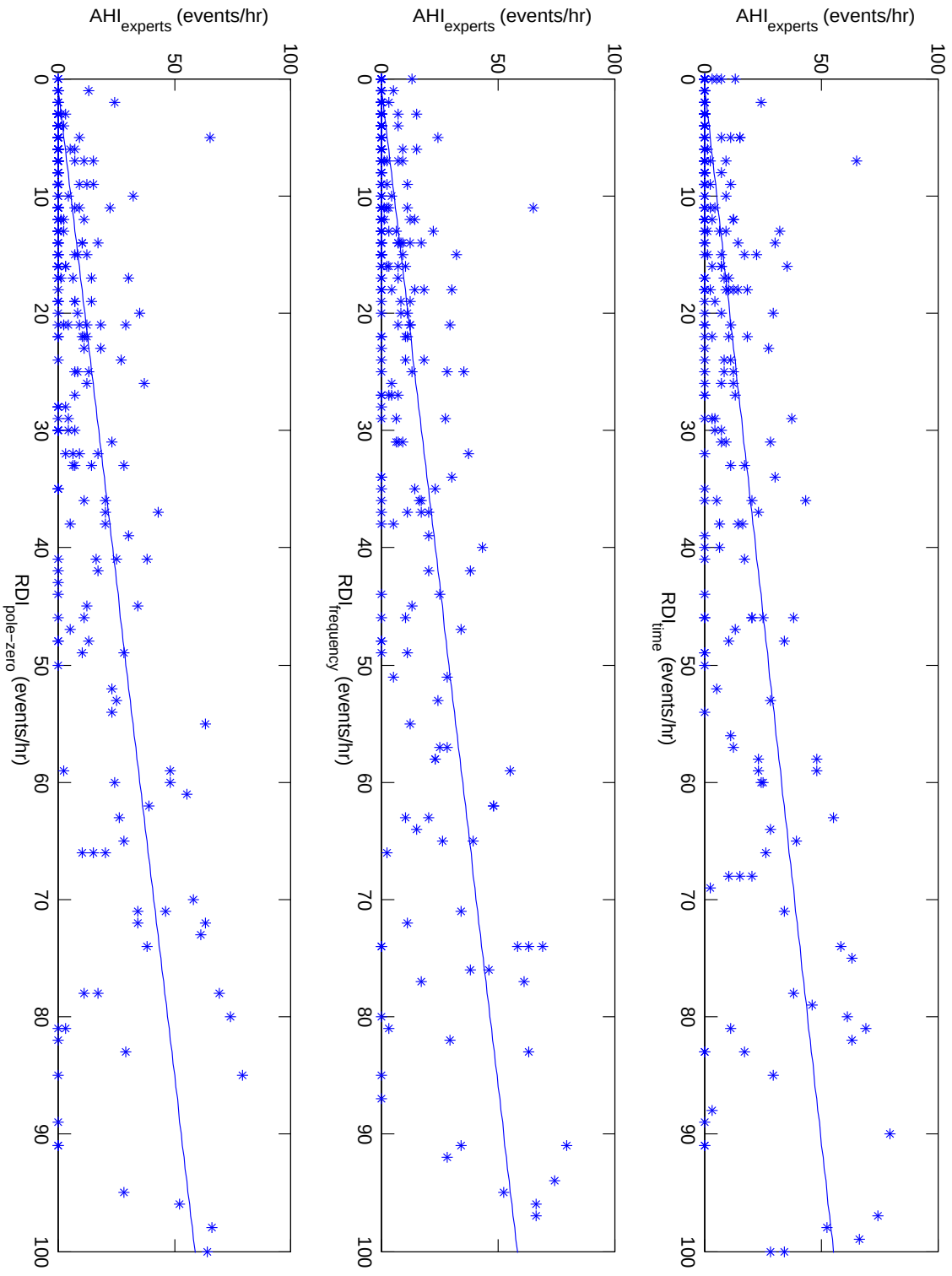


Figure 9.1: Linear regression analysis using the indices on a per hour basis. From top to bottom: a comparison of manual analysis ($\text{AHI}_{experts}$) and time domain analysis (RDI_{time}); frequency domain analysis ($\text{RDI}_{frequency}$), and pole-zero analysis ($\text{RDI}_{pole-zero}$). Indices are marked events per hour.

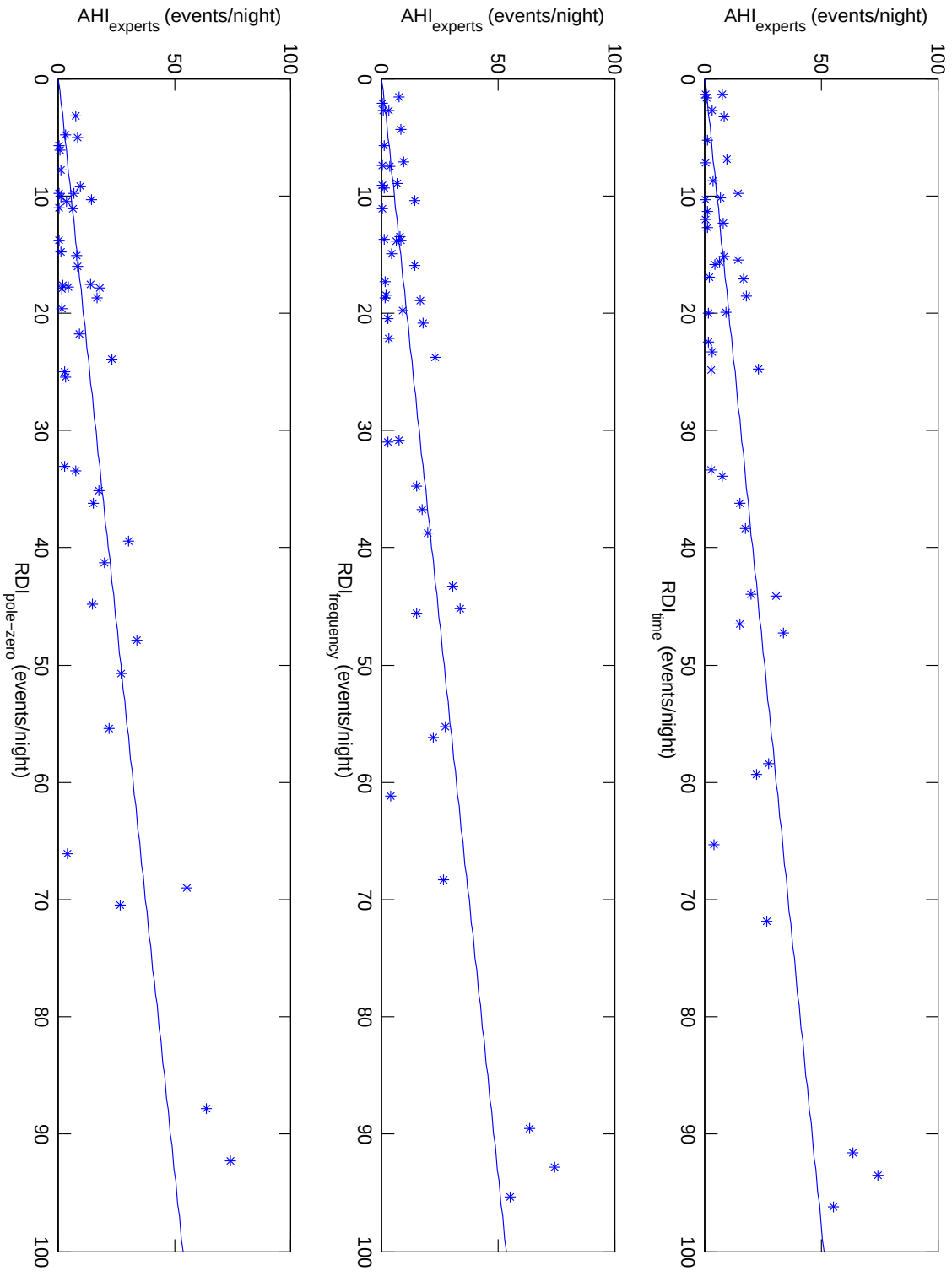


Figure 9.2: Linear regression analysis using the indices for the entire night. From top to bottom: a comparison of manual analysis ($\text{AHI}_{experts}$) and time domain analysis (RDI_{time}); frequency domain analysis ($\text{RDI}_{frequency}$); and pole-zero analysis. Both indices are in events per night.

9.7 Overnight analysis using a histogram-based RDI

In Section 8.4, the distribution of dominant frequencies in the form of the histogram was discussed, and a clear separation between apnoeic and normal breathing frequencies was observed in all four patient examples (see Figures 8.7 through 8.10). As may be seen from these figures, as OSAHS severity increases (from normal to severe), the distribution of apnoeic frequencies predominates - suggesting another RDI based on the histogram distribution values. First, the total number of occurrences within the apnoeic region of interest (*apnoeic distribution*) and normal breathing region of interest (*normal breathing distribution*) are computed. Two indices are evaluated: the ratio of the area of the apnoeic distribution to the area of the normal breathing distribution (Equation 9.2), and the area of the apnoeic normalised according to the total number of segments used in the analysis (Equation 9.3). Using a bin width of 0.5, the apnoeic distribution lies between bins 7 and 37; and the normal breathing distribution lies between bins 73 and 289. Figure 9.3 shows the regression lines fitted to the sleep index variables computed using these two histogram-based metrics. The top panel of Figure 9.3 shows the regression line for the RDI based on the ratio of histogram areas (apnoea to normal breathing - see Equation 9.2), with a resultant correlation coefficient of $\rho = 0.67$ - the lowest reported so far, due to the outliers, or RDI values greater than 5 events per hour. These outliers are due to extremely severe patients who, in addition to having apnoeic activity distributed throughout the majority of the night, have very little normal breathing. For these patients, the normalised distribution within the apnoeic region of interest does not result in such dramatic effects. (Note that the fitted regression line is somewhat misleading, due to the presence of the outliers, but has been included in the plot nonetheless for illustrative purposes.) The bottom panel of Figure 9.3 is the equivalent plot for the RDI based on the normalised apnoeic distribution (Equation 9.3), with $\rho = 0.83$. The latter correlation coefficient is comparable to those derived using the RDI of Equation 9.1.

$$RDI_{hist-apnoea/breath} = \left(\frac{\text{Area of apnoeic distribution}}{\text{Area of normal breathing distribution}} \right) \times 100\% \qquad (9.2)$$

$$RDI_{hist-apnoea} = \left(\frac{\text{Area of apnoeic distribution}}{\text{Total number of segments used in analysis}} \right) \times 100\% \qquad (9.3)$$

9.8 Optimisation of the RDI threshold values

The AHI definition is defined in terms of cutoff values which best differentiate patients into the four categories (normal, mild, moderate, severe) - see Table 3.1. To find the equivalent cut-off values for the computed RDI values, an optimisation routine is applied to the 47-patient training database. The routine is similar to ROC analysis: for each subject of interest, a range of threshold values is investigated from 1 to 100, in integer increments. (These values are selected empirically, based on the observed values of RDI for all four categories of patients: normal, mild, moderate, severe.) For the RDI values based on histogram areas, the range of threshold values for the two RDIs are 0.01 to 15, with increments of 0.01 for the first index, and 0.01 to 1, with increments of 0.01 for the second index. The threshold value which yields the highest classification accuracy is optimal for the training dataset. As four patient categories must be differentiated, three passes of the algorithm are required. The first-pass aims to differentiate between normal and mild patients. The second and third-passes differentiate between mild and moderate patients, and moderate and severe patients, respectively.

