Non-contact monitoring of respiration in the neonatal intensive care unit

João Jorge

St Hilda's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

Supervised by

Professor Lionel Tarassenko

Declaration

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

_____

(*João G. Jorge*)
Oxford,
January 25, 2019

Every breath you take,
(...)
I'll be watching you.

— The Police
Every Breath You Take (1983)

To my parents

Acknowledgements

This thesis would not have been possible without the support of a great number of people.

Above all, I would like to thank my doctoral supervisor, Professor Lionel Tarassenko. I will always be grateful for his mentorship. His genius, knowledge and enterprise have been an inspiration to me, and the example he has provided will stick with me long past my time at Oxford.

Special thanks are due to the postdocs, docs and predocs in the Biomedical Signal Processing group. In particular, I had a great slice of good fortune for sharing an office and ideas with Dr Mauricio Villarroel, Sitthichok Chaichulee, Hamish Tomlinson, Alessandro Guazzi, and Jonathan Daly. To Mauricio, thank you for being ready to talk each and every one of the 3,499 times that I turned around in my chair and called your name over the past five years. Thank you for believing that I could make it and to Alessandro for being able to emphasise (and empathise).

A great deal of the work in this thesis was undertaken in collaboration with the clinical teams at the Dialysis Unit at the Churchill Hospital and the Neonatal Intensive Care Unit at the John Radcliffe Hospital. Thanks are due to Dr Chris Pugh, Sheera Sutherland, Dr Clare MacEwan, Sara Davis, Sharon Garrett, Sheula Barlow, Dr Gabrielle Green, Dr Kevin McCormick, and last but never least, Dr Matthew Frise. I would also like to acknowledge all the volunteers who took part in our studies.

There are countless friends to whom I am indebted for their support during the past five years, but I feel particularly compelled to name the following: Catarina do Coito e Silva, Joana Pinto, Eleanor Greer, Rory Geogheghan, Tom Partrige, Johanna Schnurr, Thais Roque, Flora Kennedy McConnell, Sam Dockree, and, of course, the original Hildabeasts, Benjamin Schaper, Lauren Mason, and Colleen Curran. There are no bounds to the tremendous amount of gratitude I owe you all.

To the Principal, fellows, staff and students of St Hilda's College, Oxford, thank you for welcoming me and for making me feel at home. I would like to express my gratitude to my college tutor, Professor Alison Noble, for the support and kindness shown to me. To the members of St Hilda's Middle Common Room, in particular those who had the imponderable joy (or misery) of serving committees with me; it was a pleasure.

A big thank you also to the members of the MCR Presidents' Committee for instilling great ideas and even greater mischief.

I would like to acknowledge both Fundação para a Ciência e Tecnologia and EPSRC-UK for the grants that supported my research over the last five years.

Abstract

Camera-based monitoring is an emerging sensing modality which has attracted widespread interest due to its potential to offer comfortable and concurrent measurements of the cardiorespiratory vital signs: the cardiac pulse rate, respiratory rate and peripheral capillary oxygen saturation.

In recent years, several methods have been proposed for the processing of visible and infrared video data and subsequent extraction of the respiratory rhythm. However, due to the low signal-to-noise ratio (SNR) of these signals, their application has so far been limited to adult subjects and controlled research environments. In this thesis, we aim to improve the signal processing aspects of non-contact respiratory measurements by developing a novel technique for extracting this signal in a clinical setting. The main contribution of this work is a multi-channel source separation method which combines information from multiple image domains and uses *a priori* information about the distinctive temporal structure of the signal of interest to improve the quality of the respiratory signal extracted.

Considering the neonatal breathing component as an extreme case of poor SNR in video signals, we present a case study of the extraction of respiratory signals from video signals recorded from a CCD camera placed above an incubator inside which critically-ill infants are nursed. Evaluation results were obtained from recordings on 30 neonatal patients nursed in the Neonatal Intensive Care Unit at the John Radcliffe Hospital, Oxford, UK. The respiratory signals extracted using the methodology described compare favourably against those derived from direct and indirect sources during periods of manual breath counts. In addition, we show that a classification algorithm which uses features of these signals to identify periods of cessation of breathing can successfully distinguish between true and artefactual decreases in respiratory rate in a range of critically low values of this vital sign.

Contents

List of Figures

Nomenclature

Roman Symbols

e	base of the natural logarithm
f	frequency
Fi	Fraction of inspired gas a participating in gas exchange in the alveoli
$\mathcal{G}_{xx}$	Power spectral density function of signal x
$\mathcal{G}_{xy}$	Cross-spectral density function of signal x and y
$\mathcal{H}$	Hjorth complexity descriptor
I	Image intensity
i	unit imaginary number $\sqrt{-1}$
p	p-value
k	discrete frequency
p	partial pressure of species a
$\mathcal{P}_x$	Power of signal x
V_s	Stroke volume
V_t	Tidal volume

Greek Symbols

α	Activity
ϵ_a	Molar extinction coefficient of species a $(\mathrm{m}^{-1}\mathrm{M}^{-1})$
Γ_s	Spectral purity
γ^2	Magnitude-squared spectral coherence
Ω	angular frequency
π	mathematical constant $\simeq 3.14\ldots$
φ_{nasal}	nasal airflow pressure
ψ	Camera-derived respiratory signal
$\bar{\Omega}_n$	n^{th} order spectral moment
τ_e	expiratory constant
λ	wavelength
τ_i	inspiratory constant

Acronyms

ABP	Arterial blood pulse
AGC	Automatic gain control
AM	Amplitude modulation of a signal
ANSI	American National Standards Institute
AOP	Apnoea of prematurity
AR	Auto-regressive
AUC	Area under the curve
BP	Blood pressure
bpm	breaths $\min^{-1}$
BVP	Blood volume pulse

BW   Baseline wander of a signal
CCD   Charge-coupled device
CIE   Commission Internationale de l'Eclairage
CMD   Complex demodulation
CMOS   Complementary metal-oxide-semiconductor
CNN   Convolutional neural network
COBE   Cessation of breathing event
CRF   Case report form
CUREC   University of Oxford Central University Research Committee
CWT   Continuous wavelet transform
DFT   Discrete Fourier transform
DSLR   Digital single-lens reflex
ECG   Electrocardiography
EDR   ECG-derived respiration
EEG   Electroencephalography
EM   Expectation-Maximisation
EVM   Eulerian video magnification
FFT   Fast Fourier transform
FM   Frequency modulation of a signal
FN   Number of false negatives
FoV   Field of view
FP   Number of false positives
FPGA   Field Programmable Gate Array
fps   frames per second
g.a.   Gestational age (in weeks)
GMM   Gaussian mixture model
HDA   High Dependency Area
HFOV   High-frequency oscillation ventillation
IH   Isocapnoeic hypoxia
PH   Poikilocapnoeic hypoxia
pH   potential of hydrogen. It is approximately the negative of the base 10 logarithm
        of the molar concentration of hydrogen ions
HR   Heart rate
HRV   Heart rate variability
HVR   Hypoxic ventilatory response
IBI   Interbreath intervals
ICA   Independent component analysis
IP   Impedance pneumogram
IQR   Inter-quantile range
L*a*b* CIE L*a*b* perceptually-uniform colour space
LAN   Local Area Network
LF   Left ventricule of the heart
LWIR   Long wavelength infrared window defined as electromagnetic radiation with a
        wavelength comprised between 8 and 14 μm
MAE   Mean absolute error
MAP   Maximum *a posteriori* probability
MIT-BIH Collection of recordings of multiple physiologic signals from subjects monitored
        in Boston's Beth Israel Hospital Sleep Laboratory
ML   Maximum likelihood

