

# **Wearable In-Ear Pulse Oximetry: Theory and Applications**

Harry J. Davies

Supervisor: Professor Danilo P. Mandic

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy

Communications and Signal Processing Group  
Department of Electrical Engineering  
Imperial College London, April 2022



## **Statement of Originality**

I hereby declare that this thesis and the work described is my own and that everything else is appropriately referenced and attributed.





## Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.



## Publications

### Journal articles:

- **H. J. Davies**, I. Williams, N. S. Peters, and D. P. Mandic, “In-Ear SpO<sub>2</sub>: A Tool for Wearable, Unobtrusive Monitoring of Core Blood Oxygen Saturation,” *Sensors*, vol. 20, no. 17, p. 4879, Aug 2020.
- **H. J. Davies**, P. Bachtiger, I. Williams, P. L. Molyneaux, N. S. Peters, and D. P. Mandic, “Wearable In-Ear PPG: Detailed Respiratory Variations Enable Classification of COPD,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 7, pp. 2390-2400, July 2022.
- **H. J. Davies**, I. Williams, G. Hammour, M. C. Yarici, M. J. Stacey, B. M. Seemungal, and D. P. Mandic, “In-Ear SpO<sub>2</sub> for Classification of Cognitive Workload,” *IEEE Transactions on Cognitive and Developmental Systems*, 2022 (In press).

### Conference papers:

- **H. J. Davies**, T. Nakamura, and D. P. Mandic, “A transition probability based classification model for enhanced n1 sleep stage identification during automatic sleep stage scoring,” In Proc. of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 3641–3644.
- T. Nakamura, **H. J. Davies**, and D. P. Mandic, “Scalable automatic sleep staging in the era of big data,” In Proc. of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 2265–2268.
- **H. J. Davies**, I. Williams, and D. P. Mandic, “Tracking Cognitive Workload in Gaming with In-ear SpO<sub>2</sub>”, In Proc. of the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2022. (to appear)
- E. Occhipinti, **H. J. Davies**, G. Hammour and D. P. Mandic, “Hearables: Artefact removal in Ear-EEG for continuous 24/7 monitoring”, In Proc. of the IEEE International Joint Conference on Neural Networks. (IJCNN), 2022. (to appear)

### Book chapters:

- M. C. Yarici, **H. J. Davies**, T. Nakamura, I. Williams, and D. P. Mandic, “Hearables: In-Ear Multimodal Brain Computer Interfacing”, In: C. Guger, B. Z. Allison, and M. Tangermann, “Brain-Computer Interface Research: A State-of-the-Art Summary 9”, Springer International Publishing, 2021, pp. 79–87.

#### Articles on arxiv:

- **H. J. Davies**, G. Hammour, H. Xiao and D. P. Mandic, “An Apparatus for the Simulation of Breathing Disorders: Physically Meaningful Generation of Surrogate Data” arXiv preprint arXiv:2109.06699, Sep 2021.

## Abstract

Wearable health technology, most commonly in the form of the smart watch, is employed by millions of users worldwide. These devices generally exploit photoplethysmography (PPG), the non-invasive use of light to measure blood volume, in order to track physiological metrics such as pulse and respiration. Moreover, PPG is commonly used in hospitals in the form of pulse oximetry, which measures light absorbance by the blood at different wavelengths of light to estimate blood oxygen levels ( $\text{SpO}_2$ ). This thesis aims to demonstrate that despite its widespread usage over many decades, this sensor still possesses a wealth of untapped value. Through a combination of advanced signal processing and harnessing the ear as a location for wearable sensing, this thesis introduces several novel high impact applications of in-ear pulse oximetry and photoplethysmography. The aims of this thesis are accomplished through a three pronged approach: rapid detection of hypoxia, tracking of cognitive workload and fatigue, and detection of respiratory disease.

By means of the simultaneous recording of in-ear and finger pulse oximetry at rest and during breath hold tests, it was found that in-ear  $\text{SpO}_2$  responds on average 12.4 seconds faster than the finger  $\text{SpO}_2$ . This is likely due in part to the ear being in close proximity to the brain, making it a priority for oxygenation and thus making wearable in-ear  $\text{SpO}_2$  a good proxy for core blood oxygen. Next, the low latency of in-ear  $\text{SpO}_2$  was further exploited in the novel application of classifying cognitive workload. It was found that in-ear pulse oximetry was able to robustly detect tiny decreases in blood oxygen during increased cognitive workload, likely caused by increased brain metabolism. This thesis demonstrates that in-ear  $\text{SpO}_2$  can be used to accurately distinguish between different levels of an N-back memory task, representing different levels of mental effort. This concept was further validated through its application to gaming and then extended to the detection of driver related fatigue. It was found that features derived from  $\text{SpO}_2$  and PPG were predictive of absolute steering wheel angle, which acts as a proxy for fatigue.

The strength of in-ear PPG for the monitoring of respiration was investigated with respect to the finger, with the conclusion that in-ear PPG exhibits far stronger respiration induced intensity variations and pulse amplitude variations than the finger. All three respiratory modes were harnessed through multivariate empirical mode decomposition (MEMD) to produce spirometry-like respiratory waveforms from PPG. It was discovered that these PPG derived respiratory

waveforms can be used to detect obstruction to breathing, both through a novel apparatus for the simulation of breathing disorders and through the classification of chronic obstructive pulmonary disease (COPD) in the real world.

This thesis establishes in-ear pulse oximetry as a wearable technology with the potential for immense societal impact, with applications from the classification of cognitive workload and the prediction of driver fatigue, through to the detection of chronic obstructive pulmonary disease. The experiments and analysis in this thesis conclusively demonstrate that widely used pulse oximetry and photoplethysmography possess a wealth of untapped value, in essence teaching the old PPG sensor new tricks.

## Acknowledgements

First and foremost, I would like to thank Professor Danilo Mandic for his support and guidance as my PhD supervisor. Your vast knowledge and wisdom combined with your rigorous approach to research has helped me immensely in my growth as a researcher over the past 3 years.

Secondly, I must express my sincere gratitude for my medical collaborators - Professor Nicholas Peters, Dr Philip Molyneaux, Dr Barry Seemungal and Dr Mike Stacey. All of whom provided me with a wealth of medical insights and as a consequence the confidence to expand my research through medical applications.

I would like to thank my examiners Professor Timothy Constandinou and Professor Jonathon Chambers for their insightful comments which have greatly benefited this thesis.

I would like to give special thanks to Dr Takashi Nakamura who gave me valuable advice in the first few months of my PhD, giving me a solid foundation from which my research has developed. I must also give thanks to Ghena Hammour, who is a brilliant problem solver and has always been there if I needed advice. On a similar note, thank you to all of the lab and the CSP group as a whole, whom I have had the pleasure of working with over the course of my PhD.

To all of my friends in both Malvern and London, and my countless friends from Barbell club, you helped me achieve balance especially during times where I did not think balance was possible and for that I will be forever grateful.

To Granny Vonny and Roger, thank you for your endless support to which I owe a great deal of my success. To Aunty Pam, thank you for your continued enthusiasm towards my work. To Emma, thank you for being a big sister that I can look up to. To Aya, thank you for all the special moments we have shared.

Above all else, thank you Mum and Dad for your love and for your belief in me.





# Contents

<b>Statement of Originality</b>	<b>i</b>
<b>Copyright Declaration</b>	<b>iii</b>
<b>Publications</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and objectives . . . . .	1
1.2 Thesis organisation and contributions . . . . .	3
<b>2 Background theory</b>	<b>7</b>
2.1 Principles of photoplethysmography . . . . .	8
2.1.1 Photoplethysmography and pulse oximetry . . . . .	8
2.1.2 The position of PPG . . . . .	10
2.1.3 Hearables: the ear as a recording site for brain and cardiac signals . . . .	12

2.2	In-ear PPG hardware . . . . .	12
2.2.1	Initial prototype board . . . . .	12
2.2.2	In-ear PPG and the Bioboard . . . . .	13
2.2.3	Adhesive PPG . . . . .	14
2.2.4	The sensitivity of SpO <sub>2</sub> readings and skin colour . . . . .	15
2.3	Empirical mode decomposition . . . . .	16
2.3.1	Univariate EMD . . . . .	17
2.3.2	Multivariate EMD . . . . .	19
2.4	Machine learning . . . . .	21
2.4.1	Random forest . . . . .	21

## **I In-ear pulse oximetry for the detection of hypoxia 26**

### **3 In-ear SpO<sub>2</sub> as a tool for wearable monitoring of core blood oxygen 27**

3.1	Introduction . . . . .	28
3.2	Materials and methods . . . . .	30
3.2.1	Hardware . . . . .	30
3.2.2	Experimental protocol . . . . .	30
3.2.3	Extraction of the SpO <sub>2</sub> signal . . . . .	32
3.2.4	Data Analysis . . . . .	33
3.3	Results . . . . .	35
3.3.1	Resting oxygen comparison . . . . .	35

3.3.2	Blood Oxygen Delay . . . . .	35
3.3.3	Photoplethysmogram amplitude . . . . .	37
3.4	Discussion . . . . .	38
3.4.1	Resting SpO <sub>2</sub> . . . . .	38
3.4.2	SpO <sub>2</sub> delay . . . . .	39
3.4.3	Photoplethysmogram amplitude . . . . .	40
3.5	Chapter conclusions . . . . .	40
<b>4</b>	<b>Adhesive ear-SpO<sub>2</sub> to monitor exercise in patients with breathing disorders</b>	<b>42</b>
4.1	Introduction . . . . .	43
4.1.1	The six minute walk test . . . . .	43
4.1.2	The problem of motion artefacts . . . . .	43
4.2	Methods . . . . .	44
4.2.1	Hardware . . . . .	44
4.2.2	Experimental protocol . . . . .	44
4.2.3	Signal processing and analysis . . . . .	45
4.3	Results . . . . .	46
4.4	Discussion and conclusion . . . . .	47
<b>II</b>	<b>Classification of cognitive workload and fatigue</b>	<b>48</b>
<b>5</b>	<b>In-ear SpO<sub>2</sub> for the classification of cognitive workload</b>	<b>49</b>
5.1	Introduction . . . . .	50

5.1.1	Physiological methods for cognitive workload tracking . . . . .	50
5.1.2	Protocols for inducing changes in cognitive workload . . . . .	51
5.1.3	The brain, oxygen and cognitive workload . . . . .	53
5.2	Methods . . . . .	54
5.2.1	Hardware . . . . .	54
5.2.2	Experimental protocol . . . . .	54
5.2.3	Signal processing . . . . .	55
5.2.4	Feature extraction . . . . .	56
5.2.5	Classification and evaluation . . . . .	58
5.3	Results . . . . .	59
5.3.1	Change in blood oxygen, heart rate, breathing rate . . . . .	59
5.3.2	Classification . . . . .	62
5.4	Discussion . . . . .	64
5.5	Chapter conclusions . . . . .	67
<b>6</b>	<b>Tracking cognitive workload in gaming with in-ear SpO<sub>2</sub></b>	<b>69</b>
6.1	Introduction . . . . .	69
6.1.1	Physiological monitoring in gaming . . . . .	70
6.2	Methods . . . . .	71
6.2.1	Hardware . . . . .	71
6.2.2	Experimental protocol . . . . .	71
6.2.3	Signal processing . . . . .	72

6.2.4	Analysis metrics . . . . .	73
6.3	Results and discussion . . . . .	73
6.4	Chapter conclusions . . . . .	76
<b>7</b>	<b>In-ear PPG for the detection of driver fatigue</b>	<b>77</b>
7.1	Introduction . . . . .	78
7.1.1	What is fatigue? . . . . .	78
7.1.2	How can we assess fatigue? . . . . .	78
7.2	Methods . . . . .	80
7.2.1	Experimental protocol . . . . .	80
7.2.2	Driving simulator . . . . .	81
7.2.3	Data acquisition . . . . .	82
7.2.4	Analysis . . . . .	83
7.2.5	Prediction . . . . .	85
7.3	Results . . . . .	86
7.3.1	Subjective and performance based fatigue . . . . .	87
7.3.2	Physiological fatigue . . . . .	87
7.3.3	Classifier prediction . . . . .	90
7.4	Discussion . . . . .	90
7.5	Chapter conclusions . . . . .	92

<b>III</b>	<b>Detection of obstructive lung disease</b>	<b>93</b>
<b>8</b>	<b>An apparatus for the artificial simulation of breathing disorders</b>	<b>94</b>
8.1	Introduction . . . . .	95
8.1.1	Changes to breathing with obstruction and restriction . . . . .	95
8.1.2	Artificial changes to breathing resistance . . . . .	96
8.2	Apparatus design . . . . .	99
8.3	Results and discussion . . . . .	100
8.4	Chapter conclusions . . . . .	102
<b>9</b>	<b>In-ear PPG for the classification of COPD</b>	<b>103</b>
9.1	Introduction . . . . .	104
9.1.1	Respiration and photoplethysmography . . . . .	104
9.1.2	In-ear photoplethysmography . . . . .	106
9.1.3	Change in breathing with COPD . . . . .	107
9.2	Materials and experimental design . . . . .	108
9.2.1	Hardware . . . . .	108
9.2.2	Experimental design . . . . .	108
9.3	Signal processing . . . . .	110
9.3.1	Extraction of respiratory modulations from PPG . . . . .	110
9.3.2	Respiratory power comparisons: ear vs finger . . . . .	111
9.3.3	Empirical mode decomposition for respiration . . . . .	112

9.3.4	Classification of COPD . . . . .	114
9.4	Results . . . . .	116
9.4.1	Spectral power of in-ear PPG vs finger PPG . . . . .	116
9.4.2	Classification of COPD from in-ear PPG . . . . .	117
9.5	Discussion . . . . .	122
9.5.1	Spectral power of in-ear PPG vs finger PPG . . . . .	122
9.5.2	Classification of COPD from in-ear PPG . . . . .	122
9.6	Chapter conclusions . . . . .	124
<b>10</b>	<b>Conclusion</b>	<b>125</b>
10.1	Conclusions and applications . . . . .	125
10.2	Advice to future researchers . . . . .	127
10.3	Future work . . . . .	127
<b>A</b>	<b>Probabilistic transition net for automatic sleep staging</b>	<b>132</b>
A.0.1	Introduction . . . . .	133
A.1	Transition model of sleep . . . . .	134
A.1.1	Previous modelling of sleep stage transitions . . . . .	135
A.1.2	Transition state features . . . . .	135
A.1.3	Transition probability model . . . . .	135
A.1.4	Pruned transition probability model . . . . .	137
A.2	Implementation . . . . .	137

A.2.1	Preprocessing and feature extraction . . . . .	137
A.2.2	Classifiers . . . . .	138
A.2.3	Transition probability classification . . . . .	138
A.3	Results . . . . .	139
A.4	Chapter conclusions . . . . .	141
<b>B</b>	<b>Example ethics application</b>	<b>143</b>
B.1	Project summary . . . . .	143
B.2	Participant recruitment . . . . .	145
B.3	Informed consent . . . . .	145
B.4	Ethical Summary . . . . .	146
B.5	Mitigation of ethical issues . . . . .	146
B.5.1	Collection of sensitive information . . . . .	146
B.5.2	Researcher in position of authority . . . . .	147
B.5.3	The potential to induce stress . . . . .	147
B.5.4	Use of non-CE approved electronics . . . . .	147
B.5.5	Incidental findings . . . . .	148
B.6	Mitigation of risks . . . . .	148
B.7	Confidentiality and Management of personal and other research data . . . . .	148
	<b>References</b>	<b>149</b>



# Nomenclature

AC	Alternating current
AdaBoost	Adaptive boosting
ANOVA	Analysis of variance
CFS	Chadler fatigue scale
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
DC	Direct current
DT	Decision tree
ECG	Electrocardiography
EEG	Electroencephalography
EMD	Empirical mode decomposition
EMG	electromyography
FDA	United States Food and Drug Administration
FEV <sub>1</sub>	Forced expiratory volume in one second
fMRI	Functional magnetic resonance imaging
fNIRS	Functional near-infrared spectroscopy
FVC	Forced vital capacity
FWHM	Full-width-half-maximum
GOLD	Global initiative for chronic obstructive lung disease

---

HF	High frequency heart rate variability
HMM	Hidden Markov model
HRV	Heart rate variability
IMF	Intrinsic mode functions
IPF	Idiopathic pulmonary fibrosis
IQR	Interquartile range
KNN	K-nearest neighbours
LDA	Linear discriminant analysis
LED	Light-emitting diode
LF	Low frequency heart rate variability
MEMD	Multivariate empirical mode decomposition
MLP	Multilayer perceptron
NA-MEMD	Noise assisted multivariate empirical mode decomposition
PERCLOS	Percentage of eyelid closure over the pupil
PF	Pulmonary fibrosis
PPG	Photoplethysmography
PSD	Power spectral density
PSG	Polysomnography
REM	Rapid eye movement
RF	Random forest
RIIV	Respiration induced intensity variation
RT	Reaction time
SAMME	Stagewise Additive Modeling using a Multi-class Exponential loss function
SCL	Skin conductance level
SpO <sub>2</sub>	Estimate of blood oxygen percentage
SVM	Support vector machine

$T_I$	Inspiratory time
$T_{TOT}$	Total breathing time
VR	Virtual reality



# List of Tables

2.1	A summary of PPG sensor positions for pulse variations, respiratory variations, artefacts and wearability. . . . .	11
2.2	A summary of classification outcomes for several classifiers applied to the four category cognitive workload dataset outlined in Chapter 5. . . . .	24
3.1	Summary of mean ( $\pm$ standard deviation) SpO <sub>2</sub> delay values (seconds). . . . .	37
3.2	Summary of different SpO <sub>2</sub> qualities for the finger, earlobe and ear canal. . . . .	38
5.1	Summary of features used for the classification of cognitive workload with 10-fold cross validation. . . . .	57
7.1	Summary of features used for the classification of fatigue induced steering wheel angles when driving. . . . .	84
9.1	Summary of the features used for the classification of COPD. . . . .	114
A.1	Comparison of performances of the standard (R) and transition classification (T) with different classifiers. . . . .	139



# List of Figures

1.1	A bar graph the showing the quantity of papers, indexed by google scholar, that utilise in-ear PPG for each year between 2005 and August 2022. . . . .	2
2.1	An illustration of reflectance photoplethysmography (PPG). . . . .	9
2.2	An illustration of the absorbance of oxygenated haemoglobin and deoxygenated haemoglobin for red and infrared wavelengths of light. . . . .	10
2.3	The first stage prototype in-ear photoplethysmography sensor. . . . .	13
2.4	The in-ear photoplethysmography sensor. . . . .	13
2.5	The adhesive behind the ear photoplethysmography sensor. . . . .	15
2.6	An illustration of the empirical mode decomposition (EMD) sifting process. . . .	18
2.7	An illustration of the multivariate empirical mode decomposition (MEMD) sifting process in three dimensions. . . . .	19
2.8	A diagram explaining the decision tree for classification. . . . .	22
3.1	Experimental setup for the investigation of blood oxygen delay. . . . .	31
3.2	A flow diagram outlining the sequence of participants' breathing. . . . .	32

3.3	Examples of the alternating current (AC) and direct current (DC) waveforms within a PPG measurement, from both the right ear canal and the right index finger. . . . .	33
3.4	Exemplar SpO <sub>2</sub> recording following the breath hold protocol. . . . .	34
3.5	Mean resting SpO <sub>2</sub> levels taken for each individual subject for the right ear canal and the right index finger. . . . .	35
3.6	The absolute blood oxygen delay from the button press to the finger, to the ear and the relative delay between the ear and the finger, shown for every subject. .	36
3.7	Boxplots of the mean absolute delays, for all subjects, between the right ear canal and the right index finger. . . . .	36
3.8	The mean resting maximum to minimum AC amplitude of the photoplethysmography signal for all subjects. . . . .	37
4.1	A comparison of mean resting SpO <sub>2</sub> levels for patients with various breathing disorders, between the in-ear and the finger. . . . .	45
4.2	Comparisons in SpO <sub>2</sub> changes between a wearable adhesive based behind the ear pulse oximeter and hospital transmission based pulse oximeters during 6 minute walk tests. . . . .	46
5.1	Illustration of the recording of in-ear SpO <sub>2</sub> during an N-back task. . . . .	54
5.2	Overview of the pulse and respiratory signals recorded from in-ear photoplethysmography. . . . .	56
5.3	Box plots of the relative change in SpO <sub>2</sub> , heart rate and breathing rate from the in-ear sensor, for each N-back stage. . . . .	60



5.4	Scatter plots of median relative change in SpO <sub>2</sub> and median heart rate against N-back level and proportion of errors, with trend lines and correlation coefficients and p values superimposed. . . . .	61
5.5	Two-dimensional contour plots of the kernel density estimates for mean SpO <sub>2</sub> , and relative change in SpO <sub>2</sub> , for the 0-back and 3-back tasks. . . . .	62
5.6	Classification accuracy of SpO <sub>2</sub> based cognitive workload estimation. . . . .	63
5.7	Feature importance for the 10 most significant features out of the 21 features considered during 10-fold cross validation. . . . .	64
6.1	Distribution of SpO <sub>2</sub> during Geometry Dash game play relative to the resting interval before. . . . .	74
6.2	Bar graph showing SpO <sub>2</sub> of boss fights in Dark Souls III relative to the easier game play before. . . . .	74
6.3	Box plots of the relative change in SpO <sub>2</sub> , heart rate and breathing rate from the in-ear sensor, for the three games Geometry Dash, Dark Souls III and Five Night's at Freddy's. . . . .	75
7.1	Custom built driving simulator, with Vision Racer seat and pedals, a Logitech G27 steering wheel and a FOVE virtual reality headset. . . . .	81
7.2	Results of subjective and performance related measures of fatigue across 7 subjects, with p-values from one-way ANOVA shown in the titles. . . . .	86
7.3	Boxplots of physiological features derived from in-ear PPG across 7 subjects undergoing an hour long monotonous drive in a driving simulator, with p-values from one-way ANOVA displayed in the titles of the plots for each feature. . . . .	88
7.4	Plots of classifier probability across the duration of each trial for both random forest and support vector machine classifiers, against true absolute steering wheel angle. . . . .	89

7.5	Plots of Pearson's correlation coefficient between classifier probability and true absolute steering wheel angle, against window length. . . . .	90
8.1	Block diagram and physical realisation of the proposed breathing disorder simulation apparatus. . . . .	96
8.2	Tidal breathing results from using the apparatus with 4 different inspiration to expiration obstruction ratios across 10 subjects. . . . .	97
8.3	Example flow volume loops and PPG waveforms during maximally forced breathing for a single subject across 5 different expiratory tube diameters . . . . .	98
8.4	Calculated FEV <sub>1</sub> /FVC ratios across 6 expiratory tube diameters from 25mm to 3mm with an inspiratory tube diameter fixed at 25mm, plotted for 5 different subjects. . . . .	99
9.1	Normalised power spectral density (PSD) between 0.15 and 0.4Hz of each respiratory mode from in-ear PPG. . . . .	110
9.2	Noise assisted multivariate empirical mode decomposition (NA-MEMD) of in-ear PPG for a subject breathing with an atypical 1:3 inspiration to expiration time ratio. . . . .	113
9.3	The log <sub>10</sub> of the respiratory power spectral density ratio between the ear and finger, for each respiratory mode. . . . .	117
9.4	Exemplar in-ear PPG respiratory waveforms and their corresponding probability density estimates. . . . .	118
9.5	Boxplots of COPD related features extracted from in-ear PPG recordings. . . .	119
9.6	An exemplar in-ear PPG derived COPD respiratory waveform, with the corresponding COPD features calculated with a sliding 500 sample (8 second) window.	120
9.7	Classification of COPD from in-ear PPG respiratory waveforms. . . . .	121

10.1 Further investigation of in-ear SpO <sub>2</sub> for cognitive workload with virtual reality (VR) applications. . . . .	130
A.1 Percentage of current states of sleep given the previous state, obtained from the manually scored hypnograms in the CCSHS data set. . . . .	134
A.2 Probability density estimates of different sleep stages for scale 1 permutation entropy. . . . .	136
A.3 Pruning table, showing connections that are included in the pruned model (P) and connections that have been removed (X). . . . .	137
A.4 The relabelling protocol for the data, showing the transitions that have been pruned in grey. . . . .	138
A.5 Flow diagram of the proposed transition probability implementation. . . . .	140
A.6 Average cross validated confusion matrix results for all classifiers. . . . .	141



# Chapter 1

## Introduction

### 1.1 Motivation and objectives

Consumer wearable devices promise to revolutionise the healthcare system, both by empowering users with knowledge and control over their own health as well as by providing doctors with valuable 24/7 physiological data that cannot be obtained through patient visits. The quantity of wearables sold increased almost 20-fold between 2014 and 2021, from 28.8 million units per year to 533.6 million units per year [1], a growing proportion of which is attributable to ear-based wearables [2]. With this incredible explosion in health data availability, comes an immense potential for life saving diagnostics, but also a considerable responsibility to understand the accuracy and precision of these diagnostics.

Photoplethysmography (PPG), the technology relied upon by smart watches to measure pulse, has been employed as a non-invasive tool to measure the pulse since the 1930s when it was published by Dr. Alrick Hertzman [4]. It works simply by emitting light into the skin and then either measuring the intensity of light transmitted through the tissue (transmittance photoplethysmography) or the intensity of light reflected by the tissue (reflectance photoplethysmography). When more blood is present, more light is absorbed, resulting in less light reflected back to the sensor. Thus, PPG effectively measures the pulse. In the 1970s, almost 40 years after the emergence of photoplethysmography, Dr. Takuo Aoyagi utilised PPG at two different wavelengths

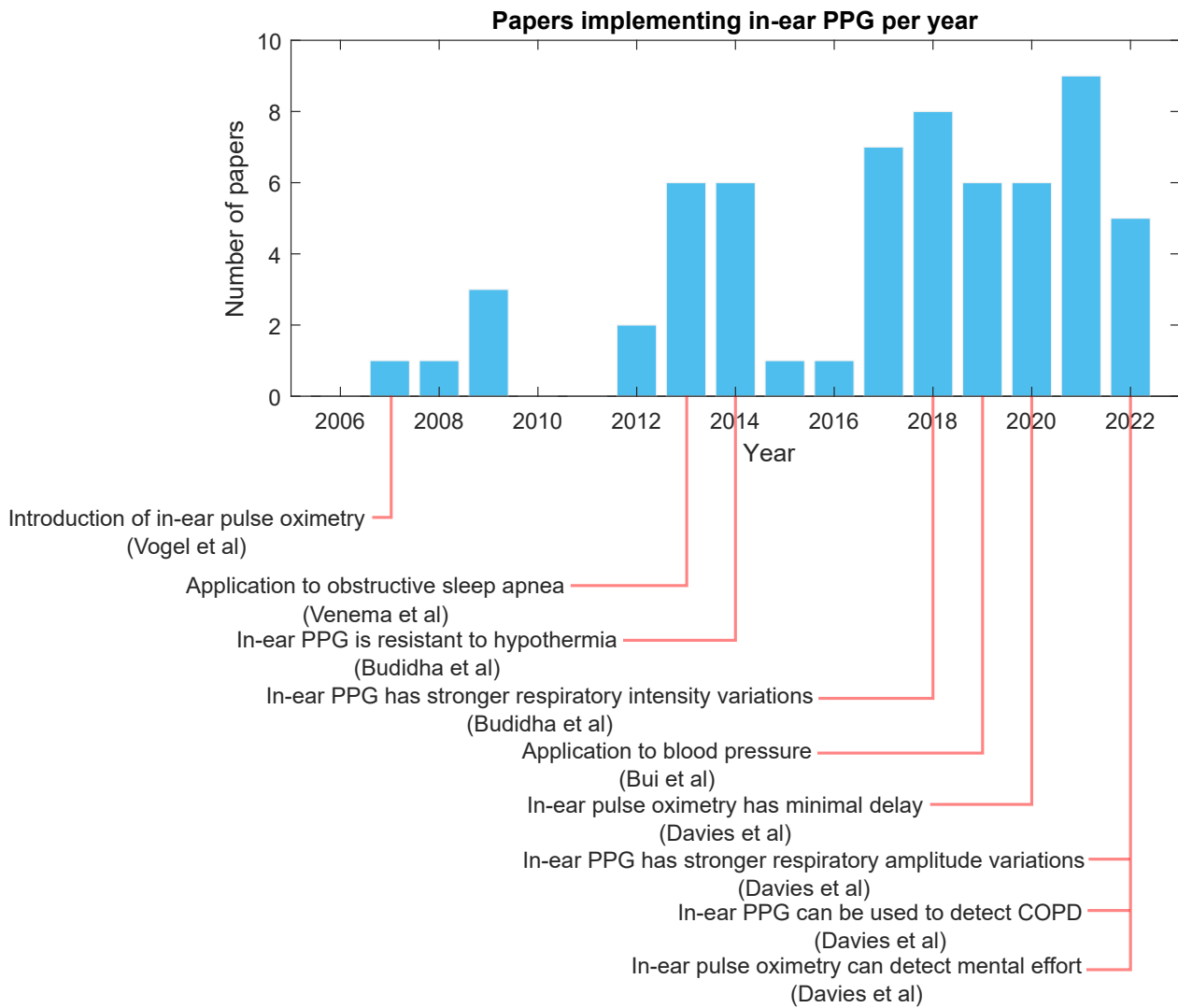


Figure 1.1: A bar graph the showing the quantity of papers, indexed by google scholar, that utilise in-ear PPG for each year between 2005 and August 2022. Significant time points are labeled, such as the introduction of in-ear pulse oximetry in 2007 [3].

of light (red and infrared) to non-invasively estimate blood oxygen levels [5]. This invention, termed pulse oximetry, is now widely used in hospitals in the form of a finger clip to measure pulse and blood oxygen percentage ( $SpO_2$ ) simultaneously. Photoplethysmography has further exploded in popularity over the past decade through the rise of modern smart watches, which rely on PPG to estimate pulse rate and in some cases respiration rate. In the past few years, several companies, including Bose, have integrated heart rate tracking with headphones, and Samsung recently filed a patent on earbuds with optical biometric sensing, suggesting that a migration to ear-based health tracking is around the corner.

The concept of an in-ear pulse oximetry sensor was introduced in the literature in 2007 by

Vogel et al [3]. For the years between 2007 and 2012, research on in-ear PPG focused mainly on validation against more commonly used sensor positions such as the wrist and the finger. In-ear pulse oximetry was then applied to measure oxygen desaturations that occur in patients with obstructive sleep apnea in 2013 by Venema et al [6]. Important qualities of the in-ear recording location then emerged in 2014 and 2018 with Budidha et al discovering that in-ear PPG is both resistant to vasoconstriction that occurs with hypothermia [7] and that it has stronger respiratory induced intensity variations than the finger [8]. In-ear PPG was then applied to blood pressure prediction, through the combination of PPG with an inflatable tube, in 2019 by Bui et al [9]. These developments are summarised in Fig 1.1. Still, compared with other wearable PPG locations such as the wrist, the in-ear is far less researched. As of August 2022, there are 5-fold more papers relating to wrist PPG than in-ear PPG indexed by google scholar. The comparative lack of research on in-ear PPG, in light of the prediction that there will soon be a migration from wrist to in-ear based consumer health sensors, shows a clear gap for further investigation and forms a strong motivation for this thesis.

The primary objective of this thesis is to show that despite being widely investigated and utilised for many decades, both pulse oximetry and PPG still possess inherent untapped value. Through the use of state of the art signal processing methods in combination with the advantageous sensor location of the ear, this thesis sheds light on previously unexplored capabilities of pulse oximetry from the robust classification of cognitive workload to the detection of chronic obstructive pulmonary disease (COPD).

## 1.2 Thesis organisation and contributions

### Background theory (chapter 2)

The biological and engineering principles of photoplethysmography and pulse oximetry are explained in depth. Different recording sites are compared with respect to pulse variations, respiratory variations, wearability and the prevalence of motion artefacts. The effect of skin pigmentation on bias in pulse oximetry readings is explained, as well as the mitigation of this

issue in the context of this thesis. The development of in-ear pulse oximetry hardware for this project is expanded upon and documented, from the early prototype stages to the Bluetooth enabled *Bioboard*. Empirical mode decomposition (EMD), a data driven method for dividing signals up into physiologically meaningful modes, is explained in depth for both the single variate algorithm and multivariate algorithm. The random forest (RF) algorithm, used heavily in this thesis for the classification of cognitive workload, chronic obstructive pulmonary disease and driver fatigue, is explained in depth. An example of the machine learning classifiers applied to physiological data is presented, and classifiers different classifier performances are compared.

### **Part I: In-ear pulse oximetry for the detection of hypoxia (chapters 3 and 4)**

Wearable in-ear SpO<sub>2</sub> was investigated with respect to finger SpO<sub>2</sub> during rest and breath holds, with the discovery of a 12.4 second delay on average between in-ear recordings and finger recordings. Moreover, a novel adhesive ear-SpO<sub>2</sub> sensor is introduced for continuous wearable SpO<sub>2</sub> measurements during exercise. It is applied to patients with breathing disorders during six minute walk tests and compared to hospital transmission based pulse oximetry on both the earlobe and the finger.

#### Relevant publications:

- H. J. Davies, I. Williams, N. S. Peters, and D. P. Mandic, “In-Ear SpO<sub>2</sub>: A Tool for Wearable, Unobtrusive Monitoring of Core Blood Oxygen Saturation,” *Sensors*, vol. 20, no. 17, p. 4879, Aug 2020.

### **Part II: Classification of cognitive workload and fatigue (chapters 5, 6 and 7)**

In-ear SpO<sub>2</sub> was introduced as a novel tool for the wearable measurement of changes in cognitive workload. With in-ear SpO<sub>2</sub> it was possible to detect minor decreases in blood oxygen that likely occur from the brains increased consumption of oxygen during increased mental effort. This principle was used to automatically classify stages of memory task based on features derived from in-ear SpO<sub>2</sub>. It was then further applied to the detection of cognitive workload in gaming, with a comparison temporally across a level of a game and comparisons across different types



of game. Moreover, the same sensor was then applied to predict driving induced fatigue. A classifier trained on features from in-ear SpO<sub>2</sub> and in-ear PPG was successfully able to predict periods where there were large absolute steering wheel angles and thus predict when a driver was more likely to make fatigue related mistakes.

Relevant publications:

- H. J. Davies, I. Williams, G. Hammour, M. C. Yarici, and D. P. Mandic, “In-Ear SpO<sub>2</sub> for Classification of Cognitive Workload,” IEEE Transactions on Cognitive and Developmental Systems, in press, 2022.
- H. J. Davies, I. Williams, and D. P. Mandic, “Tracking Cognitive Workload in Gaming with In-ear SpO<sub>2</sub>”, In Proc of the 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2022. (To appear)

### **Part III: Detection of obstructive lung disease (Chapters 8 and 9)**

In-ear PPG was investigated with respect to finger PPG for the strength of respiratory modulations. It was found that in-ear PPG had stronger venous intensity variations induced by respiration which was in-line with existing literature. Moreover, in-ear PPG was also found to have stronger pulse amplitude variations than finger PPG. A novel method for producing detailed respiratory waveforms from PPG was developed, using all three respiratory modes of PPG through multivariate empirical mode decomposition (MEMD). Using this method in combination with physiologically informed features, it was possible to classify chronic obstructive pulmonary disease for the first time from PPG-derived respiratory waveforms. This discovery was further validated through the development of a novel tube-based apparatus for the simulation of obstructive breathing disorders in healthy subjects, enabling the physically meaningful generation of surrogate data.

Relevant publications:

- H. J. Davies, P. Bachtiger, I. Williams, P. L. Molyneaux, N. S. Peters, and D. Mandic,

“Wearable In-Ear PPG: Detailed Respiratory Variations Enable Classification of COPD,”  
IEEE Transactions on Biomedical Engineering, vol. 69, no. 7, pp. 2390-2400, July 2022.

- H. J. Davies, G. Hammour, H. Xiao and D. P. Mandic, “An Apparatus for the Simulation of Breathing Disorders: Physically Meaningful Generation of Surrogate Data”, arXiv preprint arXiv:2109.06699, Sep 2021.

CONTENTS	KEY FINDINGS
<b>Part I: Detection of Hypoxia</b> <div> <div>Chapter 3 In-ear SpO<sub>2</sub> as a tool for wearable monitoring of core blood oxygen</div> <div>Chapter 4 Adhesive ear-SpO<sub>2</sub> to monitor exercise in patients with breathing disorders</div> </div>	<ul style="list-style-type: none"> <li>• Wearable in-ear SpO<sub>2</sub> shows non-inferiority with the finger for resting blood oxygen measurements.</li> <li>• In-ear SpO<sub>2</sub> responds 12 seconds faster on average than the finger for measures of hypoxia.</li> <li>• Adhesive ear-SpO<sub>2</sub> shows non-inferiority with the finger for detecting drops in blood oxygen during six minute walk tests.</li> </ul>
<b>Part II: Classification of Cognitive Workload and Fatigue</b> <div> <div>Chapter 5 In-ear SpO<sub>2</sub> for the classification of cognitive workload</div> <div>Chapter 6 Tracking cognitive workload in gaming with in-ear SpO<sub>2</sub></div> <div>Chapter 7 In-ear PPG for the detection of driver fatigue</div> </div>	<ul style="list-style-type: none"> <li>• In-ear SpO<sub>2</sub> decreases reliably in response to increased cognitive load.</li> <li>• With the in-ear SpO<sub>2</sub> signal it is possible to accurately classify levels of N-back memory task.</li> <li>• The in-ear SpO<sub>2</sub> response to cognitive workload is further evidenced in gaming tasks.</li> <li>• In-ear SpO<sub>2</sub> and photoplethysmography features are capable of predicting steering wheel angle deviations from driver related fatigue.</li> </ul>
<b>Part III: Detection of Obstructive Lung Disease</b> <div> <div>Chapter 8 An apparatus for the artificial simulation of breathing disorders</div> <div>Chapter 9 In-ear PPG for the classification of COPD</div> </div>	<ul style="list-style-type: none"> <li>• A novel tube based apparatus allows for the simulation of obstructive breathing disorders in healthy subjects.</li> <li>• PPG waveforms reliably vary in response to artificial obstruction.</li> <li>• In-ear PPG shows much larger intensity and amplitude based respiratory variations than finger PPG.</li> <li>• Detailed in-ear PPG respiratory waveforms enable the classification of chronic obstructive pulmonary disease.</li> </ul>

# Chapter 2

## Background theory

### Chapter overview

In this chapter, the principles behind photoplethysmography and blood oxygen estimation from infrared and red light based photoplethysmography are explained in detail. Moreover, a review of PPG sensor positions in the literature and their various benefits and downfalls are examined with respect to arterial pulse variations, venous respiratory variations, wearability and motion artefacts. The progression in hardware from prototype towards a portable Bluetooth in-ear SpO<sub>2</sub> sensor are detailed, and linked to their implementation in different chapters of this thesis. The sensitivity of the specific PPG chip that has been used is examined in relation to true blood oxygen saturation values as per the United States Food and Drug Administration guidelines. This is then related to the effects of skin pigmentation on SpO<sub>2</sub> calculation and it is explained that for the key findings of this thesis, which rely on relative estimation of SpO<sub>2</sub> and not absolute values, the bias related to skin pigmentation would not impact the results. Next, the signal processing and machine learning methods implemented in this thesis, namely empirical mode decomposition for extraction of respiratory variations and the random forest for classification of cognitive workload and chronic obstructive pulmonary disease, are explained in depth along with the rationale for choosing them over other methods.

## 2.1 Principles of photoplethysmography

### 2.1.1 Photoplethysmography and pulse oximetry

The “pleth” in plethysmograph stems from the Greek word “*plethusmos*” meaning “increase”, which in the case of plethysmograph means the measurement of increases in blood volume [10]. Photoplethysmography (PPG) is the non-invasive measurement of these changes in blood volume through changes in light absorption. Photoplethysmography can work with two different methods; the first being the measurement of transmission of light through tissue by placing a light-emitting diode (LED) on one side of the tissue and a photodetector on the opposite side of the tissue to measure the light transmitted through. This is referred to as transmittance photoplethysmography and usually involves a clip on sites such as the finger or earlobe. The second method involves the placement of an LED and a photodetector in close proximity on the same side of the tissue, visualised in Fig 2.1, and thus in this case the photodetector measures the light reflected back by the tissue. This is referred to as reflectance photoplethysmography and allows the sensor to be placed anywhere on the body with unobstructed skin.

When more blood is present, more light is absorbed by the tissue, and in turn less light is transmitted through the tissue or reflected back to the sensor. There are multiple ways in blood volume changes, the most obvious of which being pulsatile changes in arterial blood due to heart beats. There are however also changes in both venous and arterial blood volume that occur due to breathing. Photoplethysmography is extremely effective at measuring these volume changes, making it a useful tool for the measurement of pulse and respiration.

Photoplethysmography can be measured with a broad range of light wavelengths, but commonly utilises either green light or red and infrared light. Importantly, the absorbance of blood with respect to light changes depending on both the properties of the blood and the wavelength of light. A useful example of this is the absorbance spectra of red and infrared light for oxygenated haemoglobin and deoxygenated haemoglobin, shown in Fig 2.2. Haemoglobin is a protein which forms one of the building blocks of red blood cells and allows red blood cells to transport oxygen from the lungs to tissues throughout the body. Oxygenated haemoglobin

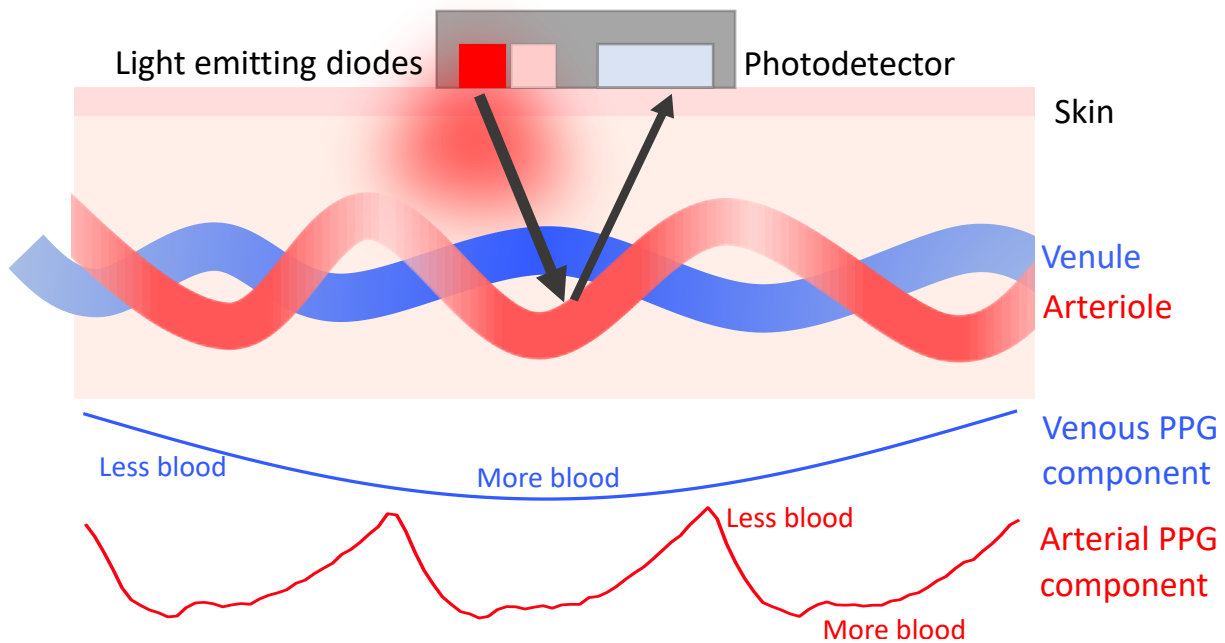


Figure 2.1: An illustration of reflectance photoplethysmography (PPG). Light emitting diodes (LEDs) emit light through the skin into the vascular tissue beneath it. Depicted in the form of blood vessels are the venous blood, with a darker blue indicating more blood, and the arterial blood, with a darker red indicating more blood. A photodetector is placed next to the LEDs to measure the light that is reflected back. Depicted beneath are the light intensity waveforms measured by the photodetector, including the venous component and the arterial component. When more blood is present (a darker colour) more light is absorbed and thus less light is reflected back to the sensor. In this example the arterial component is a real world in-ear PPG waveform, with its characteristic carotid artery pressure waveform.

absorbs a higher proportion of infrared light than red light and deoxygenated haemoglobin absorbs a higher proportion of red light than infrared light. When there is a higher absorbance of a particular wavelength in the blood, the change in amplitude of the light reflected back to the sensor is even greater with each pulse. This means that the recorded pulse amplitude of red light is increased when blood oxygen decreases, and decreased when blood oxygen increases; the opposite being true of infrared light. Given this, when both infrared and red light PPG is used simultaneously, the ratios of amplitude can be used to estimate blood oxygen percentage ( $\text{SpO}_2$ ). Using PPG to measure  $\text{SpO}_2$  is referred to as pulse oximetry.

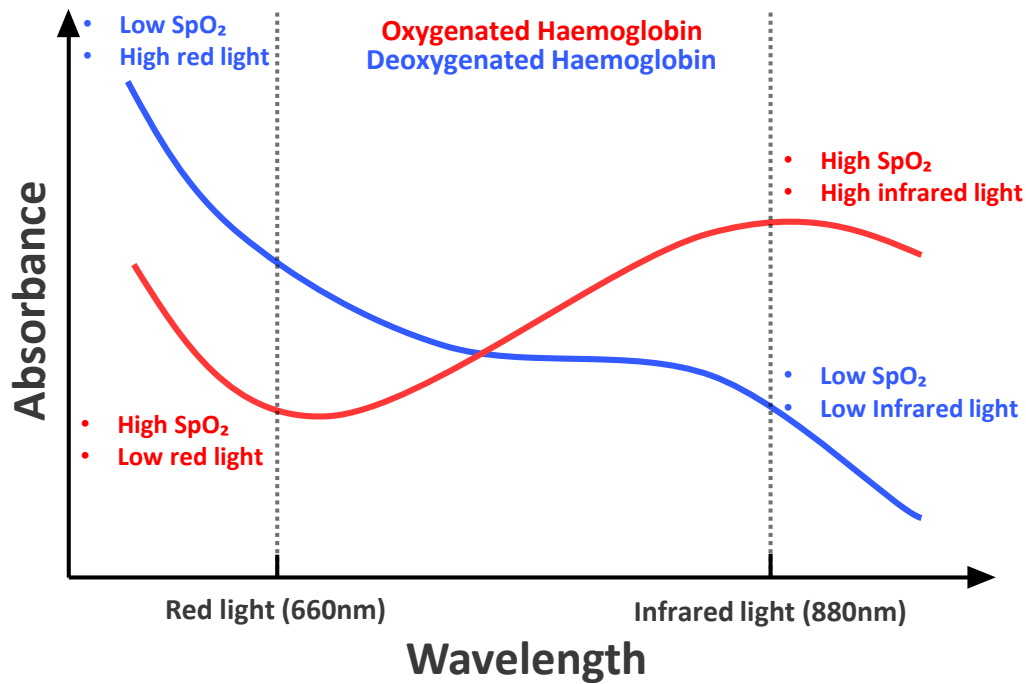


Figure 2.2: An illustration of the absorbance (also known as the extinction coefficient) of oxygenated haemoglobin and deoxygenated haemoglobin for red (660nm) and infrared (880nm) wavelengths of light. It is highlighted that when deoxygenated haemoglobin is dominant (low blood oxygen), red light absorbance is highest and infrared light absorbance is lowest, and when oxygenated haemoglobin is dominant (high blood oxygen) red light absorbance is lowest and infrared light absorbance is highest. This principle forms the basis for estimating blood oxygen (SpO<sub>2</sub>) with photoplethysmography at red and infrared wavelengths of light, known as pulse oximetry.

### 2.1.2 The position of PPG

A reflectance photoplethysmography sensor can be placed on any area of the body where skin is accessible, but not all PPG sensor positions are created equal and different areas of the body have very different outcomes for the PPG signal. Positioning of a reflectance PPG sensor is a trade off between four major factors:

- Strength of pulse variations (Arterial)
- Strength of respiratory variations (Venous)
- Wearability
- Prevalence of motion artefacts

One of the most common PPG positions is the finger with transmittance via a finger clip. Whilst the finger position provides strong pulse variations, it gives poor respiratory intensity variations [11] which are based on variations in venous blood volume. In contrast, the forearm provides poor pulse variations (arterial blood volume changes) and the strongest respiratory intensity variations [11], likely due to the proximity to large veins. The ear location tends to have a balance of good respiratory variations and pulse variations [8] [12]. Artefacts are another consideration, with positions such as the wrist being corrupted by some of the most frequent artefacts, including typing, writing, general hand movements and general arm movements. The ear artefacts are different but still common, corresponding to talking, chewing and most facial expressions. Positions such as the shoulder and forehead [13] have the least common artefacts, such as upper arm movements in the case of the shoulder and eyebrow movements in the case of the forehead. Finally, for PPG one of the most important considerations is wearability. Sensor positions such as the wrist (watch based) and the ear canal (headphone based) have the highest wearability. Positions such as the forehead and finger (clip based) have some of the lowest wearability, given the obstruction to daily life in the case of the finger clip, and the potential stigma that could come with wearing PPG device on the forehead in the case of forehead PPG. It is important to note that as devices get miniaturised, wearability issues diminish. The *Oura Ring* [14] is a good example of this, as it packages a PPG sensor into a ring whilst maintaining a week long battery life, thus making the finger a viable wearable PPG location. These comparisons are summarised below in Table 2.1. One of the aims of this thesis is to more thoroughly investigate the in-ear recording site for the PPG modality and to harness the advantages to expand on the capabilities of PPG sensors.

Position	Pulse	Respiration	Artefacts	Wearability
Finger [11, 12]	Best	Poor	Common: Typing, writing, hand movements	High (ring [14]), Low (clip)
Wrist [11]	Good	Adequate	Common: Typing, writing, hand movements	High
Forearm [11]	Poor	Best	Common: lower arm movements	Moderate
Shoulder [11]	Good	Good	Uncommon: whole arm movements	Low
Ear [8, 12]	Good	Good	Common: Talking, chewing, facial expressions	High
Forehead [11, 13]	Good	Adequate	Uncommon: eyebrow expressions	Low

Table 2.1: A summary of PPG sensor positions across four major dimensions, namely the strength of pulse variations, the strength of respiratory variations, the presence of artefacts and the wearability. All four of these factors should be considered when designing a PPG sensor for a specific purpose.