For the time domain analysis, the optimal RDI threshold values turn out to be 13, 32, and 44 events per hour. For the frequency domain analysis results, the equivalent cut-off values are: 14, 32, and 46 events per hour. For AR-modelling with amplitude criterion, the optimal RDI threshold values are, in this case: 16, 32, and 43 events per hour. Finally, for the histogram-based results, the optimal RDI threshold values are: 0.16,

0.44, 0.73 ($\text{RDI}_{hist-apnoea/breath}$) and 0.15, 0.31, 0.41 ($\text{RDI}_{hist-apnoea}$). These determined cutoff values are applied to the 44-patient test dataset.

The confusion matrices of Tables 9.7 through 9.11 give the classification results using these cutoff values, presented in terms of the four patient categories. The leading diagonal shows when there is agreement between the manual and automated analyses; i.e., when normal, mild, moderate, and severe patients are classified correctly. The off-diagonal elements of the matrices show the misclassifications. As there is inevitably overlap between categories (i.e., RDI cut-off values are not perfect), it is important that the off-diagonal elements fall close in number to their nearest diagonal element. For example, for time domain analysis, the perfect classifications (diagonal) are for 10 normal, 4 mild, 3 moderate, and 6 severe patients. Note how no severe patients (according to the experts) are classified as normal by the automated analysis, and no normal patients (according to the experts) are classified as severe by the automated analysis. The cost of patient misclassification is not balanced with respect to the presented confusion matrices. In Table 9.7, four patients are categorised as mild, even though the experts scored them as moderate. Based on the automated result, these four patients could be treated with nCPAP therapy for their 'mild' case OSAHS. As their actual diagnose was moderate, the administered nCPAP treatment administered could be beneficial. However, in the off-diagonal case whereby the automated analysis categorised the patient as normal, the experts scored them as severe. Normal cases of OSAHS are not treated further. The cost of this misdiagnose could be deadly; for example, the patient could be allowed to drive a car and fall asleep at the wheel. Therefore, the time saving benefit of utilising an automated system must be measured against the cost of misdiagnoses.

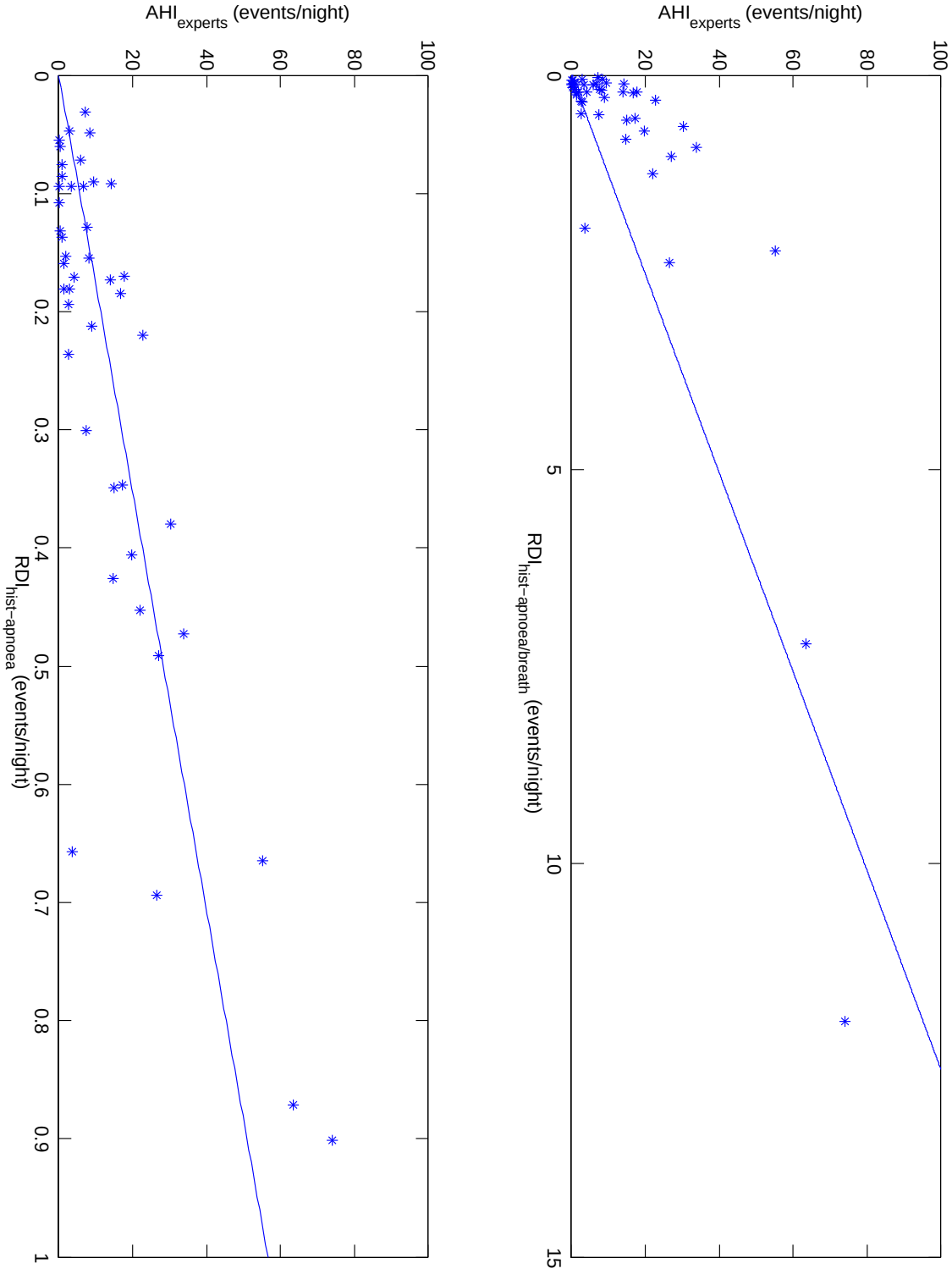


Figure 9.3: Linear regression analysis using the indices for the entire night using histogram data from AR-modelling with amplitude criterion analysis. From top to bottom: a comparison of manual analysis ($\mathrm{AHI}_{experts}$) and the RDI based on the ratio of histogram distributions ($\mathrm{RDI}_{hist-apnoea/breath}$); a comparison of manual analysis ($\mathrm{AHI}_{experts}$) and the normalised apnoeic distribution ($\mathrm{RDI}_{hist-apnoea}$). Both indices are in events per night.

		Automated			
		NORMAL	MILD	MODERATE	SEVERE
Manual	NORMAL	10	5	1	0
	MILD	5	4	1	1
	MODERATE	1	4	3	3
	SEVERE	0	0	0	6

Table 9.7: Time domain analysis confusion matrix for the 44-patient test dataset, with optimal RDI thresholds as computed on the 47-patient training dataset.

		Automated			
		NORMAL	MILD	MODERATE	SEVERE
Manual	NORMAL	10	6	0	0
	MILD	7	3	0	1
	MODERATE	0	5	4	2
	SEVERE	0	0	2	4

Table 9.8: Frequency domain analysis confusion matrix for the 44-patient test dataset, with optimal RDI thresholds as computed on the 47-patient training dataset.

		Automated			
		NORMAL	MILD	MODERATE	SEVERE
Manual	NORMAL	10	6	1	0
	MILD	5	3	4	0
	MODERATE	1	1	3	1
	SEVERE	0	0	3	5

Table 9.9: AR-modelling with amplitude criterion analysis confusion matrix for the 44-patient test dataset, with optimal RDI thresholds as computed on the 47-patient training dataset.

		Automated			
		NORMAL	MILD	MODERATE	SEVERE
Manual	NORMAL	8	5	1	0
	MILD	7	4	4	0
	MODERATE	1	1	3	1
	SEVERE	0	0	3	5

Table 9.10: AR-modelling with amplitude criterion ($\text{RDI}_{hist-apnoea/breath}$) analysis confusion matrix for the 44-patient test dataset, with optimal RDI thresholds as computed on the 47-patient training dataset.

		Automated			
		NORMAL	MILD	MODERATE	SEVERE
Manual	NORMAL	10	6	1	0
	MILD	6	4	4	0
	MODERATE	0	0	3	1
	SEVERE	0	1	3	5

Table 9.11: AR-modelling with amplitude criterion analysis ($\text{RDI}_{hist-apnoea}$) confusion matrix for the 44-patient test dataset, with optimal RDI thresholds as computed on the 47-patient training dataset.

9.9 RDI result comparison with previous work

There are significant differences between this work and others reported in literature. For example, Vázquez *et al.* produced an oximeter derived RDI based on that was highly correlated ($\rho = 0.97$) with the PSG derived AHI for 246 patients with suspected OSAHS. For each patient, the RDI was created by the retrospective analysis of the SpO_2 signal in the time domain whereby an event was marked when the percentage of SpO_2 signal fell below the baseline, the latter defined as the average of the SpO_2 values for five minutes preceding the marked event. The patients' entire night's sleep record (SpO_2 signal) was classified as OSAHS or non-OSAHS based on the computed RDI values. There are two problems with this study: first, the definition of AHI cut-off values (see Section 9.1) and arousals differ from the standards used today [4]. Second, only subjects with "PSG-AHI" values greater than 10 were investigated; i.e., normal and normal-to-mild subjects were not considered. As previously mentioned, any diagnostic test typically performs less well on patients with borderline disease [80]. Therefore, the results of a study with moderate to severe subjects can be tuned for that work, in terms of correlation coefficients and classification accuracies.

Three other studies report correlation values between automated RDIs and expert-scored AHIs. First, Lévy *et al.* report a correlation of 0.72 in their study of the 'Delta Index' (see 4.8.1, [11]) and AHI values [63]. Olson *et al.* show a Spearman correlation coefficient of 0.71 in their study of the Δ-index and AHI values [80]. Magalang *et al.* show a Pearson correlation coefficient of 0.83 between predicted and actual AHI values in their study of the Δ-index combined (in their prediction model) with other oximetry indices such as desaturation events per hour and cumulative time spent below various percentages of SpO_2 signal [66]. These latter methods yield results lower than those reported here, with the exception of the most recent work by Magalang *et al.* which utilises several oximetry indices at once in order to increase correlation with AHI values. In their 516-patient database, the patients were approximately partitioned equally into

OSAHS severity categories (according to their AHI).

Another major difference between this work and others reported in literature, is that results are presented here on a *per-category* basis (normal, mild, moderate, severe). Zamarrón *et al.* (2003) studied the SpO_2 signal in the frequency domain for 300 patients showing clinical symptoms of OSAHS [121]. In the frequency domain, cyclical activity (apnoea) was shown to correspond with a spectral peak in the periodogram of SpO_2 signal (entire night's sleep) between 30 and 70 seconds [121]. No attempt was made to identify patients according to OSAHS severity. It is assumed that the 94% sensitivity and 82% specificity reported reflects the main spectral peak of apnoeic activity for all 300 patients. The interpretation of reported results, such as these, is open to doubt. It is also difficult to compare the data from any of these sleep studies with this work as all of the studies are based on different patient databases, with different definitions of apnoea and hypopnoea.

9.10 Summary

The choice of the ultimate analysis method for identifying apnoeic activity in patients with OSAHS is made according to two factors, both with respect to the expert scores: classification results (Se, Sp, Acc), and correlation coefficient value (ρ). This Chapter describes the results from three comparable methods of analysis of the SpO_2 signal: time domain analysis with σ_{SpO_2}, frequency domain analysis with $P_{ave,apnoea}$, and AR-modelling with amplitude criterion. The overall classification values, as already discusses, are comparable - see Table 9.1. Furthermore, in comparison to the expert-derived AHI, time, frequency, and AR-modelling with amplitude criterion analyses are found to be highly correlated with ρ values varying from 0.82 to 0.86 for overnight analysis and from 0.78 to 0.82 for hourly analysis. Overall, all three types of analyses yield comparable classification accuracy, correlation coefficient, and OSAHS diagnostic results.