MRE   Mean relative error
MR   Magnetic resonance
NHS   National Health Service
ICU   Intensive Care Unit
NICU   Neonatal Intensive Care Unit
NIR   Near infrared radiation
OSA   Obstructive sleep apnoea
PCA   Principal component analysis
PPG   Photoplethysmography
QRS   The combination of three of the graphical deflections seen on a typical electrocar-
       diogram
RAM   Random-access memory
RDIMM Registered memory module
RGB   Red, green and blue colour model
RGB-D Integrated system consisting of a RGB and depth sensing units
RIAV   Respiratory-induced amplitude variations (of PPG)
RIFV   Respiratory-induced frequency variations (of PPG)
RIIV   Respiratory-induced intensity variations (of PPG)
RIP   (Respiratory) inductance plethysmography
RMSE   Root-mean-square error
ROC   Receiver operating characteristic
ROI   Region of interest
rPPG   remote photoplethysmography
RR   Respiratory rate
RSA   Respiratory sinus arrhythmia
RV   Right ventricule of the heart
SATA   Serial AT attachment. A motherboard interface to storage devices
SDR   Spectral destribution ratio of a signal
SIDS   Sudden infant death syndrome
SNR   Signal-to-noise ratio
spO_2 Peripheral oxygen saturation
SQI   Signal quality index
SSF   Slope sum function
STFT   Short-term Fourier Transform
SVD   Single value decomposition
SVM   Support vector machine
τCA   Periodic component analysis
THM   Traube-Hering-Mayer waves
TN   Number of true negatives
TP   Number of true positives
TTN   Transient tachypnoea of the newborn
UI   User interface
vPPG   video photoplethysmography
WHO   World Health Organisation

Chapter 1

Introduction

Identifying the problem

The ability of the newborn infant to adapt to the extra-uterine environment is critical to their survival. While all body systems undergo important physiological changes in the first days of life, arguably none is more critical to survival than the adaptation of the foetal lungs to extra-uterine gas exchange [Sinha and Donn, 2006].

The newborn infant is, therefore, vulnerable to a range of respiratory disorders unique to this period of early life. In addition to delayed or mal-adaptation to extra-uterine life, respiratory disorders can also arise from existing conditions, such as surgical or congenital anomalies, or from acquired conditions, such as pulmonary infections occurring either pre- or post-delivery.

Overall, it is estimated that between 2.2 % [Rubaltelli et al., 1998] to 6.7 % [Kumar and Bhat, 1996] of all births are complicated by a respiratory disorder. In fact, respiratory conditions are the most common reason for admission to a neonatal unit in both term and preterm infants [Pramanik et al., 2015]. One study reported that, with the exception of infants with syndromes and those with congenital or surgical conditions, 33.3 % of all neonatal admissions at > 28 weeks of gestation had respiratory conditions as their primary reason for admission [Parkash et al., 2015]. A further study estimated that 20.5 % of all neonatal admissions showed signs of respiratory distress [Qian et al., 2010]. Evidence of increasing rates of neonatal admissions due to respiratory conditions, possibly driven by increased rates of caesarean section delivery, can also be found [Ersch et al., 2007].

The clinical signs of respiratory distress are, therefore, important to recognise. Most conditions have some manifestation in the breathing pattern, namely in the morphology or frequency of the respiratory signal. The continuous monitoring of this vital sign during neonatal sleep is thus of crucial importance both in (a) intensive care units, and in (b) home environments.

Conventional contact devices based on impedance pneumography (IP) present clear shortcomings in these scenarios. Not only do they often require cumbersome equipment but they can also cause infant stress and even pain due to the electrodes damaging the infant's fragile skin. While IP is a convenient method if patients are already monitored by electrocardiography (ECG), it is susceptible to inaccurate readings due to a number of factors including poor electrode placement, motion artefacts, and physiological events unrelated to breathing (such as coughing, or crying) which nevertheless cause thoracic movements [Brouillette et al., 1987; Warburton D, Stark AR, 1977]. The high prevalence of noise and high false alarm rates means that respiratory signals are still largely disregarded in neonatal intensive care units.

In an effort to address these issues, a large body of research has been devoted to emerging monitoring modalities which do not require contact sensors to be physically attached to the patient, such as thermal imaging [Garbey et al., 2007], microwave-based [Baltag, 2008] or radar-based [Li et al., 2009] sensors. Although these are technically non-contact approaches, they all require specialist instrumentation and complex set-up procedures.

Under the paradigm of remote sensing, the latest decade has seen the emergence of numerous approaches to the measurement of vital signs using visible-light video cameras. Although several studies in neonates have been reported, most have focused on the extraction of heart rate alone [Aarts et al., 2013; Fernando and Haan, 2015; Klaessens et al., 2014; Mestha et al., 2014; Scalise and Bernacchia, 2012; Villarroel et al., 2014], with few approaches aimed at measuring respiratory parameters [Janssen et al., 2016; van Gastel et al., 2016; Villarroel et al., 2014] in this population. The added complexity of neonatal measurements may be partly explained by two factors; firstly, the immaturity of the brain mechanisms governing the respiratory rhythm in neonates can result in irregular breathing. As a consequence, the range of resting respiratory rates detected in neonates (40 - 120 bpm) is much wider than that found in adults (12 - 20 bpm) [Guyton et al., 2012]. In addition, due to the small amplitude of abdominal and chest tidal excursions (typically in the order of a few millimetres [De Groote et al., 1997]), the signal-to-noise ratio (SNR) of breathing motion-related changes in pixel intensity recorded from the infant body surface by a camera monitor are notoriously low for typical recording distances (~ 1 m). Both factors render the problem of accurate RR detection much more challenging in this patient group.

Objectives

The objective of this research was to improve the signal processing aspects of camera-based derivation of respiration in premature infants nursed in the neonatal unit. The contributions of this thesis are threefold.

Signal denoising

The spatial redundancy of the camera sensor means that algorithms for the estimation of vital signs, such as heart rate (HR), respiratory rate (RR) and peripheral oxygen saturation (spO$_2$), often extract time series from multiple regions of interest (ROIs) in the image as a strategy to boost SNR. The deconvolution of these mixtures of physiological information and contaminating noise (from ambient lighting or motion) is an essential step in any vital sign extraction sequence. While in previous studies not much attention has been paid to the denoising aspects of this approach, we propose a new methodology for modelling the signal of interest (in our case, the respiratory signal) which uses *a priori* information to extract this signal from an array of multi-ROI video signals.