### 2.1.3 Hearables: the ear as a recording site for brain and cardiac signals

Hearables, otherwise known as smart headphones, are earbuds which have the capability of monitoring neural and physiological function. The concept of Hearables has risen in popularity in part due to ear-EEG technology [15] [16]. Electroencephalography (EEG) refers to the monitoring of the electrical signals of the brain, which aggregate in specific frequency bands due to the rates of firing of neurons. Whilst scalp EEG involves multiple electrodes positioned across the scalp, the ear-EEG provides a wearable solution through a single channel ear based sensor. This has applications to drowsiness detection [17], automatic sleep staging [18], user authentication through biometrics [19], measuring evoked potentials [20] and many more. With two in-ear sensors, this can be extended to monitoring electrocardiography (ECG) through ear pieces, otherwise known as ear-ECG [21]. In the future, an integration of ear-EEG and in-ear pulse oximetry will allow for full polysomnography (clinical grade sleep monitoring) from a single in-ear sensor.

## 2.2 In-ear PPG hardware

The hardware in this section and throughout this thesis was developed thanks to the tireless work of Dr Ian Williams.

### 2.2.1 Initial prototype board

For the initial exploration of in-ear photoplethysmography, the in-ear sensor consisted of a long prototype board with the MAX30101 chip by Maxim Integrated (San Jose, California, United States), decoupling capacitors and level shifting circuitry that enabled digital communication between 1.8 V and 3 V domains, shown in Fig. 2.3(a). This prototype board was embedded inside a foam earplug as shown in Fig. 2.3(b), and, although it was comfortable, the end



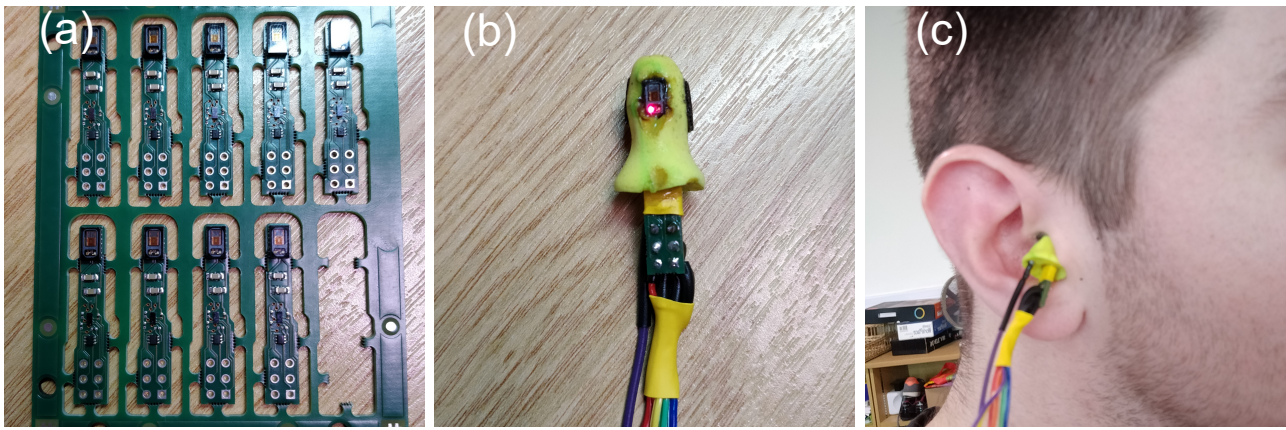


Figure 2.3: The first stage prototype in-ear photoplethysmography sensor. (a) The initial PPG sensor printed circuit board. (b) The PPG sensor within a viscoelastic foam earplug. (c) The in-ear PPG prototype being worn in the ear.

result was large which lead to some obstruction of hearing. The prototype in-ear photoplethysmography is shown in Fig. 2.3(c). This initial prototype was used to evaluate in-ear PPG through measurement of pulse transit time (PTT) between the heart and the ear, detection of an arrhythmia and comparison with chest electrocardiography for measurements of heart rate variability (HRV) [22].

### 2.2.2 In-ear PPG and the Bioboard

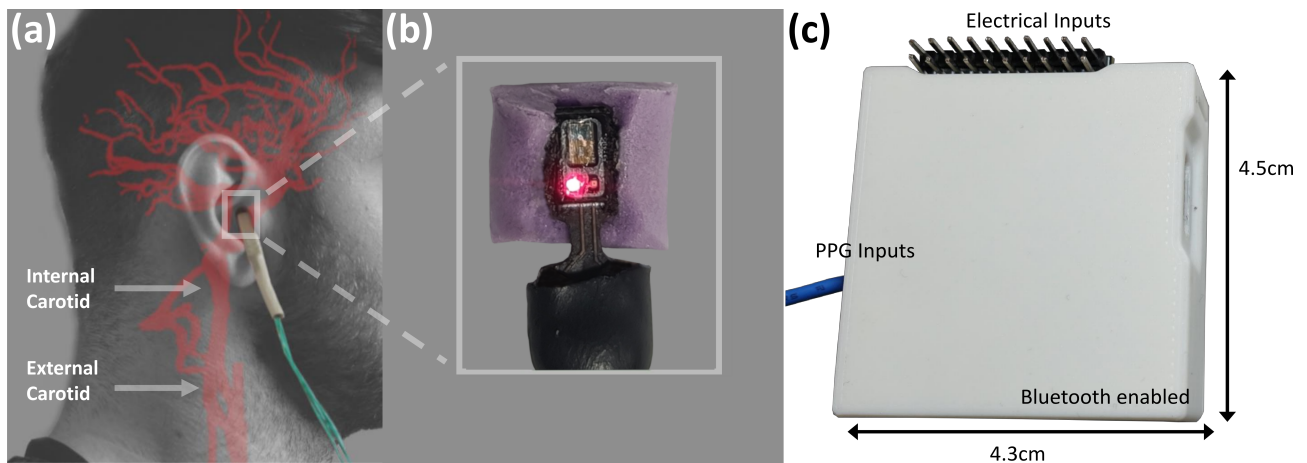


Figure 2.4: The in-ear photoplethysmography sensor. (a) The sensor placement within the ear canal, with the major arteries supplying the brain and the ear highlighted. (b) Zoom-in of the pulse oximetry sensor, with a form factor of a viscoelastic memory foam earbud. (c) The Bioboard, enabling the simultaneous recording of multiple electrical inputs and photoplethysmography, with the option of wireless Bluetooth recordings or wired recordings via micro-USB.

The next step of development saw most of the circuitry condensed within a device named the Bioboard, that has dual electrical and PPG recording capabilities, along with an accelerometer and microphone to record movement. Some of the PPG circuitry was migrated away from the in-ear position and to the Bioboard, allowing the in-ear sensor to be miniaturised and resulting in a form factor much closer to an in-ear headphone as shown in Fig. 2.4(a,b). The MAX30101 digital PPG chip continued to be used, consisting of red (660 nm) and infrared (880 nm) light emitting diodes as well as a photo-diode to measure the reflected light. The PPG chip has dimensions of 5.6 mm  $\times$  3.3 mm  $\times$  1.55 mm, with the entire in-ear “Hearables” sensor having dimensions of 19 mm  $\times$  10 mm  $\times$  7 mm with a medium sized earbud. The earbud used was identical to the buds used in consumer in-ear headphones, shown in Fig. 2.4(b). This ear plug had sizes of small, medium and large to improve comfort of use for each user.

The Bioboard, shown in Fig. 2.4(c) greatly enhanced portability of the in-ear photoplethysmography owing to its relatively small size (4.3cm by 4.5cm). The Bioboard is generally clipped to clothing at the point of the shoulder or the neck, and has the option to send data via Bluetooth or through a micro-USB connection. The in-ear photoplethysmography sensor shown in Fig 2.4 was used to monitor cognitive workload and respiratory diseases. For experiments detailed in chapters 3, 5 and 9, the MAX30101 sensor was used in conjunction with a large data-logging circuit board and data was stored on an SD card. For the experiments detailed in chapters 4, 6, 7 and 8, the Bioboard was utilised, allowing for the data to be visualised on a computer monitor during the recordings.

### 2.2.3 Adhesive PPG

Given that SpO<sub>2</sub> derived from pulse oximetry relies on accurate measures in pulse amplitude for its calculation, it is extremely sensitive to motion artefacts. When motion artefacts are sparse, such as when a subject is sitting still, these artefact regions can be removed and interpolation can fill in the gaps. However, during tasks involving continuous movement, such as walking, the majority of the amplitude envelope for the SpO<sub>2</sub> calculation can become corrupted. With this in mind, we introduced an adhesive backed behind the ear sensor for the continuous measurement

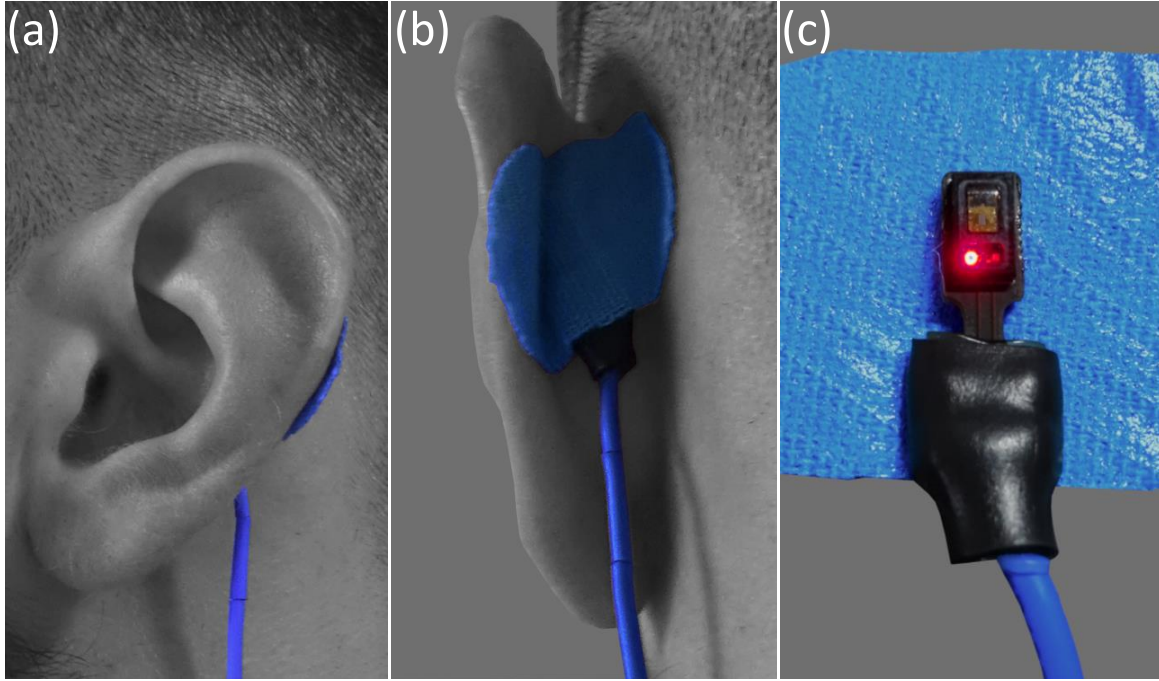


Figure 2.5: The adhesive behind the ear photoplethysmography sensor. (a) The view of the sensor from the side, showing that the sensor is hidden and discrete. (b) The view of the sensor from behind, showing the attachment of the sensor behind the ear. (c) Zoom-in of the reflectance pulse oximetry sensor, on top of an adhesive plaster. The chip itself has dimensions of 5.6mm x 3.3mm x 1.55mm.

of blood oxygen in scenarios such as the 6 minute walk test. The discrete form factor and the behind the ear position are highlighted in Fig. 2.5(a,b). The sensor board backed with an adhesive plaster is shown in Fig. 2.5(c). The principle of an adhesive secured PPG sensor is that movement between the sensor and skin is significantly reduced even during full body motion. With a strong enough adhesive and small enough form factor, motion between the sensor and the skin could conceivably be reduced to the point where it is practically non-existent during all artefact inducing movements. This adhesive behind the ear PPG sensor was used in conjunction with the Bluetooth enabled Bioboard for highly portable recordings.

#### 2.2.4 The sensitivity of $\text{SpO}_2$ readings and skin colour

Implementations of the MAX30101 chip by Maxim Integrated have been approved by the United States Food and Drug Administration (FDA) for  $\text{SpO}_2$  measurements from both the wrist and finger. In terms of error, the FDA guidelines as of 2018 specifically state that the

overall root mean square error for blood oxygen estimates must fall below 3.5% for reflectance pulse oximetry [23]. It is important to note that this applies to full systems and not just the chip, and that the in-ear system presented in this thesis has not been validated as per FDA regulations. Whilst we do broadly discuss agreement between absolute values of finger pulse oximetry and our in-ear pulse oximeter, the key  $\text{SpO}_2$  related findings in this thesis are the  $\text{SpO}_2$  delay and in-ear  $\text{SpO}_2$  decreases with increased mental effort, both of which depend on relative values of  $\text{SpO}_2$  and not accurate absolute values.

It is well known that skin colour can bias pulse oximetry readings. Both melanin and deoxygenated haemoglobin have a high absorbance of red light, and therefore darker skin can change the path at which the light takes through tissue. This can result in a higher  $\text{SpO}_2$  estimate for subjects with darker skin. This bias is known to occur primarily when blood oxygen saturation is low [24], which can mean that those with darker skin are less likely to receive life saving oxygen when it is needed. Given this, FDA guidelines (which the MAX30101 sensor adheres to) state that a pulse oximetry device must be tested on a subject pool where at least 15% of subjects have darker skin pigmentation [23]. In this thesis, diversity of skin colour was present in hospital cohorts where low blood oxygen saturation was examined, but sample sizes were too small to draw conclusions on the effects of skin colour on in-ear  $\text{SpO}_2$  estimation. Notably, the conclusions that this thesis makes relating to in-ear  $\text{SpO}_2$ , namely the  $\text{SpO}_2$  delay and relative  $\text{SpO}_2$  change under cognitive workload, would be unchanged by the effects of skin colour on absolute values. Whilst it is a step forward that the FDA ensures approved pulse oximeters are tested on darker skin pigmentation, the effects of bias are still likely within the tolerance margins. It is my opinion that future pulse oximetry sensors should consider auto-calibration to a user's skin pigmentation, and that this would reduce the likelihood of bias.

## 2.3 Empirical mode decomposition

Empirical mode decomposition (EMD) is a data driven (empirical) method for splitting a signal (decomposition) into components (modes) that can vary temporally in amplitude and frequency

[25]. These components are referred to as intrinsic mode functions (IMFs). Because of its data driven nature, EMD can decompose signals that are non-linear and non-stationary, making it particularly useful for physiological signals such as electroencephalography [26] [27] [28]. This is in contrast to the Fourier transform; the projection of a signal on to sine and cosine basis functions which are assigned different weights to represent the different frequency components in the signal. The Fourier transform works best for linear and stationary data.

### 2.3.1 Univariate EMD

---

**Algorithm 1** Empirical mode decomposition [25]

---

**Input:**  $f(x)$

- (i) Define a proto-IMF as  $f_i(x) = f(x)$  and let  $i = 1$ ;
  - (ii) Identify all local extrema of  $f_i(x)$ ;
  - (iii) Interpolate the maxima  $Env_{max}(x)$  and minima  $Env_{min}(x)$  of  $f_i(x)$ ;
  - (iv) Find the mean envelope,  $Envelope_i(x) = (1/2)(Env_{max}(x) + Env_{min}(x))$ ;
  - (v) Determine the residual:  $Residual_i(x) = f_i(x) - Envelope_i(x)$ ;
  - (vi) Let  $f_i(x) = Residual_i(x)$ ;
  - (vii) Repeat steps (ii)-(vi) until  $Residual_i(x)$  satisfies the IMF criteria;
  - (vii) Let  $f_{i+1}(x) = f_i(x) - Residual_i(x)$  and increment  $i$ ;
  - (viii) Repeat steps with (ii)-(vii) until  $f_{i+1}(x)$  is a monotonic residue or trend.
- 

Within the single channel EMD algorithm, listed under Algorithm 1, firstly a proto-IMF is defined as the original signal. All local extrema are extracted and interpolated separately to give a signal envelope. The mean of these upper and lower envelopes is calculated and removed from the proto-IMF, leaving a residual signal. The proto-IMF is then redefined as the residual. These steps of envelope removal are repeated until the proto-IMF satisfies the following conditions: i) the number of extrema and zero crossings must differ by at most one (giving an oscillatory signal) and ii) at any point the mean value of the maxima and minima envelope must be zero. When the proto-IMF satisfies these conditions it represents a valid IMF, and is removed from the original signal. The steps are repeated again on the signal with the IMF removed, and IMFs are continually extracted (sifting) until a monotonic residue or trend is left. This algorithm is further demonstrated in Fig. 2.6, in which a signal  $f(x)$  is decomposed into 3 intrinsic mode functions.

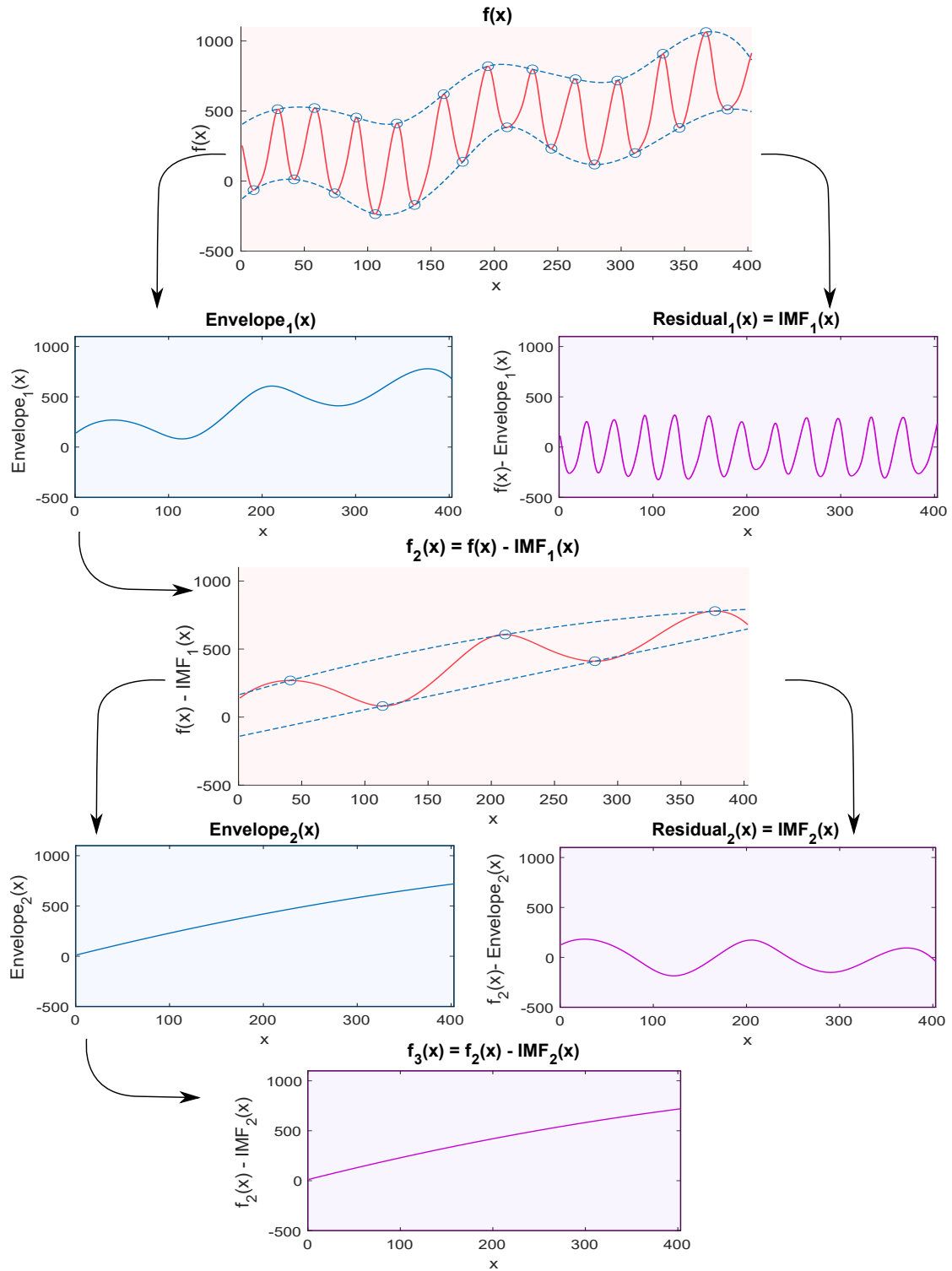


Figure 2.6: An illustration of the empirical mode decomposition (EMD) sifting process, decomposing a signal in red into its three intrinsic mode functions (IMFs) in purple. A signal  $f(x)$  (top, red) is split into its mean envelope  $Envelope_1(x)$  (left, blue) and a residual  $Residual_1(x)$  (right, pink) which satisfies the IMF criteria. The mean envelope then becomes the new signal  $f_2(x)$  and the process is repeated, resulting in another sinusoidal IMF and a linear trend which is the final IMF.

### 2.3.2 Multivariate EMD

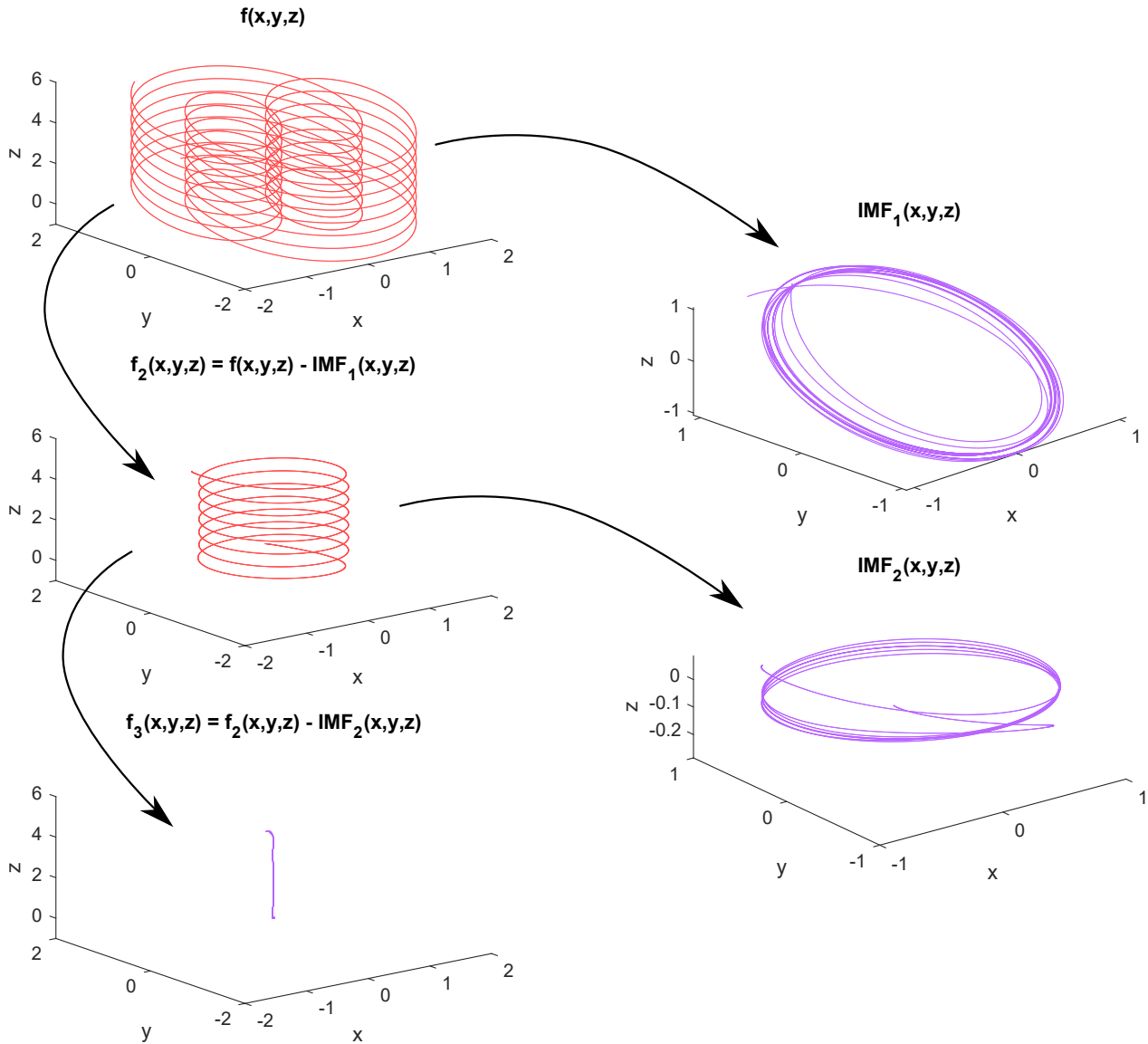


Figure 2.7: An illustration of the multivariate empirical mode decomposition (MEMD) sifting process in three dimensions, decomposing a signal in red into its three intrinsic mode functions (IMFs) in purple. A signal  $f(x, y, z)$  (top, red) is split into its mean envelope  $f_2(x, y, z)$  (middle, red), and a residual  $IMF_1(x, y, z)$  (right, purple) which satisfies the IMF criteria. The process is then repeated on  $f_2(x, y, z)$ , resulting in another IMF and a linear trend which is the final IMF.

Multivariate empirical mode decomposition (MEMD) takes the univariate EMD algorithm and generalises it to multi-channel data [25]. There are no conditions on the number of extrema and zero crossings for an IMF in the MEMD algorithm as extrema cannot be properly defined in multivariate signals [29], but the condition on the mean envelope being zero at any point still



stands. Multivariate EMD works by firstly generating a hypersphere of  $Q$  uniformly sampled points, based on a low discrepancy Hammersly sequence. Discrepancy refers to a lack of balance between two sets, and in this case low discrepancy means that the number of samples in a set will be proportional to the size of the set. If two sets of the same size are selected, they should therefore have a similar number of samples. In other words, low discrepancy quasi-random sequences have a roughly even distribution of sample points, which is favourable to truly random sequences in which there can be large gaps with no samples [30]. The  $Q$ -point Hammersly sequence is used to project the multivariate signal onto  $Q$  single dimensions. The envelope is then extracted for each of these single dimensional projections and the mean is taken across the  $Q$  direction vectors [25]. This multivariate sifting process is repeated and IMFs are extracted according to the satisfaction of the MEMD IMF condition. A common application of MEMD is to scalp-based electroencephalography data given the many channels and the non-stationarity of EEG [31] but it has also been applied in numerous other scenarios, from material fault diagnosis [32] to soil water prediction [33]. The multivariate EMD sifting process is demonstrated in three dimensions in Fig. 2.7.

Multivariate empirical mode composition can encounter the problem of mode mixing, in which two IMFs that are close in frequency and amplitude mix together. One way of avoiding this problem is by adding noise channels to increase the degrees of freedom that MEMD has for extracting IMFs. This technique is referred to as noise assisted multivariate empirical mode decomposition (NA-MEMD) [31].

In this thesis, NA-MEMD was applied successfully to extract respiratory variations from PPG in the context of classifying chronic obstructive pulmonary disease (COPD) and these results are explained in depth in chapter 9. Empirical mode decomposition can also be used to estimate instantaneous frequency, by performing the Hilbert transform on each IMF independently [34]. There are many other methods for estimating instantaneous frequency, a notable example being the commonly used smoothed Wigner-Ville distribution [35][36] which essentially computes the Fourier transform of the local auto-correlation of a signal, and is smoothed to minimise the cross-term interference (the result of the correlation of two separate signal components). The motivation for employing MEMD in this thesis was not its application to instantaneous



frequency. Instead, MEMD was chosen for its ability to extract physically meaningful modes in the time domain, allowing specific time domain characteristics to be extracted such as time spent breathing in as a function of a complete breathing cycle. Whilst the vast majority of research and consumer products focus heavily on extracting respiratory rate from PPG, it was physically meaningful time domain features that enabled our success in classifying COPD.

## 2.4 Machine learning

Many machine learning techniques are data hungry; a statement that is especially true of deep learning techniques. The vast majority of the work in this thesis is at the proof of concept stage, involving small self-recorded data sets without access to so called big data. To this end, the focus throughout this thesis has been on the engineering of physically informed features in combination with machine learning classifiers that have a strong ability to generalise, such as the random forest.

### 2.4.1 Random forest

#### Decision tree

A decision tree classifier [37] uses several decision "nodes" to split data with the goal ending in only pure leaf nodes, defined as nodes which contain only one class of data. An example of this is shown in Fig. 2.8, which shows three boundaries for classification resulting from a tree with three decision nodes. Decision trees commonly select and learn these decision nodes based off maximising information gain, otherwise known as a reduction in entropy from the decision node (parent) to the subsequent nodes (children). Entropy is defined as follows:

$$Entropy = \sum_{i=1}^K -p_i \log(p_i) \quad (2.1)$$

where  $p_i$  corresponds to the probability of class  $i$  and  $K$  is the number of classes.

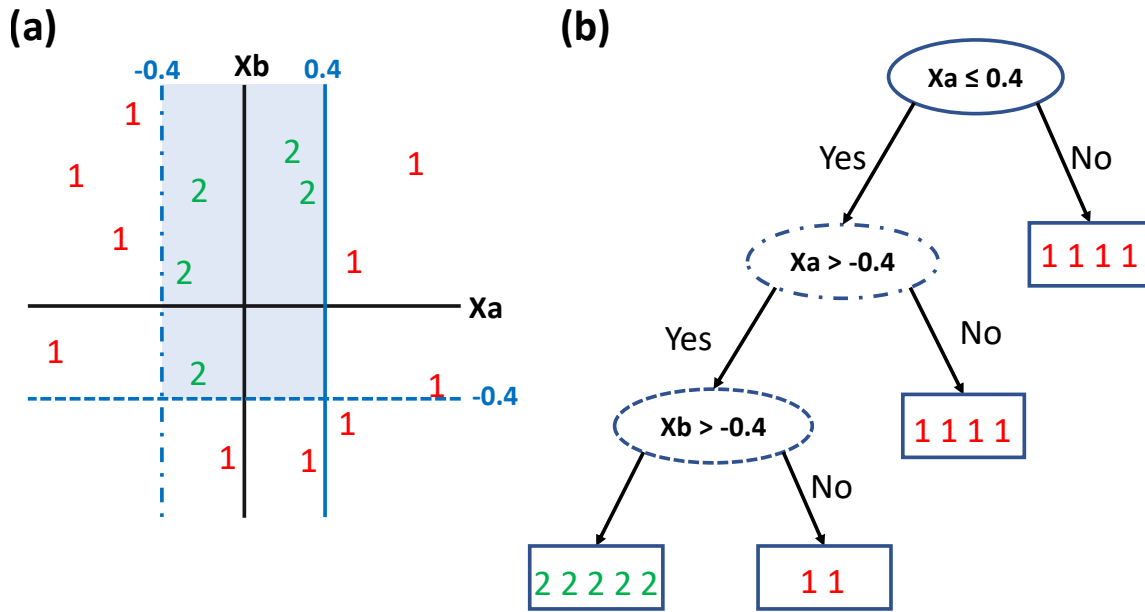


Figure 2.8: A diagram explaining the decision tree for classification. (a) The resulting classification boundaries of a decision tree trained to discriminate between class 1 (red) and class 2 (green). All data points that fall within the blue shaded area will be classified as class 2. (b) The decision tree that produces the classification boundaries, with three decision nodes.

The information gain for a given decision node is the difference between the entropy of the parent node and the weighted sum of entropy from the resulting child nodes, therefore defined as follows:

$$IG = Entropy(Parent) - \sum w_n \times Entropy(Child_n) \quad (2.2)$$

where  $w_n$  corresponds to the relative number of data points associated with the child node  $Child_n$ .

The maximum information gain is 1, which would correspond to a parent node having an entropy of 1 and both child nodes having an entropy of 0. This would occur if the class distribution was equal in the parent node and the child nodes both had a single class, thus going from no discrimination between classes to perfect discrimination.

It is important to note that decision trees are greedy, which is to say that once information gain has been maximised the algorithm does not retrace its steps and reassess previous nodes. Generally decision trees are very likely to overfit to training data, resulting in great performance on training data but poor performance on new data.

## Random forest

A random forest is a collection of many random decision trees [38]. It is less sensitive to training data than a single decision tree, making it less prone to overfitting. The first aspect of randomisation is in its use of a random subset of features to create each decision tree, ensuring that different trees see different features and thus trees have less correlation with each other. Random forests also use bootstrapping followed by aggregating, collectively known as *bagging*. Bootstrapping involves the random selection of data, where duplicates can occur. In this case each tree sees different data, further reducing the chance of overfitting to the data. When testing, a classification decision is made based on the most frequent decision across all trees, otherwise known as aggregating.

## Adaptive boosting

Adaptive Boosting (AdaBoost) [39] is a boosting algorithm that weights the output of several weaker classifiers to perform a classification [40]. The weak classifiers used in AdaBoost are usually decision trees with a single split, known as stumps, but many other base classifiers can be used with the AdaBoost algorithm. There are two key parts to the AdaBoost algorithm; firstly, base classifiers are weighted so that some have more say over the classification than others and secondly, each new classifier is evaluated and weighted considering the errors of the previous classifier. AdaBoost starts with the base classifier that best classifies the data on its own and then works through the rest of the base classifiers.

The weighting for each classifier is calculated as follows:

$$W_i = \frac{1}{2} \log \left( \frac{1 - error_i}{error_i} \right) \quad (2.3)$$

where  $W_i$  corresponds to the weight for the classifier, and  $error_i$  corresponds to the total error (proportion of incorrectly classified samples) of the classifier. For example, if a classifier with two classes has an error of 0.5, corresponding to random performance, its weighting would be zero as it contributes nothing to the overall classification. Moreover, if a classifier reliably

performs worse than random, its weighting would be negative as we want to do the opposite to what the classifier tells us.

If a base classifier performs well but for a few data points, then the data points that it fails to classify will have an increased weighting in the overall dataset. When bootstrapping is performed on the data to create samples to evaluate the next base classifier, the data points with increased weight are more likely to occur multiple times and thus have more influence over the subsequent base classifiers weighting. Aggregation of the weighted results of all base classifiers occurs to make the final classification for a given data point.

Whilst decision trees and commonly stumps are used in conjunction with AdaBoost, there is scope for using random forests themselves as the base classifier, with slightly improved accuracy over a random forest alone [41]. Given the aggregation of several forests, stability of estimation is also improved, with the downside of added computational time.

## Summary

Classifier	RF	DT	MLP	LDA	KNN	SVM
Test Accuracy	<b>90.20%</b>	77.86%	77.94%	54.11%	71.69%	71.94%
Train Accuracy	100.00%	100.00%	85.97%	55.00%	85.91%	78.17%
Test Accuracy $\sigma$	2.91%	3.48%	7.05%	2.15%	3.39%	3.12%
Train Time (s)	0.477	0.030	1.044	0.004	0.000	0.973
Test Time (s)	0.004	0.000	0.001	0.000	0.008	0.027

Table 2.2: A summary of classification outcomes for several classifiers applied to the four category cognitive workload dataset outlined in Chapter 5, all calculated across 10 fold cross validation. Included is the mean testing accuracy, the mean training accuracy, the standard deviation of the testing accuracy, the mean training time and the mean testing time.

In this thesis a range of classifiers from random forest (RF), random forest with AdaBoost, support vector machine (SVM), linear discriminant analysis (LDA) and multilayer perceptron (MLP) were applied to several problems including automatic sleep staging [42, 43], classification of cognitive workload [44], prediction of steering wheel angle with the onset of fatigue in driving and the classification of chronic obstructive pulmonary disease [12]. In all cases the random forest and boosted random forest performed best out of the classifiers tried, likely due to their strong ability to generalise, which is especially important on small datasets where there is a

greater chance of overfitting. An example of these results compared across the aforementioned classifiers and also the decision tree (DT) and K-nearest neighbours (KNN) classifiers, applied to the four stage memory task classification problem detailed in Chapter 5, is provided in Table 2.2. It is clear that both the RF and DT classifiers possess sufficient degrees of freedom to easily discriminate between different classes in the training data, shown in their perfect training accuracy. Decision trees are well suited to problems involving physiological variables, as they are set up to capture the interactions between them. If heart rate is already high in one subject, then perhaps the relative increase in stress of a cognitive workload task won't lead to a further increase in heart rate, or if absolute SpO<sub>2</sub> is low in another subject, then perhaps there will not be a minor decrease in in-ear SpO<sub>2</sub> with increased cognitive load. These interactions are further influenced by changes in breathing and sympathetic tone. The sequential hierarchical decisions built into the decision tree easily capture these dependent interactions between physiological features, where other classifiers can struggle. In the example shown in Table 2.2, the application of ensemble learning across many trees in the random forest reduces the chance of overfitting and significantly increases training accuracy, making it a natural choice for physiological data sets such as this one.

# Part I

## In-ear pulse oximetry for the detection of hypoxia

# Chapter 3

## In-ear SpO<sub>2</sub> as a tool for wearable monitoring of core blood oxygen

### Chapter overview

The non-invasive estimation of blood oxygen saturation (SpO<sub>2</sub>) by pulse oximetry is of vital importance clinically, from the detection of sleep apnea to the recent ambulatory monitoring of hypoxemia in the delayed post-infective phase of COVID-19. In this chapter, the goal was to investigate the feasibility of SpO<sub>2</sub> measurement from the ear canal as a convenient site for long term monitoring, and perform a comprehensive comparison with the right index finger—the conventional clinical measurement site. During resting blood oxygen saturation estimation, a root mean square difference of 1.47% was found between the two measurement sites, with a mean difference of 0.23% higher SpO<sub>2</sub> in the right ear canal. Using breath holds, we observe the known phenomena of time delay between central circulation and peripheral circulation with a mean delay between the ear and finger of 12.4 s across all subjects. Furthermore, lower photoplethysmogram amplitudes from the ear canal are documented and ways to mitigate this issue are suggested. In conjunction with the well-known robustness to temperature induced vasoconstriction, this makes conclusive evidence for in-ear SpO<sub>2</sub> monitoring being both convenient and superior to conventional finger measurement for continuous non-intrusive monitoring in both clinical and everyday-life settings.

### 3.1 Introduction

One of the major roles of blood is to supply oxygen to tissues throughout the body. This is achieved through the protein haemoglobin within red blood cells, which has a high affinity to oxygen. Thus, as blood passes through capillaries in the lungs, the haemoglobin in red blood cells binds to oxygen which is subsequently pumped through arteries via the heart and transported to various tissues. The maintenance of a high arterial blood oxygen saturation is therefore extremely important, as otherwise tissues cease to be adequately supplied. The term blood oxygen saturation specifically refers to the proportion of haemoglobin in the blood that is carrying oxygen [45], and is given by

$$\text{Oxygen Saturation} = \frac{HbO_2}{HbO_2 + Hb}, \quad (3.1)$$

where  $Hb$  refers to haemoglobin not bound with oxygen and  $HbO_2$  refers to haemoglobin bound to oxygen.

Arterial blood oxygen saturation is typically measured using pulse oximetry which gives a percentage estimate (SpO<sub>2</sub>). It has been established that those with a healthy respiratory system typically exhibit SpO<sub>2</sub> values of 96–98% at sea level [46]. According to the World Health Organisation, hypoxia is defined as a blood oxygen saturation level of less than 94%, while a blood oxygen level of less than 90% may indicate the need for clinical action [47]. Hypoxic SpO<sub>2</sub> readings are a sign of hypoxia without breathlessness in COVID-19 patients [48, 49] where the major respiratory failure peaks 10 days after initial infection [50]. Indeed, in the case of COVID-19 it is strongly recommended that patients receive supplemental oxygen if their SpO<sub>2</sub> reading falls below 90% [51]. Hypoxic SpO<sub>2</sub> levels may also occur in other examples of respiratory failure [52] and during breathing obstruction which is common in sleep apnea [53].

In practice, the SpO<sub>2</sub> levels are calculated indirectly, through photoplethysmography (PPG), the non-invasive measurement of light absorption (usually red and infrared) through the blood. In short, when more blood is present, less light is reflected, so given a pulsatile increase in blood volume with each heart beat, PPG effectively measures the pulse. Depending on the level of blood oxygen saturation, the PPG measurements also experience a change in the ratio of light absorbance between the red and infrared light. Namely, the extinction coefficient of oxygenated haemoglobin with red light ( $\approx 660$  nm) is lower than it is for deoxygenated haemoglobin, and the



reverse is true for infrared light ( $\approx 880\text{--}940\text{ nm}$ ) [54]. Thus, a simultaneous measurement of the absorbance of both infrared and red light allows for an estimation of blood oxygen saturation. Calculation of blood oxygen saturation, referred to as the ‘ratio of ratios’ method, will be fully explained in the Methods section.

Reflective PPG can be measured from any site with skin that has vasculature [11], but is, for convenience, commonly measured from the fingers (usually the index finger), wrist and earlobe. This convenience comes with some disadvantages pertaining to 24/7 continuous measurements, which include the intrusive nature of measurements given that living life with a finger or earlobe clip is problematic, and also poor sensor skin contact and motion artefacts in the case of the wrist.

Other devices such as wearing a helmet for behind the ear [55] or a headband for the forehead [56], whilst providing good quality signal, giving a proxy for core blood oxygen rather than peripheral and being less obstructive to daily life, are less suitable for 24/7 use as they can be perceived as stigmatising which decreases adherence to wearing the device in the patient and consumer populations.

On the other hand, the recent interest in the development of Hearables [16], has promoted the ear canal as a preferred site for the measurement of vital signs in eHealth technology. Indeed, the ear canal represents a unique opportunity for physiological measurements, due to its proximity to the brain (ear-EEG [57]), the property of the ear canal to act as a shield from external electrical noise (nature-built Faraday cage [16]), and the general fixed position of the head wrt. vital signs and neural function, as the head does not move much in daily life unlike sites such as the wrist. Owing to these desirable properties, the ear-canal has been established as a feasible wearable site by ear-EEG [57, 18] and ear-ECG [16].

When it comes to ear-PPG, unlike the finger PPG signal and the earlobe signal, it has been shown that the ear canal offers a photoplethysmogram which is stable and resistant to changes in blood volume which occur during hypothermia [58]. This is because peripheral areas of the body experience restricted blood flow during the cold, whereas the ear canal, being a narrow cavity surrounded by skin, maintains internal blood flow levels. Additionally, this makes the ear canal a preferred site for accurate core body temperature measurement. The PPG from the ear canal has also been shown to be far more sensitive than earlobe and finger PPG to amplitude variations that arise from respiration, thus allowing for a better measurement of respiration

rate [8]. Furthermore, a significant delay has been observed between earlobe/behind-the-ear pulse oximetry, and pulse oximetry on the hand or the foot for detection of hypoxemia (low levels of blood oxygen) [59, 55], which is primarily caused by the distance from the core blood supply. Although never previously shown, a similar fast response time was hypothesised from the ear canal given that its vasculature is supplied by the carotid artery, as is the brain. The robustness of the PPG signal from the ear canal, the faster expected SpO<sub>2</sub> response time to changes in blood oxygen levels, the opportunity for a proxy to brain oxygen saturation levels and the potential for relatively non-intrusive 24/7 ambulatory monitoring of patients with breathing disorders form the motivation for this research into ear canal pulse oximetry. The aim of this study is therefore to investigate the feasibility, establish technical characteristics, and identify advantages in the recording of SpO<sub>2</sub> from the ear canal, compared with the most commonly used site, the right index finger.

## 3.2 Materials and methods

### 3.2.1 Hardware

The hardware used in this chapter is detailed in chapter 2, section 2.2.2, with the data from the in-ear sensor being stored directly on an SD card.

### 3.2.2 Experimental protocol

The recordings detailed in this chapter were performed under the 881/MODREC/18 ethics approval, and all subjects gave full informed consent. The participants in the recordings were 14 healthy subjects (7 males, 7 females) aged 19–38 years. Two PPG sensors were used per subject, one secured within the right ear canal and the other to the right index finger; both sensors recorded simultaneously. The subjects were seated in front of a monitor during the recording where a video guided them on when to breathe normally, as well as when to exhale and hold their breath. The position of the sensors, monitor and recording device relative to the subject is shown in Fig. 3.1. This video included a built in countdown, so that at every stage of the protocol, depicted in Fig. 3.2, the subject knew the amount of time left. The experiment lasted for 435 s, consisting of 120 s of normal breathing, 3 repeats of 5 s of exhaling and

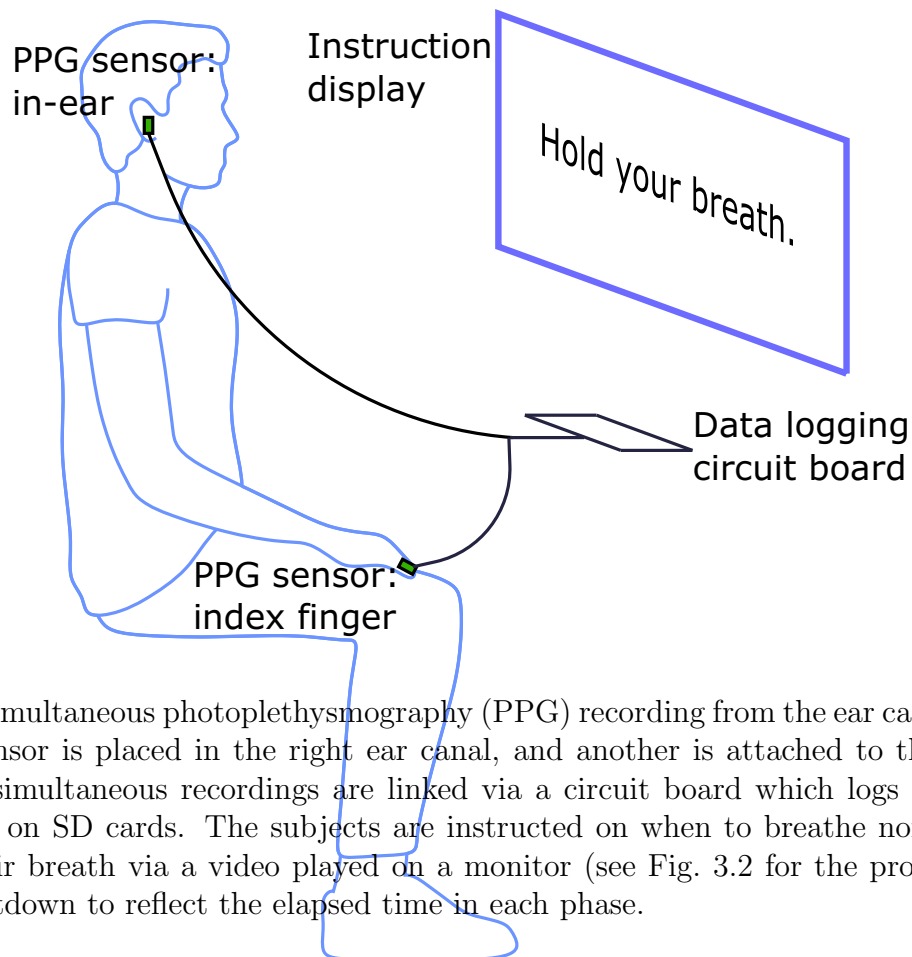


Figure 3.1: Simultaneous photoplethysmography (PPG) recording from the ear canal and finger. One PPG sensor is placed in the right ear canal, and another is attached to the right index finger. The simultaneous recordings are linked via a circuit board which logs the respective data streams on SD cards. The subjects are instructed on when to breathe normally, exhale and hold their breath via a video played on a monitor (see Fig. 3.2 for the protocol), with a built in countdown to reflect the elapsed time in each phase.

20 s of breath holding with 60 s of normal breathing between, and a final 120 s of normal breathing, as depicted in Fig. 3.2. The exhale before the breath hold was included because if the oxygen from the lungs is expelled before holding one's breath, there is a sharper desaturation in blood oxygen than if a subject did not exhale first, thus allowing for clearer comparisons of sensitivity between the different measurement sites. Furthermore, in a real world scenario, such as obstructive sleep apnea, the lungs would not be filled with oxygen before the obstruction to inspiration occurred, making the conventional breath hold, where a subject inhales before holding their breath, less realistic. Subjects were also told to hold down a button, connected to the circuit board, from the moment they started exhaling until they stopped the breath hold. This was necessary, in order to know precisely when the subject was holding their breath, as reaction times would add uncertainty if the instructional display was used as the ground truth.

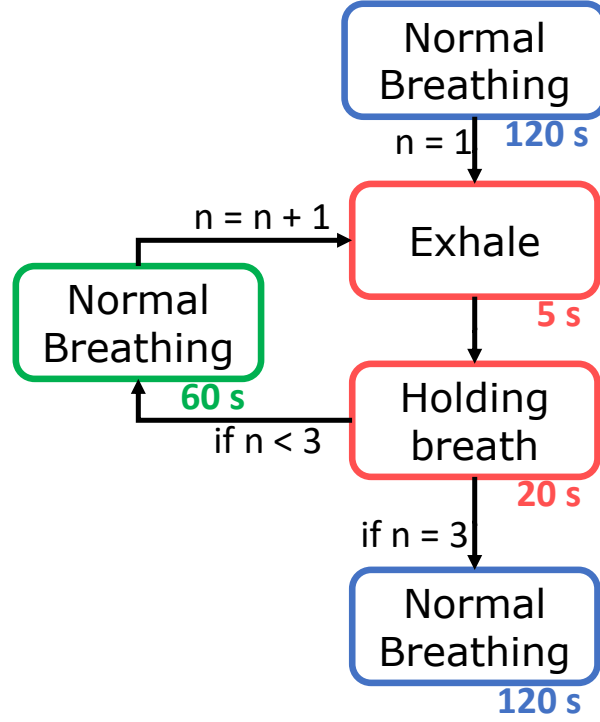


Figure 3.2: Flow diagram of the experiment, outlining the sequence of participants' breathing. The duration of each stage is provided beneath in seconds. The exhale and breath holding stages are repeated three times, the 60 s normal breathing is repeated twice, and 120 s of normal breathing occurs both at the start and at the end of the experiment.