Chapter 10

Conclusions

The aim of the work described in this thesis was to investigate the automated quantification of OSAHS severity using the SpO_2 signal acquired from the non-invasive pulse oximetry sensor. There were three main goals of this work, the first being to establish objective, reproducible, and automated procedures to quantify apnoeic activity on a minute-scale; the second being to develop signal processing algorithms for detecting apnoeic activity accurately on this time scale; the third being to create an RDI comparable to that derived by experts which is able to summarise the entire an night's sleep record (on both a per-hour and an overnight basis). Section 10.1 gives an overview of the research and draws conclusions from the results. Section 10.2 suggests some areas for further work.

10.1 Overview and conclusions

For the clinical diagnosis of OSAHS, a variety of signal processing techniques have been proposed to detect apnoeas using the SpO_2 signal (as well as combinations of the SpO_2 signal, airflow, HRV, and the EEG). Chapter 4 described the existing methods used for detecting apnoeic activity from the SpO_2 signal. The main conclusions of this extensive review were three-fold. First, clinicians and sleep researchers have highlighted the SpO_2 signal for its potential usefulness in the clinical diagnosis of OSAHS over the past twenty

years.

Second, there is no consensus amongst sleep researchers as to the choice of oxygen desaturation feature - such as percentage of desaturation - which optimises event detection. Eight years following the study completed by Yamashiro *et al.* in 1995, Magalang *et al.* showed that oxygen saturation variability (as measured by taking the average of absolute differences of oxygen saturation values within successive 12-second intervals, or 'Delta Index') provides an oxygen desaturation index with a similar level of diagnostic accuracy as found with PSG [66], [119]. However, for their 416-patient study, **ten** event indices were employed to create AHI values to classify patients as OSAHS or non-OSAHS, including: desaturation events per hour (using 2%, 3%, and 4% dips in the SpO_2 signal), cumulative time spent below (90%, 88%, 86%, 84%, 82%, 80%) saturation of the SpO_2 signal, and oxygen saturation variability. Third, it is difficult (if not impossible) to compare the results from any of these sleep studies with each other, as the patient databases, and basic definitions (apnoea, hypopnoea) are not the same throughout (see Section 10.1.1).

One of the main aims of this thesis has been to establish whether a method using the SpO_2 signal alone can provide reliable information as to the extent of apnoeic and normal breathing activity present in a patient with suspected OSAHS. In Chapter 5, the preliminary processing of SpO_2 signals was described in preparation for the automated assessment of apnoeic events for comparison with expert scores. Two sets of databases were used for developing and testing the automated algorithms. The four-patient **MIT-BIH Polysomnography Database** was found to be an inadequate database as it consisted of four patients who suffer from severe OSAHS only; whereas, the **Bristol Database**, consisting of 91 patients with varying degrees of apnoeic severity, was divided for training purposes ($\sim$50% of the patients) and subsequent evaluation on an independent test set of patients (the other $\sim$50%). It should be noted that subsequent analysis and results were based on this *one database* and the opinion of at least *one expert*. The degree to which our expert opinion agrees with others in the field is unknown; i.e., the Bristol Database is only accessible for diagnosis, research, and subsequent result publication for

the clinicians and polysomnographers at the Bristol Hospital - with the present exception of this collaboration. Rescoring the sleep signals is impossible, as other sleep clinics do not have access to these patients' data. However, our confidence in the established sleep laboratory at Bristol and experience of our polysomnographer(s)/expert(s) is high. Thus, should the signals be scored again another expert, we would expect overall agreement in general, with some disagreement in the most ambiguous areas of the signals. (It should also be noted that the MIT-BIH PSG Database is an exception to the above, whereby the Database is made accessible on the internet to any sleep clinic or person that wishes to make use of the data for sleep research.)

Chapters 6 and 7 proposed methods of time and frequency domain analyses of the segmented SpO_2 signal. The results showed that:

- The 120-second window is optimal for SpO_2 signal segmentation such that classification accuracy (the correct detection of apnoeic and normal breathing segments) is the highest in comparison to the results found using a 30-second, 60-second, and 90-second window.

- A 120-second window is the highest resolution reported for the identification of apnoeic activity to date, in comparison with sleep studies reported in the literature.

- In the time domain, the time resolution is not severely affected by window size; i.e., using smaller window lengths will not drastically reduce classification accuracies.

- In the frequency domain, the spectral resolution is entirely dependent upon window size; i.e., using longer window lengths results in higher frequency resolution and will affect classification accuracies.

- In the time domain, a 0.778 units s.d. threshold is optimal for dividing segments into apnoeic and normal breathing classes, corresponding to a 2.2% dip in the SpO_2 signal.

- In the frequency domain, the average power found within the apnoeic region of interest is a better metric than the ratio of average power (apnoea to normal breathing).
- Comparable classification sensitivities, specificities, and accuracies are obtained using both time and frequency domain methods.

The potential for extracting more information out of the SpO_2 segments was discussed in Chapter 8, in which a combination of AR-modelling and an amplitude criterion was investigated. The overall and per-category results were found to be comparable to those of Chapters 6 and 7. In conclusion:

- The use of AR-modelling is successful in identifying 25% of data segments as either definitely apnoea or normal breathing.
- The remaining 75% of data segments require the application of an amplitude criterion for classification into apnoea or normal breathing.
- A combination of AR modelling and amplitude criterion does not give a significant improvement in the accuracy of apnoeic and normal breathing activity detection, which may be due to the larger window size utilised (120-seconds).
- The overall accuracy in identifying apnoeic and normal breathing activity in patients with OSAHS is 81.2%, 82.1%, and 80.0% - using time domain, frequency domain, and AR-modelling with amplitude criterion analyses respectively.
- Each automated method is over-sensitive in the identification of apnoeic activity overall, and on a per-category basis (normal, mild, moderate, severe).

As the percentage of False Positives is also proportional to the severity of the patient condition, the over-sensitivity of automated methods should not affect the categorisation of the patient by the automated system.

In Chapter 9, a method to summarise the entire night's sleep record was investigated. First, a simple RDI was proposed for comparison with the AHI values created by experts.

The RDI was used to compute the number of apnoeic events detected both hourly and overnight. Per-hour analysis of the SpO_2 signal gave correlation coefficients between 0.78 and 0.82. For overnight analysis, this rose slightly to a range of 0.82 to 0.86. Furthermore, an RDI based on the non-overlapping distributions of dominant pole angles, as illustrated in the histograms of Chapter 8, gave correlation coefficients of $\rho = 0.82$ (using the area of the apnoeic distribution), but $\rho = 0.67$ (using the ratio of the areas of the apnoeic to normal breathing distributions) due to severe patient outliers. To conclude:

- The identification of individual apnoeic segments via methods described in Chapters 6 through 8 leads to RDI values which are highly correlated with AHI values.
- The diagnosis of apnoeic severity is possible through the optimisation of RDI threshold values (using True Positives and True Negatives), thus distinguishing between OSAHS severity categories.
- The clinical viability of including such automated methods in OSAHS diagnoses must be measured against the potential for these computer methods to misdiagnose patients - which can be extremely dangerous, even deadly, should moderate-severe patients be misdiagnosed as normal-mild and left without adequate medical treatment.

10.1.1 Contributions

The contributions of this thesis may be paraphrased as follows:

- The results of using time domain, frequency domain, and AR-modelling with amplitude criterion methodologies have been presented altogether for the first time on the same (Bristol) database.
- Various window sizes, from 30 to 120 seconds, were utilised; whereas other published studies only use large window sizes (such as 15 minutes), a single (overnight) window, or no specified window size whatsoever.

- An extensive range of patients was examined (Bristol Database); i.e., the experts categorised 33 patients in the database as 'normal' (AHI: 0-<5 events per hour); 23 patients as 'mild' (AHI: 5-14 events per hour); 23 patients as 'moderate' (AHI: >14-30 events per hour); and 12 patients as 'severe' (AHI: >30 events per hour).

- The detection of apnoeic activity was investigated on the overall dataset, but also, for the first time, on a *per-category* basis.

- In addition to achieving overall classification accuracies above 80.0%, a high correlation was found between the automated-derived RDI and expert-derived AHI values, using *all three* aforementioned analysis methods.

In the context of other work, as previously mentioned, it is difficult to compare results accordingly. For example, in 1996, Lévy *et al.* used an oxygen saturation variability based index to detect apnoeic events in subjects with sleep apnoea syndrome, ruling out the diagnosis of OSAHS in 301 screened patients with only three negative test results [63]. Overall, Lévy *et al.* reported a sensitivity of 98% and specificity of 46% - values respectively ∼20% greater and ∼35% less than those sensitivity and specificity values reported in this thesis (using any of the three described methods). In 2000, Vázquez *et al.* produced an oximeter derived RDI that was highly correlated ($\rho = 0.97$) with the PSG derived AHI for 246 patients with suspected OSAHS. The AHI was created by a single trained observer using the R & K criteria for staging sleep, and the AASM definitions for identifying apnoeas and hypopnoeas. For each patient, the RDI was created by the retrospective analysis of the SpO_2 signal whereby an event was marked when the percentage of oxygen saturation value dropped below baseline, defined as the average of the SpO_2 signal of the five minutes preceding the marked event. Patients were classified as OSAHS or non-OSAHS based on their new RDI values. The almost perfect correlation reported between AHI-RDI values is undermined by the lack of available information regarding patient database utilised; i.e., the severity of each patient is unknown. If all 246 patients suffered from severe OSAHS, a high correlation coefficient would indeed be

expected; therefore, a comparison between ρ values (from their study and this thesis) is not entirely possible.

10.2 Further work

10.2.1 Obstructive vs. central apnoeas

This thesis has investigated the feasibility of detecting apnoeic and normal breathing activity from the non-invasive SpO_2 signal. Although it is shown that comparable classification results are achieved using the time, frequency, and AR-modelling with amplitude criterion analyses, the method with the most potential is the latter. Using the dominant pole tracking identification technique, described in Chapter 8, it may be possible to differentiate between obstructive and central apnoeas (see Section 3.4). In their 1996 study, as introduced in Section 4.8.1, Shiomi *et al.* reported a significant increase in very low frequency activity (0.008-0.04 Hz, or 25-120 seconds) within the HRV and SpO_2 signals for 12 patients with OSAHS, and a subsequent decrease upon treatment with prosthetic mandibular advancement [102]. According to their study, the presence of high-frequency signals after the onset of very low-frequency signals indicates the occurrence of inspiratory efforts during obstructive apnoeas. The absence of high-frequency signals indicates the occurrence of central apnoeas [102]. Thus, the tracking of dominant frequencies on a per segment basis may be ideal for this purpose.

10.2.2 Apnoeas and hypopnoeas

An argument for the differentiation of apnoeas and hypopnoeas can also be made accordingly. A hypopnoea is defined as a reduction in airflow (Section 3.4); i.e., a desaturation-resaturation cycle is typically visible in the SpO_2 signal and breathing continues. Therefore, hypopnoeas are ideally represented by **both** the apnoeic and normal breathing frequencies - or the 'mix' class of segments. In Section 8.3, it was noted that 75.4% of the

data segments for the 47-patient Bristol test database were labelled as 'mixed' - a majority of which may potentially represent hypopnoeic activity. To differentiate between apnoeas and hypopnoeas, the distance from the pole to the unit circle could be explored further. The closer a pole lies to the unit circle, as measured by pole radius r, the stronger its resonant frequency is; i.e., a study of pole strength values with respect to class (apnoea, hypopnoea) would first be required. The exclusive use of the SpO_2 signal to differentiate between SDB events has also been touched upon in [78]. Netzer *et al.* showed how the length of oxygen desaturation in the SpO_2 signal can help to distinguish desaturations due to COPD (Chronic Obstructive Pulmonary Disorder[1]) from those caused by obstructive apnoeas or hypopnoeas. Specifically, the desaturations secondary to COPD tended to last much longer and have a much lesser degree of slope in the SpO_2 waveform. (This differentiation is important for the diagnosis of OSAHS where COPD is present, which is also known as the 'overlap syndrome' [78].)