Validation against gold standards

Although a considerable amount of published works relating to the task of deriving respiration from video are available, none of these evaluate the performance of their methods against the gold standard of manual respiratory assessments. In this thesis, we assess the accuracy of interbreath intervals (IBI) and RR estimates from respiratory signals derived using our proposed method against those provided by manual breath counts by clinical staff.

In addition, the performance of methods developed in this work is compared against both direct and indirect contact-based methods described in the literature. To our knowledge, no study has been published which analyses the reliability of RR estimates provided by indirect methods based on electrocardiography (ECG) and photoplethysmography (PPG) in neonatal patients.

Detection of cessations of breathing

Of as much clinical importance as the frequency of the respiratory cycle is the ability to detect periods during which breathing is absent. This motivates the final objective of determining the accuracy of methods to detect cessations of breathing using information contained in the derived camera-based signal.

Thesis overview

In the remainder of this chapter, we give a brief overview of the following chapters. The main contributions of the research described in this thesis are presented in Chapters 7 and 8 while the image and video processing methods upon which they rely are developed in Chapters 5 and 6. The conceptual links between these chapters are shown in Figure 1.1.
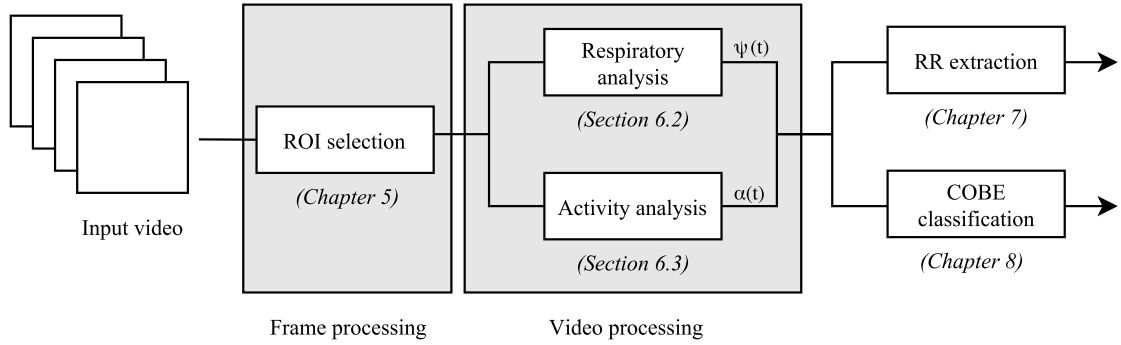
Figure 1.1

Processing pipeline for neonatal video data. First, the region of interest containing the skin cluster is segmented from the raw video footage (Chapter 5); then, a global activity signal $\alpha(t)$ and a camera-derived respiration waveform $\psi(t)$ are derived (Chapter 6). Their application to the derivation of respiratory rate and the detection of period of cessation of breathing (COBE) is described in Chapters 7 and 8, respectively.

Chapter 2 introduces the relevant background. Some elements of respiratory anatomy and physiology are briefly reviewed. We review the epidemiologic, diagnostic features, and management of common neonatal respiratory conditions. The current challenges associated with the respiratory monitoring of neonates are introduced. An overview of the state-of-the-art in non-contact monitoring of respiration is given with a focus on evidence for visible and near-infrared light (NIR) monitoring of neonates.

In order to evaluate our methods, suitable datasets are required for comparison of their performance with gold standards. The description of the two datasets on which the methods presented in this thesis were validated can be found in **Chapter 3**. The first dataset — the Hypercapnoea dataset — is part of a study to assess the performance of contact and non-contact devices in monitoring the vital functions of healthy young adults under non-atmospheric concentrations of O_2 (hypoxia) and CO_2 (hypercapnoea). The second dataset — the NICU dataset — was used to investigate the feasibility of monitoring the respiration of premature infants in critical care. Both datasets were collected over the course of this doctoral thesis. The aim of the studies described here was to determine the feasibility of non-contact camera-based monitoring of physiological parameters including, but not restricted to, those investigated in this thesis. Thus, the protocol design for these studies received input from members of the research teams in the Biomedical Signal Processing group (BSP), the Robbins group in Respiratory Physiology, and the Neonatal Unit at the John Radcliffe Hospital, Oxford. The technical specifications of data collection and storage were drawn up and implemented by the BSP team, of which I am a member. The data analysis reported in this thesis is entirely my own research.

The classical signal processing methods for the processing of the physiological signals used throughout this thesis are presented in **Chapter 4**. This chapter reviews previous

literature on extracting respiration information from biomedical signals acquired using the contact-devices currently available in most neonatal units. The best-performing methods in the literature are implemented for comparison. A reproducible evaluation procedure for assessing the performance of these methods is proposed. Finally, the performance of these methods on the Hypercapnoea dataset is presented for benchmarking.

The process of using a video camera to compute estimates of vital signs in any environment begins with (a) the detection of the monitored subject in the field of view and (b) the selection of an ROI from which vital signs can be estimated. Good performance on these tasks is crucial to the successful estimation of vital signs, even more so in the complex visual scenes of a hospital setting. **Chapter 5** presents the necessary image processing tools developed in this thesis for image segmentation and validates them on a subset of still images from the NICU dataset. The proposed method is used to focus the video analysis of Chapters 7 and 8 onto the frame area occupied by the monitored infant.

Chapter 6 gathers the video processing methods to be applied in Chapters 7 and 8. In this chapter, we develop methods for deriving (a) the camera-based respiratory signal ψ and (b) a global activity signal α. We also review the general eigenvalue decomposition problem, which may be considered as the theoretical basis of the developed source separation method. The applications of this method are investigated using simulated and real signals.

Chapter 7 is concerned with the measurement of both IBIs and RR in preterm infants during periods of normal breathing. Following the ideas of Chapter 6, we extract ψ and compare RR and IBI estimates against those measured using the gold standard of breath detection provided by the attending clinical staff. The performance of the methods developed is compared with that of methods described in the literature based on (a) IP, (b) the ECG-derived respiratory rhythms, and (c) the PPG-derived respiratory rhythms also implemented.

Chapter 8 investigates detection methods for cessation of breathing in preterm infants using input features derived from ψ. A review of studies carried out using existing neonatal apnoea monitors which employ contact sensors is presented. It is noted that apnoea monitors tend to have a very high false alarm rate. Preliminary results of this chapter were presented in [Jorge et al., 2017].

In **Chapter 9**, we summarise our findings and discuss the strengths and weaknesses of our approach. We also discuss open research problems in the field and possible directions for future studies.

Chapter 2

Relevant background

Contents

2.1   Chapter summary

This chapter introduces key anatomical and physiological traits of neonatal patients with particular attention paid to the respiratory system and the conditions which affect this system during early life. We review current monitoring practices in neonatal care. The current challenges associated with the respiratory monitoring in this setting are introduced and the need for non-invasive methods is discussed. We describe both direct and indirect methods. This knowledge is required in later chapters for the development of techniques based on the cardiovascular interplay and the interpretation of the extracted respiratory signals. Finally, the state-of-the-art in visible and NIR light monitoring of neonates is reviewed and recent developments in the field of camera-based sensing of respiration are systematised.