### 3.2.3 Extraction of the $\text{SpO}_2$ signal

The ratio of absorbance of infra-red to red light within the PPG sensor changes depending on the proportion of haemoglobin that is oxygenated in the blood. This change can be quantified through the so called ratio of ratios metric [60], given by

$$R = \frac{\frac{AC_{red}}{DC_{red}}}{\frac{AC_{infrared}}{DC_{infrared}}} \quad (3.2)$$

An empirically derived linear approximation can then be used to calculate an  $\text{SpO}_2$  value as a proxy to oxygen saturation. Using the manufacturer's suggested calibration [61], the  $\text{SpO}_2$  value was calculated as

$$\text{SpO}_2 = 104 - 17R. \quad (3.3)$$

To obtain the alternating current (AC) components within the PPG measurements, the raw signals were firstly band-pass filtered between 1 Hz and 30 Hz. Peak detection was then

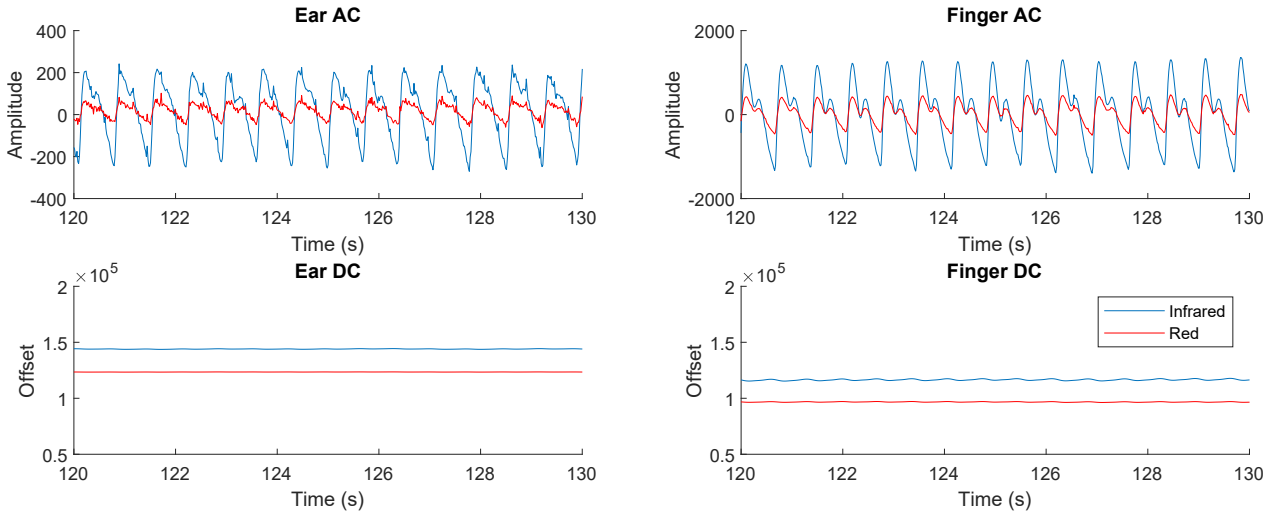


Figure 3.3: Examples of the alternating current (AC) and direct current (DC) waveforms within a PPG measurement, from both the right ear canal and the right index finger. The amplitude of PPG is recorded arbitrary units of light intensity, proportional to the count of photons reflected back to the photo-diode. For simplicity, this is referred to as ‘amplitude’. The amplitude is lower in the right ear canal for both the red and infrared signals, but the signal is clear and usable. The maximum to minimum amplitude was designated as the AC signal amplitude, which is divided by the DC signal in the ratio of ratios metric to calculate  $\text{SpO}_2$ .

performed on the infrared and red AC filtered signals to find their peaks and troughs; this was achieved using the inbuilt MATLAB by MathWorks (Natick, Massachusetts, USA) function `findpeaks`, with a minimum peak prominence of 150 arbitrary units for the infrared signal and 30 arbitrary units for the red signal. The same procedure repeated on the same signals, but scaled by  $-1$ , to find the troughs. Next, the peak values and trough values were separated and interpolated, before their absolute values were added together to give a constant estimate of the AC amplitude. The direct current (DC) components were obtained by low-pass filtering the raw signals at 0.01 Hz. These waveforms are shown in Fig. 3.3.

### 3.2.4 Data Analysis

#### Resting $\text{SpO}_2$ Comparison

The resting  $\text{SpO}_2$  was calculated over the 60 s section of normal breathing before the first exhale began. This time was ascertained from the timing of the first button press, explained in the Experimental Protocol subsection. The  $\text{SpO}_2$  signal from the finger and ear were averaged individually over this 60 s window to give a resting blood oxygen comparison for each of the

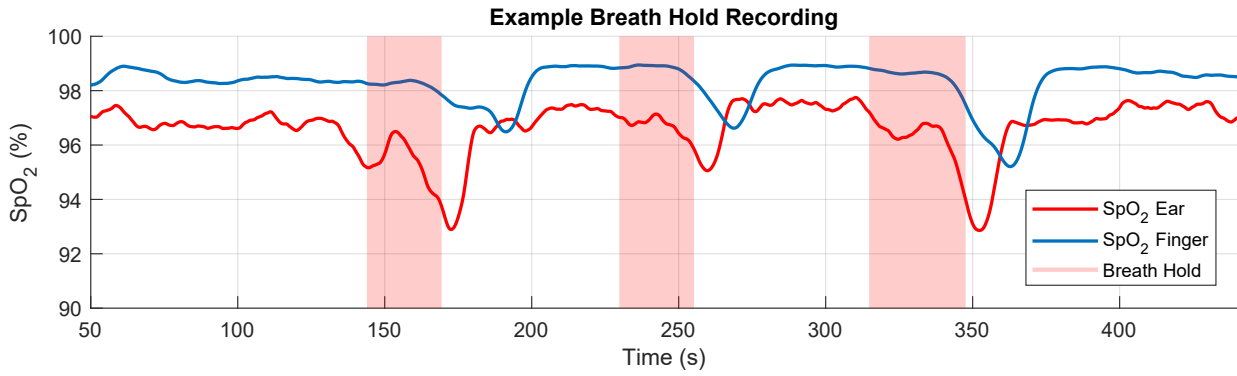


Figure 3.4: Exemplar  $\text{SpO}_2$  recording following the breath hold protocol (see Fig. 3.2). The breath holds (starting with the exhale and finishing at the end of the breath hold) are designated with the shaded regions. The  $\text{SpO}_2$  recorded from the right index finger is denoted by the solid blue line, and the  $\text{SpO}_2$  recorded from the right ear canal by the solid red line. The three significant drops in  $\text{SpO}_2$  levels correspond to the three breath holds. The breath holds also vary slightly in length as they adhere to when the subject holds down a button rather than the set time of the instructional display, thus mirroring the true breath hold duration more accurately.

measuring sites for every subject. The root mean square difference in average resting  $\text{SpO}_2$  was then calculated across all subjects, as well as the mean difference across all subjects.

### $\text{SpO}_2$ Delay

The blood oxygen estimation delay was calculated using the button release point as the marker for minimal blood oxygen, corresponding to the point at which the breath hold ends. The time between this point of minimal blood oxygen and the first trough of the  $\text{SpO}_2$  waveform for the ear and the finger was then used to calculate the  $\text{SpO}_2$  delay for the ear, the finger and then the relative delay between the ear and the finger. Three measurements for delay were taken for each measuring site, one for every consecutive breath hold as shown in Fig. 3.4. The mean of these three recorded delays was taken to give an average delay for each measurement site for each subject. The range, on a per subject basis, was taken as a measure of variability within the recordings for each subject. The distribution of mean delays was analysed against age and sex, whereby a paired sample  $t$ -test was employed to compare male and female delays while Pearson's correlation coefficient was utilised to determine if subject age was correlated to delay.

### 3.3 Results

For rigour, and to demonstrate non-inferiority of in-ear pulse oximetry to finger pulse oximetry, the experimental results span three major aspects related to the feasibility of in-ear PPG; these are the comparison of resting oxygen levels against standard finger PPG, the respective delays in detection of hypoxic events and the corresponding PPG signal amplitudes.

#### 3.3.1 Resting oxygen comparison



Figure 3.5: Mean resting SpO<sub>2</sub> levels taken for each individual subject for the right ear canal (red circles) and the right index finger (blue crosses). The mean was taken across the 60 seconds of SpO<sub>2</sub> data before the first breath hold sequence. The area within the two green dotted lines represents the healthy blood oxygen range of 94–100%.

The difference in resting SpO<sub>2</sub> between the right ear canal and the right index finger across all subjects had a root mean square value of 1.47% with a mean difference of 0.23% higher saturation in the ear. The distribution of resting values across subjects is provided in Fig. 3.5, and shows that all resting SpO<sub>2</sub> values were within the physiologically healthy range of 94–100%, and therefore there was a complete agreement between the ear canal and finger as measuring sites on whether or not a subject had ‘healthy’ blood oxygen levels.

#### 3.3.2 Blood Oxygen Delay

The mean relative time delay per subject between the ear and finger pulse oximetry, in other words the time it took from detecting minimal blood oxygen in the ear to detecting minimal blood oxygen in the finger, ranged from 4.18 s to 24.2 s. The mean relative delay across all 14 subjects was 12.4 s, with a mean relative delay of 9.67 seconds for the female subjects

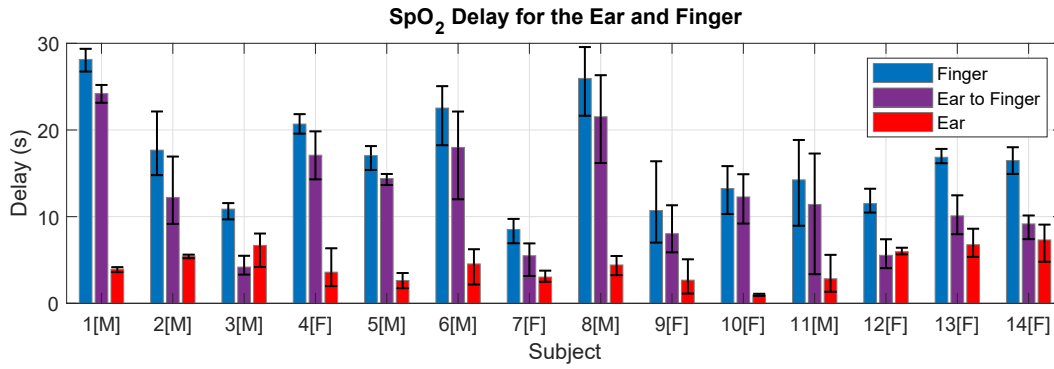


Figure 3.6: The absolute blood oxygen delay from the button press to the finger (blue), to the ear (red) and the relative delay between the ear and the finger (purple), shown for every subject. Male subjects are designated with [M] and female subjects with [F]. The delay is calculated from the time that the breath hold stops (true minimum blood oxygen) to the time at which the measuring site detects a minimum in  $\text{SpO}_2$ . Error bars are included to show the range of the three delay measurements for individual subjects.

and a mean relative delay of 15.13 s for the male subjects. The mean female relative delay was significantly lower than the mean male relative delay ( $P = 0.03$ ). The  $\text{SpO}_2$  delay values for all subjects are shown in Fig. 3.6, highlighting the large inter-subject variability in  $\text{SpO}_2$  delay and moreover the large inter-trial variability in the ear canal vs finger delay for many individual subjects.

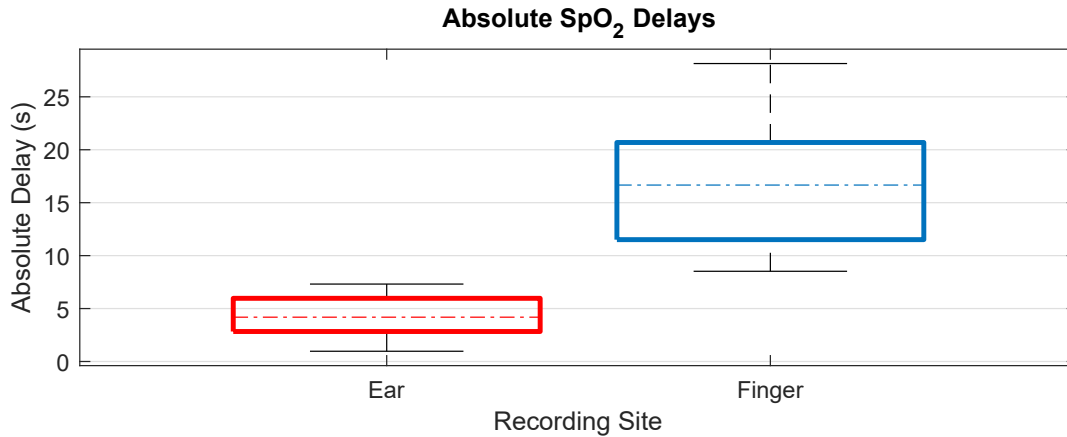


Figure 3.7: Boxplots of the mean absolute delays, for all subjects, between the right ear canal and the right index finger. The top and bottom of each coloured box represent respectively the upper and lower quartiles, the dotted lines represent the median, and the whisker lines extending out of the box represent the range. In our data, there is no overlap between the range of absolute delay from the ear canal and the absolute delay from the right index finger.

Across our 14 subjects there was no overlap between the absolute  $\text{SpO}_2$  delay from the right ear canal and the right index finger, as shown in Fig. 3.7. The absolute delay from the right ear canal had a mean of 4.35 s, with a range of 0.97–7.31 s. The absolute delay from the right



index finger had a mean of 16.75 s, with a range of 8.52–28.14 s. When separated by sex, the mean absolute delay from the finger was 14.0 s for the female subjects and 19.5 s for the male subjects ( $P = 0.03$ ). The mean absolute delay from the ear canal was 4.3 s for females and 4.4 s for males, as summarised in Table 3.1. Furthermore, there was no correlation found between age and oxygen delay across the participants.

	Relative	Finger	Ear Canal
Female	$9.70 \pm 4.07$	$14.00 \pm 4.25$	$4.33 \pm 2.37$
Male	$15.13 \pm 6.76$	$19.49 \pm 6.28$	$4.36 \pm 1.42$
Total	$12.40 \pm 6.06$	$16.75 \pm 5.88$	$4.35 \pm 1.88$

Table 3.1: Summary of mean ( $\pm$  standard deviation) SpO<sub>2</sub> delay values (seconds).

For rigour, the SpO<sub>2</sub> delay between the left and right ear canal was also examined for a single subject, giving a mean delay of 0.46 s from the left ear canal to the right ear canal across the three breath holds, suggesting that measuring SpO<sub>2</sub> from the left ear would yield a similar response time to the right ear, if not slightly faster.

### 3.3.3 Photoplethysmogram amplitude

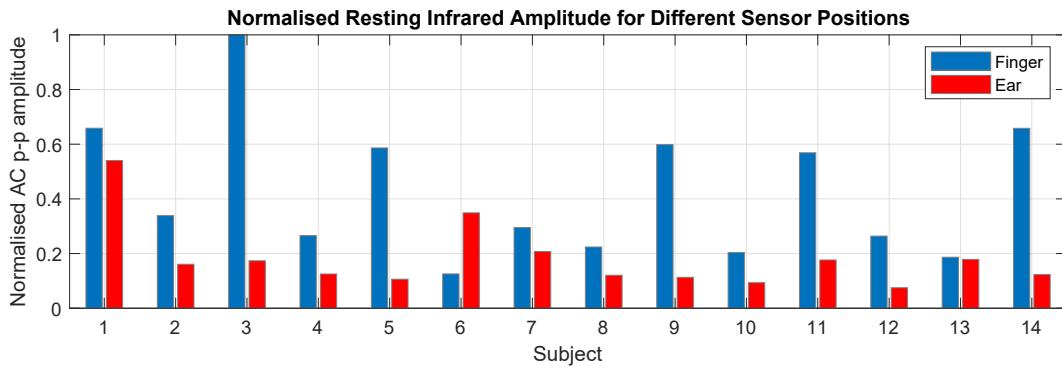


Figure 3.8: The mean resting maximum to minimum AC amplitude of the photoplethysmography signal for all subjects. These values were measured from the 60 s of filtered PPG data before the first breath hold and were normalised to the maximum mean AC amplitude across all subjects.

It was observed that 13 out of 14 subjects had a higher AC PPG amplitude from the right index finger than the right ear canal with the mean finger PPG amplitude 2.35 times higher than for the ear canal. Fig. 3.8 presents the individual normalised resting infrared PPG amplitude for the finger and ear canal results for all 14 subjects. In the case of finger amplitude, males had a mean finger amplitude of  $1.28 \times 10^3$  with a standard deviation of  $0.76 \times 10^3$ , compared with

females who had a mean finger amplitude of  $0.90 \times 10^3$  and a standard deviation of  $0.49 \times 10^3$ . In the case of ear amplitude, males had a mean ear amplitude of  $0.59 \times 10^3$  with a standard deviation of  $0.40 \times 10^3$ , compared with females that had a mean finger amplitude of  $0.33 \times 10^3$  and a standard deviation of  $0.12 \times 10^3$ . Neither sex difference between finger amplitudes and ear amplitudes was statistically significant, with  $P = 0.14$  and  $P = 0.05$  respectively. Importantly, even in the lowest amplitude cases, the in-ear PPG signals were effective for peak detection and thus usable for the calculation of SpO<sub>2</sub>.

Placement	Fast Response	Robust to Temperature
Finger	No	No
Earlobe	Yes [59]	No
<b>Ear canal</b>	<b>Yes</b>	<b>Yes [58]</b>

Table 3.2: Summary of different SpO<sub>2</sub> qualities for the finger, earlobe and ear canal.

## 3.4 Discussion

### 3.4.1 Resting SpO<sub>2</sub>

Blood oxygen saturation measurements from the right ear canal and the right index finger were comparable at rest, with a minimal systematic offset between the two, indicated by the mean difference of 0.23% higher resting SpO<sub>2</sub> in the right ear canal, as shown in Fig. 3.5. The variability between the two measurement sites, indicated by the root mean square difference of 1.47%, is to be expected, given that mean SpO<sub>2</sub> values were shown to vary as much as 0.9% even across fingers on the same hand [62]. The blood oxygen saturation delay between the ear and the finger may also be a source of error when using the same absolute 60 s time window for comparison, although given that the window was long in comparison with the SpO<sub>2</sub> delay and the fact that resting values stay fairly constant, this source of error should be minimal. Both sensor positions were in 100% agreement concerning the patients having healthy SpO<sub>2</sub> levels of 94 to 100%. The established similarity in the resting oxygen saturation values with the commonly used right index finger location are a first step towards validating the ear canal as an alternative measurement site for pulse oximetry.

### 3.4.2 SpO<sub>2</sub> delay

Our results indicate a mean oxygen saturation delay between the right ear canal and the right index finger of 12.4 s. In other words, a steep drop in SpO<sub>2</sub> is detected on average 12.4 s faster from the right ear canal than it is from the right index finger, with two subjects exhibiting over 20 s faster detection from the right ear canal. The faster response from the ear-canal strongly recommends the incorporation of in-ear SpO<sub>2</sub> measurements into so-called Hearables, as there are many situations where fast SpO<sub>2</sub> response time is critical, such as improving outcomes during neonatal resuscitation [63] and closed loop automatic oxygen therapy [64]. Moreover, this opens the avenue to aid comparison of sleep stage scoring and sleep apnea events by aligning changes in SpO<sub>2</sub> to the correct 30 s sleep epoch [65].

A large inter-subject variability was found in both the absolute oxygen delay to the right ear canal and the absolute oxygen delay to the right index finger. On average, the female subjects had lower relative oxygen delay between the ear and the finger, caused by the much smaller absolute delay to the right index finger in combination with an almost identical absolute delay to the right ear canal. We hypothesise that this could be related to males being taller on average and therefore having longer arms, which would give oxygenated blood further to travel from the heart to the fingers. Variations in height and arm length may also explain some of the large inter-subject variability in absolute SpO<sub>2</sub> delays. With a taller subject we would expect on average a longer distance for blood to travel to both the ear canal and the fingers, but no correlation was found between the absolute delay to the ear and absolute delay to the finger. One relevant factor was average heart rate across the trials, which was found to have a significant moderate positive correlation with absolute delay to the ear ( $R = 0.56$ ,  $P = 0.03$ ), but was not correlated with absolute delay to the finger. Importantly, age was not found to be a correlating factor to oxygen delay in our subjects. Future work must consist of investigating the efficacy of in-ear SpO<sub>2</sub> as well as the SpO<sub>2</sub> delay in older cohorts that may have underlying respiratory conditions, and not just young healthy subjects. The natural question arises, could SpO<sub>2</sub> delay be an indicator of the health of a patients circulatory system?

Whilst not as precise as arterial blood gas analysis, this is the first attempt to measure the blood oxygen delay to the ear itself. Importantly it should be noted that although the ear canal responds faster to changes in blood oxygen and much more closely resembles core blood oxygen than peripheral sites such as the finger, it is still not instantaneous, with our results showing a

mean blood oxygen delay to the ear of 4.35 s.

### 3.4.3 Photoplethysmogram amplitude

The in-ear PPG in this study was found to have consistently smaller amplitudes from the ear canal when compared with the finger. The three reasons for a lower PPG amplitude are: (1) less vascular density of the tissue (2) poorer sensor contact with the skin and (3) arterial pulse pressure amplitude. Considering that the ear and fingers are both common PPG placement sites, given the high density of arterioles and capillaries [66, 67], the most likely explanations for a lower PPG amplitude are therefore sensor placement issues within the ear canal and a lower carotid arterial pulse amplitude compared to the radial artery in the arm.

The large variability in ear canal sizes was somewhat mitigated by different available earbud sizes, but even with the smallest earbud some subjects found it difficult to insert the sensor fully into the ear canal. Some subjects with wider ear canals also noted that the sensor became looser during the trial. These issues did not affect the conclusive nature of this feasibility study and will be addressed in future with an improved sensor design, such as by employing ear-hooks, commonly used with sports headphones, to stabilise the position of the sensor within the ear.

Even disregarding placement issues, a lower PPG pulse magnitude may be expected from the ear canal given a higher pressure wave amplitude in the radial artery of the wrist leading to the fingers, compared to the carotid artery leading to the ear canal [68]. This is also evidenced by the broader peak from the ear canal, shown in Fig. 3.3, which is a characteristic of the pressure wave found in the carotid artery.

Importantly, despite a lower signal amplitude, all data recorded from the ear canal was functional for peak detection of the AC PPG signal and therefore the SpO<sub>2</sub> calculation.

## 3.5 Chapter conclusions

There is a growing need for improved non-intrusive wearable SpO<sub>2</sub> recording across many different scenarios, from the monitoring of sleep apnea to the tracking of vital signs of outpatients with a threat of respiratory deterioration, such as those with general breathing disorders or

those recovering from COVID-19. To this end, the possibility of estimating blood oxygen saturation from the ear canal was investigated in healthy subjects and it has been comprehensively demonstrated that the ear can be considered non-inferior compared with conventionally used finger pulse oximetry, with a significantly faster response time averaging 12.4 s. The results have indicated that the favourable speed of in-ear pulse oximetry, in conjunction with the previously documented advantage of resilience to changes in circulation associated with environmental temperature changes (summarised in Table 3.2), may offer significant clinical advantages and a proxy to brain oxygen levels. Given its privileged position on the human body and a fixed distance to vital signs during most daily activities and while sleeping, the ear canal may even be a superior site for measurement of  $\text{SpO}_2$  in the scenario of 24/7 continuous monitoring. In an era of urgency for enhanced tracking of hypoxia and increasing use of smart devices for health monitoring, this proof of concept study provides a compelling argument for the integration of ear canal pulse oximetry into current state-of-the-art Hearables [16].

# Chapter 4

## Adhesive ear-SpO<sub>2</sub> to monitor exercise in patients with breathing disorders

The recordings in this chapter were performed with the help of Professor Nicholas Peters, Dr Patrik Bachtiger and Dr Manisha Gandhi at the National Heart and Lung Institute and Dr Philip Molyneaux, Dr Mairi MacLeod and Dr Suhani Patel at the Royal Brompton Hospital.

### Chapter overview

An ability to monitor respiratory diseases with a non-invasive wearable sensor would provide immense value for disease management, detection of exacerbations and even the recovery of patients in the aftermath of COVID-19. To this end, we monitor SpO<sub>2</sub> in patients with respiratory diseases, both at rest with our wearable in-ear sensor and during six minute walk tests with an adhesive based ear sensor. Our results show minimal bias between wearable reflectance pulse oximetry sensors and hospital grade transmission sensors, with 0.24% lower mean SpO<sub>2</sub> in the wearable sensor at rest, and 0.48% less of a mean drop in SpO<sub>2</sub> from the wearable sensor during the 6 minute walk tests. Variability was high with root mean square differences of 2.76% and 2.75%, but this is expected given a multitude of factors from positional variability to circulation at the probe site. This work paves the way for a future adhesive “SpO<sub>2</sub> patch” for 24/7 unobtrusive monitoring of respiratory disorders, for detection of exacerbation and improvements in chronic respiratory disease management.

## 4.1 Introduction

Patients with breathing disorders, such as severe chronic obstructive pulmonary disease (COPD) or idiopathic pulmonary fibrosis (IPF) can suffer from hypoxic levels of blood oxygen during exercise or even at rest in extreme cases. With in-ear pulse oximetry showing promise for the rapid detection of low blood oxygen in healthy subjects, the next logical step was to validate it clinically in patients that experience low blood oxygen without the need for a breath hold.

### 4.1.1 The six minute walk test

Six-minute walk tests [69] are commonly performed in respiratory wards to measure the impact of exercise on patients with respiratory disorders, with the collection of data on blood oxygen saturation, heart rate, subjective exertion in terms of breathlessness [70] and total distance walked. Blood oxygen percentage and heart rate, usually measured via a hospital grade transmission pulse oximetry finger or earlobe probe, are recorded at the start, end and at every 1 minute interval between. Breathlessness is also recorded on a scale from 1-10 at the same time and distance is noted down at the end of the walk. If blood oxygen levels get low in the absence of symptoms then it is not recommended that the doctor or clinician stop the walk test [71] but patients themselves are allowed to stop at any time.

Currently blood oxygen is sampled every minute, but a continuous measure of blood oxygen, such as from our ear probe, would provide clinicians with richer information on the dynamics of blood oxygen and the interplay between blood oxygen percentage and heart rate.

### 4.1.2 The problem of motion artefacts

Reflectance pulse oximetry is very sensitive to motion artefacts, with tiny movements between the skin and the sensor having the potential to completely corrupt segments of data. Calculation of heart rate is generally robust to motion artefacts given that it relies purely on pulse timing and not amplitude information, which is why reflectance photoplethysmography in devices such as smart watches provide a continuous measure of heart rate. On the other hand, the calculation of SpO<sub>2</sub> relies on the amplitude of each pulse, which is extremely sensitive to motion artefacts.

This is why smart devices that do measure blood oxygen, such as the Apple watch, advise users to remain still over a period of 15 seconds for a single calculation of blood oxygen. In the case of in-ear pulse oximetry and smart watches, artefacts are generally too great whilst walking for continuous SpO<sub>2</sub> measurements to be accurate. To this end, we implemented an adhesive based behind the ear pulse oximeter for the walk tests, which sufficiently limited the movement between the sensor and the skin, allowing for a continuous measure of SpO<sub>2</sub> and heart rate.

## 4.2 Methods

### 4.2.1 Hardware

For resting recordings, our general in-ear SpO<sub>2</sub> sensor was used [72]. For the 6 minute walk tests, an adhesive backed behind the ear PPG sensor was used, as detailed in chapter 2, section 2.2.3 and visualised in Fig. 2.5. The photoplethysmography chip is the MAX30101 by Maxim Integrated (San Jose, CA, USA) with red (660nm) and infrared (880nm) light emitting diodes as well as a photo-diode to measure the reflected light. The in-ear sensor was connected to a small circuit board attached to the subjects clothing which transmitted the data wirelessly via Bluetooth to a laptop.

### 4.2.2 Experimental protocol

The recordings in this chapter were performed under NHS research ethics approval 20/SC/0315. For the resting recordings, blood oxygen was recorded at rest over a period of 2 minutes, with both our in-ear sensor [72] and a hospital grade finger clip sensor simultaneously. For the six minute walk tests, subjects walked back and forth between two cones placed 30 meters apart. The subjects either walked for 6 minutes, or stopped early due to breathlessness or discomfort. During the walk, blood oxygen was recorded with our adhesive-based ear-SpO<sub>2</sub> sensor, alongside a hospital grade earlobe or finger probe. Values from the hospital probes were recorded at the start and end of the walk, as well as once every minute during the walk.



### 4.2.3 Signal processing and analysis

SpO<sub>2</sub> was extracted based on the ratio of red to infrared light, otherwise known as the ratio of ratios [72]. Artefacts were removed and interpolated based on an assumed Gaussian distribution of regular peak magnitudes. Motion artefacts were removed by calculating a z-score for the peak magnitude of the current peak, based on the sample mean and standard deviation of the 10 peaks before. A z-score threshold was determined for each individual case and if the z-score of the current peak was above this threshold it was removed and linear interpolation was used.

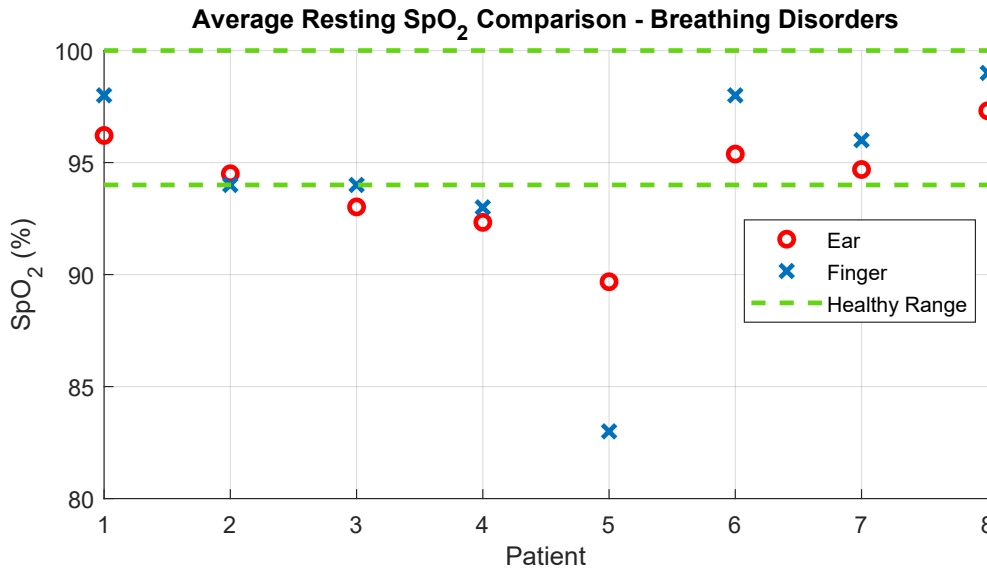


Figure 4.1: Mean resting SpO<sub>2</sub> levels for patients with various breathing disorders. Results are shown for each individual patient for the wearable in-ear probe (red circles) and the hospital finger probe (blue crosses). The mean was taken across the 120s seconds of SpO<sub>2</sub> data for each sensor position. The area within the two green dotted lines represents the healthy blood oxygen range of 94-100%.

In the case of the resting comparisons the mean SpO<sub>2</sub> was calculated using the whole 2 minute period, for both the hospital grade finger probe and our continuous in-ear probe. For the 6 minute walk tests the difference in SpO<sub>2</sub> between the start and the end of the walk was compared for both our wearable behind the ear probe, and the hospital clip finger and ear probes.

The root mean square difference between the wearable ear-SpO<sub>2</sub> and the hospital probes was calculated to assess variability in measurements, and the mean difference was calculated to assess if there was bias between the two measurements.

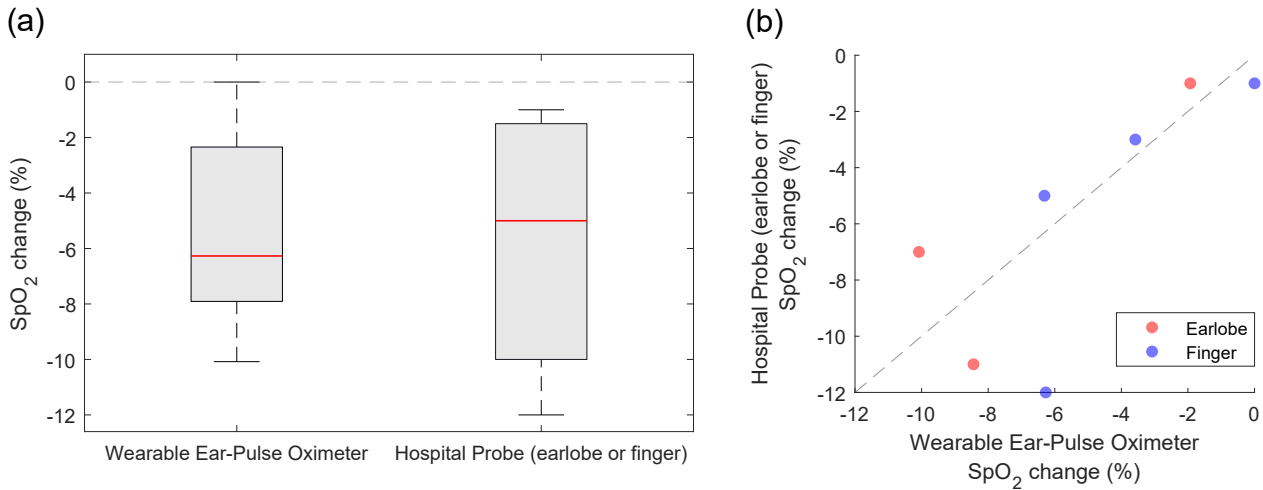


Figure 4.2: Comparisons in SpO<sub>2</sub> changes between a wearable adhesive based behind the ear pulse oximeter and hospital transmission based pulse oximeters positioned either on the ear or finger, for patients with breathing disorders undergoing six minute walk tests. (a) Boxplots for the change in SpO<sub>2</sub> from the start to the end of the walk test, for the wearable ear-pulse oximeter (left) and the hospital probe (right) across 7 patients. (b) Scatter plot showing change in SpO<sub>2</sub> induced by the six minute walk test for the wearable ear-pulse oximeter against change in SpO<sub>2</sub> for the hospital probe, positioned either at the earlobe (red) or the finger (blue), for each of the 7 patients.

### 4.3 Results

The resting hypoxia results, presented in Fig. 4.1, show a rank order agreement between the hospital grade finger probe and the wearable in-ear probe. The root mean square difference was 2.76% when including all results, and 1.53% when removing patient 5 which saw an abnormally large difference in recorded values. The mean difference between the two probes was -0.24%.

The six minute walk test results, visualised in Fig. 4.2, show broad agreement in exercise induced blood oxygen desaturations between the wearable adhesive ear-SpO<sub>2</sub> probe and both the ear and finger based hospital probes. There was a root mean square difference of 2.75% between the wearable ear reflectance probe and the hospital transmission probes and 0.48% less decrease in SpO<sub>2</sub> from the wearable probe, although this is likely due to two large decreases of 12% and 11% disproportionately lowering the mean hospital value. Notably, Fig. 4.2(a) shows that the median decrease in SpO<sub>2</sub> measured with the wearable probe is 1.27% greater than that measured with the hospital probe.

## 4.4 Discussion and conclusion

On real world data, wearable ear-based pulse oximetry shows minimal bias but high variability when compared with hospital transmission based probes. High variability is somewhat expected from different probe sites, given that there are also notable differences in blood oxygen saturation recordings from fingers on the same hand. For resting comparisons, in the case of patient 5 specifically, the finger transmission probe shows far lower values. Whilst the reason for this cannot be known for certain, it is highly likely that circulation was poor in this patients fingers which lead to an abnormally low reading on the finger pulse oximeter. For rigorous conclusions to be drawn future work should compare with an invasive arterial line as the gold standard, but it should be noted that in the case of walk tests this is not feasible.

Importantly, the wearable adhesive probe sufficiently mitigates motion artefacts, allowing for a continuous recording of  $\text{SpO}_2$ . This however is not a perfect solution; artefacts can still dominate if the adhesive is not secure and if there is less blood perfusion to the back of the ear when the patient is cold. Future research could focus on recovering a more robust photoplethysmography waveform in the presence of artefacts, rather than just discarding the artefacts. One possible method is leveraging the two different wavelengths of PPG (red and infrared) with multi-channel empirical mode decomposition. If artefacts are slightly different between the two due to a different diode position, then the shared information that is extracted should be the pulse waveform in both channels. Another method would rely on the characteristic PPG waveform being robust in it's characteristics, similar to electrocardiography (ECG). It is feasible that a recurrent neural network could learn the waveform features and thus extract the underlying pulse waveform whilst removing noise [73]. In the future, patients with chronic respiratory disorders could be provided with an adhesive “ $\text{SpO}_2$  patch” for 24/7 unobtrusive monitoring of respiration, pulse and blood oxygen, similar to a glucose monitoring in patients with diabetes. It is likely that this data would enable an early warning system for respiratory exacerbations.

## Part II

# Classification of cognitive workload and fatigue

# Chapter 5

## In-ear SpO<sub>2</sub> for the classification of cognitive workload

### Chapter overview

The brain is the most metabolically active organ in the body, which increases its metabolic activity, and thus oxygen consumption, with increasing cognitive demand. This motivates us to question whether increased cognitive workload may be measurable through changes in blood oxygen saturation. To this end, we explore the feasibility of cognitive workload tracking based on in-ear SpO<sub>2</sub> measurements, which are known to be both robust and exhibit minimal delay. We consider cognitive workload assessment based on an N-back task with randomised order. It is shown that the 2 and 3-back tasks (high cognitive workload) yield either the lowest median absolute SpO<sub>2</sub> or largest median decrease in SpO<sub>2</sub> in all of the subjects, indicating a robust and measurable decrease in blood oxygen in response to increased cognitive workload. This makes it possible to classify the four N-back task categories, over 5 second epochs, with a mean accuracy of 90.6%, using features derived from in-ear pulse oximetry, including SpO<sub>2</sub>, pulse rate and respiration rate. These findings suggest that in-ear SpO<sub>2</sub> measurements provide sufficient information for classification of cognitive workload over short time windows, which promises a new avenue for real time cognitive workload tracking.

## 5.1 Introduction

Cognitive workload is defined as the level of mental effort undertaken by an individual in response to a task. The mental effort is usually related to working memory, and thus corresponds to the utilisation of brain resources [74]. Cognitive workload affects almost every task-related aspect of our daily lives, from general learning to driving to internet browsing. The ability to accurately measure cognitive load would yield manifold benefits, as too little cognitive workload leaves us vulnerable to distraction, whereas too much cognitive workload makes us prone to making mistakes. Depending on the task a person is engaged in, these mistakes can be more benign, such as less efficient studying, through to life threatening as is the case with driving and the possibility of fatal accidents. The ability to accurately measure and predict cognitive workload would therefore make possible personalised task adaptation, together with the associated benefits on an individual and a societal level, from increasing productivity to decreasing the likelihood of mistakes. Classification of cognitive workload therefore promises immense benefit in diverse areas ranging from driver safety to augmenting human capability with closed loop brain computer interface.

### 5.1.1 Physiological methods for cognitive workload tracking

It is natural to attempt to track cognitive workload based on scalp electroencephalography (EEG), with examples including the classification of skilled vs bad driver performance [75] and the prediction of performance in working memory tasks [76]. Scalp EEG has proven effective at discerning the relevant brain activity changes that arise from changes in cognitive workload, with Johannesen et al achieving 84% accuracy when it came to classifying performance in different stages of a working memory task [76]. Scalp EEG is obtrusive and thus impractical for daily life applications, while discrete and non-stigmatising wearable solutions are still being developed, such as Hearables [16] and ear-EEG [57] [18].

In recent years, eye gaze tracking has become a useful tool for estimating cognitive workload, such as in classification of cognitive workload as well as predicting correctness in an N-back task whilst in a driving simulator [77] [78]. Wang et al achieved an F1 score of 0.71 when distinguishing low and high levels of cognitive load during an N-back task, with 5 second windows of gaze information, and this increased to an F1 score of 0.88 when using 20-second

windows [77]. However, the ways to measure gaze and pupil dilation inevitably involve cameras; these are generally fixed and positioned to track the face and eyes and can be embedded into glasses for wearable gaze tracking.

Other sensing modalities relevant for the estimation of cognitive load include electrocardiography (ECG) and photoplethysmography (PPG) and the use of the corresponding heart rate metrics to classify cognitive workload in a range of scenarios, including driving whilst performing an N-back memory task [79], taking maths tests of varying difficulty [80] and when engaging in a partially automated task with a machine based component [81]. Using ECG data, Tjolleng et al achieved an accuracy of 82% when distinguishing between three levels of cognitive load based on an N-back task [79]. Whilst ECG and PPG are both less obtrusive in daily life than scalp EEG, and offer a wearable solution to cognitive workload tracking, it remains unclear whether the documented increases in heart rate are associated with the stress of performing well during higher cognitive workload tasks [21] [82], or indeed the increased cognitive workload itself. Namely, heart rate is known to correlate strongly with stress level whilst driving, as well as skin conductivity (sweat level) [83]. For the purpose of rigorous cognitive workload tracking, it is therefore important to consider tasks whereby the aspect of stress that a maths test or driving may cause is reduced, whilst still maintaining the ability to vary cognitive workload.

Functional magnetic resonance imaging (fMRI) reliably measures changes in blood flow to different regions of the brain, and due to this it can be considered a gold standard for detecting changes in cognitive workload [84]. However, MRI scanners are not portable or wearable, and thus using fMRI to track cognitive workload does not make sense in real world scenarios. Moreover, due to the strong magnetic field required for the MRI scanner to work, it is not feasible to validate wearable cognitive workload trackers during simultaneous fMRI.

### 5.1.2 Protocols for inducing changes in cognitive workload

#### Arithmetic tasks

Arithmetic tasks are great for inducing constant cognitive workload and can be adapted to have many difficulty levels, examples of which include stating if numbers are divisible by a divisor that increases in size [85] or adding two numbers with an increasing number of digits [86]. A

major downside to using arithmetic tasks is that they also tend to induce stress. Participants are likely to feel judged on their mathematical ability and therefore may put increased pressure on themselves to perform well. An example of this is a paper which used ECG to classify the increased cognitive workload of a maths test provided to students, but used heart rate and heart rate variability metrics (primarily associated with stress) to distinguish these scenarios [80].

### **Real-world of simulated virtual reality tasks**

Real-world and simulated virtual reality task allow for cognitive workload tests that are specific to a specific vocation and thus have real-world applicability. Examples of this include driving [75] and surgery [87]. Importantly, VR tasks are immersive which limits external stimuli and distractions and may lend itself to a more consistent cognitive workload output. With specific tasks however, they are usually only useful in subjects which have prior training with that task, such as air traffic controllers using an air traffic control simulator [74].

### **The N-back task**

The N-back task involves a subject recalling something, such as a number [88] or letter [74], that they saw or heard N-steps back depending on the stage of N-back that the subject is in. For example, for a sequence of numbers 8, 7, 3, 5, where 5 was given last, the correct 0-back response is 5, 1-back is 3, 2-back is 7, and 3-back is 8. Generally, the N-back task is only performed to a maximum level of 3-back. The N-back task is good at inducing cognitive workload given that it provides 4 clear levels of workload and that there is usually a marked subjective increase in difficulty between 0-back and 3-back. Moreover, N-back tasks can easily be introduced as a secondary task that is performed at the same time as a more specific primary task [77] [89] [79]. The downside of the N-back task is that subjects can become easily distracted, or may have to mentally refresh and for example wait for 3 new windows to pass in the middle of a 3-back task. This lends itself to areas of low cognitive workload in regions which were designed to induce consistently higher cognitive load. Despite the issue of distraction, N-back tasks are excellent for inducing quantifiable levels of cognitive load, require minimal training and likely induce less stress than tasks such as driving or arithmetic tests. For these reasons, our version



of an N-back task was implemented in this chapter.

### 5.1.3 The brain, oxygen and cognitive workload

The brain is the most metabolically active organ in the human body. At rest, the brain consumes 20% of the body's oxygen [90] and this percentage increases with increased cognitive demand. Oxygen restriction has significant effects on cognitive function; for example, less oxygen delivery to the brain has been observed in those with memory impairments [91] [92]. Moreover, the administration of oxygen, through the breathing of supplemental oxygen and the associated increase in blood oxygen, has been shown to result in a significantly better memory performance and faster reaction times [93] [94] [95] [96].

Functional near-infrared spectroscopy (fNIRS), a tool for measuring oxygenation of tissue and thus oxygen consumption, has shown increases in oxygen consumption of the brain with an increase in cognitive workload in drivers [97]. Furthermore, fNIRS has helped to detect increased oxygenation of specific brain regions (such as the left inferior frontal gyrus, involved in language processing) with an increase in the difficulty of a letter based N-back memory task in pilots [74]. This motivates us to investigate whether these changes in oxygen consumption are also observable in spectral analysis of blood, or if they manifest themselves through changes in breathing rate or breathing magnitude.

Given that the brain is the most metabolically active organ in the body and that it increases oxygen consumption with cognitive workload, we here hypothesise that increased cognitive workload may be measurable through blood oxygen saturation. Considering that wearable in-ear pulse oximetry provides a robust SpO<sub>2</sub> signal with minimal delay, we set out to answer whether in-ear pulse oximetry can be used to accurately classify different levels of cognitive workload, and furthermore can this classification be performed in an almost real-time fashion?

## 5.2 Methods

### 5.2.1 Hardware

The hardware used in this chapter is detailed in chapter 2, section 2.2.2, with the data from the in-ear sensor being stored directly on an SD card.

MATLAB 2018a by MathWorks (Natick, MA, USA) was used to create a graphical user interface which refreshed four single digit numbers every 5 seconds on a screen in front of the subject. The MATLAB program also communicated with an Arduino Uno by Arduino (Somerville, MA, USA) with each refresh, which in turn communicated with the data logging circuit board with an electrical pulse to align the PPG data to each 5 second window.

### 5.2.2 Experimental protocol

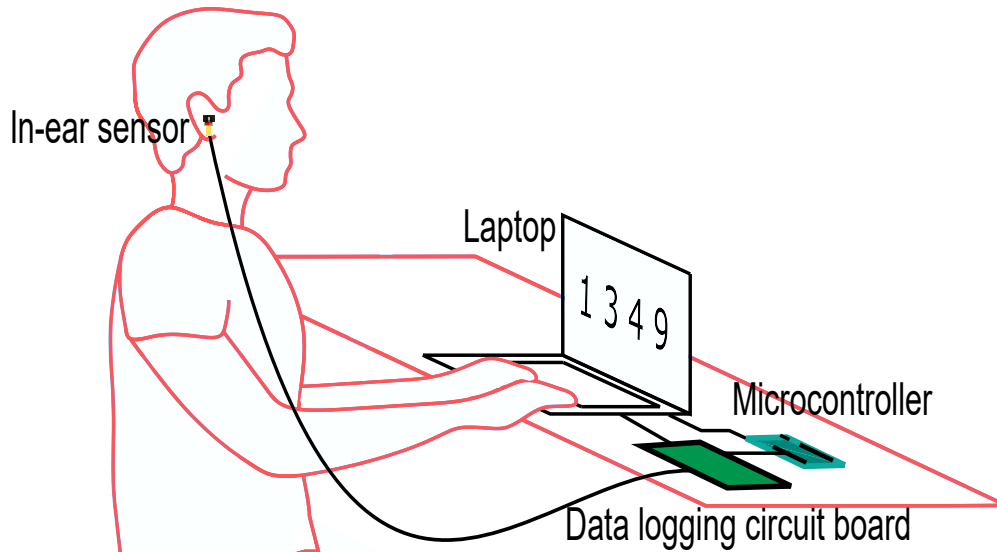


Figure 5.1: Illustration of the recording of in-ear  $\text{SpO}_2$  during an N-back task. The  $\text{SpO}_2$  sensor links to a circuit board which logs the data stream and also accepts input from the microcontroller. The four single digit numbers displayed on the laptop refresh every 5 seconds, and communicate this refresh time to the microcontroller, which in turn sends an electrical pulse to the circuit board to align the task with the physiological data.

The recordings in this chapter were performed under the Imperial College London ethics committee approval JRCO 20IC6414, and all subjects gave full informed consent. The participants in the recordings were 10 healthy subjects (5 males, 5 female) aged 22 - 29 years. A single

PPG sensor was used per subject and was safely secured within the right ear canal. The subjects were seated in front of a monitor during the recording where a MATLAB graphical user interface updated with 4 randomly generated single digit numbers every 5 seconds, as shown in Fig. 5.1. Subjects were asked to count the number of odd numbers and, depending on the N-back trial, they were tasked with entering the current number of odd numbers using the keyboard (0-back), the previous number (1-back), the number 2 steps back (2-back), or the number of odd numbers 3 steps back (3-back). Each trial lasted for 5 minutes and 40 seconds (68 5 second epochs), with 6 epochs used for calibration, leaving 62 epochs for analysis. Four trials were performed by each subject, corresponding to the four levels of N-back task that were presented in a quasi-randomised order. Each subject was given between 5 and 10 minutes rest between trials, and allowed to practice until they were confident with the tasks before the recordings started.

### 5.2.3 Signal processing

Extraction of the SpO<sub>2</sub> was performed according to Chapter 3, subsection 3.2.3. The peak detection procedure of the AC infrared troughs was also used to calculate pulse rate, given that the PPG peak from the ear canal is broader than the peak from the finger (a characteristic of the pressure wave found in the carotid artery) and would thus give a noisy pulse rate estimate. An example of the photoplethysmography pulse signal from the ear is shown in Fig. 5.2a.

Fluctuations in the baseline of ear-PPG due to inspiration and expiration have been evidenced as far stronger from the ear-canal than from the finger [12] [8]. For the calculation of respiration rate, the raw PPG signal was first band-pass filtered between 0.2Hz and 30Hz, followed by a moving average filter with a 150 sample window (corresponding to 2.4 seconds). Peak detection was performed using the MATLAB function *findpeaks* with a minimum peak prominence of 10, to give respiration peaks. The difference of the timings of these peaks was then used to give a breathing interval, shown in Fig. 5.2b. The inverse of the interval signal was then multiplied by 60 to give breathing rate (in breaths per minute). The amplitude values of the respiration peaks were also used as an estimate of breathing amplitude. No epochs of data were discarded, even in the presence of motion artifacts.