10.2.3 Event duration

Clinically, patients with severe OSAHS may experience *both* a higher incidence of events and events of longer duration over the course of the night - see Section 3.6. With the time domain analysis described in this thesis, it would be possible to identify event duration in the apnoeic segments by measuring the peak-to-trough interval, and plotting the distribution of event duration during the night using a histogram. Alternatively, it would be possible to compute the period (and subsequently the duration) of the desaturation-resaturation cycles using the dominant frequency information for each segment as provided by the AR-modelling with amplitude criterion technique. The distributions of duration could then be compared with the distribution of event duration from the Expert Scores, and the correlation between event duration and apnoeic severity could then be explored further.

[1]COPD is a lung disease in which the airways are partially obstructed; i.e., damaged, making it difficult to breathe.

10.2.4 Other work

Prior to any further research work, however, it would be useful to extend the scope of the data available in size, to ensure statistically meaningful results. When collecting the data, it is recommended that more stringent rules be applied for scoring such that the expert scores may be easily replicated in other clinics to ensure a more uniform analysis in future; i.e., some clinics have more resources than others in terms of the signals they may acquire during PSG. Particular attention would also be required to ensure that OSAHS severity is addressed; i.e., automated methods need to be tested on patient databases including the full spectrum of OSAHS (from normal to severe cases). It would also be interesting to investigate the relationship between the initial, at-home, screening tests using pulse oximetry (e.g., 'Download 2001' software used in Bristol) on patients with suspected OSAHS and the resultant PSG records. For example, how do the initial screening (at-home) test results compare with those test results found using PSG in the sleep clinics?

As experts score events in real-time, the methods presented in this thesis may benefit from the use of either adaptive windows for the identification of apnoeic activity - with window sizes changing according to apnoeic activity duration (which changes throughout the night). The 'optimal' 120-second window found in this work is only optimal (as previously discussed) within the set of four window sizes examined (30-second, 60-second, 90-second, 120-second). The advantage of combining results from several different window sizes is such that a more representative picture of apnoeic activity may be provided. Additionally, merging of automated scores has been investigated with success in [17,125], which may prove useful in future. Lastly, the optimisation of the RDI threshold values could be based on False Positives and False Negatives, instead of True Positives and True Negatives (or a combination thereof). For example, one idea is to compare the mean σ_{SpO_2} values for segments classified as FP and FN, and determine any potential correlation to apnoeic severity. (Optimisation of the computed RDI threshold values could follow thereafter.)

Analysis of airflow signal

Establishing the relationship between airflow cessation/resumption and oxygen desaturation/resaturation would be most advantageous in the accurate identification of apnoeic events. According to the AASM, the definition of an apnoea is the complete or near complete cessation of airflow that lasts for at least 10 seconds or longer [3] - as described in Section 3.4. Thus, the identification of apnoeic activity is possible using either the oronasal airflow or respiratory effort signals acquired during PSG. First, the chosen airflow signal is (low-pass) filtered to acquire the envelope of the waveform. Next, a threshold is applied whereby waveform values falling below correspond to apnoea, and values falling above correspond to normal breathing. With an optimal filter choice and threshold value set using ROC analysis on a training dataset, this type of analysis could be employed and tested on an independent dataset - such as that used in this thesis. The difficulty with the envelope-based method is such that the airflow signals do not always flatten to zero during apnoeic activity, according to definition, but reduce to some low level which is determined subjectively by experts. Also, this low level constantly shifts over time. Despite such limitations, further understanding of breathing (or lack thereof) in patients with OSAHS is imperative to establish the quantitative relationship between the airflow and SpO_2 signals.

10.3 Home monitoring using the SpO_2 signal

In this thesis, the SpO_2 signal has been highlighted for its potential usefulness in the clinical diagnosis of OSAHS. Using pulse oximetry at home for monitoring of OSAHS has four major advantages. First, pulse oximetry is readily available and relatively inexpensive. Second, the equipment is simple to use and fairly unobtrusive; i.e., the hardware sits on the tip of the finger only. Third, the home environment is more conducive to a 'typical' night's sleep for subsequent analysis [6]. (For a patient suffering from some degree of sleep disorder, a hospital visit may prove disruptive.) Fourth, recordings can be repeated at

no extra cost (both monetary and time). For example, if the hardware fell off during the course of the night, the recording may proceed on the following day without further delay (instead of a new booking at a sleep clinic). The major disadvantage of home monitoring using the SpO_2 signal is such that one signal may not be enough, or at least is not enough at the present time, to differentiate between apnoea types. For example, two signals are required to distinguish central apnoea from obstructive apnoea. Furthermore, and for future consideration, some sleep disorders may not be altogether detectable using the SpO_2 signal alone. For example, in patients with periodic limb movements (PLMs), the movement does not always create artefact in the SpO_2 signal nor do the SpO_2 levels necessarily change throughout the course of the night (oxygen levels in the blood may remain stable). Full-night PSG benefits from the inclusion of multiple channels of information for a more complete diagnosis. However, several methods have been proposed - such as in this work - which make use of the SpO_2 signal to automatically quantify apnoeic behaviour and do so in high agreement with the experts. Techniques developed in this thesis have potential for home scoring of the OSAHS, especially if robustness can be proven on a more extensive patient group.

Appendix A

Pulse Oximetry: Background

A.1 Haemoglobin as oxygen carrier

In 1864, Stokes discovered that the coloured pigment in blood was also its carrier of oxygen [62]. That same year, Hoppe-Seyler purified the pigment and called it haemoglobin [77]. One red blood cell contains approximately 265 million haemoglobin molecules [20, 25]. Haemoglobin is composed of four haeme subunits, each containing a polypeptide chain (globin) and one ferrous (Fe^{2+}) iron atom, and plays an essential role in transporting the necessary amount of oxygen to the body. Normal oxygen transport to the body includes the following processes: binding, diffusion, release [96].

First, gases are not particularly soluble in blood (which is composed mostly of water); haemoglobin, however, provides a binding mechanism that allows oxygen to be transported through the blood [96]. One haemoglobin unit can carry one molecule of oxygen. As haemoglobin is composed of four haeme subunits, one haemoglobin molecule can carry, or be bound to, four molecules of oxygen. Once a completely deoxygenated haemoglobin molecule binds with one oxygen molecule, it has a greater affinity for the second oxygen molecule - and each additional oxygen molecule [20]. This binding process takes place in the relatively high oxygen pressures of the alveoli. When all four iron atoms are oxidised, the haemoglobin molecule is referred to as oxygenated haemoglobin, or **oxyhaemoglobin,**

HbO_2. The process is impaired in two common forms of the haemoglobin molecule: carboxyhaemoglobin (HbCO) and methaemoglobin (metHb) [62]. HbCO is formed when carbon monoxide, CO, is bound to the haemoglobin molecule; metHb is formed when the blood is exposed to various drugs or oxidising agents, and the ferrous iron (Fe^{2+}) in the Hb molecule is converted to (Fe^{3+}) [28].

Second, the transfer of HbO_2 to tissues occurs through diffusion, or the net movement of particles from areas of higher partial pressure to areas of lower partial pressure[1] [20]. Diffusion occurs over a distance of approximately 1 mm; thus, HbO_2 does not diffuse until it reaches the capillaries with a wall thickness equal to one cell, even if the HbO_2 passes through oxygen depleted tissue. Specifically, HbO_2 diffuses through the cell membrane into the interstitial fluid of the cell (whilst carbon dioxide diffuses from the cell into the blood). The overall process of ventilation, or gas exchange, also requires diffusion. After oxygenated blood has been circulated through the body, it is brought back to the lungs through arterial capillaries to receive oxygen and rid itself of carbon dioxide through diffusion in the alveoli. Blood is then reoxygenated and recirculated through the body [96].

Third, when the partial pressure of oxygen in the surrounding tissue falls below 20 mmHg, the oxygen is released to the nutritional vessels. For example, during exercise, oxygen is released from storage in myoglobin. Myoglobin is a respiratory pigment found in muscles, and is responsible for the reddish brown colour of muscle cells. (Its affinity for oxygen is greater than that of haemoglobin [20, 49].) Once the first oxygen molecule is released from the haemoglobin molecule, it is more likely to release the second oxygen molecule. The removal of oxygen is called reduction, and a haemoglobin molecule completely devoid of oxygen is referred to as deoxygenated haemoglobin, or **reduced haemoglobin, Hb**.

[1]Partial pressure is defined as the pressure that would be exerted by one of the gases in a mixture (i.e., oxygen) if it occupied the same volume on its own (expressed in mmHg) [1].

A.2 History of oximetry

The first instrument to measure oxygen saturation in human blood, by transilluminating it with coloured light, was built by Matthes in 1935. Such instrumentation gave rise to the term *in vivo* oximetry, where oxygen saturation could be analysed in the blood directly. (Prior to the 1930's, blood samples had to be withdrawn out of the body for *in vitro* analysis only.) Instruments of this type could reliably follow trends in oxygen saturation, but could not make absolute measurements and were incapable of distinguishing between arterial, venous, and capillary blood [62]. Matthes was the first to use two independent measurements (wavelengths of light) to measure the two unknowns: oxygenated haemoglobin molecule (HbO_2) and deoxygenated haemoglobin molecule (Hb) [5, 32, 79].

Attempts to exclude venous and capillary from arterial blood focused on the following two methods: arterialisation of the blood, and taking a 'bloodless' reading. Millikan was the first to coin the term *oximeter* to describe 'a spectrophotometric instrument for measuring continuously the oxygen saturation in arterial blood without puncture of the vessel' [75]. In 1942, Millikan attempted to measure the amount of oxygen saturated in haemoglobin at the ears of pilots, in order to control the amount of oxygen they received whilst flying [32, 84]. The overall project sought to counteract the loss of consciousness experienced by pilots during dog-fights in World War II. The compact and lightweight instrument included a small earpiece with red and green filters, and an electrode that required the skin to be heated to 43°C in order for the blood to arterialise; i.e., the arterial blood content in tissues is increased as a result of thermal dilation of vessels and a subsequent increase in capillary blood flow [77]. Unfortunately, this process posed a risk of burns and the instrument needed to be recalibrated every four hours [28].

In 1949, Wood made an improvement to Millikan's device by adding an inflatable balloon to the earpiece in order to compress the tissue, cut off the circulation to the ear, and acquire a 'bloodless' reading from the earlobe [5, 32]. This absence of blood, or zero-setting, was used for device calibration; however, the device's success was short-

lived as Wood later showed that there were inconsistencies in the photocells used for light detection, showing the device to be inaccurate [32, 62, 84].

In 1964, Shaw created a self-calibrating, 8 wavelength ear oximeter to measure oxygen saturation. In 1970, the device became the Hewlett Packard (HP) ear oximeter, only to be further adapted with fiberoptic catheters by Oximetrics [5, 62]. HP produced a multi-component ear model composed of m independent light-absorbing substances (skin, tissues, HbO_2, Hb). Simultaneous equations were written for the total absorbance of light by the m substances at each measurement wavelength. Empirical calibration coefficients were derived from a number of studies on a sample of volunteers [110]. In the late 1970's, more vigorous testing took place, and inaccuracies were demonstrated at low SaO_2 values (less than 70%) [62, 118]. In addition, the bulk of the equipment, high cost of the instrument, and need for measurements at 8 different wavelengths (and regular calibration) meant the device never found regular clinical use.

In 1971-2, Aoygi realised that light transmission in human tissue varied with arterial pulsations. He postulated that cardiac synchronous variations in light intensity reaching the detector were entirely attributable to blood volume changes in small arteries and arterioles; therefore, comparing the pulsating portions of red/infrared signals could yield absorption data for arterial blood alone. A patent, filed by the Nikon Kohden Company of Japan, described the successful isolation of the relative absorption of red and infrared light by arterial blood only, with the absorption of venous blood and other tissues being constant or slowly varying [62, 77]. Furthermore, the Nikon device was modified, using microprocessor control and human volunteer calibration data, to produce the first commercially available pulse oximeter with clinically acceptable accuracy (1975) [62].