2.2 Elements of respiratory physiology

The aim of this section is to provide the reader with a brief introduction to the respiratory system: its structure, function, and interaction with the cardiovascular system.

Our aim is not to provide a complete picture of this system, but instead to focus on aspects of functional anatomy and respiratory physiology relevant to the understanding of the principles of the clinical measurement of respiration and the mechanisms of respiratory pathology addressed in this thesis. This summary is grounded in the textbooks by Guyton et al. [2012]; Hansen and Koeppen [2002]; Tortora and Derrickson [2014].

2.2.1 Structure and function

The functions of the respiratory system may be divided into four primary subfunctions [Guyton et al., 2012].

(a) Pulmonary ventilation *i.e.* the maintenance of (a) a pressure gradient, for the inflow and outflow of air between the atmosphere and lung alveoli, and (b) a concentration gradient, for gas diffusion to take place;

(b) Gas diffusion at lung level *i.e.* the transport of oxygen (O_2) and carbon dioxide (CO_2) via gaseous exchange between the lung alveoli and the network of pulmonary capillaries;

(c) Gas diffusion at tissue level *i.e.* the transport of oxygen to tissues for the metabolism of fats, carbohydrates, and proteins and the removal of the carbon dioxide that results from metabolic cell reactions;

(d) Regulation of ventilation *i.e.* the control systems that regulate the above mechanisms.

As the emphasis of this thesis is on external respiration, or breathing, *i.e.* the physical flow of air into or out of the lungs, the remainder of this section focuses primarily on the anatomical structures that support functions (a) and (d) of the respiratory system.

Anatomy

The respiratory tract is formed by the following structures (illustrated in Figure 2.1):

the upper airways refer to the parts of the tract which lie outside the thorax. These include the nose and nasal passages, the paranasal sinuses, the pharynx, and the portion of the larynx above the vocal folds;

the lower airways composed of the portion of the larynx below the vocal folds and the respiratory tree, a term used to refer to the branching structure of the airways
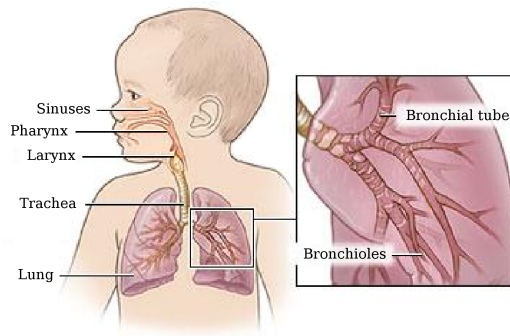
supplying air to the lungs (starting at the trachea and terminating in the alveoli). The lungs are divided into lobes; the left lung is composed of the upper and lower lobes, the right lung is composed of the upper, middle and lower lobes.

An infant's respiratory system is similar to that of an adult, however, some structures differ in size, position, or maturation[1].

Histology

Histologically, the respiratory system is divided into the conducting portion and the respiratory portion.

The air passages between the nasal cavities and the terminal bronchioles form the *conducting portion* of the respiratory system, which transports air to the *respiratory portion*. The former portion is lined with cuboidal or columnar epithelium. These cell layers have an important role in warming the air to body temperature, humidifying it to saturation, removing dust through nasal hairs and mucosal surfaces, and facilitating coughing due to the flexible dorsal membrane of the trachea.

The respiratory portion is intrapulmonary and consists of respiratory bronchioles, alveolar ducts, alveolar sacs and alveoli. The respiratory bronchioles, whose epithelium is ciliated, lead into alveolar ducts (surrounded by smooth muscle, elastin and collagen).

These, in turn, lead into the alveolar sacs, which contain several alveoli, the basic units of gas exchange in the lung. Each alveolus consists of a round-shaped epithelial layer and an extracellular matrix surrounded by capillaries from the pulmonary circulation. The alveoli are small and delicate structures; their epitelial layer reaches an average radius of just 0.1 mm and a wall thickness of about 0.2 μm when fully developed [Berthiaume et al., 2006].

---

[1]For example, the tongue of an infant is disproportionately large in relation to the mouth when compared to an adult's. An infant's larynx is located higher in the neck, and the ribs are found to be more horizontal with the sternal cavicular heads and the diaphragmatic domes higher than in older children and young adults [Openshaw et al., 1984].

The epithelium of the alveoli contains two main types of cells: type I pneumocytes (95% of the total alveolar area) which form a very thin diffusion barrier for gas exchange, and type II pneumocytes (5% of the total alveolar area) which secrete lung surfactant, a mixture of phospholipids combined with proteins and glycoproteins [Berthiaume et al., 2006] responsible for the decrease in surface tension of the thin alveolar walls. Surface tension accounts for most (approximately two thirds [Nkadi et al., 2009]) of the inward recoil forces in the lung, and prevents the alveoli from collapsing during exhalation.

2.2.2 Mechanics of pulmonary ventilation

Pulmonary ventilation relies on the maintenance of a pressure gradient between the pressure in the thoracic cavity (intrapulmonary pressure) and that of the environment (atmospheric pressure) for the inflow and outflow of air into the lungs.

To allow breathing, the thorax must change volume. Lungs can be expanded and contracted in two ways. The upward and downward movement of the diaphragm changes the length of the chest cavity (the contraction of the diaphragm pulls the lungs downward as it flattens, while its relaxation allows the lungs to contract). This is the preferred mechanism in normal breathing, and the one responsible for most of the change in intrapulmonary volume that accounts for tidal volume. Alternatively (and, in somes instances, additionally), the diameter of the chest cavity can be changed through the elevation or depression of the ribs, via the intercostal and abdominal musculature. Collectively, during spontaneous breathing, the processes just described act to adjust the pressure of the intrapulmonary gas, causing the inflow and outflow of air into the respiratory tree, as shown in Figure 2.2. These two phases, known as inspiration and expiration, function sequentially and are jointly referred to as the respiratory cycle.

The number of times the respiratory cycle is repeated over a unit of time sets the *respiratory* or *ventilatory rate* (RR), and is commonly expressed in breaths per minute (bpm) [Guyton et al., 2012]. The normal range of this vital sign in healthy individuals changes with age, with children generally having faster respiratory rates than adults (Table 2.1).

The *respiratory minute volume* or *ventilation* ($\dot{V}_E$) is the volume of air inspired or

Table 2.1

Normal resting respiratory rates by age [Barrett and Ganong, 2010; DeBoer, 2004].