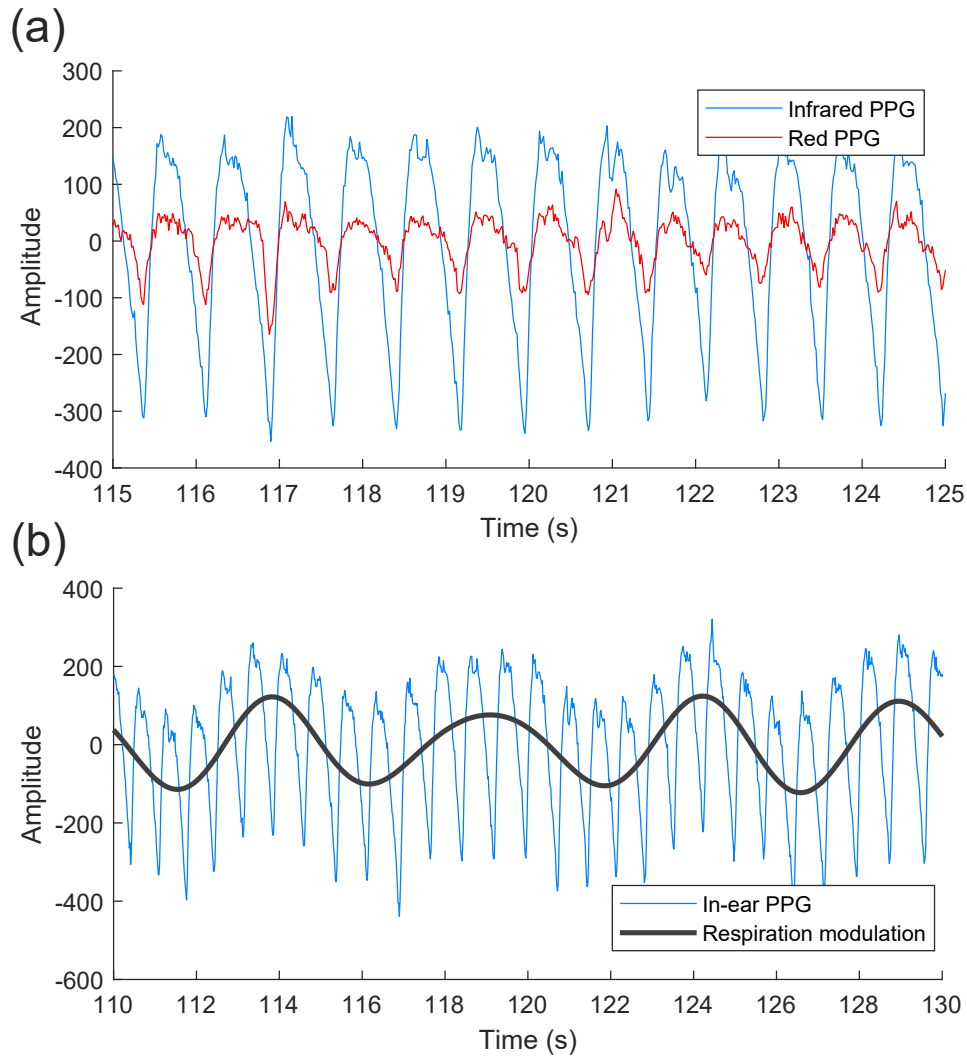


Figure 5.2: Overview of the signals recorded from in-ear photoplethysmography. (a) Both the red and infrared AC photoplethysmography signals, bandpass filtered between 1Hz and 30Hz for the calculation of heart rate and  $\text{SpO}_2$ . (b) The infrared in-ear PPG signal bandpass filtered between 0.2 and 30Hz with the respiration modulation superimposed in black.

#### 5.2.4 Feature extraction

For each 5-second epoch, 21 time domain features (13  $\text{SpO}_2$  based features, 5 pulse based features and 3 breathing based features) were extracted. Frequency-based features were not used as the 5-second window is too short for reliable heart rate variability metrics from PPG. Five features were calculated using both the 5-second epoch and the calibration data from the start of the task. This was particularly important in the case of  $\text{SpO}_2$ , as although healthy  $\text{SpO}_2$  levels generally fall within a small range of 94-100%, the changes we detected due to cognitive load were less than 1%. Whilst absolute values are adequate for testing and training on the same subject, features that are relative to a calibration period are more useful for generalising

Table 5.1: Summary of features used for the classification of cognitive workload with 10-fold cross validation.

Category	Features
SpO <sub>2</sub>	SpO <sub>2</sub> mean, relative change in SpO <sub>2</sub> , red amplitude mean, infrared (IR) amplitude mean, relative change in red amplitude, relative change in IR amplitude, red AC/DC ratio mean, IR AC/DC ratio mean, red peak prominence mean, IR peak prominence mean, red/IR AC ratio mean, red amplitude variance, IR amplitude variance.
Pulse	Heart rate mean, relative change in heart rate, pulse full-width-half-maximum (FWHM) mean, pulse width ratio <sup>†</sup> mean, pulse width ratio variance.
Breathing	Breathing rate mean, relative change in breathing rate breathing amplitude mean.

<sup>†</sup> Pulse width ratio is the ratio between the FWHM of the peak and the FWHM of the trough, giving a systolic to diastolic duration ratio.

across subjects. The 21 features used are summarised in Table 5.1.

### SpO<sub>2</sub> features

The SpO<sub>2</sub> mean was calculated based on the ratio of ratios defined in chapter 3. Infrared amplitude mean and variance were defined as the mean and variance of the infrared light peak amplitudes when the infrared signal has been band-pass filtered between 1Hz and 30Hz, and the red amplitude mean and variance were defined as the mean and variance of the red light peak amplitudes when the red signal has been filtered in the same way. Alternating current to direct current (AC/DC) ratios were defined as the mean peak amplitudes after band-pass filtering between 1Hz and 30Hz, divided by the mean of the signal low-pass filtered at 0.01Hz. Peak prominence was defined as the peak value minus the minimum of the signal, either between two peaks that had larger peak values than itself, or across the whole signal if it was the highest peak. All relative features were calculated as the feature minus the same feature calculated from the 6 calibration epochs at the start of the task.

## Pulse features

Pulse based features were calculated from the infrared light signal band-pass filtered between 1Hz and 30Hz, shown in Fig. 5.2(a). Heart rate is defined as 60 divided by the peak to peak time interval in seconds. The mean heart rate across the 5-second window and mean relative heart rate compared with the initial calibration period were used as features. The pulse width was implemented using the full-width-half-maximum (FWHM), defined as the width of the peak at half of the peak height relative to the rest of the signal [98]. Pulse width ratio is the ratio between the FWHM of the peak and the FWHM of the trough, giving a systolic (heart beating) to diastolic (heart resting) duration ratio given by

$$\frac{Width_{systolic}}{Width_{diastolic}} \simeq \frac{FWHM_{peak}}{FWHM_{trough}} \quad (5.1)$$

where FWHM is the full-width-half-maximum which was used as pulse width [98]. Both the mean and variance of the pulse width ratio were used as features.

## Breathing features

Breathing related features were calculated from the infrared light signal band pass filtered between 0.2Hz and 30Hz and moving average filtered over a window of 2.4 seconds, resulting in a similar respiratory modulation signal to the example shown in Fig. 5.2(b). The breathing rate was calculated as 60 divided by the interval in seconds between breathing modulation peaks. The mean of this across the 5-second interval was used as mean breathing rate, and relative breathing rate was calculated as the mean breathing rate of the current segment minus the breathing rate of the calibration period. Breathing amplitude mean was calculated as the mean peak amplitude of the breathing modulation signal.

### 5.2.5 Classification and evaluation

A random forest classifier with AdaBoost was employed from the publicly available scikit-learn Python toolbox [99]. For the random forest base, the number of trees was set to 50, the class weight was set to ‘balanced subsample’. For the AdaBoost framework, the random forest was

set as the base classifier, the maximum number of estimators was set to 50, the learning rate was set to 1.0 and the real boosting algorithm “SAMME-R” [100] was used.

Ten-fold cross-validation was employed on the fully shuffled data for the case of four-category classification (0-back, 1-back, 2-back, 3-back). Leave-one-subject-out cross-validation was employed on two-category classification (0-back and 3-back). All 21 features were used in ten-fold cross-validation, but only the mean SpO<sub>2</sub>, the relative change in SpO<sub>2</sub> and the mean heart rate were used in leave-one-subject-out cross-validation. In the case of 10-fold cross-validation, the maximum number of features was set to 10 while for leave one subject out cross-validation the maximum number of features was set to 3, as only 3 features were used. Class-specific accuracy and overall accuracy were used as metrics to evaluate classification performance.

Given the three categories of features used (pulse, SpO<sub>2</sub> and breathing), feature importance by means of a reduction in tree impurity was calculated for each feature. This was used to ascertain the relative contribution of SpO<sub>2</sub> derived features compared with conventionally used features such as heart rate and breathing rate.

## 5.3 Results

The mean mistake percentages across subjects for each N-back stage were  $4.0\% \pm 3.3\%$ ,  $4.8\% \pm 4.3\%$ ,  $17.1\% \pm 14.4\%$  and  $29.4\% \pm 21.1\%$  for 0-back, 1-back, 2-back and 3-back tasks, respectively. The substantial increase in mistakes between 1-back, 2-back and 3-back tasks indicates that the 3-back and 2-back tasks were difficult enough to create a meaningful increase in cognitive workload.

### 5.3.1 Change in blood oxygen, heart rate, breathing rate

The mean recorded SpO<sub>2</sub> across all subjects and trials was  $97.0\% \pm 1.7\%$ , the mean heart rate was  $76.7 \pm 11.4$  beats per minute, and the mean estimated breathing rate was  $13.5 \pm 3.3$  breaths per minute. All results therefore fell into the physiologically expected range.

We observed a decrease in median SpO<sub>2</sub>, relative to the start of the task, with every increase in N-back difficulty. The median SpO<sub>2</sub> relative to the start of the N-back task was  $+0.373\%$ ,

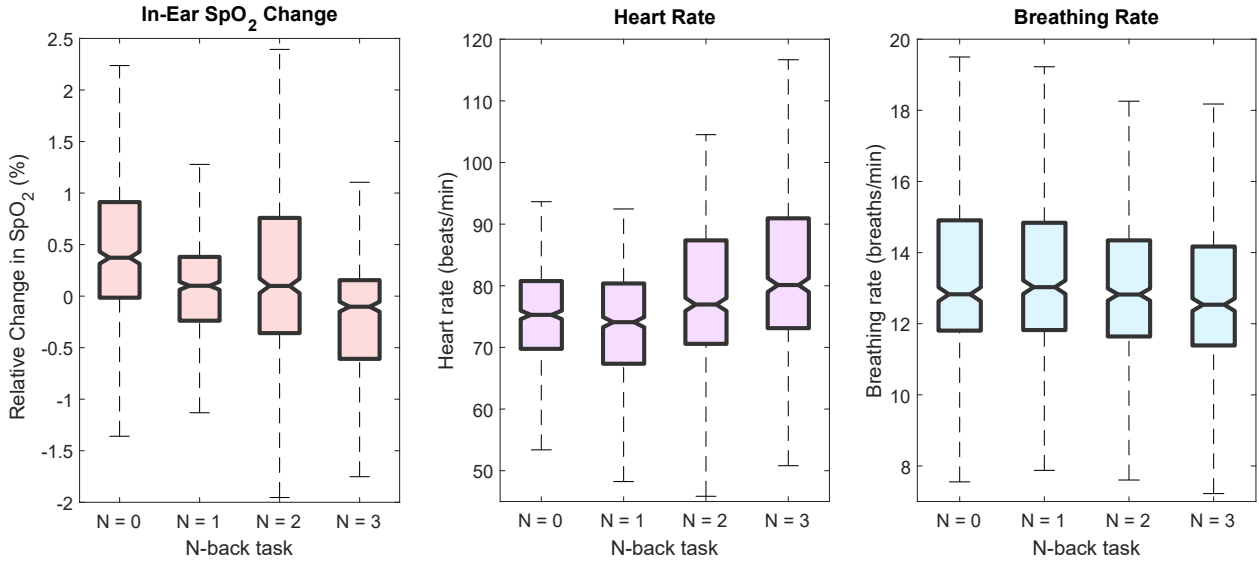


Figure 5.3: Box plots of the relative change in SpO<sub>2</sub> (left, red), heart rate (middle, purple) and breathing rate (right, blue) from the in-ear sensor, split into N-back categories and including each 5 second epoch. The top and bottom of each box represent respectively the upper and lower quartiles, the center notches of each box designate the median, and the whisker lines extending out of the box the range.

+0.101%, +0.099% and -0.102%, for 0-back, 1-back, 2-back and 3-back respectively, as shown in Fig. 5.3. Moreover, for the most difficult 3-back task either the median SpO<sub>2</sub> or median SpO<sub>2</sub> relative to the start of the task was the lowest out of all tasks in 8 out of the 10 subjects, while in the remaining two subjects this occurred for the 2-back task. Similarly, the overall median heart rate was highest in the 2-back and 3-back tasks, but on an individual subject basis the highest median heart rate only occurred in the 3-back task in 5 out of 10 subjects. A slight decrease in median breathing rate was also observed with the 3-back task.

Correlations between the task median relative change in SpO<sub>2</sub> and the task median heart rate, and both N-back difficulty and proportion of errors, were also examined. Fig. 5.4 highlights the prominence of a medium negative correlation between the median relative change in SpO<sub>2</sub> and N-back difficulty ( $r = -0.45$ ,  $p = 0.004$ ) and between the median relative change in SpO<sub>2</sub> and the proportion of errors ( $r = -0.34$ ,  $p = 0.031$ ). Similarly, there was a medium positive correlation between the median heart rate and N-back difficulty ( $r = 0.35$ ,  $p = 0.027$ ) and between the median heart rate and proportion of errors ( $r = 0.416$ ,  $p = 0.008$ ). The strongest correlation was therefore seen between a decrease in relative SpO<sub>2</sub> and increase in workload. Importantly, relative SpO<sub>2</sub> was more correlated with workload than with errors and the opposite was true of heart rate. A negative correlation was seen between median breathing rate and workload,



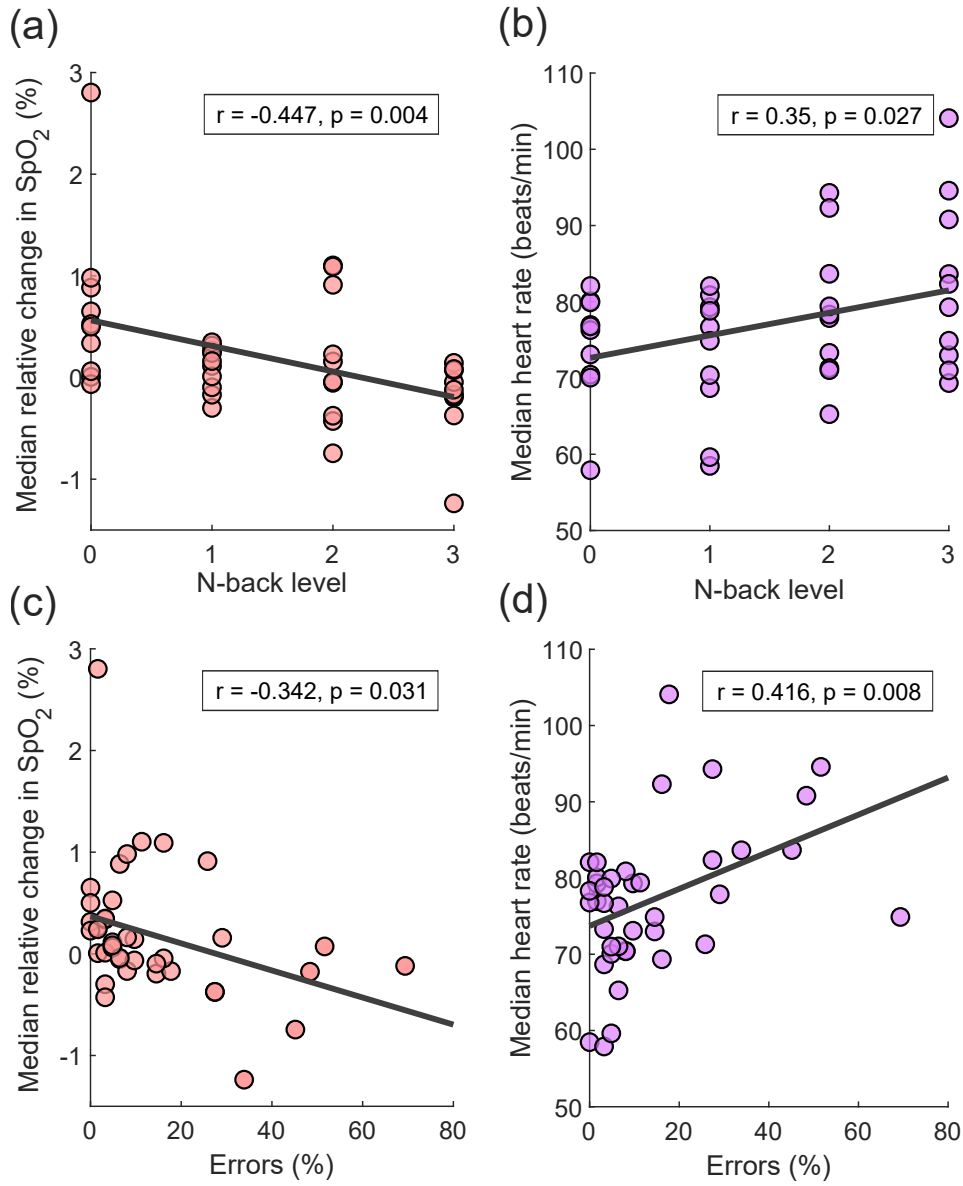


Figure 5.4: Scatter plots of median relative change in SpO<sub>2</sub> (red) and median heart rate (purple) against N-back level and proportion of errors, with trend lines (black) and correlation coefficients and p values superimposed. (a) Median relative change in SpO<sub>2</sub> against N-back level. (b) Median heart rate against N-back level. (c) Median relative change in SpO<sub>2</sub> against proportion of errors. (d) Median heart rate against proportion of errors.

but this correlation was not significant ( $r = -0.28$ ,  $p = 0.086$ ).

Fig. 5.5 further demonstrates the high separability of 0-back and 3-back tasks with SpO<sub>2</sub> features, through a two-dimensional kernel density plot of mean SpO<sub>2</sub> and relative change in SpO<sub>2</sub>.

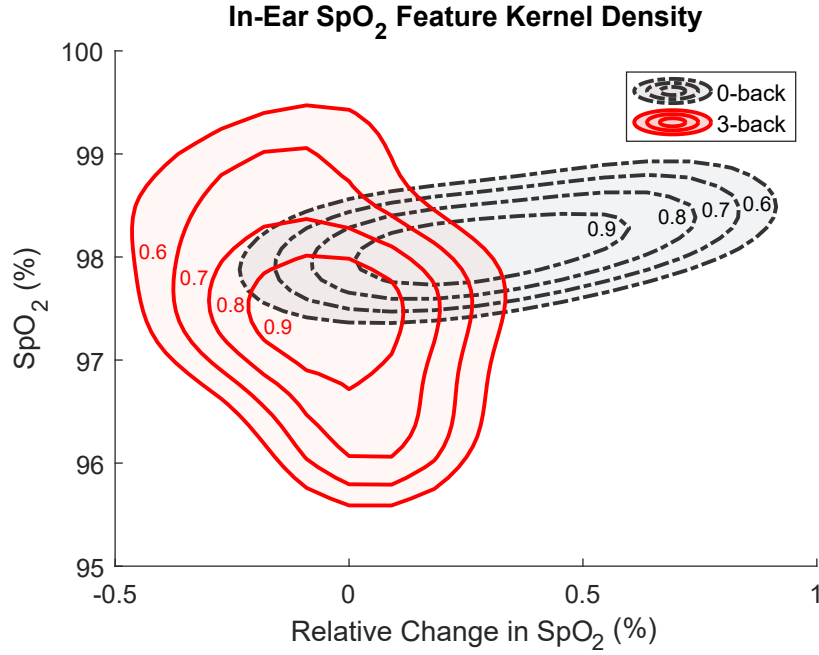


Figure 5.5: Two-dimensional contour plots of the kernel density estimates for mean  $\text{SpO}_2$ , and relative change in  $\text{SpO}_2$ , derived from all of the 5 second epochs of data. The kernel density estimates are plotted independently for two categories: 0-back (black dotted line) and 3-back (red solid line). Kernel density was normalised between 0 and 1 for each category, and the corresponding values for each contour line shown are marked within the contour lines themselves.

### 5.3.2 Classification

#### Shuffled ten-fold cross-validation

With ten-fold cross-validation, shuffled across all participants, we were able to classify the 5-second 0-back epochs with an average accuracy of 93.4%, 1-back epochs with an accuracy of 89.2%, 2-back with 89.5% and 3-back with 90.2%, giving a total average classification accuracy of 90.6%. The largest errors occurred with the miss-classification of 1-back as 3-back and vice versa, with larger errors also occurring between 0 and 1, 1 and 2, and 2 and 3. Classification accuracy was notably better for 0-back and 3-back tasks, as evidenced by the full confusion matrix averaged across 10-fold cross validation in Fig. 5.6(a).

Averaged feature importance (according to reduction in tree impurity in the random forest) across each fold for ten-fold cross-validation was calculated with the top 10 features presented in Fig. 5.7. The two most important features for classification in the case of shuffled ten-fold cross-validation were the mean heart rate and the mean  $\text{SpO}_2$ . Moreover, 6 of 10 most

(a) 10-fold Cross Validation

		Predicted N-back			
		N = 0	N = 1	N = 2	N = 3
True N-back	N = 0	93.4	3.4	1.6	1.6
	N = 1	2.4	89.2	2.7	5.6
	N = 2	3.2	4.2	89.5	3.1
	N = 3	1.3	5.5	3.1	90.2

(b) Leave One Subject Out: Binary

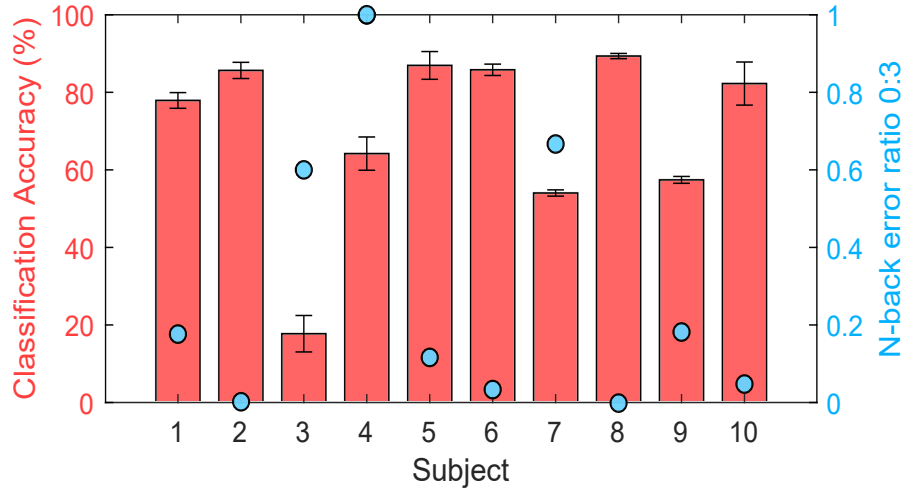


Figure 5.6: Classification accuracy of SpO<sub>2</sub> based cognitive workload estimation. (a) Mean confusion matrix for the results of ten-fold shuffled cross-validation in four-category prediction (0-back, 1-back, 2-back and 3-back). The rows correspond to the true N-back category, and the columns to the category predicted by the classifier. (b) Classification accuracy for testing on each subject, with the classifier trained exclusively on the other 9 subjects (red bars) and the N-back error ratio between the 0-back and 3-back tasks for each subject (blue circles). An error ratio of 1 means the same number of errors were made on the 3-back task as on the 0-back.

important features were derived from SpO<sub>2</sub>.

### Leave-one-subject-out cross-validation

Binary leave-one-subject-out cross validation with 3 features had varied performance, but performed well across the majority of subjects, with 6 subjects having a mean accuracy greater than 77.9%, and 4 of those subjects having an accuracy greater than 85%. The 6 subjects with highest classification accuracy made an average of 15 times more errors in the 3-back task, reflected in a low 0-back to 3-back error ratio, and the 4 subjects with the lowest classification

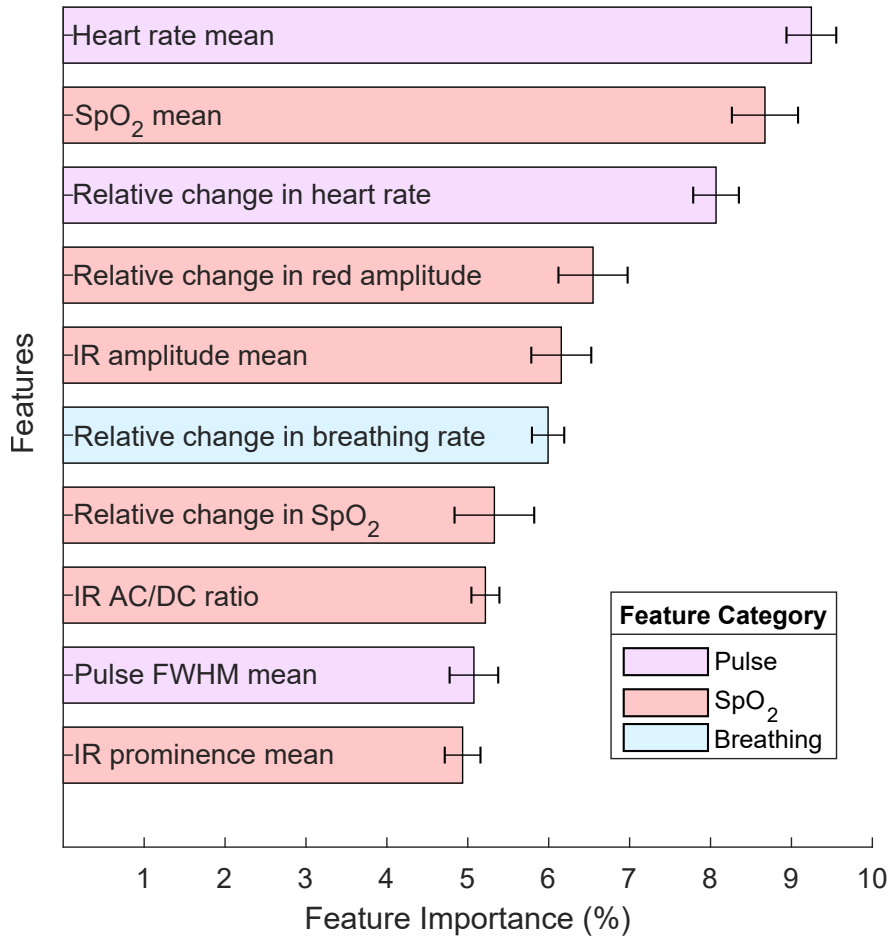


Figure 5.7: Feature importance for the 10 most significant features out of the 21 features considered during 10-fold cross validation. Feature importance was derived from the reduction in tree impurity based on the contribution of each feature to the random forest classifier. The features were split into three categories, depending on the physiological metric from which they were derived. The SpO<sub>2</sub> features are shown in red, pulse features are in purple, and breathing features are in blue.

accuracy made an average of 2 times more errors in the 3-back task, reflected in a high 0-back to 3-back error ratio. The accuracy percentages for testing on each subject, along with the 0-back to 3-back error ratios are shown in Fig. 5.6(b). Notably mean SpO<sub>2</sub> and the relative change in SpO<sub>2</sub> were the most valuable features in terms of reducing tree impurity in binary leave-one-subject-out cross validation.

## 5.4 Discussion

In general, an increase in cognitive workload led to a decrease in the measured in-ear SpO<sub>2</sub> levels. The decrease in the measured in-ear SpO<sub>2</sub> with increased cognitive workload was con-

sistent across all 10 subjects, with the lowest median relative change in SpO<sub>2</sub> or the lowest median absolute SpO<sub>2</sub> occurring in the 3-back or 2-back tasks in all subjects. This demonstrates the robustness of the in-ear SpO<sub>2</sub> response to changes in cognitive workload, compared with the commonly used metric of heart rate, where an increase in heart rate did not necessarily correspond to increased cognitive workload. As expected, errors were highly correlated with increased cognitive workload due to the increased task difficulty. Importantly, the relative change in SpO<sub>2</sub> was more correlated with the level of N-back task than it was with the proportion of errors made, and the opposite was true for the heart rate. A possible reason for this is that some subjects became stressed when making errors, thus triggering an increase in heart rate [21]. Whilst it is important to note that we did not provide live feedback to participants when they made errors, during an N-back memory task it is feasible that subjects were aware of when they have forgotten a number. In this particular experiment, the memory aspect of the task contributed more to errors than the counting of the odd numbers, as evidenced by an increase in the mean error rate from 4% to 29% between the 0-back and 3-back tasks. Critically, this highlights in-ear SpO<sub>2</sub> as a metric which tracks cognitive workload irrelevant of stress. Furthermore, the correlation between the relative change in SpO<sub>2</sub> and an increase in cognitive workload was the strongest and most significant correlation found.

The robustness of the measured SpO<sub>2</sub> changes were further reflected in the high classification accuracy. With ten-fold cross-validation, the 5-second epochs of in-ear data achieved an average classification accuracy of 90.6% across the four N-back task categories, with the two most important features for classification being the mean heart rate and the mean SpO<sub>2</sub>. The performance of leave-one-subject-out cross-validation using the two categories of 0-back and 3-back was less consistent, but was reasonably good in the majority of subjects, with an accuracy of 77.9% and above in 6 subjects, 4 of which achieved a classification accuracy of over 85%. Notably, the leave-one-subject-out evaluation was implemented with just 3 features, the mean SpO<sub>2</sub>, the mean heart rate and the relative change in SpO<sub>2</sub>. In this case, the most important features were the SpO<sub>2</sub>-derived features. Our average classification accuracy across unseen subjects with a 5-second window (70.1%) is comparable to that of gaze and pupil derived features (70.4%) [101] and the reported accuracy achieved by fNIRs when training and testing on the same subject (63.5% and 78% with 15-second to 25-second windows respectively) [102].

In general, cognitive workload tasks induce different levels of cognitive workload in different

people, which is evidenced through the large standard deviation in mistakes, suggesting highly subjective levels of difficulty. The physiological response to increased cognitive workload also varies widely between different people. The ability of leave-one-subject-out training to perform well when testing on a majority of the 10 subjects conclusively demonstrates a robustness in the SpO<sub>2</sub> response to changes in cognitive workload that becomes visible even across a few subjects. Importantly, in the subjects where the classifier performed poorly, there were comparable errors between the 3-back and 0-back tasks and the absolute errors for the 3-back task were low. This indicates that both tasks were found to be comparatively easy, and classification performance was reduced because the experiment failed to induce large changes in cognitive workload. The overlap in task difficulty across subjects was further exaggerated when comparing 0-back to 1-back, or 2-back to 3-back, making full four category leave-one-subject-out classification unfeasible.

In our data, the in-ear SpO<sub>2</sub> decrease in response to cognitive load is visible within the first two 5-second segments of increased cognitive load and this decrease tends to accumulate gradually across the trial. This is comparable to galvanic skin response which has a response time to emotion evoking stimuli in the range of 1 to 5 seconds [103] and a tonic response in the range of 10 to 100 seconds [104]. Notably, in-ear SpO<sub>2</sub> is slower than more instantaneous measures such as EEG which has a response time on the order of hundreds of milliseconds [105] and therefore it is recommended that in-ear SpO<sub>2</sub> be used to measure sustained periods of cognitive load in the period of tens of seconds and longer for maximal effectiveness, rather than to explore the cognitive load induced instantaneously by a single stimulus.

The median relative decrease found in in-ear SpO<sub>2</sub>, between a memory task difficulty of easiest to hardest, was 0.47%. If we consider a example in-ear PPG waveform with a relatively low (but not uncommon) infrared peak to peak amplitude of 150, a red peak to peak amplitude of 50, and a noise floor of 10 arbitrary units, the influence of noise would result in an average fluctuation of 1.5% in SpO<sub>2</sub> as per the calculation in Chapter 3 subsection 3.2.3. For this reason, any results pertaining to relative SpO<sub>2</sub> were always averaged over many peaks and not taken from single peak calculations. This is another reason why it is recommended that in-ear SpO<sub>2</sub> be used to measure sustained periods of cognitive workload.

Whilst it has been demonstrated that in-ear SpO<sub>2</sub> is an effective measure for distinguishing aggregate levels of cognitive load, a limitation when compared with more sophisticated measures

such as fMRI and fNIRs is that in-ear SpO<sub>2</sub> cannot distinguish the type of cognitive load, such as whether cognitive load is induced by increased memory demands (as is the case with an N-back task) or induced by audio/visual feedback or motor control. In uncontrolled environments where external stimuli are a factor, it would therefore be difficult to relate in-ear SpO<sub>2</sub> measured cognitive load changes solely to a single task.

It is also important to note that the external carotid artery supplies the ear canal with oxygen, whereas the internal carotid artery supplies the brain with oxygen. More experimentation is needed to ascertain the extent to which the observed robust decrease in in-ear SpO<sub>2</sub> is caused by the increased oxygen consumption of the brain, as opposed to other physiological factors. The impact of sympathetic tone was investigated but a change in heart rate variability metrics was not found to be predictive of an increasing cognitive load in this study. Further investigation is also needed to determine whether this SpO<sub>2</sub> response is specific to the ear canal.

## 5.5 Chapter conclusions

A proof of concept for cognitive workload estimation using a novel wearable in-ear pulse oximetry sensor has been introduced. Pulse oximetry from the ear canal has been shown to be capable of discriminating between 4 categories of cognitive workload based on an N-back task over 5-second epochs, with a mean accuracy of 90.6%. High cognitive workload in the 2-back and 3-back tasks has led to either the lowest median absolute SpO<sub>2</sub> or largest median decrease in SpO<sub>2</sub> in all of the subjects, therefore demonstrating a robust decrease in measured blood oxygen in response to increased cognitive workload. We conjecture that the decrease in measured SpO<sub>2</sub> with increased cognitive load could be related to the increased oxygen consumption of the brain under increased cognitive demands, and to this end we have examined the predictability of the change in in-ear SpO<sub>2</sub> in response to changes in cognitive workload. The consistency of the SpO<sub>2</sub> response has been further evidenced by an ability to generalise across subjects, even in a relatively small subject pool. In combination with the previously documented rapid reaction speed of in-ear SpO<sub>2</sub> measurements [72], this indicates the promise of in-ear SpO<sub>2</sub> as a tool for close to real-time cognitive workload classification. Overall, this pilot study has established in-ear SpO<sub>2</sub> as an effective tool for classification of cognitive workload, to be used alone or in combination with commonly used workload tracking equipment such as EEG and

ECG, or within the emerging multi-modal Hearables.



# Chapter 6

## Tracking cognitive workload in gaming with in-ear SpO<sub>2</sub>

### Chapter overview

The feasibility of using in-ear SpO<sub>2</sub> to track cognitive workload induced by gaming is investigated. This is achieved by examining temporal variations in cognitive workload through the game Geometry Dash, with 250 trials across 7 subjects. The relationship between performance and cognitive load in Dark Souls III boss fights is also investigated followed by a comparison of the cognitive workload responses across three different genres of game. A robust decrease in in-ear SpO<sub>2</sub> is observed in response to cognitive workload induced by gaming, which is consistent with existing results from memory tasks. The results tentatively suggest that in-ear SpO<sub>2</sub> may be able to distinguish cognitive workload alone, whereas heart rate and breathing rate respond similarly to both cognitive workload and stress. This study demonstrates the feasibility of low cost wearable cognitive workload tracking in gaming with in-ear SpO<sub>2</sub>, with applications to the play testing of games and biofeedback in games of the future.

### 6.1 Introduction

Physiological feedback promises to revolutionise how players interact with games, from more insightful information during play testing, through to adapting game parameters during game

play. It can serve to provide insight into a player's stress level, emotions, arousal level and cognitive workload (also known as mental effort in response to a task). Low cognitive workload indicates either a lack of focus and/or an easy task, while high cognitive workload indicates intense focus and/or a challenging task. Accurate tracking of cognitive workload in gaming would therefore provide valuable insights into both player effort and game difficulty.

The brain is the most metabolically active organ in the body and increases its oxygen consumption with increased cognitive workload. This has been shown through functional near-infrared spectroscopy (fNIRS) and has also been detected reliably through the measurement of minor decreases in in-ear SpO<sub>2</sub> in response to N-back memory tasks [44]. It is therefore our hypothesis that in-ear SpO<sub>2</sub> may have a similar response to cognitively demanding games which require intense audio-visual processing and coordination.

### 6.1.1 Physiological monitoring in gaming

As gaming has continued to advance over the last two decades, so has the desire for integrated biofeedback with the goal of a more rewarding and personalised user experience. Heart rate, derived from either electrocardiography (ECG) or photoplethysmography (PPG), is one of the most commonly monitored physiological metrics. It has been used to modulate game difficulty [106] and can effectively distinguish different stress responses across games (Mortal Kombat and Tetris) through the use of heart rate variability metrics [107]. Similarly, measuring perspiration from the hands through the skin conductance level (SCL) is an effective tool for estimating physiological arousal, and consequently it has been used to adapt parameters of the survival horror game Left 4 Dead [108]. A combination of heart rate, SCL and facial electromyography (EMG) may be able to achieve more detailed measures of arousal and emotion for the modulation of gameplay [109] but this requires extensive training to learn a users individual physiological responses.

Scalp electroencephalography (EEG) has shown increased theta power (indicating both increased mental workload and fatigue) in the sports game Mario Power Tennis [110] and the puzzle game Kirby's Avalanche [111]. Functional near infrared spectroscopy (fNIRs) has shown increased oxygen concentration in frontal lobe brain tissue during both fighting and puzzle genre games [112], indicating that as with N-back tasks [74] brain oxygen consumption also increases

during gaming. Whilst scalp-EEG and fNIRS give robust measures of cognitive workload, they are expensive and, largely, lack portability.

We here build upon our work on low cost wearable in-ear pulse oximetry, which has shown sensitivity in the detection of minor drops in blood oxygen level with increased cognitive workload induced by N-back memory tasks [44], to investigate its utility for tracking cognitive workload in response to gaming.

## 6.2 Methods

### 6.2.1 Hardware

The hardware used in this chapter is detailed in chapter 2, section 2.2.2. The in-ear sensor was connected to the Bioboard, attached to the subjects clothing, which transmitted the data wirelessly via Bluetooth to a laptop. The PPG was sampled at 50Hz.

### 6.2.2 Experimental protocol

The recordings detailed in this chapter were performed under the Imperial College London ethics committee approval JRCO 20IC6414, and all subjects gave full informed consent. The participants in the recordings were 7 healthy subjects (4 male, 3 female) aged 18 - 30 years. All 7 subjects played the game Geometry Dash (RobTop Games) and one subject also played Dark Souls III (FromSoftware Inc) and Five Nights at Freddy's (Scott Cawthon). Geometry Dash (GD) is a music/runner genre game where players click to jump on platforms and avoid objects. Dark Souls III (DS3) is an action role-playing genre game where players have to dodge, block and attack challenging enemies. Five Nights at Freddy's (FNAF) is a survival horror genre game which triggers jump scares when a player loses.

In GD, players repeated a level 5 times, followed by a minimum of 1 minute of rest, and repeated this process at least 5 times for a minimum of 25 trials per person. When a player completed level 1, they moved on to level 2. A total of 250 trials were recorded, 218 of which were on level 1, and 32 on level 2. With DS3 the subject repeated the first area of the game, involving

a mini-boss followed by a main boss with much easier enemies between. When the subject completed the area or died, the game was restarted. The area was repeated 8 times, resulting in 42 minutes of game play. With FNAF the subject played until losing or completing a level, followed by a minimum of 2 minutes rest and this was repeated 5 times. Rest periods were slightly longer in FNAF than in GD given that the level duration was longer and fatigue needed to be mitigated. All rest times were more than sufficient for physiological signals to return to baseline levels.

### 6.2.3 Signal processing

#### $SpO_2$

Blood oxygen percentage ( $SpO_2$ ) was calculated based on the alternating current to direct current ratios of infrared and red light [72], otherwise known as the ratio of ratios described in Chapter 3, subsection 3.2.3. Artefacts were removed by calculating a z-score for the current peak magnitude value, based on the sample mean and standard deviation of the 10 peaks before. A z-score threshold (usually 3) was determined for each individual and if the z-score of the current peak was above this threshold it was removed and linear interpolation was used.

#### Heart rate

The ear-PPG has a characteristic carotid pulse pressure waveform with troughs that are sharper than the peaks and thus peak detection was performed on the troughs of the PPG signal. Peaks were detected using the MATLAB function *findpeaks*, with a minimum peak prominence of between 30 and 100 for the infrared signal and 10 and 50 for the red signal. Sixty was divided by the inter-beat time interval to give a heart rate value in beats per minute.

#### Breathing rate

The three respiratory modes present in PPG, namely the intensity variations, pulse amplitude variations and pulse interval variations [12], were combined in 16 second intervals using noise-assisted multivariate empirical mode decomposition (NA-MEMD) [29]. Intrinsic mode functions

(IMFs) were selected if the peak frequency was between 0.18 and 0.8Hz, and subsequently summed for a clean respiratory waveform for each mode [12]. Peak detection was performed on the amplitude and interval variation modes, and the inter-peak interval was averaged across the modes to give an estimated time interval between breaths. The respiratory rate was then calculated as 60 divided by this interval, giving respiratory rate in breaths per minute.

### 6.2.4 Analysis metrics

Relative SpO<sub>2</sub>, relative heart rate and relative breathing rate were all calculated as the mean during game play minus the mean during the last 30 seconds of the resting interval. The full game play duration was used for both GD and DS3, whereas the last 3 minutes of game play was used for FNAF. This is because FNAF levels are comparatively long and the vast majority of action occurs in the final 3 minutes. Using relative SpO<sub>2</sub> instead of absolute SpO<sub>2</sub> was particularly important given that the in-ear SpO<sub>2</sub> response to cognitive workload is usually less than a percentage point, and the baseline across individuals can vary from 94-100%.

Relative SpO<sub>2</sub> across time was calculated for level 1 of GD and compared with the frequency at which players failed in each 10 second time interval. For DS3, relative SpO<sub>2</sub> was calculated for each boss fight and plotted against performance. Performance was calculated according to

$$Performance = \frac{HitsDealt - 3(HitsTaken)}{Duration} \quad (6.1)$$

where ‘HitsTaken’ is weighted higher than ‘HitsDealt’, given that being hit by the boss deals far more damage and is therefore more costly than the benefit of the player dealing a single hit.

Finally, the metrics of relative SpO<sub>2</sub>, relative heart rate and relative breathing rate were compared across the three different games genres for a single subject.

## 6.3 Results and discussion

On average, SpO<sub>2</sub> dropped relative to baseline in both level 1 and 2 of Geometry Dash as visualised in Fig. 6.1a, indicating that playing GD increased cognitive workload. Across 218

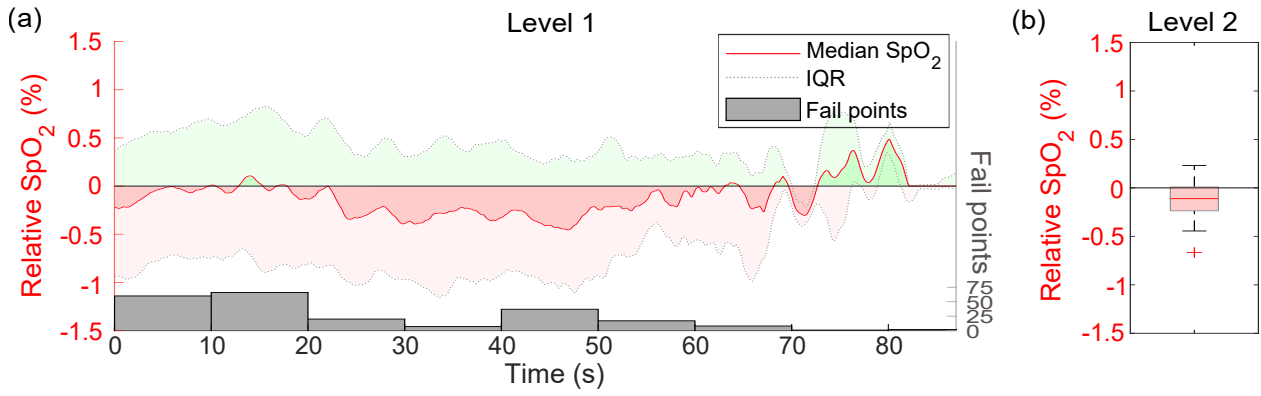


Figure 6.1: Distribution of  $SpO_2$  during Geometry Dash game play relative to the resting interval before, using 250 trials across 7 subjects. (a) Median (red solid line) and interquartile range (shaded area) of relative  $SpO_2$  during level 1 of Geometry Dash, with frequency of fails for each 10 second time window (grey). (b) Box plot of relative  $SpO_2$  during level 2 of Geometry Dash.

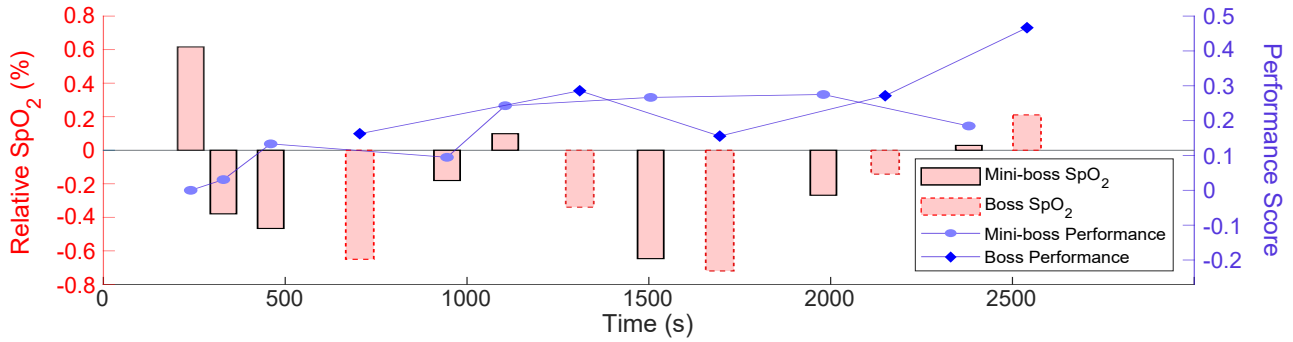


Figure 6.2: Bar graph showing  $SpO_2$  of boss fights in Dark Souls III relative to the easier game play before. Mini-boss fights are shaded pink with a solid black outline. Boss fights are shaded pink with a dotted red outline. Performance is plotted on the same graph in blue according to equation (1). Mini-boss performance is shown with light blue circles, whereas boss performance is shown with blue diamonds.

trials, the median  $SpO_2$  was consistently low between 25 and 55 seconds. The highest frequency at which players failed was between 0 and 20 seconds, which also corresponded to a period where relative  $SpO_2$  was close to zero (no change in cognitive load). A possible explanation is that a lack of focus lead to players failing early in the easier section of the level. The highest median relative  $SpO_2$  occurred at the end of the level between 70 and 90 seconds, but this was calculated using only a couple of trials given that only 2 players reached this point of the level. It is important to note whilst the in-ear response to high cognitive workload is visible within 5 to 10 seconds, it is not instantaneous and the latency varies across individuals. It is therefore best used to indicate broad time periods of increased cognitive load rather than specific time points. Out of 32 trials in level 2, the median relative  $SpO_2$  was -0.11% with an interquartile range of -0.24% to 0.01%, as shown in Fig. 6.1b. This indicates that level 2 of Geometry Dash

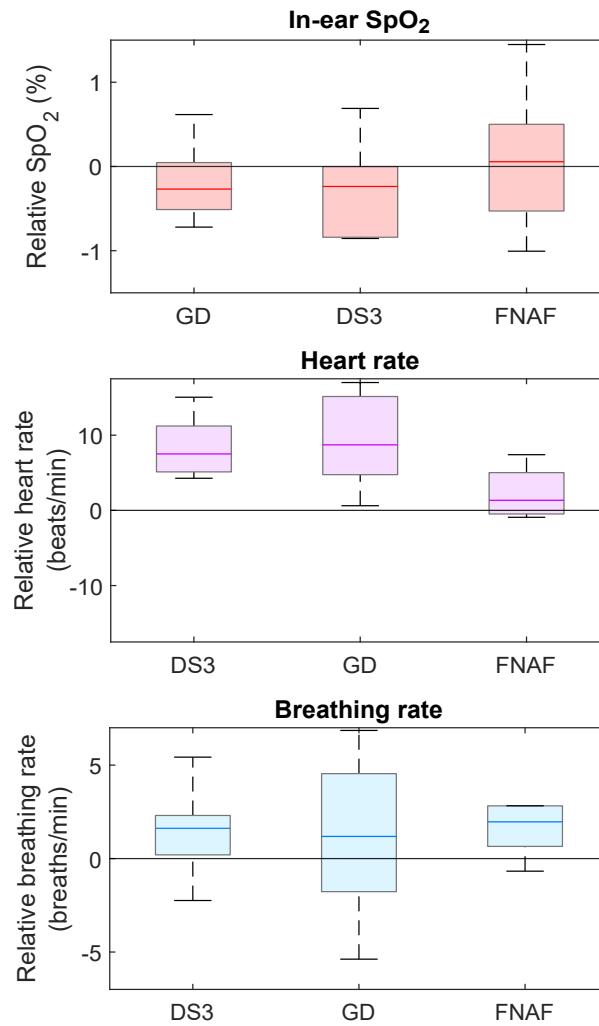


Figure 6.3: Box plots of the relative change in SpO<sub>2</sub> (top, red), heart rate (middle, purple) and breathing rate (bottom, blue) from the in-ear sensor, for the three games Geometry Dash (GD), Dark Souls III (DS3) and Five Night's at Freddy's (FNAF). The top and bottom of each box represent respectively the upper and lower quartiles, the solid center lines of each box designate the median, and the whisker lines extending out of the box represent the range.

consistently increased cognitive workload in participants.

Overall, the boss and mini-boss fights of DS3 induced a median relative SpO<sub>2</sub> of -0.24%, with an interquartile range of -0.84% to 0.00%, indicating consistent increases in cognitive load compared with the easier game play between the boss fights. As expected, performance tended to increase with subsequent attempts. There was no clear relationship between cognitive load and performance, as shown in Fig 6.2. Whilst in theory, higher cognitive load means more focus which would result in better performance, higher focus is also required if a player is less practiced at a game and thus more likely to perform badly. In the final 4 attempts, relative SpO<sub>2</sub> gradually decreases whereas performance remains high, indicating that as the subject

became more practiced at the game, they required less focus to perform well.

All three games induced consistent increases in heart rate and breathing rate, but only GD and DS3 induced a median decrease in in-ear  $\text{SpO}_2$ , as shown in Fig. 6.3. Whilst this is consistent with GD and DS3 being known for requiring high levels of mental effort, it should be noted that the change in  $\text{SpO}_2$  between groups was not statistically significant. Increases in cognitive workload are known to increase heart rate, but stress in the absence of cognitive workload can also lead to increases in heart rate. The horror game FNAF is known for intense scares but can require comparatively less focus than GD or DS3. This is reflected in a consistent increase in heart rate and breathing rate, likely due to stress, but no consistent change in in-ear  $\text{SpO}_2$ , indicating no consistent change in cognitive workload.

## 6.4 Chapter conclusions

In this feasibility study, we have employed a low cost wearable in-ear  $\text{SpO}_2$  sensor for the detection of cognitive workload in gaming. Similar to its performance in N-back memory trials [44], in-ear  $\text{SpO}_2$  has been shown to reliably respond to increased mental effort in cognitively demanding games. This is evidenced by a robust decrease in in-ear  $\text{SpO}_2$  across 250 trials of the popular music/runner genre game Geometry Dash. A similar cognitive workload response is shown in the action role playing game Dark Souls III, where we find that changes in cognitive workload levels do not necessarily correspond to changes in performance. Moreover, early results tentatively indicate that decreases in in-ear  $\text{SpO}_2$  might distinguish cognitive load on its own, whereas breathing rate and heart rate increase in response to both cognitive load and stress when gaming. In conclusion, in-ear  $\text{SpO}_2$  shows promise as a tool for physiological feedback in games of the future.