Appendix B

Comparison with Expert Scores

'Automated' refers to the output of the signal processing technique, 'manual' refers to Expert Scores (the *gold-standard*), and 'event' refers to a hypopnoea/apnoea(obstructive, central, mixed). The confusion matrix of Table B.1 shows the comparison between 'automated' and 'manual' decisions, with the metrics True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Sensitivity (Se) is the true-positive fraction, or positivity in the presence of a disease or condition (i.e., apnoea). Specificity (Sp) is the true-negative fraction, or the fraction of normal breathing (i.e., non-apnoea) segments correctly detected. There is a definite trade-off between Se and Sp as illustrated in the Receiver Operating Characteristic (ROC) curve of Figure B.1, which is a plot of Se in the y-axis versus 1-Sp on the x-axis as the decision threshold is varied. Each point on the ROC plot represents a sensitivity/specificity pair corresponding to a particular decision threshold [127]. The area under the curve (AUC) is equal to classification Accuracy, or the total percentage of correctly classified events (positive and negative results) - see Equation B.3.

Classification with no discriminatory power will yield a straight line at an angle of 45^o from the x-axis (from bottom left to top right), and an area under the curve (accuracy) of 0.5 corresponding to a random predictor yielding an equal number of true positives and false positives for a varying threshold.

$$
\begin{aligned}
TP &= \text{true positive, both manual and automated systems agree;} \\
TN &= \text{true negative, both manual and automated systems agree;} \\
FP &= \text{false positive, event only scored by automated system;} \\
FN &= \text{false negative, event only scored by manual system.} \\
Se &= P(\text{an event has been detected}|\text{an event has occurred}) \\
&\approx \frac{TP}{TP+FN} \qquad \text{(B.1)} \\
Sp &= P(\text{an event has not been detected}|\text{an event has not occurred}) \\
&\approx \frac{TN}{TN+FP} \qquad \text{(B.2)} \\
Accuracy &= \frac{TP+TN}{TN+TP+FN+FP} \qquad \text{(B.3)}
\end{aligned}
$$

		Automated	
		Apnoea	Normal Breathing
Manual	Apnoea	TP	FN
	Normal Breathing	FP	TN

Table B.1: Confusion matrix describing the metrics corresponding to the 'manual' and 'automated' decisions.

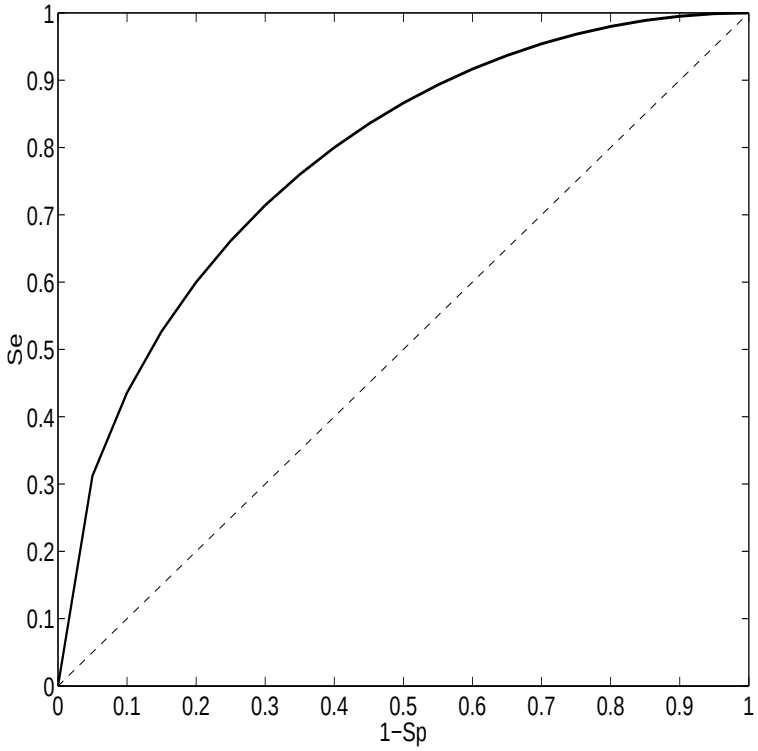


Figure B.1: A Receiver Operating Characteristic (ROC) curve, solid line, where the area under the curve (AUC) is equal to classification accuracy. Classification with no discriminatory power is shown by the dotted line at an angle of 45° from the x-axis.

Appendix C

AR Modelling & Pole-Tracking

C.1 Autoregressive Modelling

Parametric modelling is used for time series analyses, where a mathematical model is fitted to a sampled signal. If the model forms a good approximation to the signal's observed behaviour; i.e., a more compact representation of the sampled signal is found that encapsulates the salient properties, then it can be used for a wide range of applications - such as spectral estimation and feature extraction for pattern classification problems [82, 83].

Models can either be adaptive, where the free parameters are updated with the arrival of each new data sample; or non-adaptive, where the parameters are chosen to give the best fit to a sequence of data samples [82]. For non-adaptive modelling, a best-fit model may only be computed if the signal is stationary; i.e., signal characteristics - such as mean and frequency content - must not vary with time. Most biomedical signals are non-stationary (i.e., they have a time-varying frequency spectrum), but can be considered locally stationary over time intervals [82]. For such signals, either an adaptive model can be used, or the signal can be divided into sufficiently short, quasi-stationary segments and a non-adaptive model fitted to each segment [82]. The key to the performance of modelling techniques lies in the algorithms used for free parameter estimation. The most popular

algorithms include Kalman filtering (adaptive models), Levinson-Durbin and Burg (non-adaptive models).

There are three types of parametric models, all based on a mathematical rational transfer function whose exact form is determined by an estimation of the model's free parameters. Respectively, if the free parameters in the transfer function lie in the denominator, numerator, or both the numerator and denominator, the models are all-pole or autoregressive (AR), all-zero or moving-average (MA), and pole-zero or autoregressive moving-average (ARMA) models. In our study, the SpO_2 signals are non-stationary, but may be segmented into stationary epochs. The non-adaptive AR-model is used to analyse these *piecewise stationary signals* using the Burg algorithm, which is relatively simple and reliable [82].

The AR-modelling technique can be formulated in the frequency domain (as a spectral matching problem) or in the time domain (as a linear prediction problem). The time domain approach is described here, where it is assumed that the value of the current sample s_n, in a data sequence $s_{n-1}, s_{n-2}, \ldots, s_N$, can be predicted as a linearly weighted sum of the p most recent sample values $(s_{n-1}, s_{n-2}, \ldots, s_{n-p})$ [82]. The model order p (number of terms in the AR model) is generally chosen to be much smaller than the sequence length N. In addition, the selected model order must not be too low nor too high. In the former case, a low order creates an excessively smooth power spectral density (PSD) estimate where some peaks may be obscured; i.e., all resonant frequencies may not be estimated accurately. In the latter case, a high order introduces spurious peaks and spectral instability; i.e., irrelevant frequency content may be introduced into the data due to noise or "pole-splitting" [18, 51, 56]. The predicted value of s_n is denoted as $\tilde{s_n}$, expressed as follows:

$$\tilde{s_n} = -(a_1 s_{n-1} + a_2 s_{n-2} + \ldots + a_p s_{n-p}) \tag{C.1}$$

$$= -\sum_{k=1}^{p} a_{pk} s_{n-k} \tag{C.2}$$

where the weight, a_{pk}, denotes the k^{th} AR coefficient of the p^{th}-order model and $n = 1, 2, \ldots, N$. This is depicted in Figure C.1. The goodness-of-fit of this model to the data samples is dependent upon the values of the AR coefficients and to a lesser extent on the number of coefficients, or model order p.

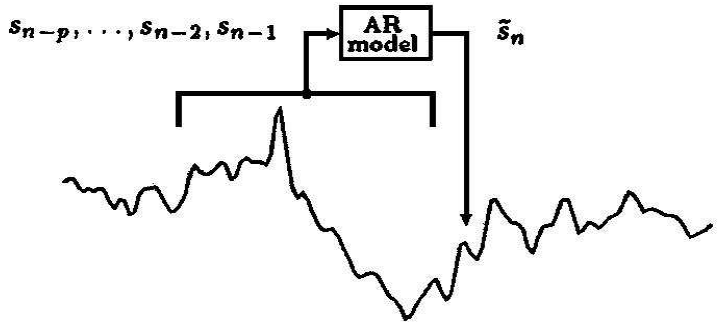


Figure C.1: Linear prediction using a pth order autoregressive model [82].

The difference between the actual and predicted coefficient values is known as the forward prediction error, e_{pn} , and may be written as:

$$e_{pn} = s_n - \tilde{s_n} \tag{C.3}$$

$$= s_n + \sum_{k=1}^{p} a_{pk} s_{n-k} \tag{C.4}$$

The crucial estimation of coefficient values is dependent upon minimisation of the prediction error power, E_p , or more specifically the mean of the squared prediction error for the entire data sequence:

$$E_p = \frac{1}{N} \sum_{n=1}^{N} (e_{pn})^2 \tag{C.5}$$

$$= \frac{1}{N} \sum_{n=1}^{N} (s_n - \tilde{s_n})^2 \tag{C.6}$$

$$= \frac{1}{N} \sum_{n=1}^{N} (s_n + \sum_{k=1}^{p} a_{pk} s_{n-k})^2 \tag{C.7}$$

Minimisation of E_p is achieved through the Burg algorithm, as described in [82]. Overall,

given a segment $(s_{n-1}, s_{n-2}, \ldots, s_{n-p})$, the Burg algorithm returns an optimal set of coefficients $(a_{p1}, a_{p2}, \ldots, a_{pp})$ along with the corresponding value of E_p . If the AR-model forms a good approximation of the segment, then it is can be used to estimate the segment's frequency spectrum, $S(f)$, given by:

$$|\tilde{S}(f)| = \frac{E_p}{\left[1 - \sum_{k=1}^{p} a_i \cos\left(\frac{2\pi k f}{f_s}\right)\right]^2 + \left[\sum_{k=1}^{p} a_k \sin\left(\frac{2\pi k f}{f_s}\right)\right]^2} \tag{C.8}$$

where f_s is the sampling frequency and $0 \leq f \leq f_s/2$ Hz [83]. The AR coefficients in the denominator determine the shape of the spectrum; whereas the E_p in the numerator acts as a scaling factor.

C.2 Pole-Tracking Technique

Equation C.4 may be rewritten as follows, where the current value of the output signal s_n is a linear combination of the p previous values of the output signal s_{n-k} for $k = 1, 2, \ldots, p$ and the current value of the input signal, e_{pn}:

$$s_n = -\sum_{k=1}^{p} a_{pk} s_{n-k} + e_{pn} \tag{C.9}$$

Thus, s_n is a linear regression on itself and e_{pn} is the error of regression; i.e., AR-modelling is identical to linear prediction. This discrete-time process (input signal e_{pn}, output signal s_n) is modelled as a rational transfer function $H(z)$ that relates the input signal to the output signal:

$$H(z) = \frac{1}{\sum_{k=1}^{p} a_{pk} z^{-k}} \tag{C.10}$$

$$= \frac{z^p}{(z - z_1)(z - z_2)\ldots(z - z_p)} \tag{C.11}$$

$$= \frac{z^p}{\mathbf{P}_1 \mathbf{P}_2 \ldots \mathbf{P}_p} \tag{C.12}$$

The denominator of $H(z)$ can be factored into p terms, where $\mathbf{P}_1\mathbf{P}_2\ldots\mathbf{P}_p$ are vectors extending from any point z in the complex z-plane to each of the p poles of $H(z)$, namely $z_1, z_2, \ldots, z_p$. Each of these poles is a complex-conjugate pair, and are the roots of the denominator $H(z)$.