Age	Normal resting respiratory range(bpm)
Birth to 6 weeks	$30 - 60$
6 months	$25 - 40$
3 years	$20 - 30$
6 years	$18 - 25$
10 years	$15 - 20$
Adults	$12 - 20$

(a)   (b)
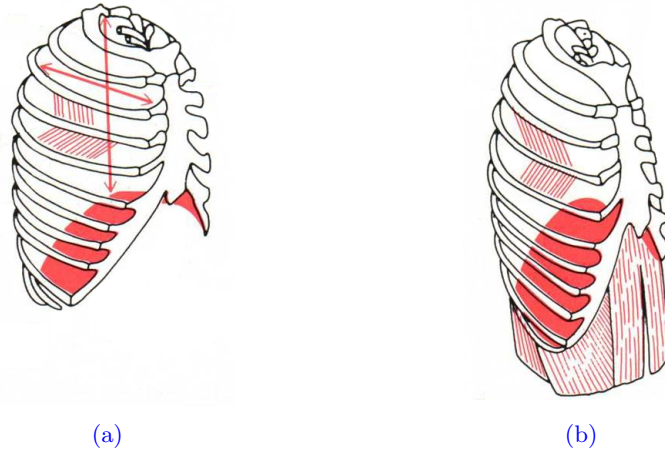
Figure 2.2

Expansion and contraction of the thoracic cage during inspiration (2.2a) and expiration (2.2b), as demonstrated by the diaphragmatic placement, rib cage elevation and function of the intercostal musculature. On inhalation, the external intercostal muscles contract and the internal intercostal muscles relax. As a result of this concerted activity of the chest musculature, the thoracic cage moves frontally and cranially. At the same time, the diaphragm contracts and flattens. These structural changes act to increase the volume inside the thoracic cavity, thereby stretching the lungs, increasing intrapulmonary volume and reducing intrapulmonary pressure (by 0 - 1 mmHg). Conversely, exhalation is largely a passive process; the intercostal muscles and the diaphragm relax and return to their resting positions. This in turn reduces the volume inside the thoracic cavity, thereby increasing the intrapulmonary pressure and forcing air out of the lungs. Due to the natural elasticity of the lungs, recoil forces tend to collapse the lungs if they were not held against the thoracic wall by the contraction of these muscles. Figures adapted from Guyton et al. [2012].

expired by the lungs per unit of time[1], usually expressed in litres per minute ($l \min^{-1}$). Under normal resting circumstances the respiratory minute volume is the product of respiratory rate and *tidal volume* (V_T) [West, 2012], *i.e.* the volume taken with each breath. The normal range of minute volume at rest is about 6 - 10 $l \min^{-1}$ in adults [Guyton et al., 2012] and 240 - 360 ml $kg^{-1} \min^{-1}$ in term infants [Null and Suresh, 2017], with a tendency for the average $\dot{V}_E$ to increase with the gain in body weight during the first nine weeks of life [Cross and Oppe, 1952].

2.2.3   Regulation and respiratory control

At rest, respiration is an involuntary act initiated at the brain stem in the respiratory centres located in the *medulla oblongata*. Ventilation is thus normally controlled by the autonomic nervous system, with only limited emotional (via the limbic system), thermal (via the hypothalamus) or voluntary (via the cerebral cortex) override.

Central and peripheral chemoreceptors, and mechanoreceptors play an important role in the regulation of respiration by providing feedback input. Whilst central chemoreceptors located at the medullar level are sensitive to pH changes [Null and Suresh, 2017],

---

[1]Thus respiratory minute volume is not volume, as its name implies, but a flow.

Respiration-induced changes in cardiac filling volumes, pressures and frequency. Respiration causes blood volume variations on both the arterial and venous sides. Respiratory variations are more pronounced in venous return than in stroke volume from the right chamber, and even more so than in left stroke volume. This is due to the greater buffering capacity of the right heart chamber and of the pulmonary circulation.

During inspiration	During expiration
Intra-thoracic pressure ↓	Intra-thoracic pressure ↑
Venous pressure ↓	LV Stroke volume ↑
Venous return ↑	HR ↓
RV Filling pressure ↑	Peripheral arterial BP ↑
RV Stroke volume ↑	Peripheral venous BP ↑
LV Stroke volume ↓	
Systolic blood pressure ↓	
Diastolic blood pressure ↓	
HR ↑	

peripheral chemoreceptors, spread over the arterial system sense arterial $[O_2]$, $[CO_2]$[1] and pH. In addition mechanoreceptors are located in the respiratory airways and the parenchyma, and act to regulate the respiration through a variety of reflex responses.

2.2.4 Interaction between respiration and circulation

Respiration and circulation are intrinsically related systems in biological organisms.

This interdependence is apparent in the coupling between the physiological variables which characterise the function of the two systems; namely, heart rate (HR), stroke volume (V_s) and blood pressure (BP) — in the case of the cardiovascular system — and respiratory rate, tidal volume (V_T) and airway pressures — in the case of the respiratory systems — to name but a few.

The inventory of interactions between these variables (henceforth referred to as cardiorespiratory variables) is wide and forms the subject of many physiology textbooks. In summary, it may be subdivided into direct pressure components, neural components and circulatory components [Barrett and Ganong, 2010]. Below we describe the two relationships between these variables which are commonly explored by indirect methods for the measurement of respiration via electrocardiography or photoplethysmography used later in this thesis.

The Traube-Hering-Mayer (THM) waves

In the late 1800s, Traube [1865] and Hering [1869] reported that the arterial blood pressure varied with respiratory frequency. The so called Traube-Hering waves, are generated when respiratory-induced variations in intrathoracic pressure are transmitted to the central veins in the manner summarised in Table 2.2. During inspiration, venous pressure is decreased and venous return is increased. The filling pressure of the right ventricle (RV) increases, and right-ventricular stroke volume increases. Blood is pooled in the lungs

---

[1][·] denotes the concentration of a chemical species in a mixture.

and the left stroke volume is decreased. Systolic blood pressure decreases, as does, to a smaller extent, diastolic blood pressure, and heart rate increases. With expiration, the blood flow to the left side of the heart is increased, left stroke volume increases, and in the periphery, both the arterial blood pressure and the peripheral venous pressure increase [van Ravenswaaij-Arts, 1993].

The Bainbridge reflex

During inspiration, the intra-thoracic pressure is decreased, thus increasing venous return to the right heart and causing the right atrium to stretch. An increase in the activity of the atrial baroreceptors leads to an elevation of heart rate by suppressing vagal activity, an association commonly referred to as the respiratory sinus arrhythmia (RSA). Conversely, during expiration, the reverse of this process results in a momentary decrease in heart rate. The reflex whereby these baroreceptors signal the medullary control centres to decrease parasympathetic tone via the vagus nerve to the heart is the so-called Bainbridge reflex [Hakumäki, 1987].

2.3 Elements of respiratory pathophysiology

The differences in lung physiology and pulmonary mechanics in neonates, particularly in those born prematurely, expose them to respiratory conditions not usually found in adults or even older children. The main respiratory disorders which afflict this population are summarised below[1] with a brief description of their epidemiology, physiological background, and clinical management.

2.3.1 Neonatal respiratory distress syndrome

Pathophysiology Also known as hyaline membrane disease or surfactant deficiency lung disease, this condition is caused primarily by inadequate production of pulmonary surfactant by structurally immature type II pneumocytes.