# Chapter 7

## In-ear PPG for the detection of driver fatigue

### Chapter overview

A system to warn drivers of a fatigue has huge potential for preventing road traffic collisions and in turn saving lives. To this end, we investigate wearable in-ear PPG for physiological signs of driver related fatigue and the ability to predict periods of fatigue through poor driving performance. Our findings across 7 subjects, each engaged in monotonous driving in a simulator for 1 hour, indicated significant increases in both subjective fatigue (Chadler fatigue scale) and performance related fatigue (absolute steering wheel angle). Moreover, we find that several physiological features derived from in-ear PPG change significantly with increased fatigue. The median in-ear SpO<sub>2</sub> decreases by a percentage point across the duration of the trial, possibly suggesting an increased cognitive load required to drive safely when fatigued. Moreover, we see an increase in both low frequency (LF) and high frequency (HF) heart rate variability power, suggesting that there may be an increased frequency of both periods of stress and periods of drowsiness with time spent driving. Absolute steering wheel angle was thresholded so that 30 second epochs above the steering threshold were labelled as fatigued epochs. Probability was extracted from classifiers trained on these epochs and correlated with true absolute steering wheel angle, yielding a Pearson's correlation coefficient of 0.42 for a 30 second window, and 0.81 when smoothed over a 5 minute window. This proof of concept study demonstrates the promise of wearable in-ear PPG for the detection of driver fatigue.

## 7.1 Introduction

Fatigue of drivers is a major contributor to death on the roads. Across all road traffic accidents in the UK in 2017, if the accident was fatal it was twice as likely to be due to fatigue when compared with non fatal accidents [113]. Furthermore, a 1995-2001 study of 15 motorways and A-roads in the UK reports that 17% of road traffic crashes and 24% of fatal road traffic crashes were caused by driver sleepiness [114]. Moreover, a recent meta analysis of studies that used surveys to assess sleepiness at the wheel reports that sleepiness behind the wheel increases the risk of motor vehicle accidents with an odds ratio of 2.51 [115].

### 7.1.1 What is fatigue?

The Oxford dictionary definition of fatigue is “extreme tiredness resulting from mental or physical exertion or illness”. In general fatigue is defined in the following ways, a subjective feeling of sleepiness, a physiological state of weakness in the muscles and central nervous system, a measurable decrease in performance or a combination of all of the above. Moreover, fatigue can be either acute from inability to recover over a short time period or chronic from a cumulative build up of acute fatigue [116]. Fatigue is not to be confused with drowsiness, which can fluctuate rapidly over a period of seconds [117], whereas mental fatigue does not [118]. This is best understood through the common experience of rest and inactivity having the potential to increase drowsiness but decrease fatigue [117]. Driver fatigue can generally be broken down into two categories: sleep/wake related mechanisms and driving factors such as length of time behind the wheel and monotonous conditions [119].

### 7.1.2 How can we assess fatigue?

#### Subjective methods

It is common to use questionnaires to assess sleepiness, such as the Epworth Sleepiness Scale [120]. However, there also exists several questionnaires to assess fatigue [121]. Many of these questionnaires focus on fatigue in general life [122] and fatigue in chronic diseases [123] [124], over long time scales. This is less useful in terms of continuous assessment of fatigue during

an activity such as driving, but the result of a questionnaire such as the Fatigue Severity Scale (FSS) [125] [123] and the Chadler Fatigue Scale (CFS) [122] can indicate the likelihood of driver fatigue and fatigue related driving accidents [126].

### **Performance based methods**

One method to assess fatigue when driving is through reaction time (RT). Ting et al assessed reaction time in a driving simulator by displaying red circular images on screen and requiring the subject to switch of an indicator when the images were presented. It was found that RT increased with driving duration, and that RT could be classified as unsafe after approximately 80 minutes of monotonous highway driving [127]. Another popular method to classify fatigue through performance is through measurements of steering wheel movement. Thiffault et al showed that the frequency of large steering wheel movements (6- 10) increased as a sign of fatigue [128]. This is related to reaction time, as when it takes longer to react to an incoming object or change in road direction, larger more abrupt steering wheel movements must be made to compensate. An issue with using performance to assess driver fatigue is that when performance starts to decline it may already be too late in terms of road traffic accident prevention. A system to indicate fatigue before the subsequent decline in performance would be safer.

### **Physiological methods**

Electroencephalography (EEG) is a reliable method to detect drowsiness, sleep, and fatigue. EEG signatures that may indicate fatigue include those seen in the sleep/wake transition N1 sleep stage, which is classified by increased theta band activity (4-7Hz) and a decrease in alpha band activity (8-13Hz) [65]. However, the presence of alpha merely indicates a relaxed wakefulness and not the ability to react fast to incoming stimuli. In fact, research shows that the frequency of alpha spindles seen in EEG increases with braking reaction time [129], likely caused by increased duration of eye blinks. In general fatigue is most commonly indicated by increases in delta (0.5-4Hz), theta and also alpha band activity [130]. It has also been reported that increases in the alpha power band can occur approximately 20s before a visual task mistake [131], indicating that with EEG it may be possible to predict decreased performance before it occurs.

Electrocardiography (ECG) has been thoroughly investigated as it pertains to driver fatigue. It has been reported that during prolonged night driving heart rate decreases [132]. Furthermore, the investigation of heart rate variability (HRV) whilst driving indicates that low frequency HRV power (0.04-0.15Hz) tends to increase as a driver becomes more fatigued, the high frequency HRV power (0.15-0.4Hz) tends to decrease [133]. These findings show an increase in sympathetic nervous system activity and a decrease in parasympathetic nervous system activity, indicating an increase in stress as the driver becomes more fatigued.

Cameras which examine gaze parameters and facial movements, such as the percentage of eyelid closure over the pupil, over time (PERCLOS) and yawn frequency have been used to indicate driver fatigue with high probability [134] [135]. Many studies which evaluate driver fatigue use PERCLOS as a ground truth for fatigue [136], and camera based commercial driver fatigue warning systems are available. Despite the high accuracy, gaze tracking, whilst useful and accurate, is perhaps unlikely to be accepted wide-scale by consumers given concerns over the camera data privacy [137].

Administering 30% oxygen to drivers has been shown to mitigate the effects of driver fatigue [138]. Moreover, a loose correlation has been shown between  $\text{SpO}_2$  and driver fatigue, but this correlation was not significant and was only discovered through driving at high altitude [139]. That being said, fatigue was only assessed subjectively in this study using the Karolinska Sleepiness Scale [140], and low resolution finger  $\text{SpO}_2$  was monitored which is known to have a large delay compared with core  $\text{SpO}_2$  values [72]. By using a performance measure of driver fatigue, such as steering wheel angle deflections, and in-ear  $\text{SpO}_2$ , we may be able to use  $\text{SpO}_2$  to predict driver fatigue in real time. Furthermore, in-ear pulse oximetry can also measure respiration rate and heart rate, and thus in-ear  $\text{SpO}_2$  can be combined with other metrics such as heart rate variability features to improve the predictive power of the single in-ear sensor.

## 7.2 Methods

### 7.2.1 Experimental protocol

The recordings detailed in this chapter were performed under the Imperial College London ethics committee approval JRCO 20IC6414, and all subjects gave full informed consent. 7

subjects (6 male, 1 female, 23-28 years old) were recruited for the recordings and asked to sign a consent form. Subjects were not screened for prior driving experience, due to the lack of skill required to operate the simulator, and were not screened for sleep disorders. The recordings were carried out at midday for a duration of 1 hour. For the duration of the hour, subjects were told to adhere to a constant speed and to stay in the middle of the lane as well as passing through the blocks which were positioned in the middle of the lane at regular intervals. Subjects were allowed a brief trial run before the experiment began, in which they familiarised themselves with the simulator.

### 7.2.2 Driving simulator



Figure 7.1: Custom built driving simulator, with Vision Racer seat and pedals, a Logitech G27 steering wheel and a FOVE virtual reality headset. Adapted from Vito Amadori et al. [78]

A Vision Racer and Logitech G27 steering wheel assembled for a custom built driving simulator shown in Fig. 7.1. This was equipped with a FOVE virtual reality headset with eye tracking

capabilities. The driving environment was designed in unreal engine and consisted of a large oval track with a series of blocks in the center of the lane that the subject had to drive through. A speedometer was also displayed in the driver's field of view so that the driver could maintain a constant speed. To enhance immersion a sunset time of day was simulated in the virtual environment with a light source consisting of dim orange light that was placed close to the horizon.

### 7.2.3 Data acquisition

#### Psychological data

The Chadler fatigue questionnaire [122] was completed by each participant directly before and after the experiment to help ascertain subjects' perception of their own fatigue before and after the hour in the driving simulator. Each of the 11 questions about fatigue, such as "do you have problems with tiredness?" and "do you have difficulties concentrating?" is answered as either "less than usual", "no more than usual", "more than usual" and "much more than usual", corresponding to scores of 0, 1, 2 and 3 respectively.

#### Performance data

Data on road position was acquired from the driving simulator, along with the angle of the steering wheel. Simulator data was sampled at approximately 60Hz.

#### Physiological data

Wearable in-ear PPG [72] [12] was recorded for the duration of the experiment, providing measures of pulse, respiration and in-ear SpO<sub>2</sub>. These were sampled at 50Hz.

### 7.2.4 Analysis

#### Psychological data

A summed Chadler fatigue scale score was taken for each individual before and after the experiment, with a minimum of zero, indicating no fatigue, and a maximum of 33 indicating extreme fatigue.

#### Performance data

Steering wheel deflections were calculated as the absolute mean corrected steering wheel angle. A small consistent steering wheel angle was required given that the track was circular, and thus this mean angle was removed before the absolute angle was used as a performance indicator. Frequency of large steering wheel angles has been shown to be predictive of driver fatigue [128].

#### Physiological data

In-ear PPG data were divided into 30 second segments and physiologically relevant features were calculated for each segment. Features were divided into 4 groups, in-ear SpO<sub>2</sub> based features, pulse based, breathing based and movement based. The SpO<sub>2</sub> features consisted of the mean SpO<sub>2</sub> and the variance of SpO<sub>2</sub>. The pulse-based features consisted of the mean heart rate, the variance of the heart rate. Four heart rate variability features were also used including the root mean square of the pulse interval signal defined as follows

$$\sqrt{\frac{1}{N} \sum_{i=1}^N (PI_i)^2} \quad (7.1)$$

where  $PI$  corresponds to the pulse interval signal, defined as a signal consisting of time difference between each pulse, sampled at the time point of each pulse and then resampled to a frequency of 4Hz.

Heart rate variability frequency domain features were also used, including low frequency power, relating to sympathetic nervous system activity [141], defined as follows

Table 7.1: Summary of features used for the classification of fatigue induced steering wheel angles when driving.

Category	Features
SpO <sub>2</sub>	SpO <sub>2</sub> mean, SpO <sub>2</sub> variance
Pulse	Heart rate mean, heart rate variance, Root mean square of pulse interval, low frequency (LF) power, high frequency (HF) power, LF/HF power ratio,
Breathing	Breathing rate mean, breathing rate variance,
Movement	Motion artefact prevalence,

$$\int_{f=0.05Hz}^{f=0.15Hz} PSD_{PI}(f) \quad (7.2)$$

and high frequency power, relating to parasympathetic nervous system activity [141], defined as follows

$$\int_{f=0.15Hz}^{f=0.4Hz} PSD_{PI}(f) \quad (7.3)$$

where in both cases  $PSD_{PI}$  corresponds to the power spectral density of the pulse interval signal sampled at the time point of each pulse and then resampled to a frequency of 4Hz.

The final heart rate variability feature implemented was the low frequency to high frequency power ratio (LF/HF ratio), relating to the balance between sympathetic and parasympathetic nervous system activity. The breathing based features consisted of the mean breathing rate and the variance of the breathing rate. The movement based feature implemented was the prevalence of motion artefacts, calculated as peaks that in the AC photoplethysmography signal that had a z-score of 4.5 or higher when compared to the mean and standard deviation of the previous 10 peaks. All of the features implemented are summarised in Table 7.1.

Features were then compared across the beginning of the trial (5-15 minutes), the middle of the trial (30-40 minutes) and the end of the trial (50-60 minutes). The first 5 minutes of data were not used in this analysis given that subjects could have heightened excitement or anxiety due to the novelty of starting the experiment. One-way ANOVA was performed across the three time windows of features across all subjects to determine if the differences in the physiological features at different time points were statistically significant.



### 7.2.5 Prediction

Given that the frequency of large steering wheel angles is predictive of fatigue [128], the assumption was made that if the mean absolute steering wheel angle exceeds a certain level in a given window that this could be labelled as fatigue. Importantly, this was chosen over percentage of eye lid closure (PERCLOS) as a ground truth, because we are distinctly interested in the fatigue associated performance decreases as these are what ultimately lead to crashes. Being fatigued does not necessarily result in an increased steering wheel angle, and an increased steering wheel angle similarly does not necessarily mean that the subject was in a state of fatigue. However, being in a state of fatigue does increase the probability that a subject will have an increased absolute steering wheel angle. To this end, classification was evaluated through the correlation between classifier probability of fatigue, and true absolute steering wheel angle.

An absolute angle threshold of 2.7 degrees was chosen, whereby mean absolute angles above the threshold resulted in the window being labelled as “fatigued” and mean absolute angles below or equal to the threshold resulted in the window being labelled as “not fatigued”. This threshold resulted in approximately 12% of 30 second data segments being labelled as “fatigued” and 88% being labelled as “not fatigued”.

Two classifiers were compared for the prediction of fatigue. The first classifier was a random forest with AdaBoost (RF) was implemented via the scikit-learn Python toolbox [99]. For the random forest base, the number of trees was set to 50, the class weight was set to “balanced subsample” and the maximum number of features was set to 5. For the AdaBoost framework, the random forest was set as the base classifier, the maximum number of estimators was set to 50, the learning rate was set to 1.0 and the real boosting algorithm “SAMME-R” [100] was used. The second was a support vector machine (SVM) implemented with a radial basis function, a kernel coefficient that scales inversely to the number of features multiplied by the variance and a regularisation parameter of 1.

Classifiers were trained and tested with 7-fold shuffled cross validation, and classifier probabilities were extracted for each epoch for comparison with true absolute steering wheel angles. Both the classifier probabilities and absolute steering wheel angles were smoothed with varying window sizes by implementing a moving average filter, from no smoothing to a window size of 0 (corresponding to the minimum epoch size of 30 seconds) to 10 (corresponding to 5 minutes).

The Pearson's correlation coefficient was calculated between classifier probabilities and steering wheel angles for each of these window sizes, to determine how predictive classifier probability was of true steering wheel angle and thus driver fatigue related performance metrics.

### 7.3 Results

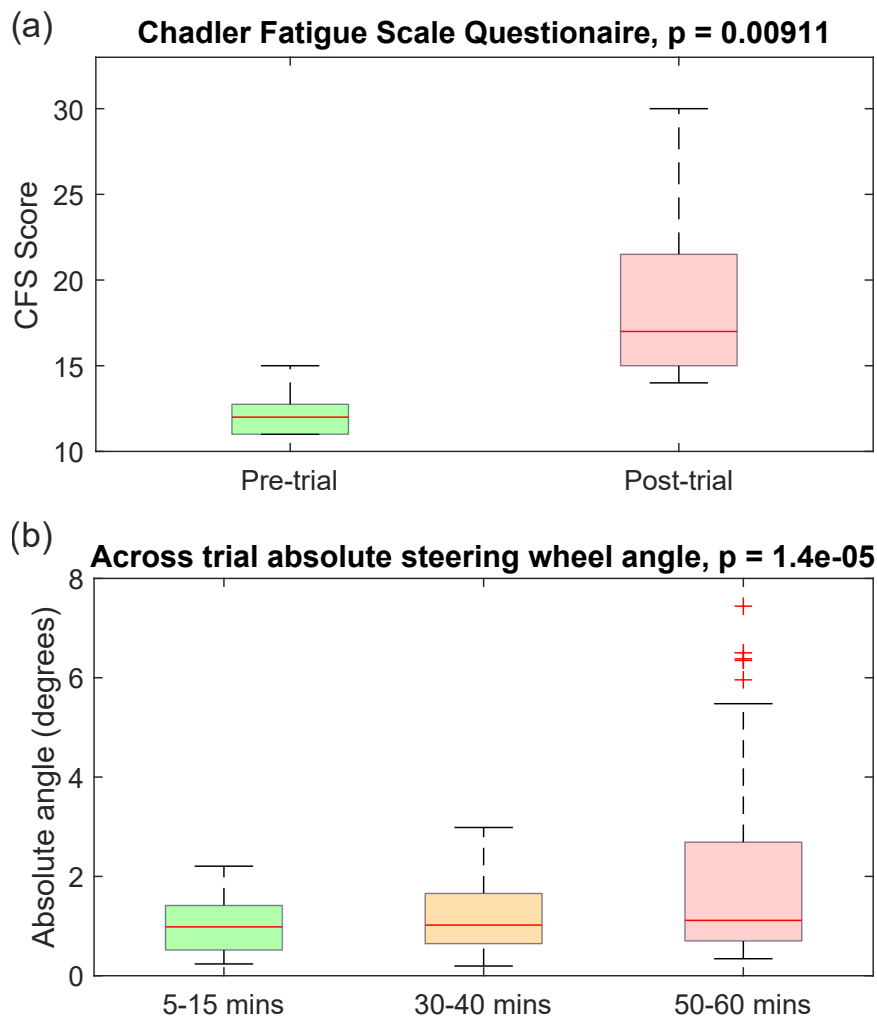


Figure 7.2: Results of subjective and performance related measures of fatigue across 7 subjects, with p-values from one-way ANOVA shown in the titles. (a) Boxplots of the Chadler fatigue scale (CFS) score across each participant before (left, green) and after (right, red) the trial. (b) Boxplots of absolute steering wheel angle across each participant 5-15 minutes (left, green), 30-40 minutes (middle, amber) and 50-60 minutes (right, red) into the trial.

### 7.3.1 Subjective and performance based fatigue

All 7 subjects reported an increase in fatigue as indicated by the Chadler Fatigue scale (CFS) questionnaire. As shown in Fig 7.2(a), the median CFS score was 12 before the trial, with an interquartile range of 11 to 12.75, indicating that the majority of answers to fatigue based questions were “no more than usual”. Importantly, post trial the median CFS score was 17, with an interquartile range of 15 to 21.5, indicating that for some fatigue related questions subjects answered “more than usual” or “much more than usual”. The increase in subjective fatigue was highly significant, with a p value of 0.009.

In addition, the probability of an increased absolute steering wheel angle increased throughout the duration of the trial. The majority of 30 second epochs had no increase in absolute steering wheel angle across the duration of the task, as to continue to navigate the track effectively steering wheel had to be relatively consistent. This is reflected in only a minor increase in the median angle from the start of the trial (5-15 minutes), to the middle of the trial (30-40 minutes) and to the end of the trial (50-60 minutes). However, as the trial went on the likelihood of occasional large deflections in steering wheel angle increased significantly, which is evidenced by the upper quartile of absolute steering wheel angle increasing from 1.4 degrees at the start of the trial, to 1.7 degrees at the middle of the trial and 2.7 degrees at the end of the trial, as shown in Fig 7.2(b). The distribution of absolute steering wheel angles across all subjects in the beginning middle and end of the trials was examined via one-way ANOVA yielding a p-value of  $1.4 \times 10^{-5}$ , showing that the recorded difference in steering wheel angle was statistically significant.

### 7.3.2 Physiological fatigue

In-ear SpO<sub>2</sub> showed a consistent decrease across the duration of the trial, with the median in-ear SpO<sub>2</sub> decreasing from 97.5% (5 - 15 minutes) to 96.4% (50 - 60 minutes). The change in in-ear SpO<sub>2</sub> was statistically significant with a p-value of  $8.3 \times 10^{-5}$ . The median heart rate showed an increase across the trial from 67.4 beats per minute (5-15 minutes) to 71.4 beats per minute (50 - 60 minutes) but this increase was not statistically significant, with a p-value of 0.051. There were minor changes in the median breathing rate, from 0.33 Hz (5-15 minutes), to 0.31Hz (30-40 minutes) to 0.34Hz (50-60 minutes). The differences in breathing

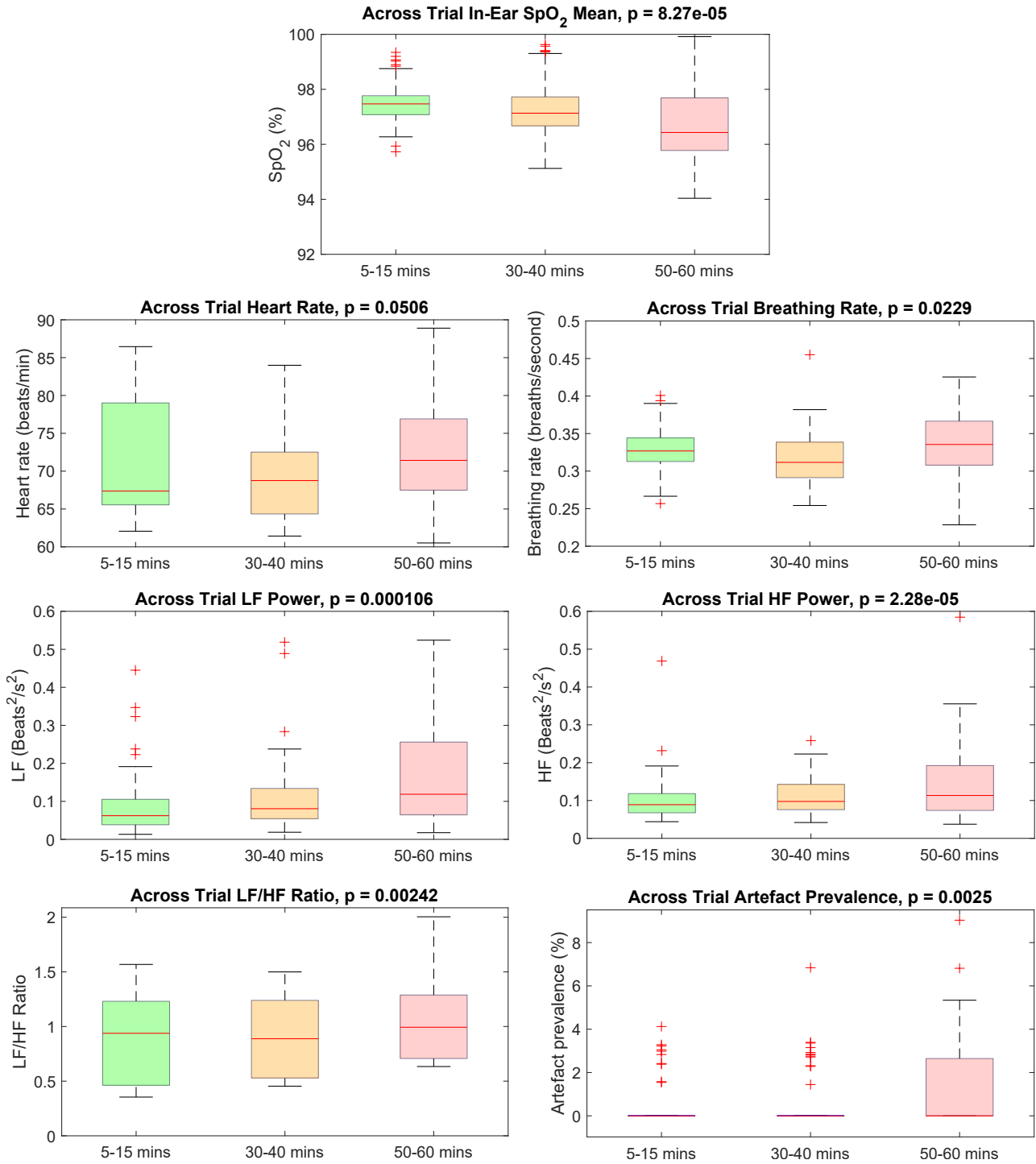


Figure 7.3: Boxplots of physiological features derived from in-ear PPG across 7 subjects undergoing an hour long monotonous drive in a driving simulator, with p-values from one-way ANOVA displayed in the titles of the plots for each feature. For each feature there is a boxplot for 5-15 minutes (left, green), 30-40 minutes (middle, amber) and 50-60 minutes (right, red) into the trial. From top to bottom, left to right, the features displayed are as follows: In-ear SpO<sub>2</sub> mean, heart rate, breathing rate, low frequency heart rate variability power (LF power), high frequency heart rate variability power (HF power), LF/HF power ratio and motion artefact prevalence.

rates across different periods of the trial were statistically significant with a p-value of 0.023 but had no consistent direction. All heart rate variability power spectral density features had statistically significant changes across the trial, with a consistent increase in low frequency (LF) power across the trial ( $p = 1.1 \times 10^{-4}$ ), a consistent increase in high frequency (HF) power ( $p = 2.3 \times 10^{-5}$ ) and a slight increase in the LF/HF ratio ( $p = 0.0024$ ). Finally, there was a statistically significant increase in the prevalence of artefacts in the PPG signal across the trial ( $p = 0.0025$ ), but, given that the vast majority of epochs had no artefacts, the median prevalence remained 0% through the trial. The distribution of physiological features across the trial is summarised in Fig. 7.3.

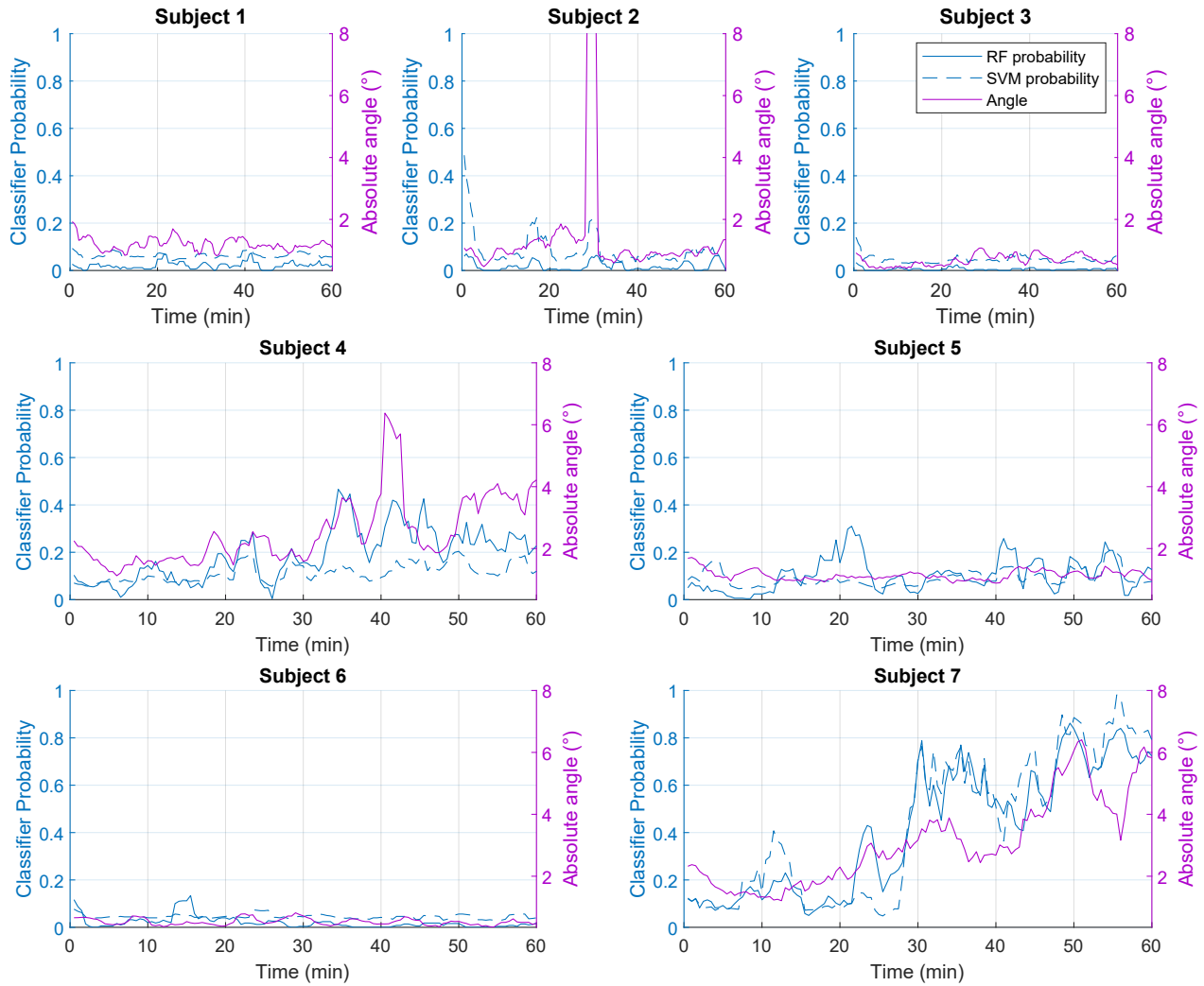


Figure 7.4: Plots of classifier probability across the duration of each trial for both random forest (RF, blue solid line) and support vector machine (SVM, blue dotted line) classifiers, against true absolute steering wheel angle (pink, solid line). The plots are smoothed over a 2.5 minute window and are shown individually for all 7 subjects.

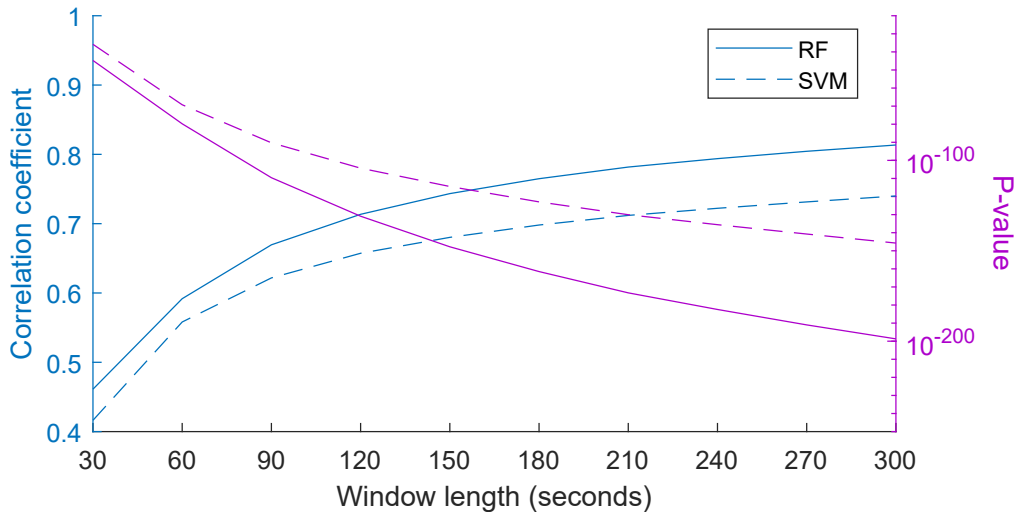


Figure 7.5: Plots of Pearson's correlation coefficient between classifier probability and true absolute steering wheel angle, against window length for both the random forest classifier (RF, blue solid line) and support vector machine classifier (SVM, blue dotted line). P-values for each correlation are also displayed in pink.

### 7.3.3 Classifier prediction

Both the random forest (RF) and support vector machine (SVM) demonstrated an ability to predict absolute steering wheel angle through probability of fatigued steering wheel angle. Probabilities for both classifiers and the true absolute steering wheel angles, smoothed over a 2 and a half minute window, for all 7 subjects, are shown in Fig. 7.4. Visually, it can be seen that when there is a consistent increase in absolute steering wheel angle, such as in subjects 4 and 7, there also an increase in classifier probability. Examination of the Pearson's correlation between predicted probability and true absolute angle yielded moderate correlations from 0.46 and 0.42 for 30 second windows with the RF and SVM classifiers respectively, up to strong correlations of 0.81 and 0.74 for 5 minute windows for the RF and SVM classifiers respectively. The Pearson's correlation coefficient and corresponding p-value for each window size and each classifier is shown in Fig 7.5.

## 7.4 Discussion

All subjects showed an increase in subjective fatigue through the Chadler Fatigue Scale, thus indicating that the experimental protocol did an adequate job of inducing fatigue. Similarly,

increases in steering wheel angle across trial correlate with what is expected of an increase in driver fatigue [128]. It is however important to note that, as illustrated in Fig. 7.4, only 2 of the 7 subjects saw consistently high absolute steering wheel angles towards the end of the trial, associated with extreme fatigue, to the point where they even crashed the vehicle.

In terms of the physiological data recorded with in-ear PPG, there was a significant increase in low frequency heart rate variability power and in the LF/HF ratio, inline with results also seen by Zhao et al [133]. However, we also saw a significant increase in HF power, which was not seen by Zhao et al. In our case this means that whilst the balance of nervous system activity became more sympathetic dominant as the driver became more fatigued, there were periods of both increased sympathetic activity and increased parasympathetic activity. A possible explanation for this is that there were periods where the driver was stressed, for example being frustrated at the task and/or making mistakes (LF power increased), and there were also periods of drowsiness (HF power increased) [141] and both of these periods of stress and drowsiness increased in frequency with time spent in the simulator.

Moreover, we saw a robust and highly significant decrease in in-ear SpO<sub>2</sub> with time spent driving. Notably, the decrease in median in-ear SpO<sub>2</sub> was a single percentage point over the duration of the experiment, and thus with a conventional discrete measure of SpO<sub>2</sub> this drop would likely have gone unnoticed. This echoes previous results with decreases in in-ear SpO<sub>2</sub> with cognitive workload [44], indicating that there could have been an increased cognitive demand associated with adhering to both a constant speed and to the center of the track when fatigued.

Both the in-ear SpO<sub>2</sub> and heart rate variability based features proved useful for the prediction of steering wheel angle. A random forest classifier trained on the physiological features achieved a strong Pearson's correlation coefficient of 0.81 between predicted probability and true absolute steering wheel angle, when smoothed over a 5 minute window. This result is particularly meaningful when considering that labelling physiological data as fatigued based on an increase in steering wheel is an assumption which doesn't always hold true. Whilst the probability of large steering wheel angle deflections is increased with fatigue, fatigue doesn't guarantee a large steering wheel angle, with the majority of angles being normal even when fatigued. Similarly, a large steering wheel angle could occur when not fatigued. Despite this, classifier probability is strongly predictive of true absolute steering wheel angle, indicating that a driver fatigue warning system based on wearable in-ear PPG is possible. It is important to note that in our

data, due to only 2 subjects showing extreme fatigue and thus consistently high steering wheel, leave one subject out training was unworkable. This model would need to be able to generalise to unseen subjects for maximal applicability in real-world driving scenarios.

## 7.5 Chapter conclusions

Features derived from in-ear PPG, including in-ear  $\text{SpO}_2$  and heart rate variability metrics, change significantly with fatigue induced by a driving simulator. Moreover, classifiers trained on these features are strongly predictive of absolute steering wheel angle, of which increases are associated with fatigue. In the real world, alerting a driver to an increased probability of fatigue related performance decreases would help to prevent road traffic collisions and ultimately save lives. These early findings indicate the promise of a wearable and discrete in-ear PPG sensor for the detection and warning of driver fatigue.



## Part III

### Detection of obstructive lung disease

# Chapter 8

## An apparatus for the artificial simulation of breathing disorders

### Chapter overview

The rapidly increasing prevalence of debilitating breathing disorders, such as chronic obstructive pulmonary disease (COPD), calls for a meaningful integration of artificial intelligence (AI) into healthcare. While this promises improved detection and monitoring of breathing disorders, AI techniques are almost invariably “data hungry” which highlights the importance of generating physically meaningful surrogate data. Indeed, domain aware surrogates would enable both an improved understanding of respiratory waveform changes with different breathing disorders, and enhance the training of machine learning algorithms. To this end, we introduce an apparatus comprising of PVC tubes and 3D printed parts as a simple yet effective method of simulating both obstructive and restrictive respiratory waveforms in healthy subjects. Independent control over both inspiratory and expiratory resistances allows for the simulation of obstructive breathing disorders through the whole spectrum of  $FEV_1/FVC$  spirometry ratios (used to classify COPD), ranging from healthy values to values seen in severe chronic obstructive pulmonary disease. Moreover, waveform characteristics of breathing disorders, such as a change in inspiratory duty cycle or peak flow are also observed in the waveforms resulting from use of the artificial breathing disorder simulation apparatus. Overall, the proposed apparatus provides us with a simple, effective and physically meaningful way to generate faithful surrogate breathing disorder waveforms, a prerequisite for the use of artificial intelligence in respiratory

health.

## 8.1 Introduction

The prevalence of obstructive breathing disorders, such as chronic obstructive pulmonary disease (COPD) and asthma, is increasing rapidly [142], whilst other breathing disorders such as the restrictive pulmonary fibrosis (PF) continue to suffer from poor clinical outcomes and a lack of treatment options [143]. Therefore, the understanding of breathing mechanics and resulting respiratory waveforms for different breathing disorders is paramount for the classification of breathing disorders, both in terms of screening and identifying their severity. To this end, we propose an apparatus for the artificial generation of obstructive breathing disorder waveforms through healthy subjects and mechanisms for reliably generating the whole spectrum of disease severities.

### 8.1.1 Changes to breathing with obstruction and restriction

Chronic obstructive pulmonary disease (COPD) is caused by an increased inflammatory response in the lungs which leads to obstructed airflow [144]. Chronic obstructive pulmonary disease encompasses both emphysema, defined by a breakdown in the elastic structure of the alveolar walls [145] and bronchitis, defined by increased mucus secretion in the lungs [146]. When we expire, the airways narrow due to reduced pressure, and thus if airway obstruction exists it is exaggerated during expiration. This explains why patients with COPD generally take longer to breath out than breathe in, and can generate higher inspiratory peak flows than expiratory peak flows. The COPD can be diagnosed with a spirometry test, which measures the ratio of volume during forced expiration in one second ( $FEV_1$ ), against forced vital capacity (FVC). Practically those with COPD usually exhibit  $FEV_1$  to FVC ratios of less than 0.7 [147], but COPD is more specifically defined by different severities. According to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) COPD severity should no longer be labelled based on spirometry alone, but spirometry can be split into 4 major categories based on a percentage of expected  $FEV_1$  for patients age, gender and height. A mild  $FEV_1$  is defined as  $\geq 80\%$  of the expected value, moderate is defined as between 50 and 79%, severe is 30%-49%

and finally very severe is defined as less than 30% [148]. The increased effects of obstruction during expiration also lead to a decreased inspiration time ( $T_I$ ) in comparison with the overall breathing time ( $T_{TOT}$ ) as it takes longer to breathe out. The ratio  $T_I/T_{TOT}$ , known as the inspiratory duty cycle, is lower in patients with COPD [149] [150].

This is in contrast to restrictive lung disease, an example of which is pulmonary fibrosis (scarring of the lungs). In this case, there is no obstruction of airways, but a restriction that applies equally to both inspiration and expiration. Whilst diagnosis of pulmonary fibrosis requires a multidisciplinary approach, such as the use of CT scans [143], spirometry tests will generally show healthy  $FEV_1/FVC$  ratios, but with a lower peak flow for both inspiration and expiration as well as a greatly reduced vital lung capacity.

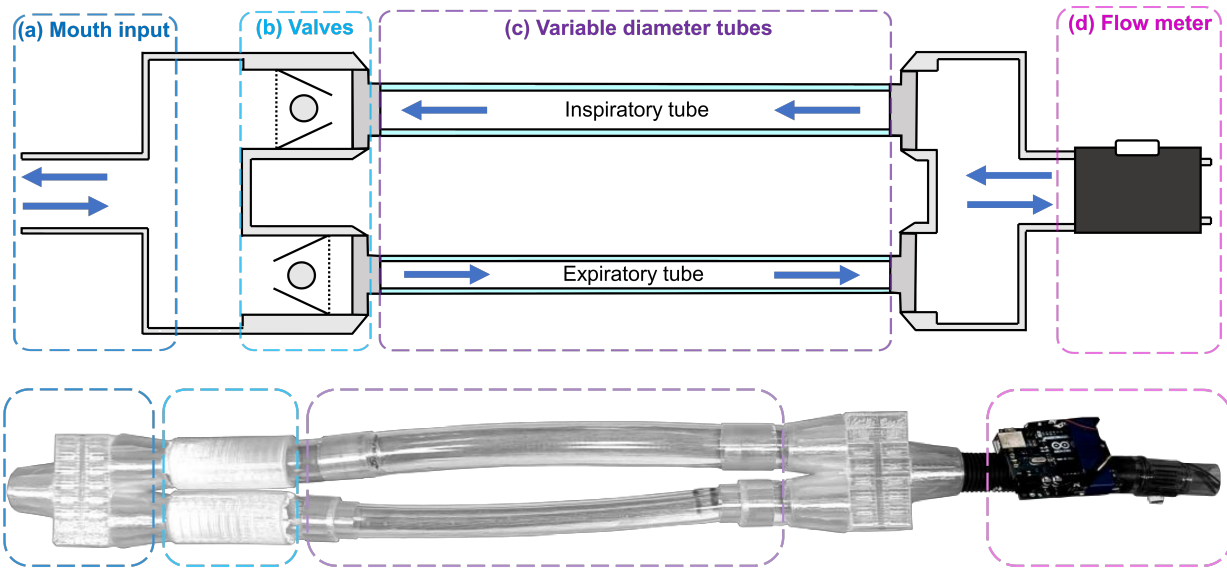


Figure 8.1: Block diagram (top) and physical realisation (bottom) of the proposed breathing disorder simulation apparatus. (a) The mouth input. (b) One-way valves in different directions for inspiration and expiration, comprised of a low density foam plug, a cone shaped funnel with a hole that is slightly smaller in diameter than the plug, and a fine mesh with allows air through but not the plug. (c) Tubes for both inspiration and expiration which can be easily swapped out for tubes of different diameter, allowing for independent control of resistances to inspiration and expiration. (d) A digital flow meter to record spirometry waveforms.

### 8.1.2 Artificial changes to breathing resistance

Resistance to breathing has been considered both to measure the strength and endurance of lungs in subjects, and also as a potential avenue to train lungs for increases in strength and endurance. A portable apparatus for collecting respiratory gas was designed in the early

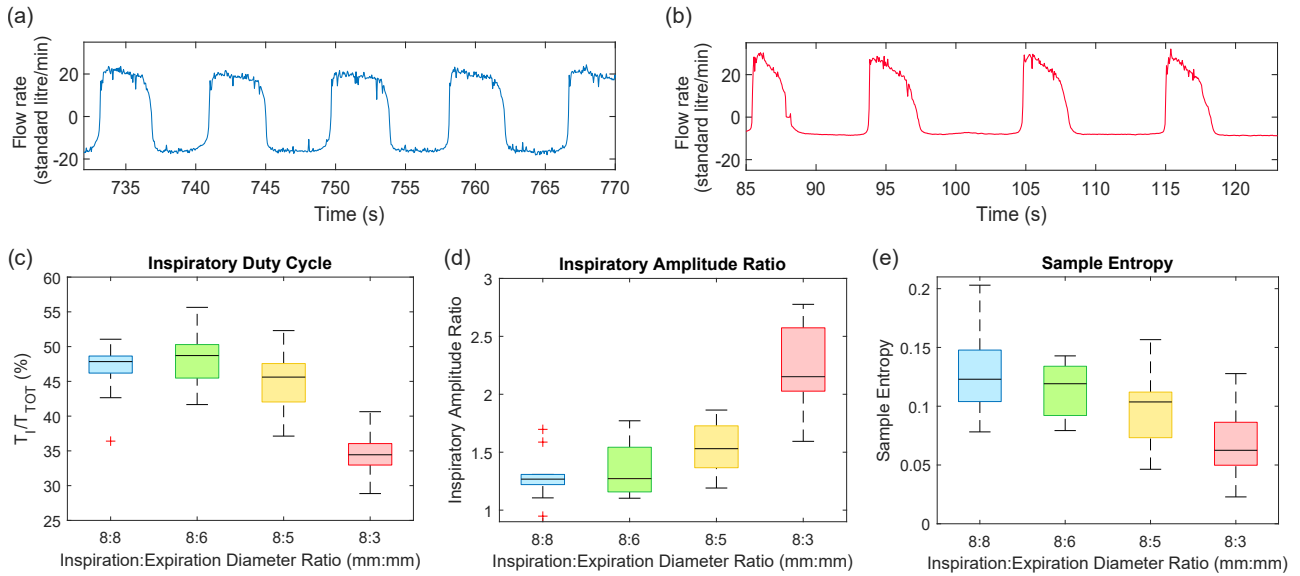


Figure 8.2: Tidal breathing results from using the apparatus with 4 different inspiration to expiration obstruction ratios across 10 subjects. (a) Exemplar spirometry waveform with an 8mm inspiratory tube and 8mm diameter expiratory tube giving a balanced obstruction ratio. Positive flow corresponds to inspiration and negative flow corresponds to expiration. (b) Exemplar spirometry waveform with an 8mm inspiratory tube and 3mm diameter expiratory tube giving an unbalanced obstruction ratio. (c) Boxplots of inspiratory duty cycle (%), referring to the proportion of overall breathing duration spent in inspiration, across 10 subjects and 4 different inspiration:expiration tube diameter ratios. (d) Boxplots of inspiratory amplitude ratio, referring to peak inspiratory flow divided by peak expiratory flow, across 10 subjects and 4 different inspiration:expiration tube diameter ratios. (e) Boxplots of sample entropy (scale 1, tolerance = 0.2) across 10 subjects and 4 different inspiration:expiration tube diameter ratios.

1970s, comprising of tubes with 32mm diameter (giving negligible resistance to breathing) and a one-way valve so that gas could be stored when breathing out, but new air would be breathed in [151]. This apparatus was adapted in the mid to late 1970s by replacing the 32mm inspiratory tube with different smaller tube diameters (14mm, 11mm or 8mm), and breathing under different inspiratory resistances was examined in endurance athletes [152]. A similar apparatus with four different inspiratory tube sizes was used to investigate the lung strength of a group of British coal miners over the age of 45 [153]. More recently, resistance has been applied to both inspiration and expiration through masks that have multiple inspiratory and expiratory valves, with the desire to train lungs for increased strength and endurance [154].

Different from the existing set-ups, the apparatus presented in this chapter is capable of providing different resistances to both inspiration and expiration independently, with the aim of simulating the respiratory waveforms of different breathing disorders.

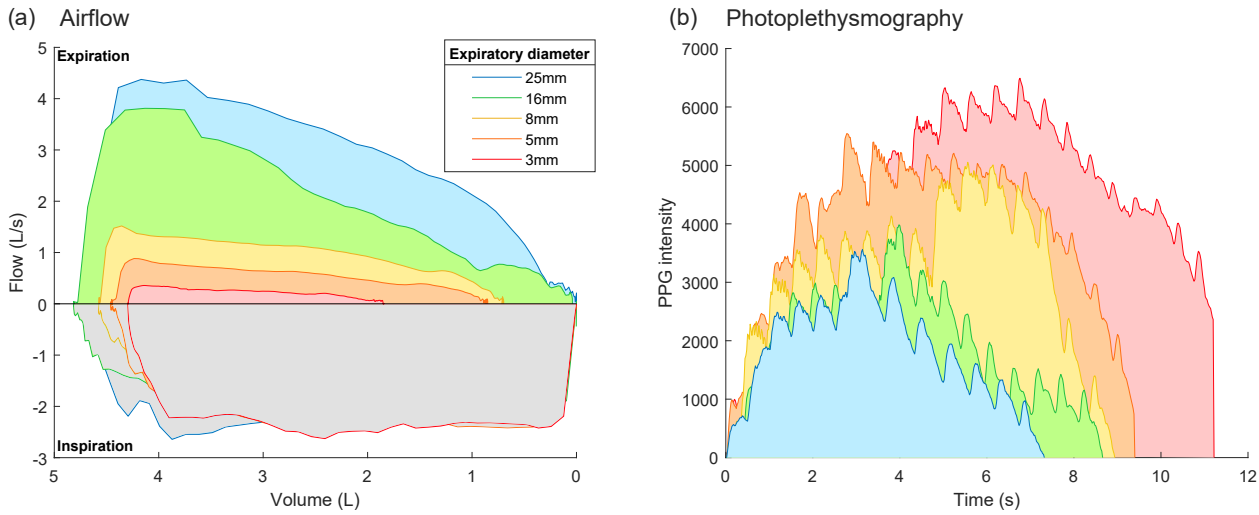


Figure 8.3: .

Example plots of maximally forced breathing for a single subject across 5 different expiratory tube diameters from 25mm to 3mm, with a fixed inspiratory tube diameter of 25mm providing low obstruction to inspiration. (a) Flow-volume loops for different expiratory tube diameters, showing a decreased flow for a given volume with a decrease in expiratory tube diameter. (b) Simultaneously recorded ear-photoplethysmography waveforms during maximally forced breathing with each tube diameter, showing an increase in both PPG intensity and duration with a decrease in tube diameter.

The so enabled simulation of breathing disorders through healthy subjects has the following benefits:

- Ability to collect vast amounts data by expanding the subject pool to include healthy individuals;
- Full control over breathing resistances for both inspiration and expiration;
- Multiple obstructive breathing disorders of different severities can be investigated on the same healthy individual, thus keeping individual physiological differences constant;
- A controlled environment makes it easier to investigate how other physiological measures vary with resistance to breathing;
- A physically meaningful way to generate surrogate breathing disorder waveform data for both training and testing machine learning models.

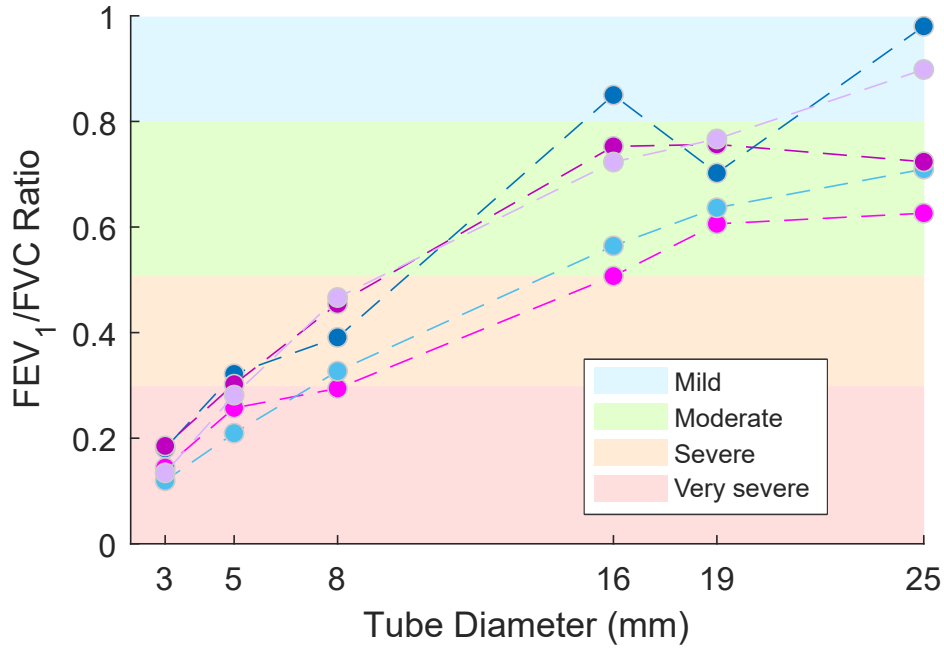


Figure 8.4: Calculated  $FEV_1/FVC$  ratios across 6 expiratory tube diameters from 25mm to 3mm with an inspiratory tube diameter fixed at 25mm, plotted for 5 different subjects. Highlighted as shaded colours are the 4 different obstruction severities at the corresponding  $FEV_1/FVC$  ratio, with blue indicating mild obstruction, green indicating moderate obstruction, orange indicating severe obstruction and red indicating very severe obstruction.