The PSD of the output signal, $P_s(z)$, is related to the PSD of the input signal $P_e(z)$, where the asterisk indicates the complex conjugate:

$$P_s(z) = H(z)H^*\left(\frac{1}{z^*}\right)P_e(z) \tag{C.13}$$

The signal driving the system, e_{pn}, is assumed to be a Gaussian white noise source with zero mean and variance σ^2; therefore, the PSD of the input signal is constant:

$$P_e(z) = \sigma^2\Delta t \tag{C.14}$$

where Δt, the sampling interval for both s_n and e_{pn}, is retained in order to specify the power in the original continuous-time signal [17]. The PSD of the output signal is then:

$$P_s(z) = H(z)H^*\left(\frac{1}{z^*}\right)\sigma^2\Delta t \tag{C.15}$$

The spectral content of the output PSD is assumed to be solely determined by the transfer function $H(z)$, as the input signal is assumed to be white noise. White noise has a constant PSD and is uncorrelated with all other signals by definition; thus, the output signal is assumed to be uncorrelated with the input signal. In many real-world applications, however, the input signal is not white and some correlation between input-output signals may exist, making it difficult to describe the input signal using an AR-model [17, 45, 56].

To evaluate the frequency response of the AR-model, $z = e^{j2\pi f\Delta t}$ can be substituted into $H(z)$ over the range $-\frac{1}{2\Delta t} \geq f \geq \frac{1}{2\Delta t}$. This is equivalent to evaluating $H(z)$ around

the unit circle in the complex z-plane:

$$P_s(f) = |H(z)|^2 \sigma^2 \Delta t \tag{C.16}$$

$$= \frac{\sigma^2 \Delta t}{|1 + \sum_{k=1}^{p} a_{pk} e^{-j2\pi f k \Delta t}|^2} \tag{C.17}$$

where $H(z) = H(e^{j2\pi f \Delta t})$. Each complex-conjugate pair of poles in $H(z)$ corresponds to one resonant frequency in $P_s(f)$. Therefore, a pth-order AR-model will have p poles and r resonant frequencies: $f_1, f_2, \ldots, f_r$ where $r = \frac{p}{2}$ when p is even and $r = \frac{p+1}{2}$ when p is odd (where the pth pole lies on the real axis). The values of each of the r resonant frequencies are given by the phase angle of the corresponding pole in the upper half of the complex z-plane.

To specify $P_s(f)$, the AR parameters $a_1, a_2, \ldots, a_p$ and σ^2 must be evaluated. The Yule-Walker equations relate the AR parameters to the autocorrelation sequences of s for lags $k = 0, \ldots, p$, which reduces the problem to the estimation of the autocorrelation sequence [17]. If N, the number of samples in the signal s, is much greater than p, the order of the AR-model, then $\tilde{R}_{ss}(k)$, a suitable biased estimate of the autocorrelation sequence of s for lags $k = 0, \ldots, p$, can be calculated as:

$$\tilde{R}_{ss}(k) = \frac{1}{N} \sum_{i=0}^{N-k-1} s_{i+k} s_i^* \tag{C.18}$$

The Yule-Walker equations can then be written as:

$$\begin{pmatrix} \tilde{R}_{ss}(0) & \tilde{R}_{ss}(-1) & \ldots & \tilde{R}_{ss}(-p) \\ \tilde{R}_{ss}(1) & \tilde{R}_{ss}(0) & \ldots & \tilde{R}_{ss}(-(-p-1)) \\ \vdots & \vdots & & \vdots \\ \tilde{R}_{ss}(p) & \tilde{R}_{ss}(p-1) & \ldots & \tilde{R}_{ss}(0) \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{C.19}$$

There are now p equations to solve for $a_1, a_2, \ldots, a_p$ and another to solve for σ^2. The Burg algorithm can be used to solve these equations recursively, while the less-efficient

method of Gaussian elimination can solve these equations in a much more straightforward manner [17]. In this Chapter, the Burg method fits the pth-order AR model to the input signal by minimising (least squares method) the forward and backward prediction errors, while constraining the AR parameters to satisfy the Levinson-Durbin recursion [73].

To specify the power at a resonant frequency f_c, the term $f = f_c$ can be substituted directly into Equation C.17. Alternatively, a graphical solution based on Equation C.12 can be used, in which $H(z)$ is estimated at A, the point on the unit circle in the complex z-plane with the phase angle of the cth-pole [53]. This point corresponds to the cth resonant frequency f_c, the power at which can be estimated using the following equation:

$$P_s(f_c) = |H(f_c)|^2 \sigma^2 \Delta t \tag{C.20}$$

$$= \frac{\sigma^2 \Delta t}{(|z - z_1||z - z_2| \ldots |z - z_p|)^2} \tag{C.21}$$

$$= \frac{\sigma^2 \Delta t}{(|\mathbf{P}_1||\mathbf{P}_2| \ldots |\mathbf{P}_p|)^2} \tag{C.22}$$

where $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_p$ now specify vectors extending from point A to each of the p poles. Figure C.2 shows five poles $z_1, z_2, \ldots, z_5$ in the complex z-plane, where (z_2, z_5) and (z_3, z_4) are complex-conjugate pairs. The point $z = A$ corresponds to the second resonant frequency, f_2, because it has a phase angle equal to that of the second pole z_2 and it is on the unit circle. As the magnitude of z_2 approaches unity, the length of the vector $\mathbf{P}_2$ approaches zero, and the power at f_2 approaches infinity (by Equation C.12). Therefore, $P_s(f)$ peaks at resonant frequencies whose corresponding poles have magnitudes near unity. Thus, the power at the resonant frequencies of output signal s can be estimated by tracking the magnitudes of the corresponding poles of its AR-model.

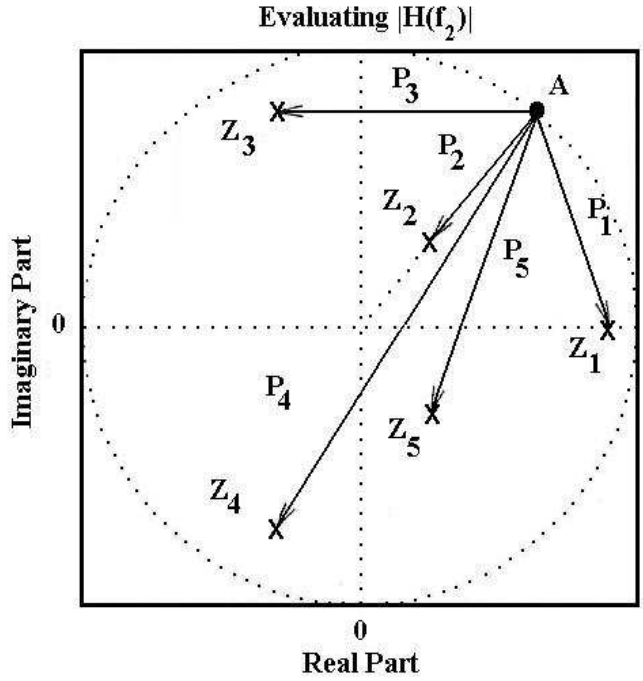


Figure C.2: Evaluating the magnitude of an AR transfer function at a resonant frequency: The magnitude of the transfer function can be specified in terms of $|\mathbf{P}_1||\mathbf{P}_2|\ldots|\mathbf{P}_5|$, where each vector $\mathbf{P}_k$ extends from point A on then unit circle with a phase angle equal to that of the second pole. As the magnitude of the second pole increases, the length of $\mathbf{P}_2$ decreases, leading to increasingly sharp peaks at the second resonant frequency in the magnitude of the transfer function and thus the PSD [17, 56].

C.3 Pole Assignment Algorithm

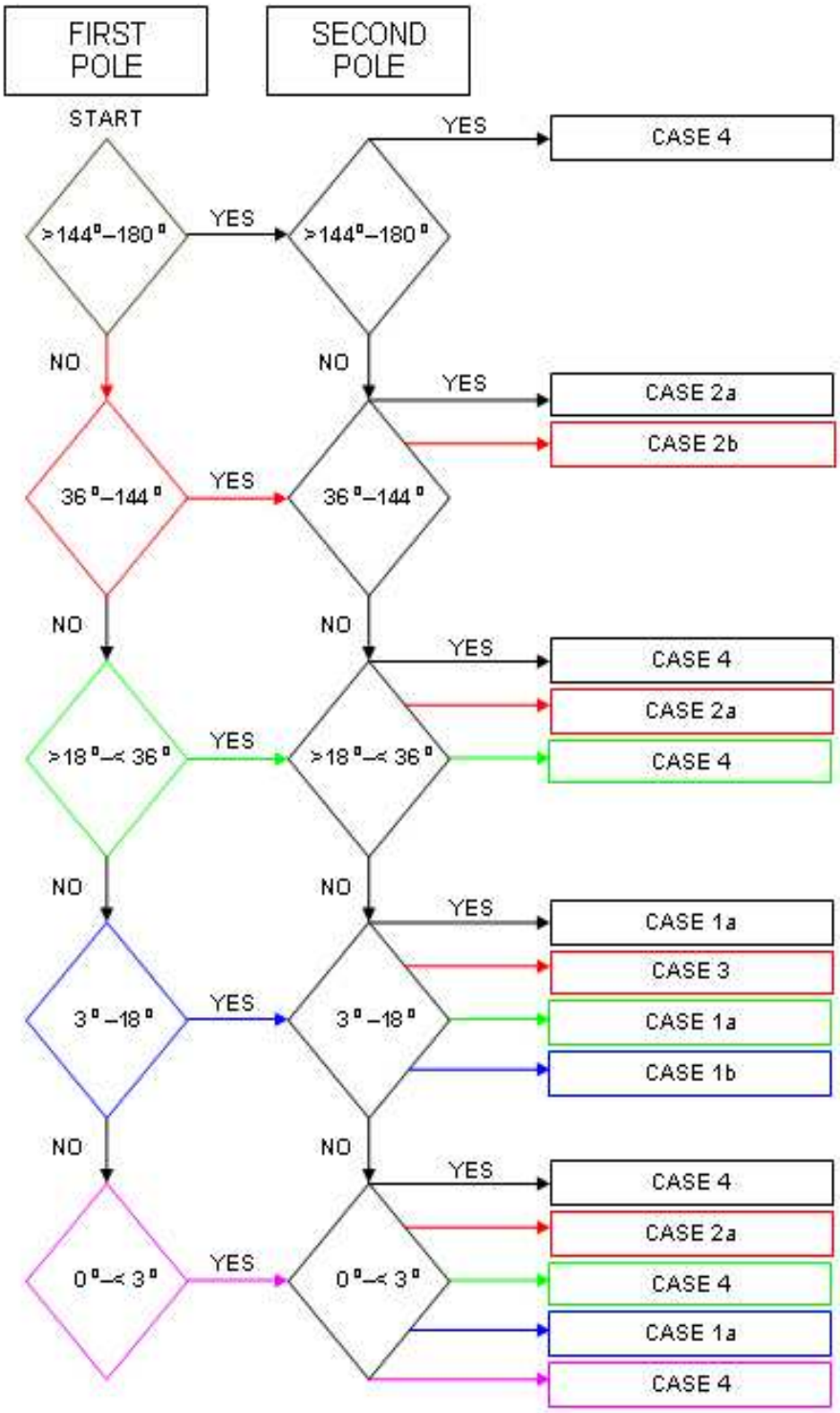


Figure C.3: Pole-zero assignment algorithm flowchart.

Bibliography

[1] In *Concise Oxford English Dictionary*, J. Pearsall, Ed. Oxford University Press, 2002.

[2] Ahrens, T., and Rutherford, K. *Essentials of Oxygenation*. Jones and Bartlett, Boston, USA, 1993.