With surfactant deficiency, a greater distending pressure is needed to maintain the integrity of the alveoli. Without sufficient airway pressure, several areas of the lungs collapse (the lungs become diffusely atelectatic [Holme and Chetcuti, 2012]), triggering inflammation and pulmonary oedema [Kramer et al., 2009]. As blood perfusing atelectatic areas of lung is effectively not oxygenated, the infant can become hypoxaemic. To aggravate this, a decrease in lung compliance increases the work of breathing. In severe cases, this causes the diaphragm and intercostal muscles to fatigue, causing a retention of CO_2 and respiratory acidosis to develop [Holme and Chetcuti, 2012].

---

[1]The neonatal infant is vulnerable to a wide range of respiratory disorders [Gallacher et al., 2016]. Emphasis was placed on the most common conditions.

Presentation   Regarding its presentation, symptoms of respiratory distress syndrome include tachypnoea (RR > 60 bpm) with rapid, strained, grunting breathing accompanied by sternal retractions and nasal flaring. As respiratory failure progresses, symptoms worsen, and may include cyanosis (a result of hypoxaemia), lethargy, irregular breathing, and apnoea [Hermansen and Lorah, 2007]. The diagnosis usually requires a chest x-ray.

Epidemioloy   Respiratory distress syndrome is the most common respiratory disorder of premature infants. As surfactant is not produced in adequate amounts until the late stages of gestation (34 to 36 weeks of gestational age[1] [Nkadi et al., 2009]), respiratory distress syndrome typically affects preterm infants below this age; it occurs in one third of infants born at 28 to 34 weeks' gestation, but occurs in less than 5 % of those born after 34 weeks' of gestation [Association, 2006]. Nevertheless, older infants with delayed lung maturation due to several aetiologies can also be afflicted [Rodriguez, 2003].

Prognosis and clinical management   If detected, the prognosis of this syndrome is positive with mortality under 10 % [Gibson and Nawab, 2015]. Treatment consists in the administration of exogenous surfactant complemented by O_2 delivery and mechanical ventilation as required. Respiratory distress syndrome resolves within 4 or 5 days, however, in severe cases the severe hypoxaemia can result in multiple organ failure and death [Gibson and Nawab, 2015].

2.3.2   Transient tachypnoea of the newborn

Pathophysiology and epidemiology   Transient tachypnoea of the newborn is a transient complication caused by a delay in the resorption of foetal lung fluid after birth. It occurs in approximately 1 % of preterm infants and 0.36 - 0.57 % of term infants [Gallacher et al., 2016]. Although its aetiology is not fully understood, it is generally agreed that transient tachypnoea has a higher incidence in neonates delivered by caesarean section [Tutdibi et al., 2010]. The increased fluid in the lungs of infants delivered in this manner leads to increased airway resistance and reduced lung compliance.

Presentation   As a result, the presentation of transient tachypnoea of the newborn is similar to that of neonatal respiratory distress syndrome, and is generally seen as a cause of or differential diagnosis to this condition. It is suspected when the infant develops tachypnoea, intracostal and subcostal retractions, grunting, nasal flaring, and possible cyanosis [Gallacher et al., 2016].

---

[1]There are three different quantities that are commonly used to describe the age of a neonate. We define them here. Gestational age (g.a.) is measured in weeks from the day of the last menstrual period of the mother. Gestational age increases through gestation but becomes fixed at birth. The chronological age is the age of the infant since birth and is sometimes given in days. The postmenstrual age is the sum of gestational and chronological ages.

Prognosis and clinical management   Transient tachypnoea of the newborn is generally a benign condition and it is spontaneously self-resolving although recovery is normally supported by oxygen delivery through non-invasive strategies [Gibson and Nawab, 2015].

2.3.3   Apnoea of prematurity

Pathophysiology   Apnoeic episodes are defined as absent airflow for a time period in excess of 20 s [Committee on Fetus and Newborn, 2003] and are conventionally classified based on the presence or absence of upper airway obstruction into either obstructive apnoeas or central apnoeas [Guyton et al., 2012].

Central apnoeas are characterised by a pause in alveolar ventilation due to a lack of diaphragmatic activity without evidence of airway obstruction. Obstructive apnoeas emerge when the respiratory effort meets an obstruction in the upper airways, resulting in chest wall motion without airflow through the lower airways. In practice, most apnoeas occur as mixed apnoeas [Dransfield et al., 1983], *i.e.* a combination of weak respiratory drive (central apnoea) and failure to maintain a patent airway (obstructive apnoea).

Regarding their aetiology, central apnoeas result from decreased nervous output from the respiratory centre, which in turn is due to the immaturity of the respiratory control system at this centre.

The immaturity of the respiratory control system [Mathew, 2011] usually spans all levels of the system reviewed in Section 2.2.3, including respiratory rhythmogenesis at the brainstem, and peripheral and central chemoreceptor responses[1]. Obstructive apnoeas, on the other hand, most often occur during sleep, when a portion of the upper airways is collapsed; the predominant site of airway obstruction is the pharynx, which shows reduced muscle tone during this period [Mathew, 2011]. In fact, in the absence of muscle activity, the pharynx of a neonate whose head is in the neutral position may become occluded even at the transmural pressures seen during tidal breathing [Wilson et al., 1980].

Presentation   In terms of presentation, whilst central apnoeas are quickly followed by a bradycardia[2] (see Figure 2.3a), obstructive apnoeas are often associated with an initial tachycardia followed later by hypoxic bradycardia (see Figure 2.3b) [Cunningham et al., 1992]. In fact, in the absence of monitors that measure airflow in the upper airways, mixed or obstructive apnoea are likely identified as episodes of bradycardia with an accompanying desaturation.

In the majority of such episodes, apnoea or hypopnoea (possibly accompanied by a

---

[1]Histologically, immaturity of the preterm brain is manifested by a decreased number of synaptic connections, dendritic arborisations and poor myelinisation [Abu-Shaweesh and Martin, 2008].

[2]Bradycardia and tachycardia are defined as a decrease or increase in HR from baseline. Regarding neonatal resuscitation, bradycardia is concerning when the HR < 100 bpm [Fleming et al., 2011].

(a) Central apnoea.