## 8.2 Apparatus design

The apparatus consists of 3D printed parts and PVC tubes. It has a single input tube which a subject breathes in and out of. This is connected to two one-way valves facing in opposite directions to switch the airflow path depending on inspiration and expiration. The valves consist of low density foam plug in a 3D printed cone shaped funnel with a hole slightly smaller than the diameter of the plug. Securing the ball in the funnel is a fine mesh in which air can pass through but the ball cannot. Depending on the orientation of the valve, either positive or negative airflow will seal the hole with the ball, thus preventing air from passing through. It is important that the plug is light so that it will move easily to the hole under low pressures.

Connected to the inspiratory valve is an inspiratory tube which can be varied in diameter, as is the case with the expiratory valve and expiratory tube. The largest tube diameter is 25mm, which is considered as very low resistance to breathing. The smallest tube diameter used is 3mm, which provides very challenging resistance to breathing. To minimise the resistance of the whole apparatus, 3D printed parts also have an internal diameter of 25mm. Both the

inspiratory and expiratory tubes are then connected to an output tube which leads into a SFM3200 digital flow meter by Sensiron (Stäfa, Switzerland) to record the breathing flow. The entire apparatus is shown in Fig. 8.1. The digital flow meter was connected to an Arduino Uno by Arduino (Somerville, MA, USA), which sampled flow values at a sampling frequency of 20Hz and displayed them on a computer monitor.

The apparatus was evaluated with tidal breathing in 10 subjects (5 male, 5 female) aged 18-30 years, across 4 different inspiration to expiration tube diameter ratios, and further evaluated with maximal forced breathing in 5 subjects (3 male, 2 female) across 6 different tube diameters for measurements of FEV<sub>1</sub>/FVC ratios.

The recordings detailed in this chapter were performed under the Imperial College London ethics committee approval JRCO 20IC6414, and all subjects gave full informed consent. Trial recordings were performed on 8 subjects (4 male, 4 female) aged 18-25 years, and included normal breathing under different resistances, as well as breathing in and out as hard as possible for both FEV<sub>1</sub>/FVC measurements and peak expiratory and inspiratory flow measurements. Photoplethysmography was recorded from the ear simultaneously [72] [12] during all recordings to gain insight into the effects of varying obstruction on thoracic pressure waveforms.

### 8.3 Results and discussion

Small decreases in the expiratory tube diameter in relation to the inspiratory tube diameter resulted in changes to tidal breathing waveforms that are typical of patients with obstructive breathing disorders such as chronic obstructive pulmonary disease (COPD). Example spirometry waveforms in Fig. 8.2(a,b) show the roughly symmetric breathing patterns when obstruction to inspiration and expiration is balanced (a) and the characteristic longer expiration time and reduced expiratory flow when obstruction to expiration is exaggerated with a tube diameter of only 3mm (b). Furthermore, these results are consistent across all 10 subjects, with Fig. 8.2(c) showing a decrease in inspiratory duty cycle (percentage of overall breathing time spent inspiring) as obstruction to expiration is increased and Fig. 8.2(d) showing an increase in the inspiratory amplitude compared with expiratory amplitude with increased obstruction. The median duty cycle of 34.4% for the 8mm:3mm inspiration to expiration diameter ratio echoes the duty cycle of COPD patients, which were found to be around 35% at rest [149]. Results also



display a gradual decrease in sample entropy with increased obstruction to expiration, shown in Fig. 8.2(e). Sample entropy is a measure of the complexity of a signal, and thus it is natural that sample entropy would decrease with increased obstruction, as obstruction decreases the degrees of freedom for breathing and in turn makes breathing patterns more predictable. Similar reductions in sample entropy have been shown in the breathing patterns of patients with COPD, with sample entropy decreasing as COPD severity increases [155].

The broad range of obstruction achievable by the apparatus is exemplified by the volume flow loops in Fig. 8.3(a), which show decreased flow for a given volume with decreased tube diameter. This is specific to expiration due to the inspiratory tube diameter being kept constant whilst the expiratory tube diameter was varied, resulting in substantial changes to the expiration side of the volume flow loop with minimal changes to the inspiration side of the volume flow loop. It should be noted that the volume flow loops in Fig. 8.3(a) also illuminate two important limitations of the apparatus. Firstly, whilst flow values are expected to decrease, overall recorded expiratory volumes should not decrease with tube diameter, given that this would not effect lung volume. The apparent reduction in recorded volume shown in the volume flow loops is due to leakage of the system at higher pressures, and could be rectified straight-forwardly with more robust materials and valves and joiners printed to more precise specifications. The second limitation is that the tube apparatus impedes breathing with a constant level of obstruction for a given tube diameter, whereas in reality as we expire the airways continue to narrow in proportion to lung volume decreasing. Obstruction in patients with COPD therefore increases further with continued expiration, resulting in a concave inflection in real-world volume flow loops that isn't captured by this apparatus.

Whilst tube diameter decreases, larger photoplethysmography (PPG) intensities are generated over a longer period of time, as shown in Fig. 8.3(b). Thoracic pressure increases as we expire to push air out of the lungs and this in turn increases venous return and fills peripheral venous beds at the site of the PPG probe. Increased PPG intensity thus reflects increased thoracic pressure and, as expected, thoracic pressure over time increases in proportion to increased obstruction simulated by smaller tube diameters, as increased pressure is required to force air through a smaller tube. Through measuring PPG, it is clear that the apparatus can simulate different thoracic pressure profiles internally based on changes in external obstruction.

The apparatus was able to achieve a wide range of  $FEV_1/FVC$  ratios across all subjects, with

an example of the varied ratios in 5 subjects across 6 different expiratory tube diameters shown in Fig. 8.4. The maximum  $FEV_1/FVC$  achieved was 0.98 with the 25mm diameter tube, and the minimum achieved was 0.12 with the 3mm diameter tube. Importantly, artificially induced  $FEV_1/FVC$  ratios were able to cover the full range of ratios across mild, moderate, severe and very severe. This indicates the promise of the tube based apparatus for simulating a full range of disease severities in each individual, and in turn vastly expanding the quantity of obstructive breathing disorder data available.

## 8.4 Chapter conclusions

We have demonstrated a simple yet effective method of simulating both obstructive and restrictive respiratory waveforms in healthy subjects with the use of a tube-based apparatus. Independent control over both in inspiratory and expiratory resistances allows for the simulation of respiratory waveforms corresponding to obstructive breathing disorders with a wide range of  $FEV_1/FVC$  ratios, from healthy values through to values seen in very severe chronic obstructive pulmonary disease. Notably, this makes it possible for obstructive breathing disorders at a range of severities to be investigated in the same individual, allowing the waveform differences due to different tube resistances to be isolated whilst individual physiological differences are kept constant. Importantly, this apparatus provides us with a physically meaningful way to generate surrogate breathing disorder waveforms for testing and training machine learning models for classification of breathing disorders. Finally, this apparatus could serve the educational purpose of illuminating the difficulties that patients with breathing disorders face, both for public health awareness and as a persuasive argument against behaviours which increase the risk of breathing disorders such as smoking.

# Chapter 9

## In-ear PPG for the classification of COPD

### Chapter overview

An ability to extract detailed spirometry-like breathing waveforms from wearable sensors promises to greatly improve respiratory health monitoring. Photoplethysmography (PPG) has been researched in depth for estimation of respiration rate, given that it varies with respiration through overall intensity, pulse amplitude and pulse interval. We compare and contrast the extraction of these three respiratory modes from both the ear canal and finger and show a marked improvement in the respiratory power for respiration induced intensity variations and pulse amplitude variations when recording from the ear canal. We next employ a data driven multi-scale method, noise assisted multivariate empirical mode decomposition (NA-MEMD), which allows for simultaneous analysis of all three respiratory modes to extract detailed respiratory waveforms from in-ear PPG. For rigour, we considered in-ear PPG recordings from healthy subjects, both older and young, patients with chronic obstructive pulmonary disorder (COPD) and idiopathic pulmonary fibrosis (IPF) and healthy subjects with artificially obstructed breathing. Specific in-ear PPG waveform changes are observed for COPD, such as a decreased inspiratory duty cycle and an increased inspiratory magnitude, when compared with expiratory magnitude. These differences are used to classify COPD from healthy and IPF waveforms with a sensitivity of 87% and an overall accuracy of 92%. Our findings indicate the promise of in-ear PPG for COPD screening in the consumer wearables market, through the long term analysis

of the balance between inspiration and expiration timing. Moreover, this could be used for unobtrusive respiratory monitoring in ambulatory scenarios, where patients respiratory patterns are assessed over the course of days and weeks.

## 9.1 Introduction

Monitoring of respiration is being integrated rapidly into consumer wearables, with respiration rate being a standard feature in many smart watches. Whilst continuous unobtrusive monitoring of respiration rate is a valuable tool in both the consumer and patient health tracking domains, much of the respiratory information that can be derived from breathing waveforms remains untapped. To this end, we explore a relatively new tool for non-invasive respiratory monitoring in the form of in-ear photoplethysmography (PPG). For rigour, we use reference spirometry as a gold standard to test the extraction accuracy of the three major respiratory modes at three different frequencies from in-ear PPG: i) intensity variations, ii) pulse amplitude variations, and iii) pulse interval variations. Furthermore, we compare the spectral respiratory power from simultaneous ear canal PPG and finger PPG recordings across the three major respiratory modes. A novel method for extracting respiratory waveforms from PPG is next presented, which involves a data driven multi-scale algorithm, empirical mode decomposition (EMD) [25]. This waveform extraction technique allows for a simultaneous intrinsic scale-wise analysis of multi-channel data, and is applied to extract resting respiratory waveforms from in-ear PPG in both healthy subjects and patients with breathing disorders, such as chronic obstructive pulmonary disease (COPD) and idiopathic pulmonary fibrosis (IPF). This analysis is shown to allow for the detection of chronic obstructive pulmonary disease from in-ear PPG.

### 9.1.1 Respiration and photoplethysmography

Photoplethysmography (PPG) refers to the non-invasive use of light to detect changes in blood volume by transmitting light through tissue and measuring the amount of light that is absorbed. When more blood is present, more light is absorbed and thus less light is reflected back to the sensor.

When we inspire, a decrease in intrathoracic pressure is created to pull air into the lungs. This

decrease in pressure is passed to the central veins and therefore central venous pressure also decreases. This, in turn, increases venous return and drains venous beds at the site of the PPG probe, thus modulating the DC component of the PPG signal [156]. Right ventricular stroke volume also increases with the increased venous flow to the heart, sending more blood to the lungs for the uptake of oxygen. In turn, left ventricular stroke volume is decreased, leading to a decreased pulse amplitude observed through the AC component of PPG [157] [158]. This is accompanied by an increase in heart rate, otherwise known as respiratory sinus arrhythmia (RSA) [159], which causes a decrease in the interval between pulses in the PPG signal. The opposite of these effects can be observed during expiration. Three major respiratory modes are therefore present in PPG during respiration [160]:

1. Respiratory induced intensity variations (RIIVs) which are generated by changes in venous pressure which modulate the DC component of photoplethysmography, and are therefore accessible directly from the raw-PPG signal.
2. Pulse amplitude variations due to changes in left ventricular stroke volume, which can be obtained from the envelope of the AC filtered PPG signal.
3. Pulse interval variations generated through respiratory sinus arrhythmia. These can be obtained by measuring the interval between consecutive pulses.

It should also be noted that there are prominent low frequency variations which also occur in the PPG signal, related to sympathetic tone. These low frequency variations usually peak at 0.1Hz [161] [157], with harmonics at 0.2Hz which can negatively impact the extraction of respiratory signals.

An extensive literature exists on estimating respiration rate from PPG, usually from the finger but also from the forearm [156], wrist [162] and multiple other body positions such as the earlobe, forehead, neck [163] and chest [164]. The PPG yields high respiration rate accuracy in all three PPG respiratory modes [165], in both healthy subjects and subjects with breathing disorders, such as chronic obstructive pulmonary disease (COPD) and asthma [166]. It should be noted that accuracy decreases at high breathing frequencies, above 0.3Hz [167], which is likely due to a low pass filter effect [168] of the transfer function from thoracic pressure to venous return, and the reality of only having a few pulses per breath to sample from for pulse interval and pulse amplitude variations at higher frequencies.

Moreover, it has been shown that errors across the three respiratory modes are comparable when it comes to calculation of respiration rate [169]. This gives a strong argument for the utility of all three modes given the presence of different artefacts, the variability across subjects and the variability across different respiration frequencies and body positions [170]. Research suggests that the spectral power of respiration in PPG is far larger from regions of the head [171], ear [8] and shoulders [11] than the finger. Furthermore, an analysis on spectral power of the PPG at different body positions, namely the forearm, wrist, finger, forehead and shoulder, indicates that the forearm had the highest respiration power but the lowest pulse power, likely due to its close proximity to large veins which may improve the power of RIIVs induced by venous pressure changes; the shoulder and forehead maintain both high respiratory power and pulse power, while the finger had the lowest respiratory power [11].

With photoplethysmography at the forearm, it has been documented that the respiration induced intensity variations (RIIVs) are effective enough to detect amplitude changes in breathing and also simulated apnea with temporary breath holds [156]. However, changes in duty cycle or higher order statistics such as skewness have not been studied in RIIVs. Moreover, despite errors in respiration rate estimation being similar across the three major PPG respiratory modes, when it comes to the magnitude of respiratory variations, they are far more pronounced in venous return than in changes in stroke volume, usually by an order of magnitude [172]. This, in addition to the fact that the sample frequency of pulse amplitude and frequency variations are limited to the pulse rate, suggests that RIIVs may be superior for detecting detailed waveform changes in respiration.

In terms of sex and age differences in respiratory signals from PPG, literature suggests that the pulse interval variations may be more pronounced in females than in males [170], and more pronounced in the young than in the elderly [173], but that there is no significant difference in the RIIVs with age or sex [174].

### 9.1.2 In-ear photoplethysmography

As previously discussed, current research suggests that PPG from the ear canal is far more sensitive to intensity variations that arise from respiration [8] than the finger PPG, which is inline with similar results for shoulder and forehead PPG [11]. However, this increase in

respiratory power has only been documented through respiratory induced intensity variations in the raw PPG signal. To address this issue, we examine the spectral power across all three PPG respiratory modes in both in-ear PPG and finger PPG. Furthermore, in-ear PPG has only been used to estimate respiratory rate [6], but has not been used for insight into the respiratory waveform itself. In this chapter, we set out to show that in-ear PPG possesses sufficient respiratory waveform information for the screening of obstructive breathing disorders.

### 9.1.3 Change in breathing with COPD

The prevalence of respiratory diseases has grown by 39% in the last 3 decades [142], with nearly 1 in 5 people in the UK having had a diagnosis of asthma, chronic obstructive pulmonary disease or another respiratory illness [175].

Chronic obstructive pulmonary disease (COPD) is a debilitating illness caused by an increased inflammatory response in the lungs which leads to obstructed airflow [144], particularly during expiration. Chronic obstructive pulmonary disease is generally diagnosed with spirometry, by measuring the ratio of volume during forced expiration in one second ( $FEV_1$ ), against forced vital capacity (FVC), whereby COPD is defined as  $FEV_1/FVC < 0.7$ . This obstruction during expiration leads to an increased respiratory rate (tachypnea), with a decreased inspiration time ( $T_I$ ) in comparison with the overall breathing time ( $T_{TOT}$ ). The ratio of  $T_I/T_{TOT}$  in COPD, otherwise known as the inspiratory duty cycle, is therefore decreased at rest and during exercise [149] [150], with values of around 0.35 seen at rest compared with 0.42 in healthy patients [149]. This change in duty cycle is a major difference seen in obstructive lung disease when compared to restrictive lung diseases such as idiopathic pulmonary fibrosis (IPF), where, due to the restriction to both inspiration and expiration, the  $FEV_1/FVC$  remains higher [176]. Pulmonary fibrosis usually has an increased respiratory rate, and lower respiratory volume, but no significant change to the duty cycle.

Other methods of classifying respiration in COPD include examining the sample entropy of flow signals [155], where the sample entropy decreased with an increasing severity of COPD due to fewer degrees of freedom with constrained breathing. Moreover, in the PPG domain, a combination of pulse, respiration rate, and  $SpO_2$  (derived from pulse oximetry) has been used to predict exacerbations in COPD patients with moderate accuracy [177], and PPG has

recently been used to estimate lung compliance in lung disease [178]. To our knowledge, there has not yet been a successful classification of COPD from PPG-derived respiratory waveforms, and to address this void we both explore the valuable respiratory properties of in-ear PPG and employ these in patients to examine and classify COPD.

## 9.2 Materials and experimental design

### 9.2.1 Hardware

The hardware used in this chapter is detailed in chapter 2, section 2.2.2. In the simultaneous in-ear PPG and finger PPG recordings, the MAX30101 photoplethysmography chip was secured to the right index finger. Both the in-ear PPG sensor and finger sensors were wired to a purpose built circuit board which stored the data on an SD card.

For simultaneous spirometry and in-ear PPG recordings, a SFM3200 flow meter by Sensiron (Stäfa, Switzerland) was used to measure breathing flow with an airtight connection to a tube into which the participant breathed, whilst a nose clip restricted nasal breathing. The SFM3200 was connected to an Arduino Uno by Arduino (Somerville, MA, USA) which recorded the flow rate values. The Arduino was also used to send out electrical pulses to the PPG recording circuit at semi-regular time intervals, so that the two data streams of PPG and airflow could be time aligned.

### 9.2.2 Experimental design

The recordings in this chapter were performed under the IC ethics committee approval JRCO 20IC6414, and the NHS Health Research Authority 20/SC/0315. All subjects gave full informed consent.

#### Simultaneous spirometry and in-ear PPG

Simultaneous in-ear PPG and spirometry was used to evaluate presence of respiratory frequencies in the three respiratory modes of intensity, pulse amplitude and pulse interval. This was



achieved across three different frequencies corresponding to slow, moderate and fast breathing rates. Furthermore, simultaneous in-ear PPG and spirometry recordings were used to evaluate the extraction of the examined breathing waveforms: i) normal breathing and ii) breathing with a duty cycle typical to severe chronic obstructive pulmonary disease (COPD). In both the cases of frequency and flow waveform, the spirometer served as the ground truth. The participant in these recordings was a healthy male aged 25 years. The participant had the in-ear photoplethysmography sensor placed in the right ear canal, whilst breathing into the spirometer. In both cases of breathing at different frequencies and with different characteristic waveforms, the subject was informed of when to inspire and expire using a timed on-screen animation. For the frequency recordings, the animation aided the subject in breathing at frequencies of 0.18Hz, 0.25Hz and 0.33Hz, corresponding to 10.8 breaths per minute, 15 breaths per minute and 20 breaths per minute, respectively. The subject adhered to each frequency for 2 minutes, with 30 seconds rest between different frequencies. For the simulated low inspiratory duty cycle breathing, the base frequency was 0.2Hz and the subject adhered to a breathing timing ratio that was 1:3 inspiration to expiration, for 120 seconds.

### **Simultaneous in-ear and finger PPG**

The participants in the recordings were 14 healthy subjects (7 males, 7 females) aged 19 - 38 years. Two PPG sensors were used per subject, the first safely secured within the right ear canal, as shown in Fig. 2.4, and the second secured to the right index finger. Subjects were in a seated position and were instructed to breathe normally for 120 seconds, whilst photoplethysmography was recorded from simultaneously from both the ear and finger. Out of the 14 subjects, 3 subjects were discarded from analysis as they showed no clear respiratory peak in the frequency domain. The 11 subjects used for analysis consisted of 6 males and 5 females aged 19 - 28 years.

### **Resting recordings in older healthy subjects and patients with breathing disorders**

The participants in these recordings were split up into 2 groups, an older healthy subset of 4 subjects (2 males, 2 females) aged 56 - 62 years, and a subset of 6 subjects with respiratory disease (3 males, 3 females) aged 53 - 88 years. Out of the patients with respiratory disease, 4

patients had chronic obstructive pulmonary disease and 2 patients had idiopathic pulmonary fibrosis. Continuous PPG was measured from the ear canal of subjects at rest for 2 minutes using our in-ear PPG sensor.

### Artificially obstructed breathing

The participants in these recordings were 6 healthy subjects (4 males, 2 females) aged 23 - 30 years. Participants were asked to breathe in through a tube of internal diameter 8mm and length 300mm, and breathe out through a tube of internal diameter 5mm and length 300mm, giving a ratio of resistance from inspiration to expiration of  $\frac{25}{64} \approx 0.4$ . The tubes were linked with one-way valves in opposite directions so that breathing would automatically switch between the two tubes when switching from inspiration to expiration [179]. Continuous PPG was measured from the ear canal of subjects whilst they breathed through the tubes for 2 minutes.

## 9.3 Signal processing

### 9.3.1 Extraction of respiratory modulations from PPG

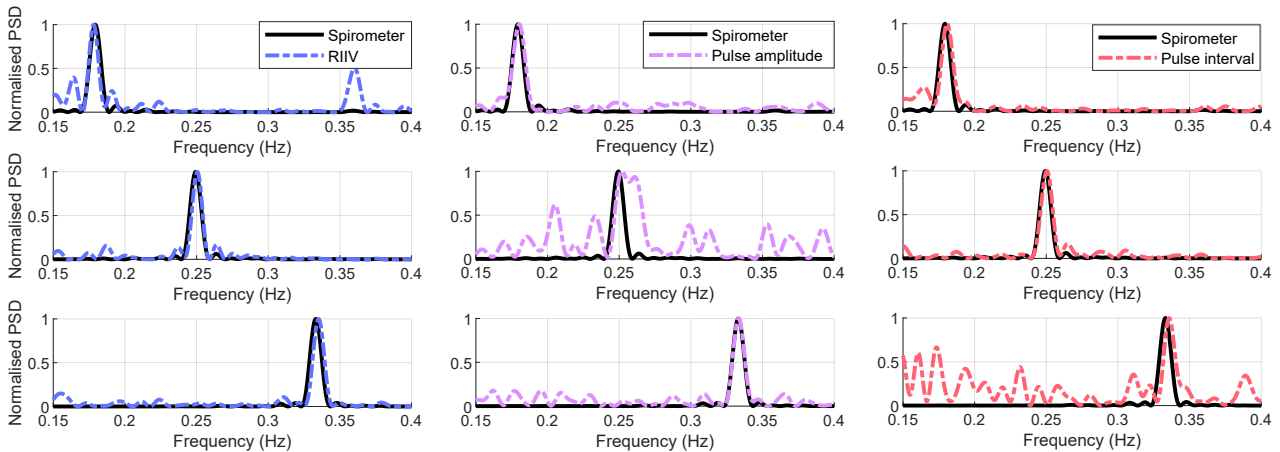


Figure 9.1: Normalised power spectral density (PSD) between 0.15 and 0.4Hz of each respiratory mode from in-ear PPG: Respiration induced intensity variations (blue, left), pulse amplitude variations (purple, middle) and pulse interval variations (red, right). The normalised PSD for each recording and respiratory mode is compared with the normalised PSD of the ground truth spirometry in each plot, which provides the true frequency distribution of respiration for that recording. The top row corresponds to recordings with a breathing rate of 0.18Hz, with the middle row and bottom row to 0.25Hz and 0.33Hz respectively.

The three major respiratory modes in photoplethysmography are: i) respiration induced intensity variations (RIIVs) which are accessible from the raw PPG, ii) pulse amplitude variations and iii) pulse interval variations, both accessible from the AC component of PPG. The extraction of the RIIV waveforms can be achieved via band-pass filtering or by adaptive methods such as empirical mode decomposition which will be described in full in the next section. However, for the purpose of spectral comparisons, the RIIV can be observed by taking the periodogram of the unfiltered PPG signal, whereby the signal is first detrended to remove the drifts and ensure that the periodogram is not biased at 0Hz.

To extract the AC component of the PPG signal, the detrended PPG was band-pass filtered between 0.9 and 30Hz. Peaks and troughs were then extracted from the pulse signal using the MATLAB by MathWorks (Natick, MA, USA) function *findpeaks*, with a minimum peak prominence of 150 arbitrary units. For the pulse amplitude variations, the envelope of the pulse signal was calculated by interpolating the peaks at 62.5Hz, and interpolating the troughs at 62.5Hz, and then summing up their absolute value. The time values of the troughs were used for the pulse interval variations as, due to the characteristic pulse waveform from the ear [72], the troughs are less sensitive to noise. The pulse interval signal was calculated as the time between consecutive pulses at the time point of each trough, and interpolated at 62.5Hz to match the sampling frequency of the PPG signal. Fig. 9.1 shows exemplar periodograms of the three respiratory modes, indicating that the normalised power spectral density has good adherence to the ground truth spirometry across all respiratory modes and across the three respiratory frequencies of 0.18Hz, 0.25Hz and 0.33Hz. The largest disparity between the estimated frequency and ground truth was an error of 0.003Hz (0.18 breaths per minute) that occurred with the pulse interval mode at 0.33Hz breathing. The periodograms were, however, taken over long time periods of 100 seconds and the recordings were performed with minimal movement, providing an ideal situation for good frequency adherence. Over shorter time periods and with motion artefacts, adherence to the ground truth periodogram would be lower.

### 9.3.2 Respiratory power comparisons: ear vs finger

From the 120 second photoplethysmography recordings of normal breathing, the two modes of *pulse amplitude variations* and pulse interval variations were extracted from the ear and finger as described in the previous subsection. The analysis considered the final 84 seconds (5250

samples) in an attempt to mitigate the conscious effects on breathing that can occur when first being told to breathe normally. All three modes were detrended to remove the mean, and power spectra of the three modes were then assessed for a shared respiratory peak. The recording was discarded if there was no clear shared peak present across modes on both the ear and finger recordings. A clear respiratory peak was observed in 11 out of the 14 recorded subjects. The normalised peak value was calculated by taking the power spectral density (PSD) value of the respiratory peak and dividing it by the sum of the power spectral density from 0Hz to 2Hz. The respiratory PSD ratio was then calculated by dividing the normalised peak from the ear by the normalised peak from the finger, and  $\log_{10}$  of the ratio was taken to make the distribution of ratios proportional in each direction. Accordingly, a negative value represented a greater relative respiratory power from the finger, a positive value represented a greater relative respiratory power from the ear, and a value of zero represented no difference between the two recording sites. This method is similar to methods previously employed in the literature, with the difference in that it focuses on a defined respiratory peak rather than on overall power in the frequency band. This choice was made as in general especially with the respiration induced intensity variations, much of the power in the respiration frequency band can be from higher harmonics of low frequency variations. Moreover, normalising by the sum of the power spectral density helped to mitigate differences in signal quality that could occur from inadequate placement of either sensor. Importantly, our analysis looked at all three respiratory modes, rather than just the respiration induced intensity variations.

### 9.3.3 Empirical mode decomposition for respiration

Empirical mode decomposition (EMD) employs a nonlinear data driven filter-bank structure to deconstruct time domain signals into data-adaptive narrow-band amplitude and frequency (time series) components, known as intrinsic mode functions (IMFs) [25]. Since the IMFs obtained by empirical mode decomposition are data driven and thus physically meaningful, empirical mode decomposition has proven effective at decomposing nonstationary and multi-scale physiological data, such as electroencephalography (EEG) into different frequency bands [26] [27] and at extracting respiration rate from PPG [180]. The EMD, multivariate EMD (MEMD) and noise-assisted multivariate EMD (NA-MEMD) algorithms are detailed in Chapter 2, section 2.3.

To extract respiratory waveforms from PPG, we employed the raw photoplethysmography sig-

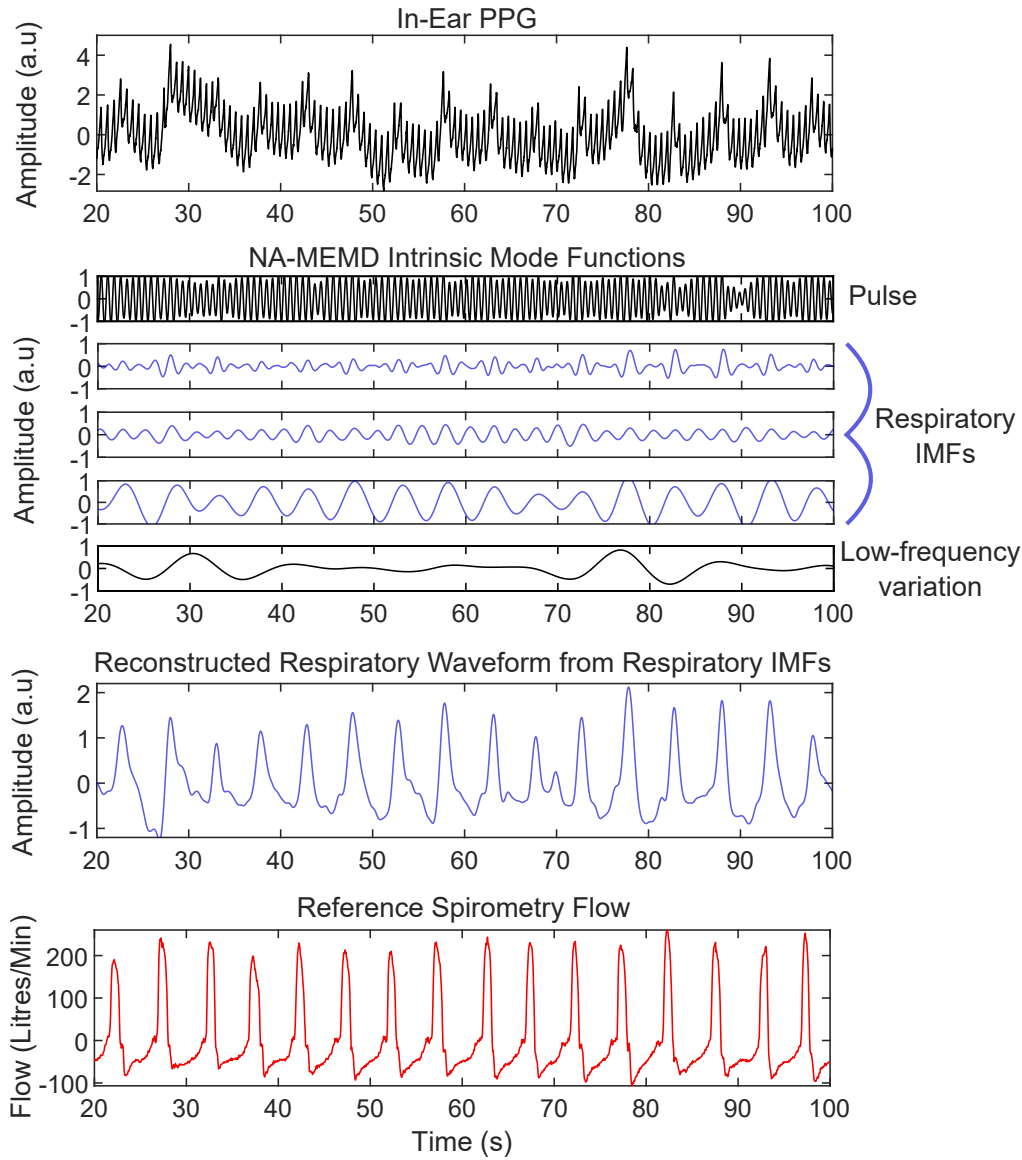


Figure 9.2: Noise assisted multivariate empirical mode decomposition (NA-MEMD) of in-ear PPG for a subject breathing with an atypical 1:3 inspiration to expiration time ratio. Indicated are the detrended in-ear PPG (top, black), the breakdown of the in-ear PPG into intrinsic mode functions (top middle) with the respiratory IMFs highlighted in blue, the reconstructed in-ear PPG respiratory waveform (bottom middle, blue) and the reference spirometry flow signal (bottom, red). For convenience, the PPG waveform is flipped so that peaks correspond to inspiration and troughs correspond to expiration.

nal, the pulse interval variations and the pulse amplitude variations as the three main channels in NA-MEMD, given that the commonality between the intensity mode, the pulse amplitude mode and the pulse interval mode is the respiratory information. Therefore, rather than using different recording channels as our inputs, our main inputs are instead all derived from the same PPG signal. Five white Gaussian noise channels are added to improve frequency localisation and reduce the mode mixing between the IMFs. This produces IMFs for each of the three respi-

Table 9.1: Summary of the features used for the classification of COPD.

Category	Features
COPD based	skewness, duty cycle, $max -  min $ , normalised $max -  min $ <sup>†</sup>
General	Standard deviation, kurtosis, spectral skewness, breathing frequency

<sup>†</sup> Normalisation corresponds to division by the standard deviation.

ratory modes. Only the IMFs for the raw PPG signal were used to reconstruct the respiratory waveform, given that the intensity based modulations are not limited in sample rate to the pulse frequency and therefore contain more detailed high frequency respiratory information. A respiratory IMF was defined as an IMF with over a third of its spectral power between 0.2Hz and 0.6Hz, allowing higher frequency respiratory detail to be captured. The respiratory IMFs were then summed to reconstruct the respiration signal. This achieved physically meaningful respiratory waveforms which vary both in amplitude and frequency across the respiratory band.

An example of NA-MEMD decomposition and reconstruction is shown in Fig. 9.2. The subject adhered to a 1:3 inspiration to expiration time ratio, similar to respiratory waveforms typical of severe chronic obstructive pulmonary disease. It can be seen that despite the lowest frequency respiratory IMF capturing the base frequency of respiration, it lacks the detail captured in the higher frequency IMFs which help to illuminate the difference in inspiration to expiration time ratio. This is an argument for using methods such as empirical mode decomposition over standard filter banks when trying to extract physically meaningful respiratory waveforms.

### 9.3.4 Classification of COPD

For the classification of chronic obstructive disease we focused extracting features based on the COPD waveform, given that we desire to be able to classify COPD not just in comparisons with healthy data but relative to other respiratory diseases. Whilst respiration frequency itself is an important feature in assessing respiratory health, both COPD and pulmonary fibrosis lead to an increase in breaths per minute, making it a poor feature for distinguishing between the two when used alone.

Six in-ear PPG recordings were used from ear finger comparisons and labelled as young and healthy. Therefore in total there were 6 young healthy subjects aged 21 - 28 years, 4 older

healthy subjects aged 56-62 years, 4 subjects with COPD aged 55 - 88 years and 2 subjects with idiopathic pulmonary fibrosis (IPF) aged 61 and 68 years. The 120 second recordings of PPG data from each subject were trimmed to remove motion artifacts occurring at the start and end of the recordings, and pulse interval and pulse amplitude variation signals were extracted. The NA-MEMD was then performed on the PPG, pulse interval and pulse amplitude signals for each subject independently. Respiratory signals were then reconstructed from the respiratory intrinsic mode functions, and 15-second epochs without the presence of motion artifacts were selected. Each 15-second epoch was rounded off so that the number of breathing cycles was an integer; in this way, 15 segments were extracted for COPD, 10 for IPF, 18 for young healthy and 16 for older healthy, resulting in a total of 59 segments. During inspiration, the raw PPG decreases in intensity, and therefore for convenience the waveforms were flipped so that peaks represented peak inspiration, and troughs represented peak expiration.

Features were chosen by accounting for the principle that COPD mainly obstructs expiration, and thus expiration takes up a larger proportion of the total breathing time, while resting inspiratory flow rates are usually higher than expiratory flow rates. The skewness, duty cycle and the difference between the maximum value and the absolute of the minimum value were extracted as COPD related features. Skewness is defined as

$$S(\mathbf{y}) = \frac{E(\mathbf{y} - \mu)^3}{\sigma^3} \quad (9.1)$$

where  $S(\mathbf{y})$  denotes the skewness of a signal  $\mathbf{y}$ ,  $E(x)$  represents the expected value of  $x$ ,  $\mu$  is the mean of the signal  $\mathbf{y}$ , and  $\sigma$  is the standard deviation of  $\mathbf{y}$ . Distributions with a longer positive tail than negative tail therefore have a positive skewness, and distributions with a longer negative tail have a negative skewness. Similarly, distributions that are symmetrical have zero skewness.

In this implementation, duty cycle was defined as time spent in inspiration divided by the total respiratory time. Given that the respiration induced intensity variation of PPG is a proxy for flow rate and assuming the mean flow rate is zero, time spent above the mean of the flipped waveform corresponds to inspiration, and time spent below the mean to expiration. Thus, inspiratory duty cycle was calculated by dividing the number of samples above the mean by the total number of samples as follows

$$D(\mathbf{y}) = \frac{\sum_{n=1}^N H(y(n) - \mu)}{N} \quad (9.2)$$

where  $D(\mathbf{y})$  corresponds to the duty cycle of a signal  $\mathbf{y}$ ,  $N$  represents the sample number,  $\mu$  is the mean of the signal  $\mathbf{y}$ , and  $H$  represents the Heaviside function. Overall, a total of 8 features were used for classification, and are summarised in Table 9.1.

For classification, features were used to train a random forest classifier, employed using the scikit-learn Python toolbox [99]. For the random forest base, the number of trees was set to 50, the class weight was set to ‘balanced subsample’ and the maximum number of features was set to 3. Binary classification was performed with the COPD data being labeled as such, and the young healthy, older healthy and IPF data being labelled as non-COPD. Both leave-one-segment-out and leave-one-subject-out cross validation methods were used. In the case of leave-one-segment-out, a summed confusion matrix was used to evaluate performance, and in the case of leave-one-subject-out the mean classifier probability for COPD was taken over all segments for that subject, and an ensemble average was taken over 5 different train-test realisations.

## 9.4 Results

### 9.4.1 Spectral power of in-ear PPG vs finger PPG

The  $\log_{10}$  ear to finger respiratory power ratios for the three respiratory modes of respiration induced intensity variations (RIIVs), pulse amplitude variations and pulse interval variations for 11 subjects are presented in Fig. 9.3(a). The boxplots presented in Fig. 9.3(b) show median  $\log_{10}$  ratio values of 0.927 for RIIV, 0.463 for pulse amplitude variations and -0.002 for pulse interval variations, corresponding to an average of an 8.5-fold increased RIIV power from the ear compared to the finger, a 2.9-fold increase in pulse amplitude power from the ear and no change in the power of the pulse interval variation between the ear and the finger. A one sample t-test rejected the null-hypothesis that the  $\log_{10}$  ratios had a distribution mean of zero in the case of the RIIV and pulse amplitude ratios ( $p = 0.0008, p = 0.01$ ) and did not reject the null hypothesis in the case of the pulse interval power ratios ( $p = 0.79$ ). When comparing the ratios and normalised respiratory peak values across the sexes (5 female, and 6 male) the



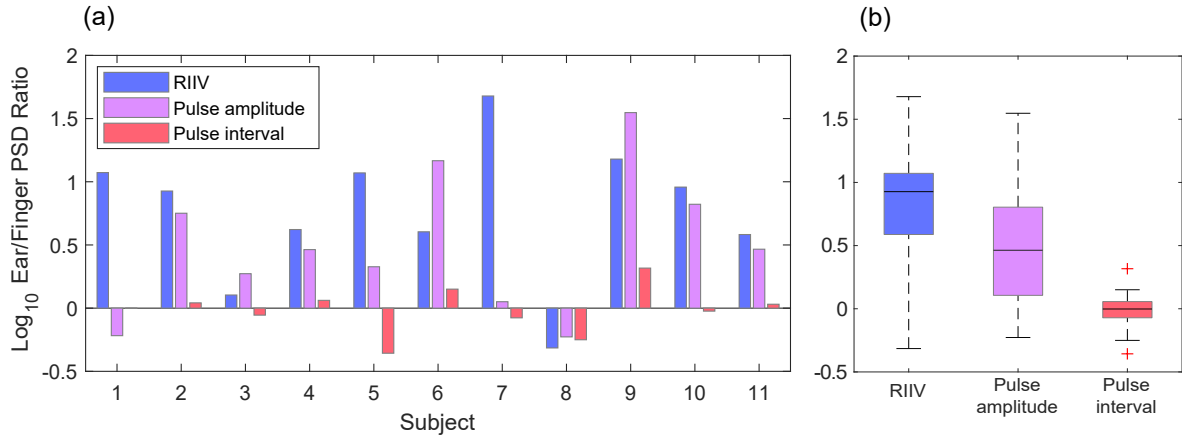


Figure 9.3: The  $\log_{10}$  of the respiratory power spectral density ratio between the ear and finger, for each respiratory mode of respiration induced intensity variations (RIIV) (blue), pulse amplitude variations (purple) and pulse interval variations (red). (a) The  $\log_{10}$  ratios for each respiratory mode presented for all 11 subjects. (b) Boxplots showing the distribution of  $\log_{10}$  ratios for each respiratory mode. Values of zero represent equal relative respiratory power for both the ear and finger, whereas values of 1 represent a 10-fold increase in relative respiratory power from the ear and values of -1 represent a 10-fold decrease in relative respiratory power from the ear.

only significant difference found was an increased pulse interval variation power in the finger in females compared with males ( $p = 0.04$ ) and whilst an increased pulse interval variation power was also seen in the ear in females, it was not significant ( $p = 0.15$ ).

### 9.4.2 Classification of COPD from in-ear PPG

The COPD data in general showed higher skewness than non-COPD data. Fig. 9.4 exemplifies that the COPD waveform is more likely to have higher amplitude inspiratory peaks when compared with expiratory troughs, thus giving the waveform distribution a positive tail and therefore a positive skew as highlighted through the histograms. The distribution of skewness between older healthy, young healthy, COPD and IPF is summarised in the boxplots in Fig. 9.5(a), and shows that overall the skewness of COPD segments tends to be higher with a median of 0.23 and an interquartile range of 0.20 to 0.48, compared with an older healthy median of -0.14 and IQR of -0.23 to 0.02, a young healthy median of -0.03 and IQR of -0.20 to 0.02 and an IPF median of -0.05 and IQR of -0.24 to 0.11. The change in duty cycle in COPD from in-ear PPG was less pronounced than the difference in skewness, but in general  $T_I/T_{TOT}$  was lower in COPD as shown in Fig. 9.5(b). The median duty cycle in COPD was 46.7%, compared with 50.0%, 51.4% and 50.5% in the older healthy, young healthy and IPF subjects,

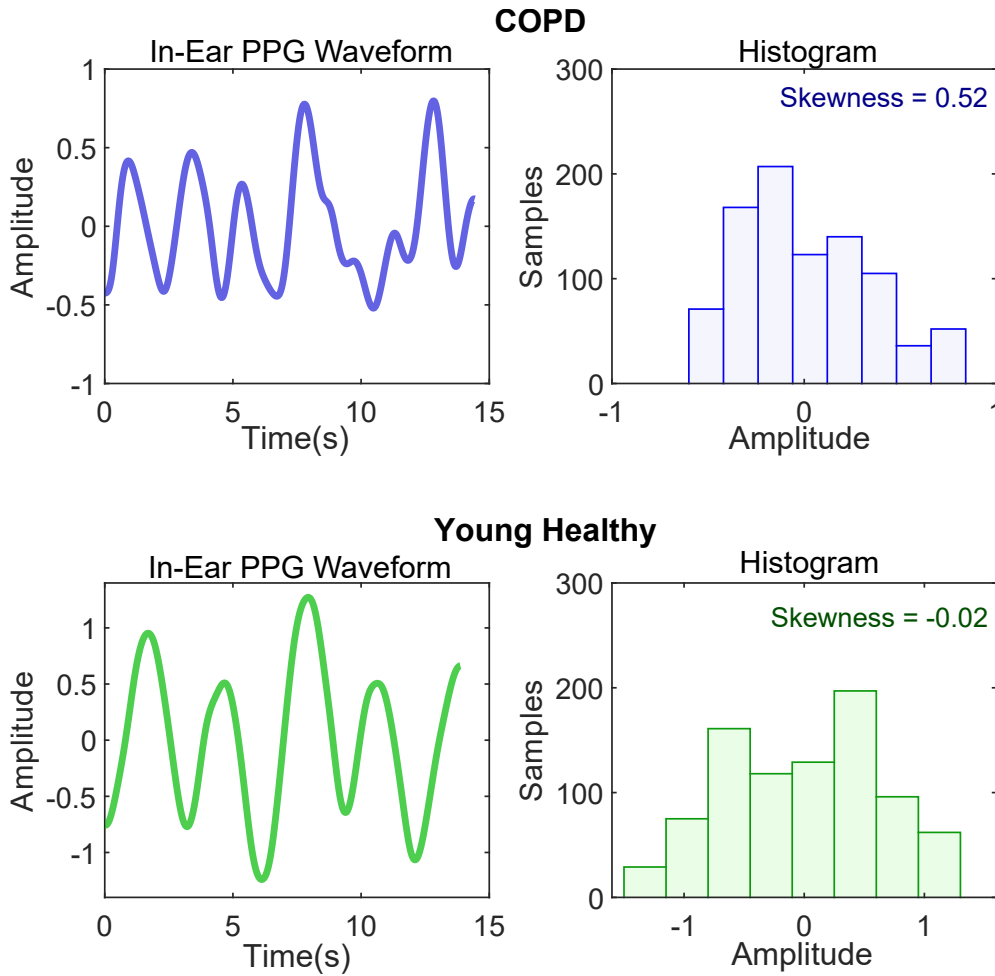


Figure 9.4: Exemplar in-ear PPG respiratory waveforms and their corresponding probability density estimates shown in the form of histograms, for the case of a patient with COPD (top, blue) and a young healthy subject (bottom, green). Skewness values for each presented waveform are shown in the histogram plots, with a skewness of 0.52 for the COPD in-ear PPG waveform and -0.02 for the young healthy in-ear PPG waveform.

respectively. The distribution of normalised difference in inspiratory and expiratory magnitude was similar to skewness, with values above zero corresponding to increased inspiratory magnitude and values below zero corresponding to increased expiratory magnitude. Fig. 9.5(c) shows that COPD skewness values trend higher with a median of 0.37 and an IQR of 0.30 to 0.86, compared with an older healthy median of -0.32 and IQR of -0.56 to -0.08, a young healthy median of -0.03 and IQR of -0.29 to 0.16 and an IPF median of 0.17 and IQR of -0.42 to 0.19. Artificial obstruction mirrored COPD across all features, with a median skewness of 0.26, a median duty cycle of 47.0% and a median normalised difference in inspiratory and expiratory magnitude of 0.41, compared to median values of 0.23, 46.7% and 0.37 in COPD.

All features were calculated on clean segments of data, with segments with motion artefacts

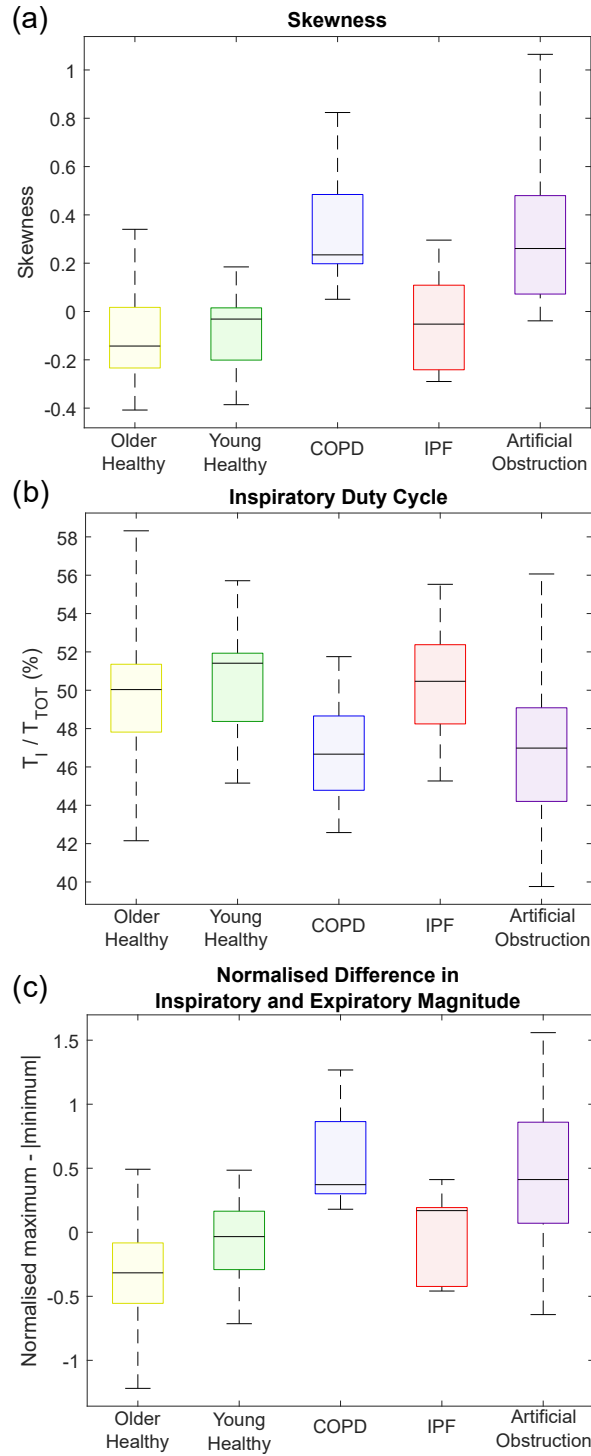


Figure 9.5: Boxplots of COPD related features extracted from in-ear PPG recordings for older healthy subjects (yellow, 16 segments), young healthy subjects (green, 18 segments), patients with COPD (blue, 15 segments), patients with IPF (red, 10 segments) and healthy subjects with artificially obstructed breathing (purple, 23 segments). (a) Boxplots of skewness defined in (9.1). (b) Boxplots of inspiratory duty cycle defined in (9.2). (c) Boxplots of the normalised difference in inspiratory and expiratory magnitude, defined in Table I as the normalised maximum from which the absolute value of the minimum is subtracted.