[3] American Academy of Sleep Medicine (AASM) Task Force. *Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research*, 22 ed., 1999. no.5, pp.667-698.

[4] American Sleep Disorders Association (ASDA) and Sleep Research Society Atlas Task Force. *EEG arousals: scoring rules and examples*, 15 ed., 1992. no.2, pp.173-184.

[5] Astrup, P., and Severinghaus, J. *The history of blood gases, acids and bases*. Munksgaard International Publishers, Copenhagen, Denmark, 1986.

[6] Bennett, J., and Kinnear, W. Sleep on the cheap: the role of overnight oximetry in the diagnosis of sleep apnoea hypopnoea syndrome. *Thorax 54* (1999), 958–959.

[7] Bennett, L., Barbour, C., Langford, B., Stradling, J., and Davies, R. Health status in obstructive sleep apnea. *Am J Respir Crit Care Med 159* (1999), 1884–1890.

[8] Bennett, L., Langford, B., Stradling, J., and Davies, R. Sleep fragmentation indices as predictors of daytime sleepiness and ncpap response in osa. *Am J Resp Crit Care Med 158*, 3 (1998), 778–786.

[9] Berry, R. Medical therapy. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 5, pp. 90–118.

[10] Berry, R. Serotonergic medications. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 15, pp. 223–234.

[11] Bloch, K. Getting the most out of nocturnal pulse oximetry. *Chest 124* (2003), 1628–1630.

[12] Bradley, T., and Phillipson, E. Sleep disorders. In *Textbook of Respiratory Medicine*, J. N. J.F. Murray, Ed. W.B. Saunders Company, Philadelphia, 2000, ch. 81, pp. 2153–2169.

[13] Burr, R. Heart rate variability. In *Advanced Electrocardiography*, S. Anderson, R. Burr, W. Downs, C. Jacobson, P. Lander, G. Massumi, D. Mirvis, J. Perry, and C. Rizo-Patron, Eds. SpaceLabs Medical, Inc., Washington, 1995, ch. 6, pp. 86–110.

[14] Carlson, J., Hender, J., Elam, M., Ejnell, H., Sellgren, J., and Wallin, B. Augmented resting sympathetic activity in awake patients with obstructive sleep apnea. *Chest 103* (1993), 1763–1768.

[15] Carskadon, M., and Dement, W. Normal human sleep: an overview. In *Principles and practice of sleep medicine*, M. Kryger, T. Roth, and W. Dement, Eds. Elsevier Saunders, Philadelphia, USA, 2005, ch. 2, pp. 13–23.

[16] Caton, R. The electric currents of the brain. *British Medical Journal 2*, 278 (1875).

[17] Cazares, S. *Automated identification of abnormal patters in the intrapartum cardiotocogram*. Doctorate of philosophy thesis, University of Oxford, 2002. Michaelmas Term.

[18] Cazares, S., Moulden, M., Redman, C., and Tarassenko, L. Tracking poles with an autoregressive model: a confidence index for the analysis of the intrapartum cardiotocogram. *Medical Engineering & Physics 23* (2001), 603–614.

[19] Chiner, E., Signes-Costa, J., Arriero, J., Marco, J., Fuentes, I., and Sergado, A. Nocturnal oximetry for the diagnosis of the sleep apnoea hypopnoea syndrome: a method to reduce the number of polysomnographies? *Thorax 54* (1999), 968–971.

[20] Clark, S. Normal oxygen transport. In *Design of Pulse Oximeters*, J. Webster, Ed. Institute of Physics Publishing, UK, 1997, ch. 1, pp. 1–12.

[21] Clifford, G. *Signal processing methods for heart rate variability analysis.* Doctorate of philosophy thesis, University of Oxford, 2002. Michaelmas Term.

[22] Cooper, B., Veale, D., Griffiths, C., and Gibson, G. Value of nocturnal oxygen saturation as a screening test for sleep apnoea. *Thorax 46* (1991), 586–588.

[23] Cooper, B., Veale, D., Griffiths, C., and Gibson, G. Value of nocturnal oxygen saturation as a screening test for sleep apnoea. *Thorax 46* (1991), 586–588.

[24] Culebras, A. Normal sleep physiology and neurobiology. In *Sleep Medicine*, T. Lee-Chiong Jr., M. Sateia, and M. Carskadon, Eds. Hanley & Belfus, Inc., Philadelphia, USA, 2002, ch. 1, pp. 1–6.

[25] Curtis, H., and Barnes, N. *Biology*, 5 ed. Worth, New York, USA, 1989.

[26] Daniels, J., Cayton, R., Chappell, M., Tjahjadi, T., and Cadosa. A fuzzy expert system for differential diagnosis of obstructive sleep apnoea and related conditions. *Expert Systems with Applications 12*, 2 (1997), 163–177.

[27] Davies, L., Gladding, P., and Goodwin, H. Download 2001. http://www.stowood.co.uk, Aug 2008.

[28] de Kock, J. *Pulse oximetry: theoretical and experimental models.* Doctorate of philosophy thesis, University of Oxford, 1991. Michaelmas Term.

[29] Deegan, P., and McNicholas, W. Predictive value of clinical features for the obstructive sleep apnea syndrome. *Eur Respir J 9* (1996), 117–124.

[30] Douglas, N., Thomas, S., and Jan, M. Clinical value of polysomnography. *Lancet 339* (1992), 347–350.

[31] Dyro, F. *The EEG handbook.* Little, Brown and Company, London, UK, 1989.

[32] Farmer, J. Blood oxygen measurement. In *Design of Pulse Oximeters*, J. Webster, Ed. Institute of Physics Publishing, UK, 1997, ch. 2, pp. 21–39.

[33] Farney, R., Walker, L., Jensen, R., and Walker, J. Ear oximetry to detect apnea and differentiate rapid eye movement (rem)and non-rem (nrem) sleep - screening for the sleep apnea syndrome. *Chest 4* (1989), 533–539.

[34] Farre, R., Montserrat, J., and Ballester, E. Importance of the pulse oximeter averaging time when measuring oxygen desaturation in sleep apnea. *Chest 21* (1998), 386–390.

[35] Findley, L., Wilhoit, S., and Suratt, P. Apnea duration and hypoxemia during rem sleep in patients with obstructive sleep apnea. *Chest 87* (1985), 432–435.

[36] Flemons, W., Littner, M., Rowley, J., Gay, P., Anderson, W., Hudgel, D., McEvoy, R., and Loube, D. Home diagnosis of sleep apnea: a systematic review of the literature. *Chest 124*, 4 (Oct 2003), 1543–1579.

[37] Fletcher, E., Costarangos, C., and Miller, T. The rate of fall of arterial oxyhemoglobin saturation in obstructive sleep apnea. *Chest 96*, 4 (Oct 1989), 717–722.

[38] Fletcher, E., Miller, J., Schaaf, J., and Fletcher, J. Urinary catecholamines before and after trachoestomy in patients with obstructive sleep apnea and hypertension. *Sleep 10* (1987), 35–44.

[39] George, C., Millar, T., and Kryger, M. Identification and quantification of apneas by computer-based analysis of oxygen saturation. *American Review of Respiratory Disease 137*, 5 (1988), 1238–1240.

[40] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C.-K., and Stanley, H. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals*, circulation 101, v.23 ed., Jun 13 2000. e215-e220, Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[41] Golpe, R., Jiménez, A., Carpizo, R., and Cifrian, J. Utility of home oximetry as a screening test for patients with moderate to severe symptoms of obstructive sleep apnea. *Sleep 22*, 7 (1999), 932–937.

[42] Gould, G., Whyte, K., Rhind, G., Airlie, M., Catterall, J., Shapiro, C., and Douglas, N. The sleep hypopnoea syndrome. *American Review of Respiratory Disease 137* (1988), 895–898.

[43] Guilleminault, C., Connolly, S., and Winkle, R. Cyclical variation of the heart rate in sleep apnoea syndrome. *Lancet 1* (1984), 126–131.

[44] Gyulay, S., Olson, L., Hensley, M., King, M., Allen, K., and Saunders, N. A comparison of clinical assessment and home oximetry in the diagnosis of obstructive sleep apnea. *American Review of Respiratory Disease 147* (1993), 50–53.

[45] Hayes, M. *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, Inc., USA, 1996.

[46] Hedner, J., Ejnell, H., Sellgren, J., Hedner, T., and Wallin, G. Is high and fluctuating muscle sympathetic nerve activity in the sleep apnoea syndrome of pathogenic importance for the development of hypertension? *Hypertens 6* (1988), 529–531.

[47] Hilton, M., Bates, R., Godfrey, K., Chappell, M., and Cayton, R. Evaluation of frequency and time-frequency spectral analysis of heart rate variability as a diagnostic marker of the sleep apnoea syndrome. *Medical & Biological Engineering & Computing 37* (1999), 760–769.

[48] Hochban, W. Skeletal facial corrections. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 12, pp. 193–209.

[49] Hole Jr., J. *Human Anatomy and Physiology*, 2 ed. Brown, Iowa, USA, 1981.

[50] Hutton, P., and Clutton-Brock, T. The benefits and pitfalls of pulse oximetry. *Brit. Med. J. 307* (1993), 457–458.

[51] Ifeachor, E., and Jervis, B. *Digital Signal Processing - A Practical Approach*, 2 ed. Prentice Hall, London, 2002.

[52] Jasani, R., Sanders, M., and Strollo Jr., P. Diagnostic studies for sleep apnoea/hypopnoea. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 4, pp. 57–85.

[53] Johnsen, S., and Andersen, N. On power estimation in maximum entropy spectral analysis. *Geophysics 43*, 4 (June 1978), 681–690.

[54] Johnson, J., and Gluckman, J. Radiofrequency tissue reduction. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 13, pp. 205–210.

[55] Johnson, J., and Gluckman, J. Tracheotomy. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 8, pp. 146–161.

[56] Kay, S., and Marple, S. Spectrum analysis - a modern perspective. *Proceedings of the IEEE 69* (1981), 1380–1419.

[57] Keyl, C., Lemberger, P., Pfeifer, M., Hochmuth, K., and Geisler, P. Heart rate variability in patients with daytime sleepiness suspected of having sleep apnoea syndrome: a receiver-operating characteristic analysis. *Clinical Science 92* (1997), 335–343.

[58] Koenig, S., and Suratt, P. Obstructive sleep apnoea: the syndrome. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 1, pp. 3–20.

[59] Kramer, K., Elam, J., Saxton, G., and Elam, W. Influence of oxygen saturation, erythrocyte concentration and optical depth upon the red and near-infrared light transmittance of whole blood. *Am. J. Physiol. 165* (1951), 229–246.

[60] Lavie, P., Pillar, G., and Malhotra, A. *Sleep Disorders: Diagnosis, management and treatment A handbook for clinicians*. Martin Dunitz Ltd., London, 2002.

[61] Lazareck, L., and Tarassenko, L. Detection of apnoeic and breathing activity through pole-zero analysis of the spo_2 signal. *Proc.28th Annual International Conference of the IEEE EMBS* (Aug-Sept 2006), 3879–3882.

[62] Leahy, M. *Biomedical instrumentation for monitoring microvascular blood perfusion and oxygen saturation*. Doctorate of philosophy thesis, University of Oxford, 1995. Trinity Term.

[63] Lévy, P., Pépin, J., Deschaux-Blanc, C., Paramelle, B., and Brambilla, C. Accuracy of oximetry for detection of respiratory disturbances in sleep apnea syndrome. *Chest 109*, 2 (Feb 1996), 395–399.

[64] Li, K., and Powell, N. Hypopharyngeal airway surgery. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 11, pp. 183–192.

[65] Loube, D., and Strohl, K. Obstructive sleep apnea syndrome: screening and diagnosis. In *Breathing Disorders in Sleep*, W. McNicholas and E. Phillipson, Eds. W.B. Saunders, UK, 2002, ch. 6, pp. 86–102.