(b) Obstructive apnoea.
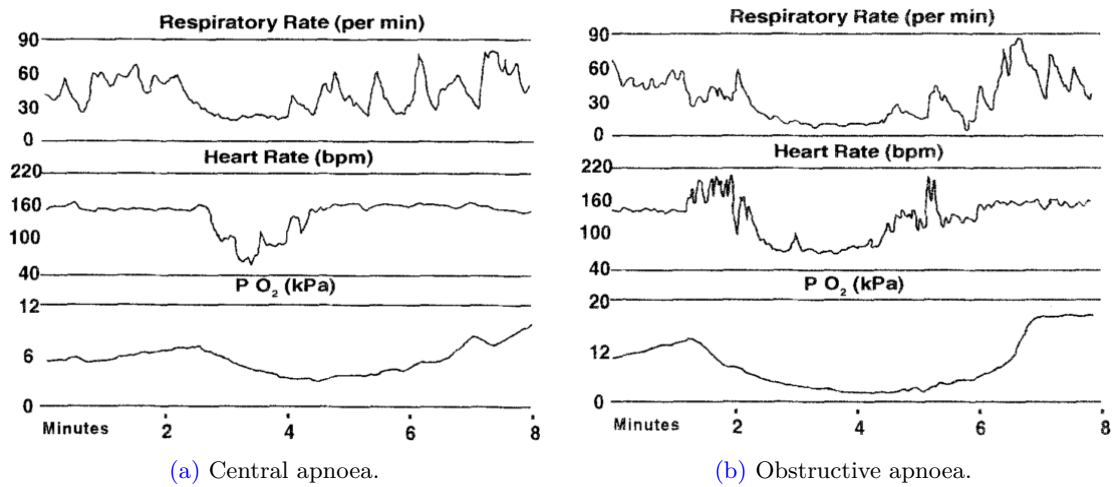
Figure 2.3

Presentation of apnoeic episodes. Respiratory rate, heart rate and partial pressure of oxygen (PO_2) are shown for 8-minute periods surrounding a central apnoea (2.3a) and an obstructive apnoea (2.3b). Reprinted from Cunningham et al. [1992].

fall in lung volume) is the initiating event, causing a fall in oxygen saturation, which in turn triggers a reflex bradycardia [Adams et al., 1997]. Other sequences may also occur; for example, during swallowing, the activation of laryngeal receptors may simultaneously trigger apnoea and bradycardia. Finally, desaturation as a primary event (for example, due to shunting) may in turn cause respiratory depression and further aggravate hypoxaemia with a reflex bradycardia. In practice, as illustrated in Figure 2.4, the relationship between neonatal apnoea, bradycardia, and desaturation is a complex one [Martin and Fanaroff, 1998].

Epidemiology   The development of respiratory control begins early in gestation but its maturation continues through weeks or months after term birth [Carroll and Agarwal, 2010]. Thus, the breathing pattern of preterm and term infants is often irregular, periodic[1], and often associated with severe apnoeic episodes.

There is little epidemiologic evidence to assess the maturation of respiratory control in premature infants. The risk for apnoea of prematurity is associated with a lower gestational age at birth as well as lower birth weights. Almost all infants born with fewer than 28 weeks of gestation suffer from apnoea. For infants born at 30 - 31 weeks, this risk falls to approximately 50 %, and, for those born at 32 - 33 weeks, the risk is furter reduced to 14 %. The risk for those born at 34 - 35 weeks is 7 % [Henderson-Smart, 1981].

Central apnoea accounts for approximately 10 % to 25 % of all cases of apnoea in

---

[1]Periodic breathing is characterised by regular, recurring cycles of breathing with a duration of 10 to 15 seconds that are interrupted by pauses of at least 3 seconds in duration [Martin et al., 2002]. Although some authors have recently hypothesized that exaggerated periodic breathing can be pathologic in newborn infants [Mohr and Vergales, 2015], it is generally considered to be a benign respiratory pattern for which no treatment is required.

NICU infants, with obstructive apnoea accounting for 10 % to 25 % and mixed for 50 % to 75 %. In each individual infant, one of these conditions tends to predominate [Stokowski, 2005].

Prognosis and clinical management   Apnoea of prematurity remains a major clinical problem in neonatal care. All newborns of fewer than 34 weeks gestational age, or less than 1800 g birth weight, should be monitored for both apnoea and bradycardia. This is done by respiratory and heart rate monitors at the cotside.

Apnoea of prematurity frequently persists beyond term gestation in infants delivered at prior to 28 weeks' of gestation. If apnoea is not managed in an appropriate medical setting, the morbidity and mortality can be significant. Thus, persistent apnoea events contribute to prolonged hospitalisation times in neonatal units.

As different types of apnoea have their own unique set of possible causes, their outcomes vary significantly and morbidity and mortality rates are difficult to assess, ranging from benign to severe illness.

The management of apnoeas to prevent sequealae is tailored to acute and chronic apneoic episodes. In acute scenarios when the infant becomes apnoeic, pale, cyanotic or bradycardic, tactile stimulation should be given. If the infant does not respond, bag and mask ventilation, along with suctioning and airway positioning, may be required [Chadha, 2010].

The chronic management of apnoea of prematurity always involves the identification of potential aetiologies and their correction. Poor respiratory control, especially in very preterm infants, might require the use of methylxanthines (such as theophylline of caffeine)

[Mathew, 2011]. Continuous Positive Airway Pressure (CPAP) is effective in treating both obstructive and mixed apnoea [Pantalitschka et al., 2009] by providing a continuous distending pressure in order to keep the airway open [De Paoli et al., 2008]. CPAP is most commonly delivered by nasal prongs, facial mask or an endotracheal tube placed in the nasopharynx. Finally, if significant apnoea persists despite the administration of pharmacotherapy and CPAP, the infant should be intubated and mechanically ventilated [Rocker and Bachur, 2010].

2.3.4   Congenital pneumonia and neonatal sepsis

Pathophysiology and presentation   Pneumonia and neonatal sepsis are associated with infection by viruses, bacteria, or fungi, the most common organism responsible being group B streptococcus [Heath et al., 2004]. Although their risk factors are multifactorial [Gallacher et al., 2016], we include both conditions in this section since they are often accompanied by signs of respiratory distress and respiratory symptoms, such as apnoea, tachypnea, grunting, nasal flaring, and intercostal retractions, are common [Simonsen et al., 2014].

Epidemiology   Congenital pneumonia is responsible for 4.5 neonatal deaths per 100 000 births per year in the UK [Tambe et al., 2015].

Clinical management   Antibiotic treatment is the main form of treatment although supportive interventions such as oxygen and mechanical ventilation may be required [Gallacher et al., 2016].

2.4   Respiration monitoring

In 1990, the World Health Organisation (WHO) established manual observation consisting of a manual breath count with a timer (to compute respiratory rate) and concomitant clinical observations (to detect abnormal breathing patterns, coughing, wheezing, grunting, asymmetries in the filling of the lungs, cyanosis, *etc.*) as the recommended procedure for respiratory monitoring [World Health Organization, 1990]. Whilst it is still the gold standard in several areas of clinical practice, there are clear shortcomings to this approach.

First, counting the number of breaths over the required amount of time (60 s, as promoted by the WHO) can be especially prone to errors in ill children, who may breathe at a rate that is upwards of 60 - 70 bpm. In practice, in busy clinical settings, healthcare workers rarely count breaths for a full minute, but instead count for only a fraction of 60 s and then scale up the number of breaths. This decreases the accuracy of the measurement by amplifying the counting error [Lovett et al., 2005]. Secondly, studies have shown that in conscious adult subjects, the awareness that one's breathing is being counted triggers

Methods for respiratory monitoring. PPG = Photoplethysmography; ECG = Electrocardiography.