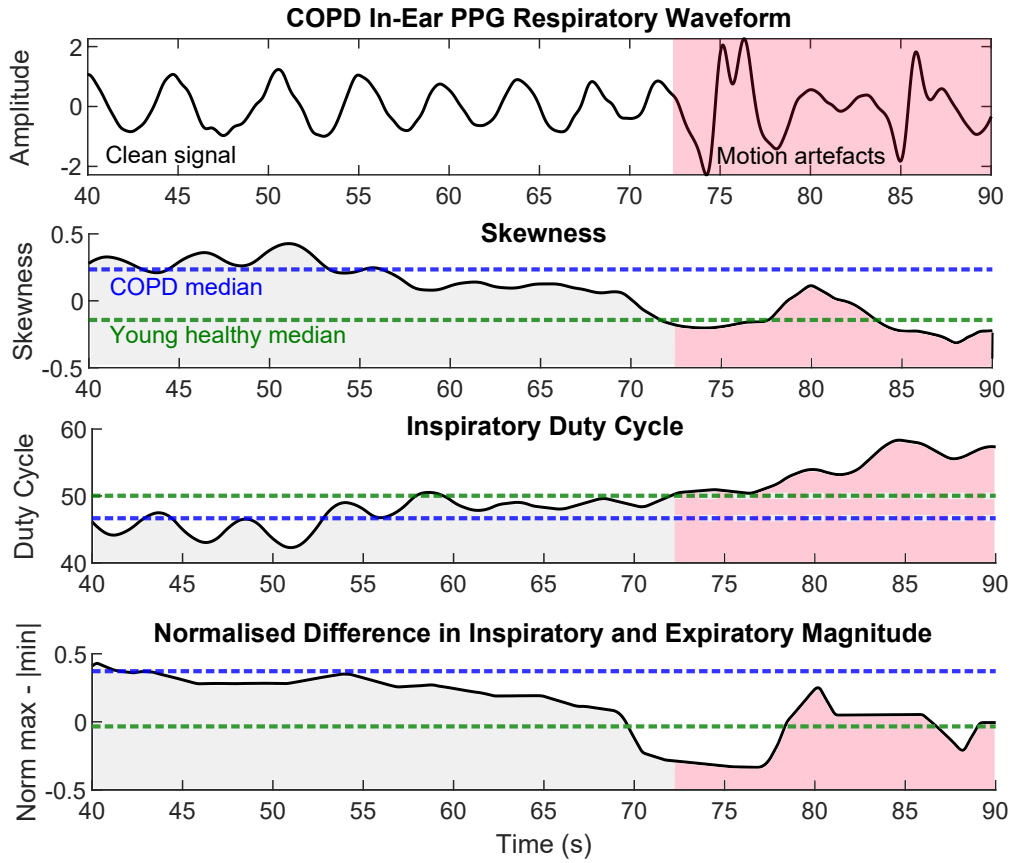


Figure 9.6: An exemplar in-ear PPG derived COPD respiratory waveform, with the corresponding COPD features calculated with a sliding 500 sample (8 second) window. For each feature, highlighted with dotted lines are medians from the COPD recordings (blue) and the medians from the young healthy data set (green). The pink shaded section highlights the presence of motion artefacts in the data, and the corresponding distortion to the calculated features.

being ignored. Fig. 9.6 highlights that motion artefacts can distort the features, with an example of feature calculation in an artefact corrupted COPD recording. It is shown that features diverge from what is expected for COPD towards what is expected for young healthy participants when a motion artifact is present, which in this case would lead to miss-classification.

Average classification accuracy for leave-one-segment-out cross validation was 92%, with class specific accuracy of 87% for COPD, and 93% for non-COPD. These results are presented in the confusion matrix in Fig. 9.7(a). Moreover, classification of COPD had a precision of 81% and an F-score of 84%. In the leave-one-subject-out cross validation results, the COPD probability was extracted from the random forest classifier and the mean was calculated across testing segments for each subject, with the ensemble average of this probability then taken over 5 realisations. The 4 COPD subjects had the 4 highest classifier probabilities of COPD, as shown in Fig. 9.7(b). In both cases of leave one subject and leave-one-segment-out, the 2 most

(a) Leave One Segment Out

		<u>Predicted</u>	
		COPD	Non-COPD
<u>True</u>	COPD	86.7	13.3
	Non-COPD	6.8	93.2

(b) Leave One Subject Out

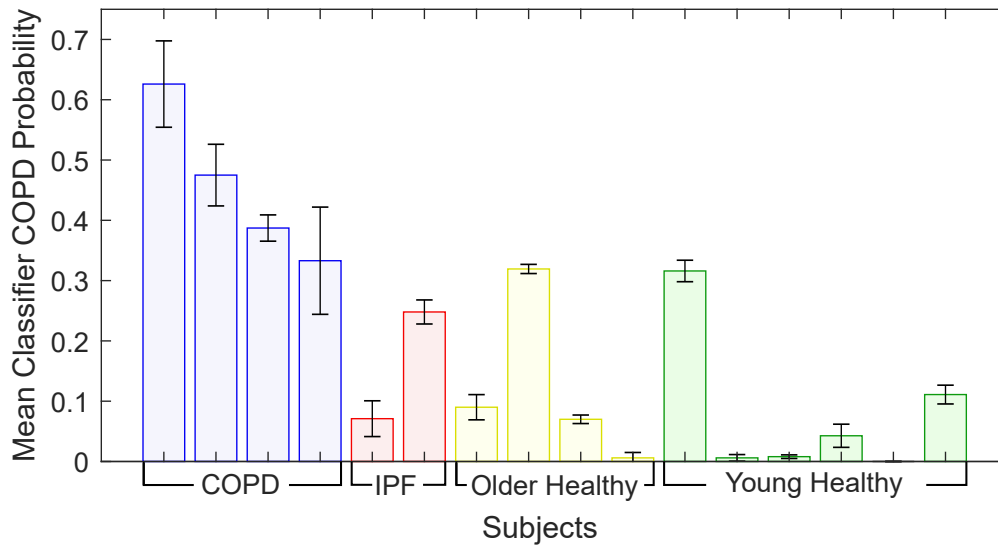


Figure 9.7: Classification of COPD from in-ear PPG respiratory waveforms. (a) The mean confusion matrix for leave-one-segment-out cross validation, with the rows corresponding to the true COPD and non-COPD allocations, and the columns corresponding to the predictions of COPD and non-COPD. (b) The mean COPD probabilities extracted from the random forest classifier across each segment for a given subject, and ensemble averaged over 5 realisations. The error bars correspond to the standard deviation of the segment mean over 5 realisations. The COPD subjects are designated in blue, IPF in red, older healthy in yellow and young healthy in green.

important features for classification, based on the reduction of tree impurity in the random forest, were the skewness and the normalised difference between the maximum and absolute minimum.

## 9.5 Discussion

### 9.5.1 Spectral power of in-ear PPG vs finger PPG

The increased respiration induced intensity variation power of roughly 8.5-fold that we see from the ear over the finger is in good agreement with Budidha *et al* [8], Nilsson *et al* [11] and Shelley *et al* [171], but is in contrast to the findings of Charlton *et al* [173]. In addition, we also see in increased spectral power of pulse amplitude variations due to respiration from the ear, again in contrast to Charlton *et al* [173], and no change in power for pulse interval variations. In the case of the work by Charlton *et al*, the authors explain that the difference may be due an increased signal to noise ratio from the finger sensor, whereas we have accounted for differences in signal quality in our calculations. One possible explanation for the increased power that we see in the amplitude variations from the ear is that the ear canal vasculature is in close proximity to the carotid artery, and therefore a more exaggerated pulse amplitude variation might be expected.

Moreover, we do not see a significant difference in RIIVs between the sexes for the ear or finger sensors which is in agreement with Nilsson *et al* [174]. Similarly we do not see a significant difference in the pulse amplitude variations between sexes from the ear and finger, and we do not see a significant difference in the pulse interval variations between sexes from the ear. We do, however, see a significantly higher respiratory power for pulse interval variations in the finger from females, which is in agreement with Li *et al* [170].

### 9.5.2 Classification of COPD from in-ear PPG

In the extracted in-ear PPG respiratory waveforms, we found an increased skewness of the data distribution towards inspiration, higher inspiratory magnitudes compared with expiratory magnitudes and decreased inspiratory duty cycle in those with COPD, compared with young healthy subjects, older healthy subjects and patients with IPF. Theoretically, these differences were expected, given COPD manifests itself in obstruction to expiration which results in a lower FEV<sub>1</sub>/FVC when testing with spirometry, and therefore in a shorter time spent inspiring and a higher peak inspiratory flow compared with expiratory flow at rest. Importantly, when breathing was artificially obstructed with tubes that restricted expiration more than inspiration,

the analysis showed the same trends as with COPD. This provides further justification that the chosen features discriminate obstructive breathing disorders. The duty cycle differences were not as pronounced from the ear-PPG as those from chest wall measurements in the literature [149] [150], and were higher for all COPD and non-COPD subjects. In the case of the recorded 1:3 inspiration to expiration example, presented in Fig. 9.2, the flow data had a inspiratory duty cycle 26%, whereas the in-ear PPG extracted waveform had a calculated duty cycle of 37%. Similarly, flow data had a skewness = 1.5 compared with a skewness = 0.9 in the in-ear PPG waveform. This is evidence for waveform differences in breathing being less pronounced from the in-ear PPG than in the ground truth air flow, and is a possible explanation for why the differences from the ear waveforms recorded in COPD patients are less pronounced than the chest wall measurements in the literature. A similar duty cycle and skewness to the extracted in-ear PPG waveforms was achieved by low-pass filtering the spirometry flow data in Fig .9.2, with a cut off frequency of 0.33Hz. This is evidence that the reduction in COPD differences shown from in-ear PPG is likely caused by the transfer function from thoracic pressure to venous return which has low-pass filter effects [168]. This effect may also be exaggerated by patients with COPD on average having higher resting respiration frequencies.

Analysis of data corrupted by artefacts, shown in Fig. 9.6, highlights that motion artefacts can indeed distort the extracted features, and thus for the classification of COPD it is recommended that artefact corrupted segments are discarded. This however, does not affect the utility of the proposed methodology, as with this method classification can be performed on data recorded at rest where artefacts are far less common than during movement such as walking. Furthermore, with only a few clean breathing cycles required to achieve an accurate prediction, the negative impact of artefacts is further reduced. Motion artefacts are generally of a broadband nature, with a peak frequency that is at least an order of magnitude higher than the peak frequency of PPG derived respiratory waveforms; this makes it straightforward to identify and reject artefact-corrupted segments.

Using features extracted from in-ear PPG waveforms segments that were less than 15 seconds long, a random forest classifier was able to distinguish between COPD and non-COPD with a specificity of 87% and overall accuracy of 92% in leave-one-segment-out cross validation. Moreover, the fact that the highest mean classifier probabilities all occurred in COPD subjects with leave-one-subject-out cross validation demonstrates that the in-ear PPG features, described in

this chapter, are robust enough to generalise across COPD patients even with only 3 training subjects. Notably, COPD was classified against the young healthy data, older healthy data and IPF patient data, which reinforces our claim that we are detecting COPD and not just differences that could occur in general breathing disorders or with age. Importantly, this indicates that the respiratory variations detected in the in-ear PPG are strong enough to preserve information that goes beyond the respiration frequency. With more subjects, the accuracy and ability of the model to generalise would likely increase further.

## 9.6 Chapter conclusions

We have demonstrated the principle and robustness of in-ear PPG as a tool for detecting respiration frequency. It has been shown to exhibit increased spectral power over the finger PPG due to respiration for both respiration induced intensity variations and pulse amplitude variations. Further, we have introduced a novel method for extracting respiratory waveforms from PPG, based on noise assisted multivariate empirical mode composition (NA-MEMD), and have proven that so extracted in-ear PPG breathing waveforms contain sufficient information to detect differences that occur with obstructive breathing disorders such as chronic obstructive pulmonary disease (COPD). This has been further validated quantitatively when classifying COPD against healthy subjects and subjects with pulmonary fibrosis, and through comparison with artificially obstructed breathing. This has indicated the promise of in-ear PPG as a means for both screening and ambulatory monitoring of patients with respiratory disorders, and as a tool for detailed breathing analysis in consumer wearables which goes beyond just respiration frequency.



# Chapter 10

## Conclusion

### 10.1 Conclusions and applications

The objective of this thesis has been to extend the capabilities of photoplethysmography and pulse oximetry, by both leveraging the advantageous location of the ear for physiological recordings and by applying state of the art signal processing methods. A wearable in-ear pulse oximetry sensor has been utilised to comprehensively investigate blood oxygen estimation from the ear canal, finding that the ear is non-inferior to the conventional finger recording site. Moreover, it has been found that the ear responded 12.4 seconds faster on average to changes in blood oxygen, likely due in part to its proximity to the brain and thus high priority for oxygenation. The fast response time of in-ear  $\text{SpO}_2$  could be applied in scenarios where a low latency blood oxygen measure is critical, such as neonatal resuscitation and closed loop automatic oxygen therapy. Low latency in-ear  $\text{SpO}_2$  would also be useful for the quick detection of sleep apnea events in polysomnography, along with the added benefit of being wearable. Furthermore, it has been demonstrated that adhesive ear-based pulse oximetry is capable of detecting exercise induced decreases in blood oxygen in patients with breathing disorders in a continuous fashion, opening the door for unobtrusive ambulatory monitoring of lung disease.

The unique position of in-ear pulse oximetry has been further examined in the context of cognitive workload. The brain is the most metabolically active organ in the body, with high oxygen demands that increase many fold with increased mental effort. It has been found that in-ear  $\text{SpO}_2$  was sensitive enough to detect these changes in oxygen demand in the blood, with

increased cognitive workload robustly leading to a decrease in in-ear  $\text{SpO}_2$ . Using this discovery, in-ear  $\text{SpO}_2$  has allowed for reliable discrimination between levels of an N-back working memory task, resulting in high accuracy classification of cognitive workload from a single wearable in-ear sensor. This concept has been explored further in relation to gaming, with games that require high mental effort showing consistent decreases in in-ear  $\text{SpO}_2$ . The detection of cognitive workload with a wearable in-ear  $\text{SpO}_2$  sensor has a clear application to the “playtesting” of games, as well as to the further study of cognitive workload in scenarios where wearability of a sensor is paramount, such as surgical training and aviation. Additionally, this concept has been extended from the tracking of cognitive workload, to the prediction of fatigue in driving. This thesis has demonstrated, through the use of a single wearable in-ear PPG sensor, that it is possible to accurately predict the likelihood of large steering wheel angle deflections when driving and, by extension, predict when a subject is fatigued. In the UK approximately 24% of fatal road traffic collisions are caused by driver sleepiness [114] and in 2016 there were an estimated 1.35 million road traffic deaths worldwide [181]. An automatic fatigue warning system utilising a cheap wearable sensor therefore has the potential to save many thousands of lives.

Photoplethysmography (PPG) also allows for monitoring of respiration, mainly through variations in venous blood volume, but also through changes in pulse magnitude and pulse interval. This thesis has shown that both respiration induced intensity variations in venous blood volume and respiration induced pulse magnitude variations are far stronger from in-ear PPG when compared conventional finger based PPG. Through the novel application of multivariate empirical mode composition, it has been shown that detailed spirometry-like respiratory waveforms can be extracted from in-ear PPG. Using this method in combination with features derived from common breathing patterns of chronic obstructive pulmonary disease sufferers, this thesis has demonstrated for the first time that it is possible to automatically detect COPD against other diseases from PPG derived respiratory waveforms. This has been further confirmed and developed through the creation of a novel tube based apparatus for the simulation of obstructive lung diseases in healthy subjects. The tube based apparatus is useful in the generation of physically meaningful surrogate data, but also has a clear application to the education and awareness of breathing disorders through the experience of obstruction to breathing. Chronic obstructive pulmonary disease is the third leading cause of death worldwide [182], with as much as 78% of sufferers remaining undiagnosed [183]. The application of wearable PPG for the detection

of COPD could significantly reduce this percentage, in turn informing users to treatment. The detection of COPD in those that are undiagnosed promises to reduce the risk of exacerbation and death [183] and thus this technology could help prevent millions of deaths worldwide.

## 10.2 Advice to future researchers

My advice for research on wearable sensors, based on my personal experience from my PhD studies, is that if you have an idea, test it in yourself as soon as possible. If it works on you, test it on other people. If it works with other people, then design a study around it. Ideas will come if you are proactive with reviewing the literature and are familiar with any gaps that are present, and the ability to test things quickly will help ensure that minimal time is wasted on things that don't work. Moreover, testing wearable sensors such as PPG on yourself is an excellent and easy way to familiarise yourself with physiological data, and this hands on approach is essential for getting to grips fast with bio-signal processing algorithms and for understanding the properties of the data that you will be working with. One thing I would emphasise is that, whilst not as important during initial testing of ideas, protocols should be extremely well thought out before recording for a larger study, as the time cost of getting something wrong and having to re-record data can be huge. In my PhD research, the first set of cognitive workload data I recorded had a sequential order of N-back tasks, and thus conclusions could not be drawn on whether the changes I recorded were related to increases in mental effort at the final stages, or due to mental fatigue built up throughout the recording. For this reason, I had to re-record this experiment with a randomised N-back order which, in the context of the COVID-19 lock-downs, took an extra 6 months to complete. In this example, a more thought out initial protocol design would have saved me a lot of time and effort.

## 10.3 Future work

### Mapping the SpO<sub>2</sub> delay

Existing literature has explored the delay between the hand and the foot for SpO<sub>2</sub> readings [59] and in this thesis we have analysed this phenomenon in depth pertaining to the ear canal [72].

We have found a large variability in oxygen delay between the ear and the finger ( $12.4 \pm 6.1$  seconds) across all subjects. Whilst some of this variability was explainable by the sex of the subject, a proxy for height differences and therefore change in distance between the ear and the finger, this cannot account for all of the variability. There is a possibility that  $\text{SpO}_2$  delay could relate in part to the cardiovascular health of an individual, as well as other physiological variables such as blood pressure. To this end, the delay from multiple sources, including both ears, should be recorded to form an individualised  $\text{SpO}_2$  delay map for each individual, at rest and post exercise. This map could then be compared to several measures of cardiovascular health, including blood pressure, in order to further investigate the role of cardiovascular health in oxygen delay.

### **Long term unobtrusive monitoring of respiratory diseases**

This thesis saw the development of an adhesive pulse oximetry patch which allows for the monitoring of  $\text{SpO}_2$  during hospital based walk tests with minimal motion artefacts. The next step is to apply this in an ambulatory scenario, with long term monitoring of patients with breathing disorders over the course of hours or days. Wearable pulse oximetry can provide continuous measures of blood oxygen that were discussed in chapter 4, along with features of detailed respiratory variations that are indicative of lung obstruction, presented in chapter 9. Moreover, the results from artificial obstruction shown in chapter 8 indicate that it is possible to assess changes in thoracic pressure that occur due to obstruction of the lungs, and that PPG alone could be used to infer important clinical parameters such as forced expiratory volume in 1 second. It is my opinion that the combination of PPG as a proxy for wearable spirometry and sensitive blood oxygen measures could provide vital information for the prediction of life threatening exacerbations that can occur in respiratory diseases. The same information could also be used to provide a ground truth for the effectiveness of drugs that are employed to treat COPD or cystic fibrosis. Deploying this technology in patients for home monitoring is the key next step for this research, and it is my opinion that this has a very real chance of both improving quality of life and saving lives of those suffering from chronic respiratory diseases.

### **Wearable pulse oximetry artefact removal with Multivariate EMD**

It is my opinion that motion artefacts present the biggest hurdle for the translation of much of the research in this thesis into real world consumer wearables. For the calculation of  $\text{SpO}_2$  specifically, precise amplitude ratios are required, making them highly sensitive to artefacts. The same is also true for the COPD detection features described in chapter 9. Significant research effort should be spent on both the mitigation of in-ear PPG artefacts (through hardware and positioning) and on artefact removal either in an on-line fashion or during post processing of data. For artefact removal, multivariate empirical mode decomposition works well when the shared information across channels is the signal of interest. In the context of pulse oximetry, both the red and infrared channels share the same pulse trace at different amplitudes. With both red and infrared channels, and channels at other wavelengths (such as green), an accurate pulse intrinsic mode function could be extracted allowing for more stable measures of  $\text{SpO}_2$  during motion artefacts. This concept could be extended further with multiple sensor positions around the ear, which would share the same pulse waveform but experience motion artefacts sufficiently differently to aid their removal.

### **Further investigation of the relationship between $\text{SpO}_2$ and cognitive workload**

This thesis has introduced the novel use of in-ear  $\text{SpO}_2$  for the classification of cognitive workload, but with this discovery come questions about its physical mechanisms. The main question remains as to whether or not the  $\text{SpO}_2$  change is present across the body, or if it is local to the ear canal. It could indeed be local, given the proximity of the ear to the brain, but an argument against this is that the ear canal is supplied by a separate branch of the carotid artery than the branch that supplies the brain. Another plausible explanation is that, given the minimal delay and the sensitivity of in-ear  $\text{SpO}_2$ , it is able to pick up tiny changes in  $\text{SpO}_2$  that might not be measurable from the finger. In order to investigate this, future work must integrate pulse oximetry sensors simultaneously at several positions on the body, as shown in Fig 10.1(a).

Moreover, one of the flaws of cognitive load experiments is that cognitive load cannot be perfectly controlled with respect to external stimuli and distraction. A potential way to combat this is through cognitive load experiments in immersive virtual reality (VR) environments. This could be achieved with a VR based N-back task, VR flight simulators, an example of which is

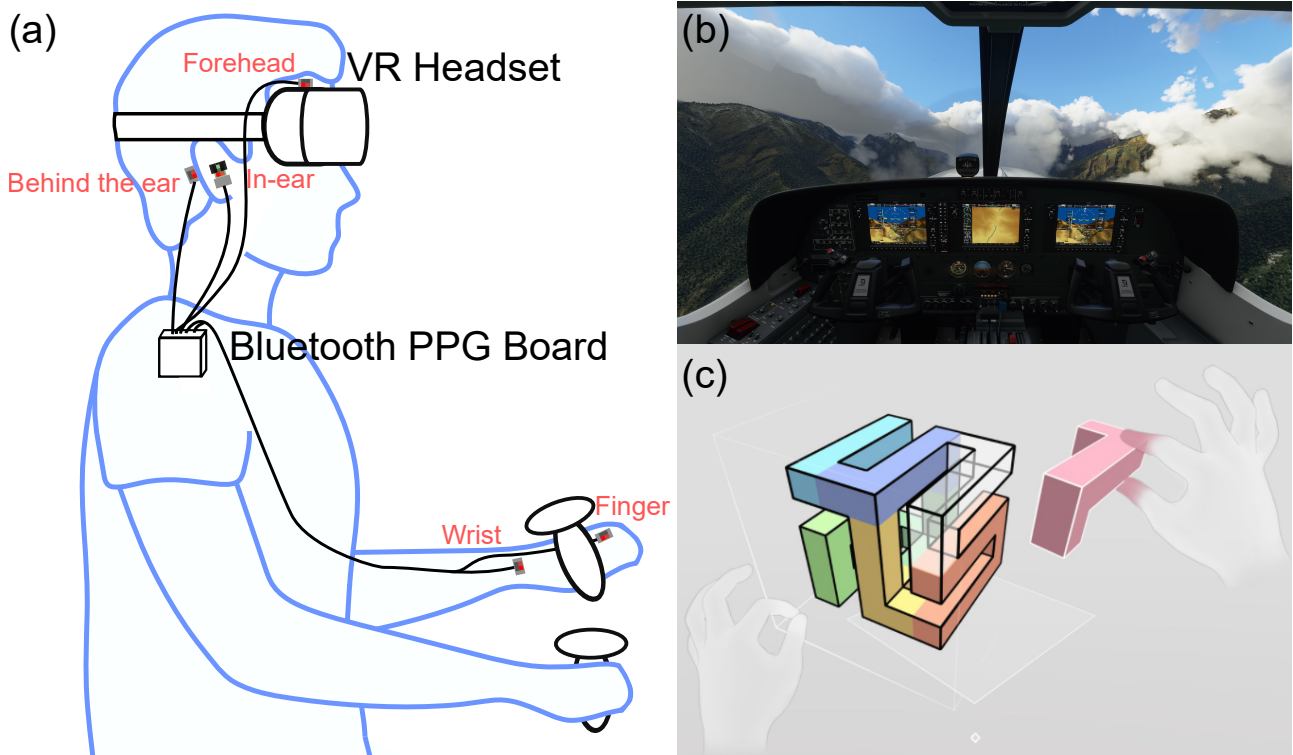


Figure 10.1: Further investigation of in-ear  $\text{SpO}_2$  for cognitive workload with virtual reality (VR) applications. (a) Experimental set-up with a subject wearing a VR headset for immersive cognitive load experiments, with PPG (labelled in red) attached simultaneously to the forehead, in-ear, behind the ear, finger and wrist, all attached to a single circuit board which will relay the PPG data wirelessly to a computer via Bluetooth. (b) An example VR flight simulator from microsoft (Credit: Microsoft, <https://news.xbox.com/en-us/2020/12/22/microsoft-flight-simulator-virtual-reality-update-available-now/>). (c) An example of puzzle solving in VR, with the game Cubism (credit: Vanbo LLC/META/Oculus, <https://www.oculus.com/experiences/quest/2264524423619421>).

shown in Fig 10.1(b), VR puzzle games, an example of which is shown in Fig 10.1(c) and even VR based surgery [87]. Future investigations of  $\text{SpO}_2$  and cognitive workload could consider the implementation of VR based tasks to induce cognitive load.

### Closing the loop: in-ear PPG for automatic detection of driver fatigue

This thesis has shown the potential of in-ear PPG, both through in-ear  $\text{SpO}_2$  and heart rate variability metrics, for the prediction of large steering wheel angles when driving. When subjects become more fatigued, the probability of large steering wheel angle deflections increases and so does the probability of crashes. Thus, by training a classifier on steering wheel angle as a label, the classifier can become a tool for automatic detection of fatigue. Whilst this currently works well with fully shuffled training, it cannot work as a fully fledged real world system until the

classifier can perform well on unseen subjects. To achieve high accuracy of leave one subject out testing, future work must carry out further recordings to increase the quantity of fatigued subject data. Future work can also experiment with individually calibrated features, similar to the features applied in Chapter 5 for the classification of cognitive workload, as calibrated features make it easier for a classifier to generalise to unseen subjects. The primary goal of future work in this area should be to close the loop by using a trained fatigue prediction classifier simultaneously with in-ear PPG in an automatic fatigue warning device, with the aim to help prevent road traffic collisions and ultimately save lives.

# Appendix A

## Probabilistic transition net for automatic sleep staging

### Summary

Automatic sleep staging provides a cheaper, faster and more accessible alternative for evaluating sleep patterns and quality compared with manual hypnogram scoring performed by a clinician. Traditionally, classification methods treat sleep stages independently of their temporal order, despite sleep patterns themselves being highly sequential. Such independent sleep stage classification can result in poor sensitivity and precision, in particular when attempting to classify the sleep stage N1, otherwise known as the transition stage of sleep which links periods of wakefulness to periods of deep sleep. To this end, we propose a novel transition sleep classification method which aims to improve classification accuracy. This is achieved by utilising both the temporal information of previous stages and treating the transitions between stages as classes in their own right. Simulations on publicly available polysomnography (PSG) data and a comprehensive performance comparison with standard classifiers demonstrate a marked improvement achieved by the proposed method in both N1 sensitivity and precision across all considered classifiers. This includes an increase in N1 precision from 0.01% to 36.75% in an MLP classifier, and an increase in both accuracy and Cohen’s kappa value in two of the three classifiers. Overall best mean performance is obtained by transition classification with a random forest classifier (RF) which achieved a kappa value of  $\kappa = 0.75$  (substantial agreement), and an N1 stage precision of 58%.



### A.0.1 Introduction

With the key role of sleep in cardiovascular health [184] and consolidation of memory [185], poor sleep quality and lack of sleep place a huge burden on society in terms of both public health and economic costs [186]. In consumer health, the demand for sleep measurement is reflected in the increasing popularity of sleep tracking apps [187] which attempt to assess sleep quality and duration; however, these do not report accuracy that is considered clinically relevant [188]. In medical practice, sleep disorders, such as obstructive sleep apnoea, can be diagnosed after the patient is observed sleeping overnight at a sleep clinic whilst polysomnography (PSG) is recorded. The PSG includes the physiological modalities of electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), airflow and blood oxygen percentage, and the 30 second traces of such data are scored by a Clinician based on the American Academy of Sleep Medicine manual [65]. Each 30 second epoch of sleep is assigned a sleep stage of either wake (W), non rapid eye movement (N1, N2, N3) or rapid eye movement (REM) [189], to yield a scored hypnogram. Stages N2 and N3 represent deeper sleep which is characterised by low frequency EEG activity, whereas REM sleep is characterised by mixed frequency EEG and lower EMG amplitude [189]. Given that the clinician must manually score many hours of data for a single recording, this method is very time consuming and expensive and thus not practically scalable to very large numbers of patients or eHealth applications. To this end, automatic sleep staging provides a way of classifying sleep stages without the involvement of a clinician, by using supervised machine learning methods. Automatic sleep staging can only be effective if its accuracy is comparable to a clinician's scored hypnogram, and should ideally be orders of magnitude faster than human scoring.

Classic machine learning approaches, such as linear discriminant analysis (LDA), support vector machines (SVM) and artificial neural networks (ANN) have traditionally been used to tackle the problem of automatic sleep staging. Deep learning solutions also exist that achieve high accuracy [190] but are computationally complex and expensive to implement. The difficulty with all classification solutions is that they tend to suffer from poor N1 accuracy, as it occupies only 5% of overall sleep duration [191] and has the lowest rate of scoring agreement between different human sleep scorers [192]. However, the N1 stage is of particular interest to overall sleep health as a high proportion of N1 represents frequent arousals that are indicative of a sleep disorder. Furthermore, N1 is also heavily related to drowsiness [17]. *This is why N1 is*

referred to as the transition stage of sleep, given that it represents the transition between wake and sleep.

Another issue is that in conventional automatic sleep stage scoring, the 5 stages of sleep are classified independently, whereas when a clinician scores sleep stages manually they also rely on the knowledge of previous epochs and look for transitions between them. To this end, we propose a new automatic sleep stage classifier by considering the transitions between stages as additional independent classes in classification. These transition classifications are then used to improve N1 classification accuracy and precision, and show clear promise to increase overall sleep stage classification accuracy.

## A.1 Transition model of sleep

		<u>Current</u>				
		W	N1	N2	N3	R
<u>Previous</u>	W	86.24	8.11	4.67	0.15	0.83
	N1	8.91	57.30	28.09	0.01	5.69
	N2	2.89	0.01	92.84	3.33	0.94
	N3	1.30	0.00	6.39	92.27	0.04
	R	2.23	0.81	1.17	0.00	95.79

Figure A.1: Percentage of current states of sleep given the previous state, obtained from the manually scored hypnograms in the CCSHS data set [193] [194].

Traditionally, automatic sleep staging is performed in the same way as the majority of classification problems, by considering the different stages as independent of previous and future stages. In reality, all stages of sleep are highly dependent on the previous stage, as shown in Fig. A.1 which gives the percentage of current states for a given previous state as obtained from scored hypnogram data. From Fig. A.1 for example, the wake (W) stage frequently transitions into N1 and N2 but rarely transitions to N3 or REM, and REM is likely to transition to W or N2, but unlikely to transition to N3 or N1. It is also clear that the most likely transition scenario is remaining in the same stage. Given this interdependence between classes, we set out to investigate whether a transition model of sleep could be used to increase classification

accuracy.

### A.1.1 Previous modelling of sleep stage transitions

Previous work includes hidden Markov model (HMM) based smoothing which uses a transition matrix from the training data and an emission matrix from the classifier to apply the well known Viterbi HMM decoding algorithm and find the most probable sequence of events [195]. Implementation of HMM smoothing was shown to significantly increase accuracy and Cohen's kappa value, on average. However, in subjects for whom sleep was highly fragmented HMM smoothing tends to eliminate this fragmentation thus decreasing accuracy and kappa values. Further work involves the combination of smoothing and a most probable path calculation using Dijkstra's algorithm [196].

### A.1.2 Transition state features

To be able to effectively predict transitions between different sleep stages, the features during transitions must be sufficiently different from the features during the main stages. Fig. A.2 highlights the differences in the probability density of feature values between the W, N1, N2 stages and the transitions between them. As expected, the distribution for the transition between W and N1 falls between the distributions for W and N1, and the same is true of N1 and N2.

### A.1.3 Transition probability model

Given the importance of transitions between stages, it is natural to ask whether a classifier can also be used to predict the probability of transitioning from one state to another. In other words, a classifier can be trained to predict 25 different classes (all 5 sleep stages and the 20 transitions between them) rather than just 5. The classifier then gives transition probabilities and class probabilities, and the final stage prediction becomes a combination of both these sets of probabilities as follows:

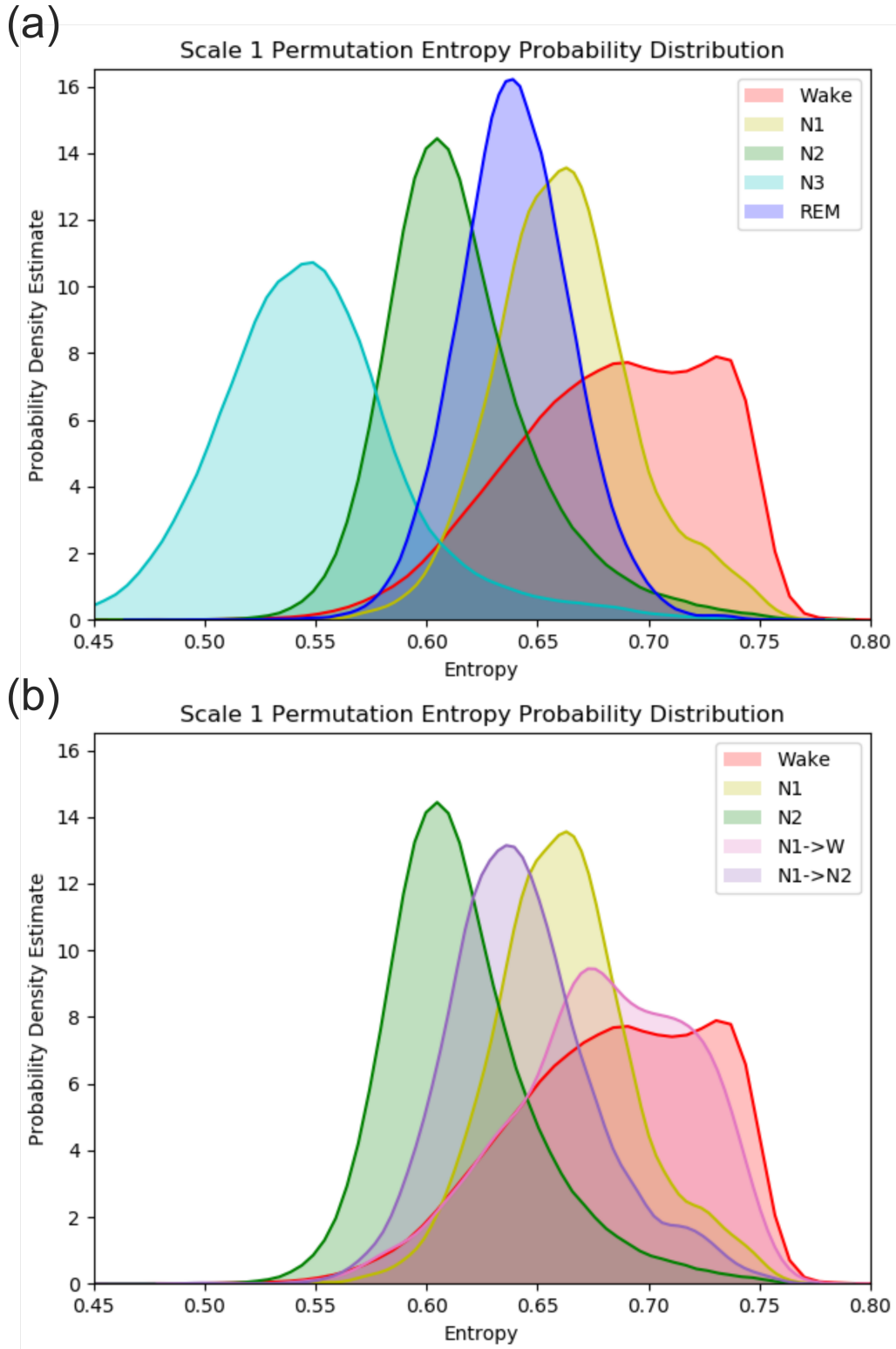


Figure A.2: Probability density estimates of different sleep stages for scale 1 permutation entropy. (a) The standard 5 class model, (b) Three of the standard classes (W, N1, N2) with the N1 to N2 and W to N1 transitions included.

$$P(C_i|C_{i-1}) = P(C_{i-1}C_i) \times P(C_i) \quad (\text{A.1})$$

where  $C_i$  represents one of the 5 sleep stages,  $C_{i-1}$  the previously classified sleep stage,  $C_{i-1}C_i$  the transition from the previous class to class  $i$ , and  $P$  the probability of the class occurring in relation to other classes. This calculation is repeated for all 5 classes and the class with the maximum probability is chosen.

#### A.1.4 Pruned transition probability model

Although the proposed model at first appears rather complex, Fig. A.1 shows that certain transitions such as N3 → N1 are extremely unlikely, and can be ignored by the model. This leaves us with a much simpler but accurate pruned transition model. Mathematically, if a connection is not present between  $C_{i-1}$  and  $C_i$  then  $P(C_{i-1}C_i) = 0$ , as shown in Fig. A.3.

		Current				
		W	N1	N2	N3	REM
Previous	W	P	P	P	X	X
	N1	P	P	P	X	P
	N2	P	X	P	P	P
	N3	P	X	P	P	X
	REM	P	P	P	X	P

Figure A.3: Pruning table, showing connections that are included in the pruned model (P) and connections that have been removed (X).

## A.2 Implementation

### A.2.1 Preprocessing and feature extraction

We used data from the Cleveland Children's Sleep and Health study [193] [194]. Bandpass filtering was implemented between [0.5]Hz and [30]Hz, and a total of 82 features were formed from the multi-scale entropy calculations and spectral edge frequencies, according to [197]. Two subjects were also removed due to a lack of REM sleep, giving a total of 513 subjects.

		Current				
		W	N1	N2	N3	REM
Previous	W	0	1	2	13	17
	N1	3	4	5	13	6
	N2	7	4	8	9	10
	N3	11	4	12	13	17
	REM	14	15	16	13	17

Figure A.4: The relabelling protocol for the data, showing the transitions that have been pruned in grey.

### A.2.2 Classifiers

To implement the transition model, training data was first relabeled to include the 5 classes of sleep and the 20 transitions between these classes. Out of 20 possible transitions, 7 were not considered as independent classes as these were very rare. This meant that many transitions, according to Fig. A.4, were simply given the same label as the current class. Different classifiers from the freely available Sklearn Python toolbox [99] were then used to classify these 18 classes, namely random forest (RF), linear discriminant analysis (LDA), and multi-layer perceptron (MLP). The hyper-parameters were set as follows: RF had 500 estimators with a max depth of 500 and a balanced class weight, LDA had the default options of a singular value decomposition solver and a tolerance of  $1 \times 10^{-4}$  and the MLP used a limited memory Broyde-Fletcher-Goldfarb-Shannon (L-BFGS) solver with a regularization parameter of  $1 \times 10^{-5}$  and hidden layer sizes of 50, 50.

### A.2.3 Transition probability classification

The first epoch, usually the wake stage, W, was classified by using the maximum probability out of W, N1, N2, N3 and REM. In the pruned transition model, these correspond to the probabilities of classes 0, 4, 8, 13 and 17 which serve as good approximations to the 5-class model, given that once in a class the most likely scenario by far is to remain in that class (Fig. A.1). For the next epoch, W now becomes the previous state. Therefore, in the pruned model there are now 3 stages to consider, W, N1 and N2. The probability of transitioning from W to N1 is given by the probability of Class 1, the probability of transitioning from W to N2

is given by Class 2 and the probability of transitioning from Class W to itself is simply the previous probability of Class 0 multiplied by the current probability of Class 0, given that it is the probability of being in Wake multiplied by the probability of being in Wake again. These probabilities are then multiplied again by the current Class Probabilities 4, 8 and 0, that is

$$P(N1_i|W_{i-1}) = P(W_{i-1}N1_i) \times P(N1_i) \quad (\text{A.2})$$

$$P(N2_i|W_{i-1}) = P(W_{i-1}N2_i) \times P(N2_i) \quad (\text{A.3})$$

$$\begin{aligned} P(W_i|W_{i-1}) &= P(W_{i-1}W_i) \times P(W_i) \\ &\approx P(W_{i-1}) \times P(W_i)^2 \end{aligned}$$

For each subsequent class, these calculations are performed again for each relevant class according to the transitions included in the pruned model (Fig. A.3).

Each method was evaluated using 5 fold cross validation to test that the results were consistent across the whole data set. The overview of this proposed sleep classification implementation is given in Fig. A.5.

### A.3 Results

Measurement	RF		LDA		MLP	
	R	T	R	T	R	T
Accuracy(%)	82.60	<b>83.39</b>	<b>79.20</b>	78.46	75.55	<b>78.75</b>
Kappa	0.744	<b>0.753</b>	<b>0.697</b>	0.688	0.645	<b>0.688</b>
N1 Sensitivity (%)	7.41	<b>20.11</b>	7.23	<b>31.31</b>	0.06	<b>11.21</b>
N1 Precision (%)	38.66	<b>58.11</b>	27.30	<b>27.91</b>	0.01	<b>36.75</b>
N1 F1 Score	0.111	<b>0.283</b>	0.099	<b>0.268</b>	0.001	<b>0.160</b>

Table A.1: Comparison of performances of the standard (R) and transition classification (T) with different classifiers.

Table. A.1 shows that random forest (RF) and multi-layer perceptron (MLP) both exhibit an increase in overall performance when equipped with transition classification, whereas linear discriminant analysis (LDA) shows a slight decrease in general performance. The MLP setup

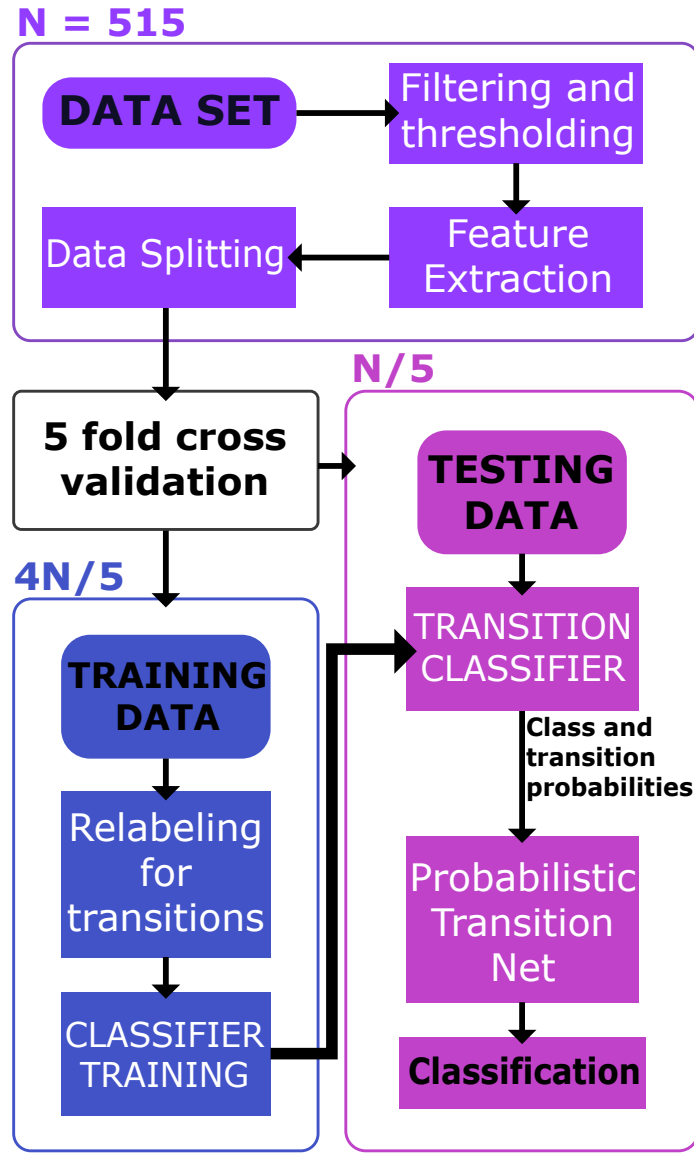


Figure A.5: Flow diagram of the proposed transition probability implementation.

shows the greatest increase in general performance, with an average increase in accuracy of 3.2% and an average increase in kappa of 0.043.

In summary, for all classifiers we observed an increase in N1 accuracy, N1 precision and N1 F1 score with the proposed transition classification. Precision is also important as it indicates that a classifier is better at recognising a particular class, rather than just blindly increasing the frequency at which it assigns a particular class label. For MLP, the increase in N1 precision was 36.74% and for RF we had an increase of 19.45%, while LDA gave a slight increase in N1 precision of 0.61%.

Although transition classification significantly boosts N1 sensitivity across all classifiers, the



A.4.

clinician

N3	0.45	0.00	15.80	83.72	0.03	N3	0.41	0.01	16.59	82.97	0.02
R	3.63	0.73	14.92	0.08	80.65	R	3.02	1.34	17.35	0.04	78.25

141

classifier prediction: RF

	W	N1	N2	N3	R
W	87.32	0.87	7.59	0.48	3.73
N1	22.92	7.41	19.20	0.06	50.40
N2	2.28	0.24	87.65	3.43	6.40
N3	0.43	0.00	15.83	83.71	0.03
R	3.67	0.76	14.95	0.08	80.55

classifier prediction: RF(T)

	W	N1	N2	N3	R
W	88.18	1.10	8.00	0.59	2.13
N1	32.01	20.11	27.08	0.05	20.74
N2	2.18	0.33	89.44	2.99	5.07
N3	0.52	0.00	16.65	82.79	0.03
R	3.12	1.37	17.29	0.03	78.19

classifier prediction: LDA

	W	N1	N2	N3	R
W	81.61	2.36	7.23	1.22	7.57
N1	21.22	7.23	13.05	0.21	58.28
N2	2.70	0.40	83.43	3.34	10.13
N3	0.47	0.01	20.69	77.97	0.86
R	4.08	1.22	11.71	0.21	82.77

classifier prediction: LDA(T)

	W	N1	N2	N3	R
W	67.93	17.02	9.22	0.87	4.97
N1	15.95	31.31	21.86	0.10	30.78
N2	2.35	0.87	84.16	2.98	9.64
N3	0.48	0.01	21.04	77.50	0.96
R	2.89	2.69	10.92	0.11	83.39

classifier prediction: MLP

	W	N1	N2	N3	R
W	77.20	0.08	10.99	1.66	10.07
N1	27.10	0.06	15.31	0.44	57.08
N2	3.66	0.01	79.11	8.85	8.37
N3	0.82	0.00	22.80	75.84	0.54
R	5.99	0.03	13.36	0.82	79.81

classifier prediction: MLP(T)

	W	N1	N2	N3	R
W	83.43	1.07	10.14	0.89	4.46
N1	48.53	11.21	18.62	0.15	21.48
N2	4.02	0.33	85.04	4.73	5.88
N3	0.70	0.00	24.48	74.61	0.21
R	7.32	2.06	13.00	0.23	77.40

Figure A.6: Average cross validated confusion matrix results for all classifiers. Left: Traditional 5 stage classification results, right: Transition classification results.

effectiveness is highly dependent on the choice of classifier. In the case of LDA, it may be that it lacks the sufficient degrees of freedom to effectively distinguish between a state of wake and transitions from wake, such as wake to N1. This theory is reinforced when the LDA confusion matrices are examined in Fig. A.6, as they show that the increase in N1 classification accuracy is accompanied with a significant decrease in wake classification accuracy.

## A.4 Chapter conclusions

By equipping sleep stage classifiers with the combination of knowledge about a previous stage and the prediction of transitions between sleep stages, we have been able to achieve increases

in overall classification accuracy as well as vast improvements in N1 stage sensitivity, F1 score and precision. The proposed method is very intuitive and can thus be adopted into many other state-of-the-art classification methods, to further improve automatic sleep staging performance.

# Appendix B

## Example ethics application

This appendix details the relevant information from our general Hearables ethics application, approved by the Imperial College Research Ethics Committee. This application was designed to encompass general physiological recordings for some of the studies on in-ear PPG and SpO<sub>2</sub> presented in this thesis, as well as studies pertaining to wearable ECG and EEG recordings.

### B.1 Project summary

Electrocardiography (ECG) and electroencephalography (EEG) are the respective non-invasive recordings of the electrical activity from the heart and the brain and are usually performed by placing electrodes on a participant's arms or chest (for ECG), or head (for EEG). Mechanoplethysmography (MPG) measures pulse non-invasively with a mechanical transducer (microphone) that detects the pulsation of blood vessels, and photoplethysmography (PPG) measures pulse non-invasively by emitting light through the skin. Blood Pressure (BP) measures the pressure of circulating blood on blood vessels; a digital blood pressure monitor will be used to measure blood pressure, or alternatively BP will be calculated digitally through the Pulse Arrival Time (PAT), a delay in peaks between the ECG and MPG/PPG. A respiration belt will be used to measure non-invasively the thoracic, abdominal respiratory movements and respiration rate. A thermometer will be used to measure skin surface temperature and an accelerometer will be used to monitor head motion. Gaze behavioural recording by gaze tracker requires no attachment on the eyes of subjects that monitors the gaze trajectories and duration. This project will

investigate the use of vital signs (heart signals, respiration, BP, temperature) and brain signals to detect cognitive parameters, such as drowsiness (i.e. sleepiness), mental workload/stress, emotion arousal, variation of concentration over time and physical fitness. The designed experiments would detect the decrements in cognitive function, arising due to sustained mental work in a controlled laboratory setting. Importantly, we are comparing the ability of wearable sensors (using the ear as our measurement site) with the current gold standard sensors such as scalp EEG, chest ECG, and finger PPG, which are all obtrusive to daily life. It is our hope that the wearable signals can give us enough information to make the assessments of cognitive load, mild stress and fatigue, and therefore medicine can begin to move away from cumbersome obtrusive physiological monitoring methods towards less obtrusive wearable solutions.

Markers for cognitive parameters will be investigated using the obtained data. Performance parameters, such as response time and error rates, during a mathematical, reading or video game task will be used to draw inferences about the participant's cognitive state. In the Brain Computer Interface (BCI) research, responses to audio, video and audiovisual stimuli shall be investigated, and optimal electrode positions will be determined for extracting relevant electrophysiological activity. The optimal parameters for the stimuli which facilitate a high BCI rate for the participants will also be considered. Traditional and novel mathematical techniques will be applied to the recorded ECG and EEG data for analysis, e.g. blind source separation for the removal of eye movement artefacts from the EEG.