[66] Magalang, U., Dmochowski, J., Veeramachaneni, S., Draw, A., Mador, M., El-Solh, A., and Grant, B. Prediction of the apnea-hypopnea index from overnight pulse oximetry. *Chest 124* (2003), 1694–1701.

[67] Magee, K., and Youngerman-Cole, S. Sleep apnea: normal and blocked upper airways. http://healthinformation.centracare.com/library/healthguide/, Aug 2006.

[68] Magee, K., and Youngerman-Cole, S. Counting respiration rate. http://www.webmd.com/hw-popup/Counting-respiration-rate, Aug 2007.

[69] Mak, V. False reassurance of pulse oximetry. *Brit. Med. J. 307* (1993), 732–733.

[70] Marcos, J., Hornero, R., Álvarez, D., Del Campo, F., and López, M. Applying neural network classifiers in the diagnosis of the obstructive sleep apnea syndrome from nocturnal pulse oximetric recordings. *Proc.29th Annual International Conference of the IEEE EMBS 7* (Aug 2007), 5174–5177.

[71] Marrone, O., Riccobono, L., Salvaggio, A., Mirabella, A., Bonanno, A., and Bonsignore, M. Catecholamines and blood pressure in obstructive sleep apnea syndrome. *Chest 103* (1993), 722–727.

[72] Mason, L. *Signal processing methods for non-invasive respiration monitoring.* Doctorate of philosophy thesis, University of Oxford, 2002. Michaelmas Term.

[73] The MathWorks Inc. *Matlab Signal Processing Toolbox*, 4 ed. USA, Dec 1996.

[74] McGrogan, N., Braithwaite, E., and Tarassenko, L. Biosleep: a comprehensive sleep analysis system. *Proc.23rd Annual International Conference of the IEEE EMBS 2* (Oct 2001), 1608–1611.

[75] Millikan, G. Oximeter, instrument for measuring continuously oxygen saturation of arterial blood in man. *Rev. Sci. Instrm. 13* (1942), 434–453.

[76] Moser, N., Berry, D., and Harbison, L. What is hypopnea, anyway? *Chest 105* (1994), 426–428.

[77] Moyle, J. *Pulse Oximetry*, 2 ed. BMJ Books, London, UK, 2002.

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

[79] Nilsson, N. Oximetry. *Physiol. Rev. 40* (1960), 1–22.

[80] Olson, L., Ambrogetti, A., and Gyulay, S. Prediction of sleep-disordered breathing by unattended overnight oximetry. *J. Sleep Research 8* (1999), 51–55.

[81] Oswald, I. *Sleeping and waking physiology and psychology.* Elsevier Publishing Company, New York, USA, 1962.

[82] Pardey, J., Roberts, S., and Tarassenko, L. A review of parametric modelling techniques for eeg analysis. *Med Eng Phys 18*, 1 (1996), 2–11.

[83] Pardey, J., Roberts, S., Tarassenko, L., and Stradling, J. A new approach to the analysis of the human sleep/wakefulness continuum. *J Sleep Research 5* (1996), 201–10.

[84] Payne, J., and Severinghaus, J. *Pulse Oximetry.* Springer, New York, USA, 1986.

[85] Penzel, T. Spectral analysis of blood pressure in patients with sleep-related breathing disorders during nrem and rem sleep. *Sleep 16* (1993), S150–S151.

[86] Pepin, J., Levy, P., Lepaulle, B., Brambilla, C., and Guilleminault, C. Does oximetry contribute to the detection of apneic events? mathematical processing of the sao_2 signal. *Chest 99* (1991), 1151–1157.

[87] Pepperell, J., Davies, R., and Stradling, J. Sleep studies for sleep apnoea. *Physiological Measurement 23* (2002), R39–R74.

[88] Pitson, D., and Stradling, J. Autonomic markers of arousal during sleep in patients undergoing investigation for obstructive sleep apnoea, their relationship to eeg arousals, respiratory events and subjective sleepiness. *J Sleep Research 7*, 1 (1998), 53–59.

[89] Rangayyan, R. *Biomedical signal analysis: A case-study approach.* John Wiley & Sons, Inc., New York, 2002.

[90] Rauscher, H., Popp, W., and Zwick, H. Quantification of sleep disordered breathing by computerised analysis of oximetry, heart rate and snoring. *European Respiratory Journal 4* (1991), 655–659.

[91] Rauscher, H., Popp, W., and Zwick, H. Model for investigating snorers with suspected sleep apnoea. *Thorax 48*, 275-279 (1993).

[92] Raymond, B., Cayton, R., and Chappell, M. Combined index of heart rate variability and oximetry in screening for the sleep apnoea/hypopnoea syndrome. *J Sleep Res 12* (2003), 53–61.

[93] Rechtschaffen, A., and Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Public Health Service, U.S. Government Printing Office, Washington D.C., 1968.

[94] Redline, S., Kapur, V., Sanders, M., Quan, S., Gottlieb, D., Rapoport, D., Bonekat, W., Smith, P., Kiley, J., and Iber, C. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. *Am J Respir Crit Care Med 161* (2000), 369–374.

[95] Rolfe, L., Olson, L., and Saunders, N. Prolonged nasal cpap therapy does not ameliorate obstructive sleep apnoea. *Aust N.Z. J. Medicine 21* (1991), 235–238.

[96] Ruchala, J. Applications of pulse oximetry. In *Design of Pulse Oximeters*, J. Webster, Ed. Institute of Physics Publishing, UK, 1997, ch. 13, pp. 227–231.

[97] Saul, J., Berger, R., Chen, M., and Cohen, R. Transfer function analysis of autonomic regulation ii. respiratory sinus arrhythmia. *Am J Physiol 256* (1989), H153–H161.

[98] Schmidt-Nowara, W. Therapy with oral appliances. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 6, pp. 119–128.

[99] Scott, D. *Understanding the EEG, An introduction to electroencephalography*. Lippincott, Philadelphia, USA, 1976.

[100] Sebald, D. Motivation of pulse oximetry. In *Design of Pulse Oximeters*, J. Webster, Ed. Institute of Physics Publishing, UK, 1997, ch. 2, pp. 13–20.

[101] Series, F., Marc, I., Cormier, Y., and La Forge, J. Utility of nocturnal home oximetry for case finding in patients with suspected sleep apnea hypopnea syndrome. *Ann Intern Med 119*, 6 (Sept 1993), 449–453.

[102] Shiomi, T., Guilleminault, C., Sasanabe, R., Hirota, I., Maekawa, M., and Kobayashi, T. Augmented very low frequency component of heart rate variability during obstructive sleep apnea. *Sleep 19*, 5 (1996), 370–377.

[103] Simon, P., Landry, S., and Leiter, J. Respiratory control during sleep. In *Sleep Medicine*, T. Lee-Chiong Jr., M. Sateia, and M. Carskadon, Eds. Hanley & Belfus, Inc., Philadelphia, USA, 2002, ch. 5, pp. 41–51.

[104] Stewards, D., Gluckman, J., and Johnson, J. Nasal obstruction and nasal surgery. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 7, pp. 129–143.

[105] Stoohs, R., and Guilleminault, C. Mesam 4: an ambulatory device for the detection of patients at risk for obstructive sleep apnea syndrome (osas). *Chest 101* (1992), 1221–1227.

[106] Stradling, J. *Handbook of sleep-related breathing disorders*. Oxford University Press, Oxford, UK, 1993.

[107] Stradling, J., and Crosby, J. Relation between systemic hypertension and sleep hypoxaemia or snoring: analysis in 748 men drawn from general practice. *BMJ 300* (Jan 1990), 75–78.

[108] Stradling, J., Pitson, D., Bennett, L., Barbour, C., and Davies, R. Variation in the arousal pattern after obstructive events in obstructive sleep apnoea. *Am J Resp Crit Care Med 159*, 130-136 (1999).

[109] Taha, B., Dempsey, J., Weber, S., Badr, M., Skatrud, J., Young, T., Jacques, A., and Seow, K. Automated detection and classification of sleep-disordered breathing from conventional polysomnography data. *Sleep 20*, 11 (1997), 991–1001.

[110] Tarassenko, L. Medical electronics. Course Notes, Pulse Oximetry, Department of Engineering Science, University of Oxford, UK, 2004. Hilary Term.

[111] Vázquez, J.-C., Tsai, W., Tsai, Flemons, W., Masuda, A., Brant, R., Hajduk, E., Whitelaw, W., and Remmers, J. Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea. *Thorax 55* (2000), 302–307.

[112] Walpole, R., Myers, R., and Myers, S. *Probability and Statistics for Engineers and Scientists*. Prentice Hall, NJ, USA, 1998.

[113] Warley, A. H., Mitchell, J., and Stradling, J. Evaluation of the ohmeda 3700 pulse oximeter. *Thorax 42* (1987), 892–896.

[114] West, P., George, C., and Kryger, M. Dynamic in vivo response characteristics of three oximeters: Hewlett-packard 47201a, biox iii, and nellcor n-100. *Sleep 10* (1986), 263–271.

[115] Williams, A., Yu, G., Santiago, S., and Stein, M. Screening for sleep apnea, using pulse oximetry and a clinical score. *Chest 100* (1991), 631–635.

[116] Woodson, B. Laser uvulopalatoplasty. In *Management of Obstructive Sleep Apnoea*, J. Johnson, J. Gluckman, and M. Sanders, Eds. Martin Dunitz Ltd., UK, 2002, ch. 10, pp. 175–182.

[117] Wright, J., Johns, R., Watt, I., Melville, A., and Sheldon, T. Health effects of obstructive sleep apnoea and the effectiveness of continuous positive airways pressure: a systematic review of the research evidence. *British Medical Journal 314*, 851-860 (1997).

[118] Wukitsch, M., Petterson, M., Tobler, D., and Pologe, J. Pulse oximetry; analysis of theory technology and practice. *J. Clin. Mon. 4* (1988), 290–301.

[119] Yamashiro, Y., and Kryger, M. Nocturnal oximetry: is it a screening tool for sleep disorders? *Sleep 18*, 3 (1995), 167–171.

[120] Young, T., and Peppard, P. Epidemiology of obstructive sleep apnea. In *Breathing Disorders in Sleep*, W. McNicholas and E. Phillipson, Eds. W.B. Saunders, UK, 2002, ch. 3, pp. 31–43.

[121] Zamarrón, C., Gude, F., Barcala, J., Rodriguez, J., and Romero, P. Utility of oxygen saturation and heart rate spectral analysis obtained from pulse oximetric recordings in the diagnosis of sleep apnoea syndrome. *Chest 123*, 5 (2003), 1567–1576.

[122] Zamarrón, C., Pichel, F., and Romero, P. Coherence between oxygen saturation and heart rate obtained from pulse oximetric recordings in obstructive sleep apnoea. *Physiol Meas 26* (2005), 799–810.

[123] Zamarrón, C., Romero, P., Gude, F., Amaro, A., and Rodriguez, J. Screening of obstructive sleep apnoea: heart rate spectral analysis of nocturnal pulse oximetric recording. *Respiratory Medicine 95* (2001), 759–765.

[124] Zamarrón, C., Romero, P., Rodriguez, J., and Gude, F. Oximetry spectral analysis in the diagnosis of obstructive sleep apnoea. *Clinical Science 97*, 4 (1999), 467–473.

[125] Zamora, M. *The study of the sleep and vigilance electroencephalogram using neural network methods.* Doctorate of philosophy thesis, University of Oxford, 2001. Hilary Term.

[126] Zamora, M., and Tarassenko, L. The study of micro-arousals using neural network analysis of the eeg. *Proc.9th International Conference on Artificial Neural Networks (ICANN 99) 2*, 470 (Sept 1999), 625–630.

[127] Zweig, M., and Campbell, G. Receiver-operating characteristic (roc) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry 39*, 4 (1993), 561–577.