Recent reports and studies have indicated that the prevalence of excessive mental workload and stress has been increasing; drowsiness and poor concentration have also been cited as growing problems in high pressured occupations. Brain computer interfacing enables biofeedback to indicate cognitive overload. Hence the potential social outcomes of this project are huge. The findings from this project will be presented at conferences and published in journals, contributing to the literature investigating the manifestation of psycho-physiological states in ECG and EEG. The findings from this project will also have a wide impact in high pressured sectors such as healthcare, teaching, the military and motor-car racing by enabling the measurement and identification of excessive cognitive load.

## B.2 Participant recruitment

Participants for this project will be recruited from the student population of Imperial College through the dissemination of participant information sheets explaining the investigation. Co-investigators will be disseminating the information sheets in person and via email.

The participant inclusion criterion is as follows:

- Participants able to provide informed consent for study participation.
- Participants aged 18 years and older.

The participant exclusion criterion is as follows:

- Participants with epilepsy.
- Participants with dyscalculia: Dyscalculia is a specific and persistent difficulty in understanding numbers which can lead to a diverse range of difficulties with mathematics.
- Participants with a history of sensitive skin.

## B.3 Informed consent

1. All participants will receive an information sheet prior to the commencement of the experiment, outlining the aims of the investigation and the data collection procedure. The experiment will then be further explained to the participant verbally, with the opportunity for them to ask questions, before being asked to sign the participant consent form.
2. It is explained on the participant information sheet that participants can withdraw from the study at any time, and their recorded data will be removed.
3. Vulnerable persons will not be recruited in this study.
4. Participants for this study will be recruited from the student population of Imperial College, and it is expected that all participants will have sufficient comprehension of the English language. However, if it is clear to the investigating researchers that a participant

does not understand the information provided, the participant will be discharged from the study.

## B.4 Ethical Summary

To the best of the PIs' knowledge the ethical issues associated with the project are minimal as it involves healthy participants and it will use either CE approved equipment or low voltage commercial sensor chips for the purposes they have been designed for – ECG, EEG, MPG, PPG, BP and respiration monitoring systems are harmless and are already commercially available for personal use (lifestyle health apps and gaming), e.g. EMOTIV Epoc, <http://www.emotiv.com/>, or g.tec, [www.gtec.at](http://www.gtec.at). Determining cognitive parameters like overload and brain computer interfacing do not involve any ethical issues, and none of the sensors/hearables are medical devices.

All devices/sensors used are commercially available and are used for their intended purpose, that is, sensing/recording from human skin. The earplugs we use are also either commercial viscoelastic earplugs or are made of a material which is intended to be used for earplugs or sensing from human skin. Therefore, the proposed Hearables in-ear sensors combine these earplugs and sensors, while maintaining their intended use/operation.

The study will involve a short arithmetic test, a light physical step exercise, facial expressions, head movements, basic daily life activities (speaking, chewing, walking) and video simulations, so no participants with a history of dyscalculia, cardiac disorders, or epilepsy will be recruited. Similarly recording EEG requires a conductive gel to be applied to the skin, so to minimize the risk of skin irritation, no participants with a history of sensitive skin will be recruited.

## B.5 Mitigation of ethical issues

### B.5.1 Collection of sensitive information

The aim of this study is to record vital signs and brain signals from participants to assess the manifestation of cognitive load and stress in the recorded signals.

To mitigate the ethical issues related to the collection of sensitive information, all collected data will be pseudo-anonymised, such that the participants' data, and resulting findings will not be identifiable. In addition, participants can ask for their data to be withdrawn from the study up until the collected data is pseudo-anonymised. If a participant requests to withdraw from the study, their data will be removed from the study dataset.

### **B.5.2 Researcher in position of authority**

The participants in this study will be recruited from the student population of Imperial College; it is therefore highly probable that the participants may be undergraduate or postgraduate students under the supervision or guidance of the researcher.

To minimise the pressure applied to the participant, care will be taken to ensure that the Principal Investigator is not present during the recruitment and conduction of the experiments.

### **B.5.3 The potential to induce stress**

The aim of this study is to assess the manifestation of stress in heart and brain signals, and stress will be induced through the use of mental arithmetic tests, step-exercises, and temperature variation (e.g. the cold-stressor test, where appropriate).

However, to ensure that the participants are not excessively stressed, the level of difficulty of the mental arithmetic test will be matched to the participant's ability, which will be estimated from the participants course of study. The step exercise will also be low-intensity (walking pace), while the cold stressor test includes putting a hand into a bucket of cold water.

### **B.5.4 Use of non-CE approved electronics**

The custom electronics uses commercial sensors that are all low voltage (3.3V maximum) and the systems have been designed in accordance with best practice and recommended circuits from the sensor datasheets.

### **B.5.5 Incidental findings**

It is possible that abnormal heart activity or oxygen levels could be discovered during our recordings, but this is not a medical investigation and thus we will not be exploring the subject's health. If any abnormality is found, the subject will be informed and then has the option to request medical tests with their general practitioner. Care will be taken not to alarm or upset the participant if a case such as this arises, as abnormal heart activity, such as arrhythmia, is very common (roughly 5% of the population) and benign in the vast majority of cases.

## **B.6 Mitigation of risks**

1. Discomfort and skin irritability may result from the use of conductive gels or plasters for participants with sensitive skin. To minimize this possibility, participants with sensitive skin will not be recruited for the study. In addition, some modern EEG systems use salty water or dry electrodes which do not irritate skin.
2. The study is non-invasive with low risk. Participants will be allowed to withdraw from the study at any time should they feel uncomfortable. Participants with a history of epilepsy will not be recruited in this study, as the experiments that present a video game or simulation with flashing visual stimulus to the participant could pose a danger to epileptics.

## **B.7 Confidentiality and Management of personal and other research data**

1. Personal data will be collected from all recruited participants.
2. Participants' ID, age, sex, and recorded signals will be collected. The key linking the subject ID and participant name will be maintained by a named researcher.
3. Personal data will not be shared outside the research team.
4. Data will be pseudo-anonymised such that only the assigned participant ID, age and sex of the participant is known to the researchers.



# References

- [1] S. Gehr and C. Russmann, “Shaping the future of cardiovascular medicine in the new era of wearable devices,” *Nature Reviews Cardiology*, vol. 19, no. 8, pp. 501–502, 2022.
- [2] A. H. Anwer, N. Khan, M. Z. Ansari, S.-S. Baek, H. Yi, S. Kim, S. M. Noh, and C. Jeong, “Recent advances in touch sensors for flexible wearable devices,” *Sensors*, vol. 22, no. 12, 2022.
- [3] S. Vogel, M. Hulsbusch, D. Starke, and S. Leonhardt, “In-ear heart rate monitoring using a micro-optic reflective sensor,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 1375–1378.
- [4] A. B. Hertzman, “The blood supply of various skin areas as estimated by the photoelectric plethysmography,” *American Journal of Physiology*, vol. 124, no. 2, pp. 328–340, Oct 1938.
- [5] T. Aoyagi, “Pulse oximetry: Its origin and development,” *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 7, pp. 2858–2859, 1992.
- [6] B. Venema, J. Schiefer, V. Blazek, N. Blanik, and S. Leonhardt, “Evaluating Innovative In-Ear Pulse Oximetry for Unobtrusive Cardiovascular and Pulmonary Monitoring During Sleep,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 1, pp. 2 700 208–2 700 208, Aug 2013.
- [7] K. Budidha and P. A. Kyriacou, “The human ear canal: investigation of its suitability for monitoring photoplethysmographs and arterial oxygen saturation,” *Physiological Measurement*, vol. 35, no. 2, pp. 111–128, Jan 2014. [Online]. Available: <https://doi.org/10.1088/0967-3334/35/2/111>

- [8] —, “In vivo investigation of ear canal pulse oximetry during hypothermia,” *Journal of Clinical Monitoring and Computing*, vol. 32, no. 1, pp. 97–107, Feb 2018.
- [9] N. Bui, N. Pham, J. J. Barnitz, Z. Zou, P. Nguyen, H. Truong, T. Kim, N. Farrow, A. Nguyen, J. Xiao, R. Deterding, T. Dinh, and T. Vu, “Ebp: A wearable system for frequent and comfortable blood pressure monitoring from user’s ear,” in *The 25th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’19. New York, NY, USA: Association for Computing Machinery, 2019.
- [10] J. B. West, “The birth of clinical body plethysmography: it was a good week,” *The Journal of Clinical Investigation*, vol. 114, no. 8, pp. 1043–1045, Oct 2004.
- [11] L. Nilsson, T. Goscinski, S. Kalman, L.-G. Lindberg, and A. Johansson, “Combined photoplethysmographic monitoring of respiration rate and pulse: a comparison between different measurement sites in spontaneously breathing subjects,” *Acta Anaesthesiologica Scandinavica*, vol. 51, no. 9, pp. 1250–1257, Aug 2007.
- [12] H. J. Davies, P. Bachtiger, I. Williams, P. L. Molyneaux, N. S. Peters, and D. Mandic, “Wearable in-ear ppg: Detailed respiratory variations enable classification of copd,” *IEEE Transactions on Biomedical Engineering*, vol. In press, 2022.
- [13] Y. Mendelson and C. Pujary, “Measurement Site and Photodetector Size Considerations in Optimizing Power Consumption of a Wearable Reflectance Pulse Oximeter,” *In Proceedings of the 2003 Annual International Conference of the IEEE Engineering in Medicine and Biology*, vol. 4, pp. 3016–3019, 2003.
- [14] H. Kinnunen, A. Rantanen, T. Kentt, and H. Koskim ki, “Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG,” *Physiological Measurement*, vol. 41, no. 4, May 2020.
- [15] D. Looney *et al.*, “An in-the-ear platform for recording electroencephalogram,” in *Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 6882–6885.

- [16] V. Goverdovsky, W. Von Rosenberg, T. Nakamura, D. Looney, D. J. Sharp, C. Papavasiliou, M. J. Morrell, and D. P. Mandic, “Hearables: Multimodal physiological in-ear sensing,” *Scientific Reports*, vol. 7, no. 1, pp. 1–10, Dec 2017.
- [17] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, “Automatic detection of drowsiness using in-ear EEG,” in *In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Jul 2018, pp. 1–6.
- [18] —, “Hearables: Automatic Overnight Sleep Monitoring with Standardized In-Ear EEG Sensor,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 203–212, Jan 2020.
- [19] T. Nakamura, V. Goverdovsky, and D. P. Mandic, “In-ear EEG biometrics for feasible and readily collectable real-world person authentication,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 648–661, March 2018.
- [20] P. Kidmose, D. Looney, M. Ungstrup, M. L. Rank, and D. P. Mandic, “A Study of Evoked Potentials From Ear-EEG,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2824–2830, Oct 2013.
- [21] W. von Rosenberg, T. Chanwimalueang, T. Adjei, U. Jaffer, V. Goverdovsky, and D. P. Mandic, “Resolving Ambiguities in the LF/HF Ratio: LF-HF Scatter Plots for the Categorization of Mental and Physical Stress from HRV,” *Frontiers in Physiology*, vol. 8, p. 360, Jun 2017.
- [22] M. C. Yarici, H. J. Davies, T. Nakamura, I. Williams, and D. P. Mandic, “Hearables: In-ear multimodal brain computer interfacing,” in *Brain-Computer Interface Research: A State-of-the-Art Summary 9*, C. Guger, B. Z. Allison, and M. Tangermann, Eds. Springer International Publishing, 2021, pp. 79–87.
- [23] *Pulse Oximeters - Premarket Notification Submissions [510(k)s]: Guidance for Industry and Food and Drug Administration Staff*, 2018. [Online]. Available: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug>

- [24] J. R. Feiner, J. W. Severinghaus, and P. E. Bickler, “Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender,” *Anesthesia & Analgesia*, vol. 105, no. 6, pp. S18–S23, 2007.
- [25] D. Looney, A. Hemakom, and D. P. Mandic, “Intrinsic multi-scale analysis: A multivariate empirical mode decomposition framework,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2173, Jan 2015.
- [26] T. M. Rutkowski, J. Dauwels, F. Vialatte, A. Cichocki, and D. P. Mandic, “Time-frequency and synchrony analysis of responses to steady-state auditory and musical stimuli from multichannel EEG.”
- [27] D. Looney, C. Park, Y. Xia, P. Kidmose, M. Ungstrup, and D. P. Mandic, “Towards estimating selective auditory attention from EEG using a novel time-frequency-synchronisation framework,” in *Proceedings of the International Joint Conference on Neural Networks*, 2010, pp. 1–5.
- [28] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, “Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain,” *BioMed Research International*, 2017.
- [29] N. Rehman and D. P. Mandic, “Multivariate empirical mode decomposition,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 466, no. 2117, pp. 1291–1302, May 2010.
- [30] H. Dai and W. Wang, “Application of low-discrepancy sampling method in structural reliability analysis,” *Structural Safety*, vol. 31, no. 1, pp. 55–64, 2009.
- [31] N. Ur Rehman, C. Park, N. E. Huang, and D. P. Mandic, “EMD via MEMD: Multivariate noise-aided computation of standard EMD,” *Advances in Adaptive Data Analysis*, vol. 05, no. 02, p. 1350007, Apr 2013.
- [32] Y. Lv, R. Yuan, and G. Song, “Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing,” *Mechanical Systems and Signal Processing*, vol. 81, pp. 219–234, Dec 2016.

- [33] W. Hu and B. C. Si, “Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition,” *Geoderma*, vol. 193-194, pp. 180–188, Feb 2013.
- [34] F.-T. Wang, H.-L. Chan, C.-L. Wang, H.-M. Jian, and S.-H. Lin, “Instantaneous respiratory estimation from thoracic impedance by empirical mode decomposition,” *Sensors*, vol. 15, no. 7, pp. 16 372–16 387, 2015.
- [35] S. Qian and D. Chen, “Joint time-frequency analysis,” *IEEE Signal Processing Magazine*, vol. 16, no. 2, pp. 52–67, 1999.
- [36] F. Hlawatsch and G. Boudreaux-Bartels, “Linear and quadratic time-frequency signal representations,” *IEEE Signal Processing Magazine*, vol. 9, no. 2, pp. 21–67, 1992.
- [37] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, 1st ed. Routledge, 1984.
- [38] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001.
- [39] J. Zhu, H. Zou, S. Rosset, and T. Hastie, “Multi-class AdaBoost,” *Statistics and Its Interface*, vol. 2, pp. 349–360, 2009.
- [40] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug 1997.
- [41] J. Thongkam, G. Xu, and Y. Zhang, “AdaBoost algorithm with random forests for predicting breast cancer survivability,” *Proceedings of the International Joint Conference on Neural Networks*, pp. 3062–3069, 2008.
- [42] H. J. Davies, T. Nakamura, and D. P. Mandic, “A transition probability based classification model for enhanced n1 sleep stage identification during automatic sleep stage scoring,” in *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 3641–3644.
- [43] T. Nakamura, H. J. Davies, and D. P. Mandic, “Scalable automatic sleep staging in the era of big data,” in *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2265–2268.

- [44] H. J. Davies *et al.*, “In-Ear SpO<sub>2</sub> for Classification of Cognitive Workload,” *arXiv preprint arXiv:2101.00647*, 2021.
- [45] A. Jubran, “Pulse oximetry,” *Critical Care*, vol. 3, no. 2, p. R11, May 1999.
- [46] G. B. Smith, D. R. Prytherch, D. Watson, V. Forde, A. Windsor, P. E. Schmidt, P. I. Featherstone, B. Higgins, and P. Meredith, “SpO<sub>2</sub> values in acute medical admissions breathing air-Implications for the British Thoracic Society guideline for emergency oxygen use in adult patients?” *Resuscitation*, vol. 83, no. 10, pp. 1201–1205, Oct 2012.
- [47] World Health Organization, *Pulse Oximetry Training Manual*, 2011.
- [48] L. Gattinoni, D. Chiumello, P. Caironi, M. Busana, F. Romitti, L. Brazzi, and L. Camporota, “COVID-19 pneumonia: Different respiratory treatments for different phenotypes?” *Intensive Care Med*, Apr 2020.
- [49] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, *Features, Evaluation and Treatment Coronavirus (COVID-19)*. StatPearls Publishing, Apr 2020.
- [50] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. L. Hesketh, L. Yang, and C. Zheng, “Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia,” *Radiology*, p. 200370, Feb 2020.
- [51] W. Alhazzani, M. H. Møller, Y. M. Arabi, M. Loeb, M. N. Gong *et al.*, “Surviving Sepsis Campaign: guidelines on the management of critically ill adults with Coronavirus Disease 2019 (COVID-19),” 2020.
- [52] R. G. Khemani, N. J. Thomas, V. Venkatachalam, J. P. Scimeme, T. Berutti, J. B. Schneider, P. A. Ross, D. F. Willson, M. W. Hall, and C. J. L. Newth, “Comparison of SpO<sub>2</sub> to PaO<sub>2</sub> based markers of lung disease severity for children with acute lung injury,” *Critical Care Medicine*, vol. 40, no. 4, pp. 1309–1316, Apr 2012.
- [53] A. Oksenberg, E. Arons, K. Nasser, T. Vander, and H. Radwan, “REM-related Obstructive Sleep Apnea: The Effect of Body Position,” *J Clin Sleep Med*, vol. 6, no. 4, pp. 343–348, Aug 2010.
- [54] M. Nitzan, A. Romem, and R. Koppel, “Pulse oximetry: Fundamentals and technology update,” pp. 231–239, Jul 2014.

- [55] B. Bradke and B. Everman, “Investigation of Photoplethysmography Behind the Ear for Pulse Oximetry in Hypoxic Conditions with a Novel Device (SPYDR),” *Biosensors*, vol. 10, no. 4, p. 34, Apr 2020.
- [56] R. Drescher, “Wearable Forehead Pulse Oximetry: Minimization of Motion and Pressure Artifacts,” *Masters Theses (All Theses, All Years)*, May 2006.
- [57] D. Looney, P. Kidmose, C. Park, M. Ungstrup, M. Rank, K. Rosenkranz, and D. Mandic, “The in-the-ear recording concept: User-centered and wearable brain monitoring,” *IEEE Pulse*, vol. 3, no. 6, pp. 32–42, 2012.
- [58] K. Budidha and P. A. Kyriacou, “Investigation of photoplethysmography and arterial blood oxygen saturation from the ear-canal and the finger under conditions of artificially induced hypothermia,” in *Proceedings of the IEEE Annual International Conference of the Engineering in Medicine and Biology Society, EMBS*, Nov 2015, pp. 7954–7957.
- [59] E. A. Hamber, P. L. Bailey, S. W. James, D. T. Wells, J. K. Lu, and N. L. Pace, “Delays in the detection of hypoxemia due to site of pulse oximetry probe placement,” *Journal of Clinical Anesthesia*, vol. 11, no. 2, pp. 113–118, Mar 1999.
- [60] T. L. Rusch, R. Sankar, and J. E. Scharf, “Signal processing methods for pulse oximetry,” *Computers in Biology and Medicine*, vol. 26, no. 2, pp. 143–159, 1996.
- [61] MaximIntegrated, “Recommended Configurations and Operating Profiles for MAX30101/MAX30102 EV Kits,” 2018.
- [62] G. Basaranoglu, M. Bakan, T. Umutoglu, S. U. Zengin, K. Idin, and Z. Salihoglu, “Comparison of SpO<sub>2</sub> values from different fingers of the hands,” *SpringerPlus*, vol. 4, no. 1, Dec 2015.
- [63] H. Baquero, R. Alviz, A. Castillo, F. Neira, and A. Sola, “Avoiding hyperoxemia during neonatal resuscitation: time to response of different SpO<sub>2</sub> monitors,” *Acta Paediatrica*, vol. 100, no. 4, pp. 515–518, Apr 2011.
- [64] E. P. Morozoff and J. A. Smyth, “Evaluation of three automatic oxygen therapy control algorithms on ventilated low birth weight neonates,” in *Proceedings of the 31st IEEE Annual International Conference of the Engineering in Medicine and Biology Society*, 2009, pp. 3079–3082.

- [65] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, R. M. Lloyd, S. F. Quan, M. M. Troester, and B. V. Vaughn, “AASM — Scoring Manual Version 2.4 The AASM Manual for the Scoring of Sleep and Associated Events RULES, TERMINOLOGY AND TECHNICAL SPECIFICATIONS VERSION 2.4,” 2017.
- [66] P. D. Mannheim, “The Light–Tissue Interaction of Pulse Oximetry,” *Anesthesia & Analgesia*, vol. 105, no. On Line Suppl., pp. S10–S17, Dec 2007.
- [67] L. S. Alvord and B. L. Farmer’, “Anatomy and Orientation of the Human External Ear,” *J Am Acad Audiol*, vol. 8, pp. 383–390, 1997.
- [68] R. Kelly, C. Hayward, A. Avolio, and M. O’Rourke, “Noninvasive determination of age-related changes in the human arterial pulse.” *Circulation*, vol. 80, no. 6, pp. 1652–1659, Dec 1989.
- [69] P. L. Enright, “The Six-Minute Walk Test,” *Respiratory Care*, vol. 48, no. 8, 2003.
- [70] G. A. Borg, “Psychophysical bases of perceived exertion,” *Medicine & Science in Sports & Exercise*, vol. 14, no. 5, pp. 377–381, 1982.
- [71] M. M. Roberts, J. G. Cho, J. S. Sandoz, and J. R. Wheatley, “Oxygen desaturation and adverse events during 6-min walk testing in patients with COPD,” *Respirology*, vol. 20, no. 3, pp. 419–425, Apr 2015.
- [72] H. J. Davies, I. Williams, N. S. Peters, and D. P. Mandic, “In-Ear SpO<sub>2</sub>: A Tool for Wearable, Unobtrusive Monitoring of Core Blood Oxygen Saturation,” *Sensors*, vol. 20, no. 17, p. 4879, Aug 2020.
- [73] K. Antczak, “Deep Recurrent Neural Networks for ECG Signal Denoising,” *arXiv:1807.11551*, 2019.
- [74] H. Ayaz, B. Willems, B. Bunce, P. Shewokis, K. Izzetoglu, S. Hah, A. Deshmukh, and B. Onaral, “Cognitive workload assessment of air traffic controllers using optical brain imaging sensors,” *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations*, pp. 21–31, 2010.
- [75] M. Hajinoroozi, Z. Mao, T. P. Jung, C. T. Lin, and Y. Huang, “EEG-based prediction of driver’s cognitive performance by deep convolutional neural network,” *Signal Processing: Image Communication*, vol. 47, pp. 549–555, Sep 2016.



- [76] J. K. Johannesen, J. Bi, R. Jiang, J. G. Kenney, and C.-M. A. Chen, “Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults,” *Neuropsychiatric Electrophysiology*, vol. 2, no. 1, p. 3, Dec 2016.
- [77] R. Wang, P. V. Amadori, and Y. Demiris, “Real-time workload classification during driving using HyperNetworks,” *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, pp. 3060–3065, Dec 2018.
- [78] P. Vito Amadori, T. Fischer, R. Wang, and Y. Demiris, “Decision anticipation for driving assistance systems,” in *In Proc of the IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2020.
- [79] A. Tjolleng, K. Jung, W. Hong, W. Lee, B. Lee, H. You, J. Son, and S. Park, “Classification of a driver’s cognitive workload levels using artificial neural network on ECG signals,” *Applied Ergonomics*, vol. 59, pp. 326–332, Mar 2017.
- [80] C. Wang and J. Guo, “A data-driven framework for learners’ cognitive load detection using ECG-PPG physiological feature fusion and XGBoost classification,” in *Procedia Computer Science*, vol. 147. Elsevier B.V., Jan 2019, pp. 338–348.
- [81] J. Zhang, S. Li, and R. Wang, “Pattern recognition of momentary mental workload based on multi-channel electrophysiological data and ensemble convolutional neural networks,” *Frontiers in Neuroscience*, vol. 11, p. 310, May 2017.
- [82] T. Chanwimalueang, L. Aufegger, T. Adjei, D. Wasley, C. Cruder, D. P. Mandic, and A. Williamon, “Stage call: Cardiovascular reactivity to audition stress in musicians,” *PLOS ONE*, vol. 12, no. 4, p. e0176023, Apr 2017.
- [83] J. A. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, Jun 2005.
- [84] Z. A. Yaple, W. D. Stevens, and M. Arsalidou, “Meta-analyses of the n-back working memory task: fmri evidence of age-related changes in prefrontal cortex involvement across the adult lifespan,” *NeuroImage*, vol. 196, pp. 16–31, 2019.
- [85] F. Putze, J.-P. Jarvis, and T. Schultz, “Multimodal recognition of cognitive workload for multitasking in the car,” 2010, pp. 3748–3751.

- [86] B. Rebsamen, K. Kwok, and T. B. Penney, "Evaluation of cognitive workload from eeg during a mental arithmetic task," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 55, pp. 1342–1345, 2011.
- [87] J. M. Tomasko, E. M. Pauli, A. R. Kunselman, and R. S. Haluck, "Sleep deprivation increases cognitive workload during simulated surgical tasks," *The American Journal of Surgery*, vol. 203, no. 1, pp. 37–43, Jan 2012.
- [88] Y. Yokota and Y. Naruse, "Phase coherence of auditory steady-state response reflects the amount of cognitive workload in a modified n-back task," *Neuroscience Research*, vol. 100, pp. 39–45, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168010215001777>
- [89] B. Mehler, B. Reimer, and J. F. Coughlin, "Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups," *Human Factors*, vol. 54, pp. 396–412, 2012, PMID: 22768642.
- [90] R. F. Butterworth, *Hypoxic Encephalopathy*, 6th ed., A. R. Siegel GJ, Agranoff BW, Ed. Lippincott-Raven, 1999.
- [91] L. Owen and S. I. Sunram-Lea, "Metabolic agents that enhance ATP can improve cognitive functioning: A review of the evidence for glucose, oxygen, pyruvate, creatine, and l-carnitine," *Nutrients*, vol. 3, no. 8, pp. 735–755, Aug 2011.
- [92] F. Eustache, P. Rioux, B. Desgranges, G. Marchal, M. C. Petit-Taboué, M. Dary, B. Lechevalier, and J. C. Baron, "Healthy aging, memory subsystems and regional cerebral oxygen consumption," *Neuropsychologia*, vol. 33, no. 7, pp. 867–887, Jul 1995.
- [93] A. B. Scholey, M. C. Moss, N. Neave, and K. Wesnes, "Cognitive performance, hyperoxia, and heart rate following oxygen administration in healthy young adults," *Physiology and Behavior*, vol. 67, no. 5, pp. 783–789, Nov 1999.
- [94] M. C. Moss, A. B. Scholey, and K. Wesnes, "Oxygen administration selectively enhances cognitive performance in healthy young adults: A placebo-controlled double-blind crossover study," *Psychopharmacology*, vol. 138, no. 1, pp. 27–33, 1998.

- [95] S. P. Kim, M. H. Choi, J. H. Kim, H. W. Yeon, H. J. Yoon, H. S. Kim, J. Y. Park, J. H. Yi, G. R. Tack, and S. C. Chung, “Changes of 2-back task performance and physiological signals in ADHD children due to transient increase in oxygen level,” *Neuroscience Letters*, vol. 511, no. 2, pp. 70–73, Mar 2012.
- [96] S. C. Chung, J. H. Sohn, B. Lee, G. R. Tack, J. H. Yi, J. H. You, J. H. Jun, and R. Sparacio, “The effect of transient increase in oxygen level on brain activation and verbal performance,” *International Journal of Psychophysiology*, vol. 62, no. 1, pp. 103–108, Oct 2006.
- [97] H. Tsunashima and K. Yanagisawa, “Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS),” *Computational Intelligence and Neuroscience*, 2009.
- [98] M. Elgendi, “On the analysis of fingertip photoplethysmogram signals,” *Current Cardiology Reviews*, vol. 8, no. 1, pp. 14–25, 2012.
- [99] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
- [100] J. Zhu, S. Rosset, H. Zou, and T. Hastie, “Multi-class adaboost,” *Ann Arbor*, vol. 1001, p. 48109, 2006.
- [101] T. Appel, C. Scharinger, P. Gerjets, and E. Kasneci, “Cross-subject workload classification using pupil-related measures,” in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research Applications*, ser. ETRA ’18. New York, NY, USA: Association for Computing Machinery, 2018.
- [102] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, “Mental workload during n-back task—quantified in the prefrontal cortex using fnirs,” *Frontiers in Human Neuroscience*, vol. 7, p. 935, 2014.
- [103] G. I. Christopoulos, M. A. Uy, and W. J. Yap, “The Body and the Brain: Measuring Skin Conductance Responses to Understand the Emotional Experience:,” *Organizational Research Methods*, vol. 22, no. 1, pp. 394–420, Dec 2016.

- [104] A. Jimenez-Molina, C. Retamal, and H. Lira, "Using Psychophysiological Sensors to Assess Mental Workload During Web Browsing," *Sensors 2018, Vol. 18, Page 458*, vol. 18, no. 2, p. 458, Feb 2018.
- [105] S. Dong, L. M. Reder, Y. Yao, Y. Liu, and F. Chen, "Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty," *Brain Research*, vol. 1616, pp. 146–156, Aug 2015.
- [106] A. Chęć, D. Olczak, T. Fernandes, and H. A. Ferreira, "Physiological computing gaming: Use of electrocardiogram as an input for video gaming," *PhyCS 2015 - 2nd International Conference on Physiological Computing Systems, Proceedings*, pp. 157–163, 2015.
- [107] A. M. Porter and P. Goolkasian, "Video games and stress: How stress appraisals and game content affect cardiovascular and emotion outcomes," *Frontiers in Psychology*, vol. 10, p. 967, Apr 2019.
- [108] M. Ambinder, "Biofeedback in gameplay: How valve measures physiology to enhance gaming experience," *Game Developers Conference*, vol. 22, pp. 27–33, 2011.
- [109] P. A. Nogueira *et al.*, "Vanishing scares: biofeedback modulation of affective player experiences in a procedural horror game," *Journal on Multimodal User Interfaces*, vol. 10, no. 1, pp. 31–62, Mar 2016.
- [110] C. Sheikholeslami *et al.*, "A high resolution EEG study of dynamic brain activity during video game play," *Proceedings of the IEEE Annual International Conference of the Engineering in Medicine and Biology*, pp. 2489–2491, 2007.
- [111] E. Pellouchoud, M. E. Smith, L. McEvoy, and A. Gevins, "Mental effort-related eeg modulation during video-game play: Comparison between juvenile subjects with epilepsy and normal control subjects," *Epilepsia*, vol. 40, no. s4, pp. 38–43, 1999.
- [112] I. Tachtsidis and A. Papaioannou, "Investigation of Frontal Lobe Activation with fNIRS and Systemic Changes During Video Gaming," *Advances in Experimental Medicine and Biology*, vol. 789, p. 89, 2013.
- [113] UK Department for Transport, "Reported Road Casualties Great Britain: 2017." [Online]. Available: <https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2017>

- [114] D. Flatley, L. A. Reyner, and J. A. Horne, "Sleep-Related Crashes on Sections of Different Road Types in the UK," 2004.
- [115] S. Bioulac, J.-A. Micoulaud-Franchi, M. Arnaud, P. Sagaspe, N. Moore, F. Salvo, and P. Philip, "Risk of Motor Vehicle Accidents Related to Sleepiness at the Wheel: A Systematic Review and Meta-Analysis," *SLEEP*, vol. 40, no. 10, 2017.
- [116] R. O. Phillips, "A review of definitions of fatigue – And a step towards a whole definition," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 29, pp. 48–56, Feb 2015.
- [117] M. W. Johns, R. Chapman, K. Crowley, and A. Tucker, "A new method for assessing the risks of drowsiness while driving," *Somnologie - Schlafforschung und Schlafmedizin*, vol. 12, no. 1, pp. 66–74, Mar 2008.
- [118] G. Borghini, L. Astolfi, G. Vecchiato, and D. Mattia, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neuroscience & Biobehavioral Reviews*, vol. 44, pp. 58–75, Jul 2014.
- [119] T. Akerstedt, C. Bassetti, F. Cirignotta, D. García-Borreguero, M. Gonçalves, J. Horne, D. Léger, M. Partinen, T. Penzel, P. Philip, and J. C. Verster, "White Paper "Sleepiness at the Wheel"," *French Motorway Companies (ASFA) and the National Institute of Sleep and Vigilance (INSV)*, 2013.
- [120] M. W. Johns, "A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale," *Sleep*, vol. 14, no. 6, pp. 540–545, Nov 1991.
- [121] G. B. Neuberger, "Measures of fatigue: The Fatigue Questionnaire, Fatigue Severity Scale, Multidimensional Assessment of Fatigue Scale, and Short Form-36 Vitality (Energy/Fatigue) Subscale of the Short Form Health Survey," *Arthritis & Rheumatism*, vol. 49, no. S5, pp. S175–S183, Oct 2003.
- [122] T. Chalder, G. Berelowitz, T. Pawlikowska, L. Watts, S. Wessely, D. Wright, and E. Wallace, "Development of a fatigue scale," *Journal of Psychosomatic Research*, vol. 37, no. 2, pp. 147–153, Feb 1993.
- [123] L. B. Krupp, N. G. LaRocca, J. Muir-Nash, and A. D. Steinberg, "The Fatigue Severity Scale," *Archives of Neurology*, vol. 46, no. 10, p. 1121, Oct 1989.

- [124] S. Hewlett, M. Hehir, and J. R. Kirwan, “Measuring Fatigue in Rheumatoid Arthritis: A Systematic Review of Scales in Use,” *Arthritis Rheum*, no. 3, pp. 429–439, Apr 2007.
- [125] P. O. Valko, C. L. Bassetti, K. E. Bloch, U. Held, and C. R. Baumann, “Validation of the Fatigue Severity Scale in a Swiss Cohort,” *Sleep*, vol. 31, no. 11, pp. 1601–1607, nov 2008.
- [126] A. Bener, E. Yildirim, T. Özkan, and T. Lajunen, “Driver sleepiness, fatigue, careless behavior and risk of motor vehicle crash and injury: Population based case and control study,” *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 4, no. 5, pp. 496–502, Oct 2017.
- [127] P.-H. Ting, J.-R. Hwang, J.-L. Doong, and M.-C. Jeng, “Driver fatigue and highway driving: A simulator study,” *Physiology & Behavior*, vol. 94, no. 3, pp. 448–453, Jun 2008.
- [128] P. Thiffault and J. Bergeron, “Monotony of road environment and driver fatigue: a simulator study,” *Accident Analysis & Prevention*, vol. 35, no. 3, pp. 381–391, May 2003.
- [129] A. Sonnleitner, M. S. Treder, M. Simon, S. Willmann, A. Ewald, A. Buchner, and M. Schrauf, “EEG alpha spindles and prolonged brake reaction times during auditory distraction in an on-road driving study,” *Accident Analysis & Prevention*, vol. 62, pp. 110–118, Jan 2014.
- [130] S. K. Lal and A. Craig, “A critical review of the psychophysiology of driver fatigue,” *Biological Psychology*, vol. 55, no. 3, pp. 173–194, Feb 2001.
- [131] R. G. O’Connell, P. M. Dockree, I. H. Robertson, M. A. Bellgrove, J. J. Foxe, and S. P. Kelly, “Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur.” *The Journal of Neuroscience*, vol. 29, no. 26, pp. 8604–11, Jul 2009.
- [132] J. B. J. Riemersma, A. F. Sanders, C. Wildervanck, and A. W. Gaillard, “Performance decrement during prolonged night driving,” in *Vigilance: Theory, Operational Performance, and Physiological Correlates*, R. R. Mackie, Ed. Boston, MA: Springer US, 1977, pp. 41–58.

- [133] C. Zhao, M. Zhao, J. Liu, and C. Zheng, “Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator,” *Accident Analysis & Prevention*, vol. 45, pp. 83–90, Mar 2012.
- [134] Q. Ji, Z. Zhu, and P. Lan, “Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue,” *IEEE Transactions on Vehicular Technology*, vol. 53, no. 4, pp. 1052–1068, Jul 2004.
- [135] T. P. Nguyen, M. T. Chew, and S. Demidenko, “Eye tracking system to detect driver drowsiness,” in *2015 6th International Conference on Automation, Robotics and Applications (ICARA)*. IEEE, Feb 2015, pp. 472–477.
- [136] S. Arefnezhad, J. Hamet, A. Eichberger, M. Frühwirth, A. Ischebeck, I. V. Koglbauer, M. Moser, and A. Yousefi, “Driver drowsiness estimation using EEG signals with a dynamical encoder–decoder modeling framework,” *Scientific Reports*, vol. 12, no. 1, p. 2650, 2022.
- [137] D. J. Liebling and S. Preibusch, “Privacy considerations for a pervasive eye tracking world,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp ’14 Adjunct*. ACM Press, 2014, pp. 1169–1177.
- [138] E. J. Sung, B. C. Min, S. C. Kim, and C. J. Kim, “Effects of oxygen concentrations on driver fatigue during simulated driving,” *Applied Ergonomics*, vol. 36, no. 1, pp. 25–31, 2005.
- [139] D. Jing, S. Zhang, and Z. Guo, “Fatigue driving detection method for low-voltage and hypoxia plateau area: A physiological characteristic analysis approach,” *International Journal of Transportation Science and Technology*, vol. 9, no. 2, pp. 148–158, Jun 2020.
- [140] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, “STOP, THAT and One Hundred Other Sleep Scales,” *Springer Publishing*, 2012.
- [141] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, “Drowsiness detection using heart rate variability,” *Medical and Biological Engineering and Computing*, vol. 54, no. 6, pp. 927–937, Jun 2016.

- [142] M. Xie, X. Liu, X. Cao, M. Guo, and X. Li, “Trends in prevalence and incidence of chronic respiratory diseases from 1990 to 2017,” *Respiratory Research*, vol. 21, no. 1, p. 49, Feb 2020.
- [143] T. E. King, A. Pardo, and M. Selman, “Idiopathic pulmonary fibrosis,” *The Lancet*, vol. 378, no. 9807, pp. 1949–1961, Dec 2011.
- [144] G. Viegi, F. Pistelli, D. L. Sherrill, S. Maio, S. Baldacci, and L. Carrozzi, “Definition, epidemiology and natural history of COPD,” pp. 993–1013, Nov 2007.
- [145] W. M. Thurlbeck and N. L. Müller, “Emphysema: definition, imaging, and quantification.” *American Journal of Roentgenology*, vol. 163, no. 5, pp. 1017–1025, 1994.
- [146] B. E. Heard, V. Khatchatourov, H. Otto, N. V. Putov, and L. Sobin, “The morphology of emphysema, chronic bronchitis, and bronchiectasis: definition, nomenclature, and classification.” *Journal of Clinical Pathology*, vol. 32, no. 9, p. 882, 1979.
- [147] M. Roman-Rodriguez and A. Kaplan, “GOLD 2021 Strategy Report: Implications for Asthma–COPD Overlap,” *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 16, p. 1709, 2021.
- [148] A. R. Patel, A. R. Patel, S. Singh, S. Singh, and I. Khawaja, “Global initiative for chronic obstructive lung disease: The changes made,” *Cureus*, vol. 11, no. 6, Jun 2019.
- [149] M. J. Tobin, T. S. Chadha, G. Jenouri, S. J. Birch, H. B. Gazeroglu, and M. A. Sackner, “Breathing Patterns: 2. Diseased Subjects,” *Chest*, vol. 84, no. 3, pp. 286–294, 1983.
- [150] H. Wilkens, B. Weingard, A. Lo Mauro, E. Schena, A. Pedotti, G. W. Sybrecht, and A. Aliverti, “Breathing pattern and chest wall volumes during exercise in patients with cystic fibrosis, pulmonary fibrosis and COPD before and after lung transplantation,” *Thorax*, vol. 65, no. 9, pp. 808–814, 2010.
- [151] J. Daniels, “Portable respiratory gas collection equipment,” *Journal of Applied Physiology*, vol. 31, no. 1, pp. 164–167, Jul 1971.
- [152] R. H. Dressendorfer, C. E. Wade, and E. M. Bernauer, “Combined effects of breathing resistance and hyperoxia on aerobic work tolerance,” *Journal of Applied Physiology*, vol. 42, no. 3, pp. 444–448, 1977.



- [153] R. G. Love, D. C. Muir, K. F. Sweetland, R. A. Bentley, and O. G. Griffin, "Acceptable levels for the breathing resistance of respiratory apparatus: results for men over the age of 45." *Occupational and Environmental Medicine*, vol. 34, no. 2, pp. 126–129, May 1977.
- [154] S. Kido, Y. Nakajima, T. Miyasaka, Y. Maeda, T. Tanaka, W. Yu, H. Maruoka, and K. Takayanagi, "Effects of Combined Training with Breathing Resistance and Sustained Physical Exertion to Improve Endurance Capacity and Respiratory Muscle Function in Healthy Young Adults," *Journal of Physical Therapy Science*, vol. 25, no. 5, pp. 605–610, May 2013.
- [155] K. K. Dames, A. J. Lopes, and P. L. De Melo, "Airflow pattern complexity during resting breathing in patients with COPD: Effect of airway obstruction," *Respiratory Physiology and Neurobiology*, vol. 192, no. 1, pp. 39–47, Feb 2014.
- [156] L. Nilsson, A. Johansson, and S. Kalman, "Respiratory variations in the reflection mode photoplethysmographic signal. Relationships to peripheral venous pressure," *Medical and Biological Engineering and Computing*, vol. 41, no. 3, pp. 249–254, May 2003.
- [157] L. M. Nilsson, "Respiration Signals from Photoplethysmography," *Anesthesia & Analgesia*, vol. 117, no. 4, pp. 859–865, Oct 2013.
- [158] D. J. Meredith, D. Clifton, P. Charlton, J. Brooks, C. W. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: A review of relevant physiology," *Journal of Medical Engineering & Technology*, vol. 36, no. 1, pp. 1–7, Mar 2012.
- [159] F. Yasuma and J. I. Hayano, "Respiratory Sinus Arrhythmia: Why Does the Heartbeat Synchronize with Respiratory Rhythm?" *Chest*, vol. 125, no. 2, pp. 683–690, Feb 2004.
- [160] P. Dehkordi, A. Garde, B. Molavi, J. M. Ansermino, and G. A. Dumont, "Extracting Instantaneous Respiratory Rate From Multiple Photoplethysmogram Respiratory-Induced Variations," *Frontiers in Physiology*, vol. 9, no. JUL, p. 948, Jul 2018.
- [161] L. Bernardi, A. Radaelli, P. L. Solda', A. J. Coats, M. Reeder, A. Calciati, C. S. Garrard, and P. Sleight, "Autonomic control of skin microvessels: Assessment by power spectrum of photoplethysmographic waves," *Clinical Science*, vol. 90, no. 5, pp. 345–355, 1996.
- [162] O. S. Hoilett, A. M. Twibell, R. Srivastava, and J. C. Linnes, "Kick LL: A Smartwatch for Monitoring Respiration and Heart Rate using Photoplethysmography," in *Proceedings*

- of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS.* IEEE, Jul 2018, pp. 3821–3824.
- [163] S. Singh, M. Kozłowski, I. García-López, Z. Jiang, and E. Rodriguez-Villegas, “Proof of Concept of a Novel Neck-Situated Wearable PPG System for Continuous Physiological Monitoring,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.
  - [164] D. Jarchi, D. Salvi, L. Tarassenko, and D. A. Clifton, “Validation of instantaneous respiratory rate using reflectance ppg from different body positions,” *Sensors*, vol. 18, no. 11, 2018.
  - [165] E. L’Her, Q. T. N’Guyen, V. Pateau, L. Bodenes, and F. Lellouche, “Photoplethysmographic determination of the respiratory rate in acutely ill patients: Validation of a new algorithm and implementation into a biomedical device,” *Annals of Intensive Care*, vol. 9, no. 1, p. 11, Dec 2019.
  - [166] D. Clifton, G. J. Douglas, P. S. Addison, and J. N. Watson, “Measurement of respiratory rate from the photoplethysmogram in chest clinic patients,” *Journal of Clinical Monitoring and Computing*, vol. 21, no. 1, pp. 55–61, Feb 2007.
  - [167] K. H. Chon, S. Dash, and K. Ju, “Estimation of Respiratory Rate From Photoplethysmogram Data Using Time-Frequency Spectral Estimation,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2054–2063, 2009.
  - [168] A. Johansson and P. Å. Öberg, “Estimation of respiratory volumes from the photoplethysmographic signal. Part 2: A model study,” *Medical and Biological Engineering and Computing*, vol. 37, no. 1, pp. 48–53, 1999.
  - [169] S. Khreis, D. Ge, H. A. Rahman, and G. Carrault, “Breathing Rate Estimation Using Kalman Smoother with Electrocardiogram and Photoplethysmogram,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 3, pp. 893–904, Mar 2020.
  - [170] J. Li, J. Jin, X. Chen, W. Sun, and P. Guo, “Comparison of respiratory-induced variations in photoplethysmographic signals,” *Physiological Measurement*, vol. 31, no. 3, pp. 415–425, Feb 2010.

- [171] K. H. Shelley, D. H. Jablonka, A. A. Awad, R. G. Stout, H. Rezkanna, and D. G. Silverman, "What Is the Best Site for Measuring the Effect of Ventilation on the Pulse Oximeter Waveform?" *Anesthesia & Analgesia*, vol. 103, no. 2, pp. 372–377, Aug 2006.
- [172] W. P. Santamore and J. N. Amoores, "Buffering of respiratory variations in venous return by right ventricle: A theoretical analysis," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 267, Dec 1994.
- [173] P. H. Charlton, T. Bonnici, L. Tarassenko, J. Alastruey, D. A. Clifton, R. Beale, and P. J. Watkinson, "Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: Technical and physiological determinants," *Physiological Measurement*, vol. 38, no. 5, pp. 669–690, Mar 2017.
- [174] L. Nilsson, T. Goscinski, A. Johansson, L. G. Lindberg, and S. Kalman, "Age and gender do not influence the ability to detect respiration by photoplethysmography," *Journal of Clinical Monitoring and Computing*, vol. 20, no. 6, pp. 431–436, Oct 2006.
- [175] N. Snell, D. Strachan, R. Hubbard, J. Gibson, E. Limb, R. Gupta, A. Martin, M. Laffan, and I. Jarrold, "Burden of lung disease in the UK; findings from the British Lung Foundation's "respiratory health of the nation" project," in *European Respiratory Journal*, vol. 48, no. suppl 60. European Respiratory Society (ERS), Sep 2016, p. PA4913.
- [176] A. M. Russell, H. Adamali, P. L. Molyneaux, P. T. Lukey, R. P. Marshall, E. A. Renzoni, A. U. Wells, and T. M. Maher, "Daily home spirometry: An effective tool for detecting progression in idiopathic pulmonary fibrosis," *American Journal of Respiratory and Critical Care Medicine*, vol. 194, no. 8, pp. 989–997, Oct 2016.
- [177] S. A. Shah, C. Velardo, A. Farmer, and L. Tarassenko, "Exacerbations in chronic obstructive pulmonary disease: Identification and prediction using a digital health system," *Journal of Medical Internet Research*, vol. 19, no. 3, p. e69, Mar 2017.
- [178] H. Yamazaki and K. Fujimoto, "A new noninvasive method for measurement of dynamic lung compliance from fluctuations on photoplethysmography in respiration," *Journal of Applied Physiology*, vol. 130, no. 1, pp. 215–225, Jan 2021.

- [179] H. J. Davies, G. Hammour, and D. P. Mandic, “An apparatus for the simulation of breathing disorders: Physically meaningful generation of surrogate data,” *arXiv preprint arXiv:2109.06699*, Sep 2021.
- [180] K. V. Madhav, M. R. Ram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, “Estimation of respiration rate from ECG, BP and PPG signals using empirical mode decomposition,” in *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2011, pp. 1661–1664.
- [181] World Health Organization, “Global status report on road safety 2018: summary,” June 2018.
- [182] World Health Organisation, “The top 10 causes of death,” Mar 2022. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [183] Y. Çolak, S. Afzal, B. G. Nordestgaard, J. Vestbo, and P. Lange, “Prognosis of asymptomatic and symptomatic, undiagnosed COPD in the general population in Denmark: a prospective cohort study,” *The Lancet Respiratory Medicine*, vol. 5, no. 5, pp. 426–434, May 2017.
- [184] A. B. Newman, F. J. Nieto, U. Guidry *et al.*, “Relation of sleep-disordered breathing to cardiovascular disease risk factors: the sleep heart health study,” *American Journal of Epidemiology*, vol. 154, no. 1, pp. 50–59, Jul 2001.
- [185] P. Maquet, “The role of sleep in learning and memory.” *Science*, vol. 294, no. 5544, pp. 1048–52, Nov 2001.
- [186] D. Hillman, S. Mitchell, J. Streatfeild *et al.*, “The economic cost of inadequate sleep,” *Sleep*, vol. 41, no. 8, Aug 2018.
- [187] A. A. Ong and M. B. Gillespie, “Overview of smartphone applications for sleep analysis,” *World Journal of Otorhinolaryngology-Head and Neck Surgery*, vol. 2, no. 1, pp. 45–49, Mar 2016.
- [188] S. Bhat, A. Ferraris, D. Gupta *et al.*, “Is there a clinical role For smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography,” *Journal of Clinical Sleep Medicine*, vol. 11, no. 7, pp. 709–15, Jul 2015.

- [189] B. Jafari, “Polysomnography,” *Clinics in Chest Medicine*, vol. 31, no. 2, pp. 287–297, Jun 2010.
- [190] S. Biswal, J. Kulas, H. Sun, B. Goparaju, and M. T. Bianchi, “SLEEPNET: Automated sleep staging system via deep learning,” *arXiv: 1707.08262v1*, 2017.
- [191] D. Shrivastava, S. Jung, M. Saadat, R. Sirohi, and K. Crewson, “How to interpret the results of a sleep study.” *Journal of Community Hospital Internal Medicine Perspectives*, vol. 4, no. 5, p. 24983, 2014.
- [192] R. S. Rosenberg and S. Van Hout, “The American Academy of Sleep Medicine inter-scorer reliability program: Sleep stage scoring,” *Journal of Clinical Sleep Medicine*, vol. 09, no. 01, pp. 81–87, Jan 2013.
- [193] National Sleep Research Resource, “Cleveland children’s sleep and health study.” [Online]. Available: <https://sleepdata.org/datasets/ccshs>
- [194] D. A. Dean, A. L. Goldberger, R. Mueller *et al.*, “Scaling up scientific discovery in sleep medicine: the national sleep research resource.” *Sleep*, vol. 39, no. 5, pp. 1151–64, May 2016.
- [195] H. Sun, J. Jia, B. Goparaju *et al.*, “Large-scale automated sleep staging,” *SLEEP*, vol. 40, no. 10, 2017.
- [196] X. Li, L. Cui, S. Tao *et al.*, “HyCLASSS: A hybrid classifier for automatic sleep stage scoring,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 375–385, Mar 2018.
- [197] T. Nakamura, V. Goverdovsky, M. J. Morrell, and D. P. Mandic, “Automatic Sleep Monitoring Using Ear-EEG,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, pp. 1–8, 2017.