Conventional methods	Camera methods
Flow-based sensors	PPG-based methods
- Pneumotachograph	
- Thermal sensors	Vision-based methods
	- Frame differencing
CO_2 sensors	- Depth sensing
	- Optical flow & motion tracking
Chest wall motion sensors	
- Impedance pneumography	
- Inductance plethysmography	
Indirect methods	
- ECG-derived respiration	
- PPG-derived respiration	

a cortical response which overrides the mechanisms that regulate respiratory rate or tidal volume at rest [Larsen et al., 2010], thus causing misleading readings. Finally, the requirement for a human observer prohibits continuous monitoring over long periods of time.

There is a wide range of devices available to measure respiration in infants (Table 2.3). Traditional devices are divided into three basic categories: flow sensors applied to the nose and/or mouth, devices that detect chest wall excursions that are applied to the chest and/or abdominal area, and devices which measure respiration indirectly, of which the electrocardiogram (ECG) and the photoplethysmogram (PPG) are prime examples.

As alluded to in Section 2.2.4, the human respiratory system performs several functions. In practice, no single device can simultaneously probe all of these functions. Hence, devices must be selected which enquire the respiratory functions relevant to a specific patient prognosis.

For example, whilst flow sensors at the nose can generally measure ventilation with high accuracy, they are inadequate systems during episodes of obstructive apnoea, where they fail to measure the flow of air into the lungs as they can only access the flow across the airways superior to the airway obstruction. Alternatively, if one wishes to assess the efficiency of the gaseous exchange processes in the lung, an assessment of the blood gas status would be required, such as that provided by a pulse oximeter or a pneumotachograph equipped with a $[CO_2]$ meter.

In summary, the choice of equipment implies a trade-off between accuracy in the assessment of respiration functions and invasiveness. In the remainder of this chapter, we discuss the advantages and disadvantages of current approaches to respiration monitoring in neonatal and paediatric care and motivate the development of non-contact solutions based on camera technology as an effective approach to this problem.

2.4.1 Current monitoring practices

2.4.1.1 Flow sensors

Exhaled air is warmer, has higher humidity and contains more CO_2 than inhaled air. These variations in pressure, temperature and gas content can be used for measuring flow rate in the upper airways.

Pneumotachography

Pneumotachographs measure flow rate as a differential pressure across a small flow resistance, a physical concept which was first used to measure respiratory airflow by [Gerrits and Muntinghe, 1955] in patients with asthmatic bronchitis.

Two types of pneumotachograph are commonly used: laminar flow and fixed orifice devices [Primiano Jr, 2009]. Laminar flow devices (Figure 2.5a) consist of a tubular compartment containing a resistive element designed to maintain laminar flow inside the lumen of the device. This can be achieved using a series of small-bore tubes arranged longitudinally (thus reducing the Reynolds number of each flow contribution) or a metallic mesh through which the gas flow must pass [Tagawa et al., 2011], although other designs exist. Due to the laminar nature of the flow, the gas pressure drop across this element can be linearly related to the flow rate through the compartment (through Poiseuille's law) with a constant determined through calibration. A heating element is often incorporated to prevent the condensation of moist gases within the flow chamber which could compromise accuracy [Mandal, 2006].

Fixed orifice pneumotachographs are placed within a tubular attachment to the respiratory airways with airflow being estimated from a pressure differential measured by transducers at either end of the device. With this class of devices, while laminar flow cannot be guaranteed, and so the measurements are not linear, they are usually sufficiently reproducible for frequency measurements.

The use of these sensors in the clinical setting is seen in both mechanically ventilated

(a)   (b)
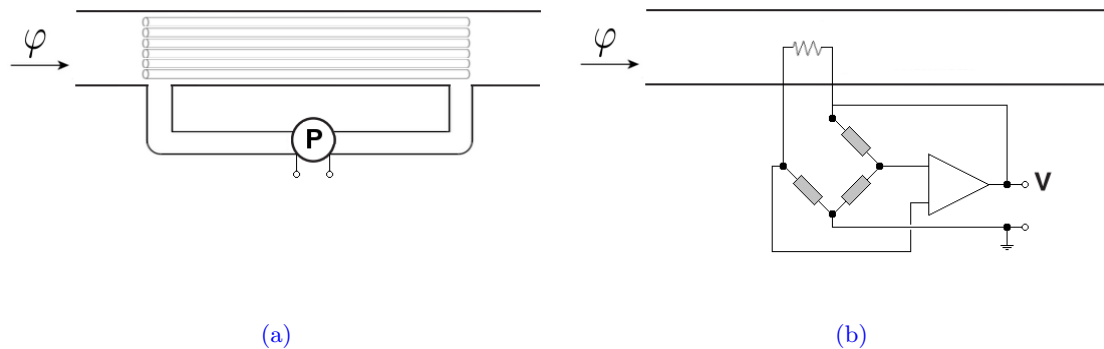
Figure 2.5
Block diagrams of (2.5a) flow pneomotachograph (laminar flow device) and (2.5b) hotwire anemometer.

subjects (attached to the endotracheal tube) since the mid-1970s [Folke et al., 2003] and more recently also in spontaneously breathing subjects (attached to a mouthpiece or a face mask).

The typical morphology of pressure flow signals measured at the opening of the nostrils (φ_{nasal}) for a spontaneously breathing subject is shown in Figure 2.6 .

Although the pneumotachograph is considered to be the gold standard for measuring flow and tidal volume [Folke et al., 2003], there are some drawbacks to its use in specific patient groups. In intubated subjects, the added resistive load placed by these devices can pose a challenge in patients unable to generate effective intrapulmonary pressure differentials, such as infants with extremely low birth weight[1] (ELBW). In spontaneously breathing patients, some authors have measured alterations to the normal breathing pattern caused by the use of nasal masks [Askanazi et al., 1980; Hirsch and Bishop, 1982].

Thermography

One of the most frequently used methods to monitor respiration is to detect airflow at the opening of the upper airways using a nasal thermistor, or a thermocouple sensor, which senses air temperature changes.

The most common form of thermal-based flow sensors is the hot-wire anemometer (Figure 2.5b). This device operates using a heated filament placed in the gas flow pathway. As the airflow passes through this chamber, the air stream cools this filament by convection. A current is applied to the hot filament so as to maintain a constant temperature (around 400 ^{o}C for platinum filaments [Mandal, 2006]). The rate of heat transfer at the filament depends on the gas flow rate and the thermal conductivity of the gas. Hence, after calibration, the gas flow rate can be calculated from the intensity of the correcting current [Yoshiya et al., 1979]. Other types of anemometers also exist based on constant current or constant voltage.

Hot-wire anemometers are usually attached to a face mask and placed proximal to

---

[1]Infants who weigh less than 1000 g at birth are extremely low birth weight infants, while those who weigh less than 15000 are very low birth weight (VLBW) infants. For reference, 90 % of infants born full term (37 weeks g.a.) weigh over 2500 g [Oken et al., 2003].

the mouth or nose, but also on the walls of the pharynx. Due to their lower resistive load [Primiano Jr, 2009], these sensors have also been incorporated in the endotracheal tube of some commercially-available ventilators. Because of their high frequency response, anemometers have also been applied in the high-frequency oscillation ventilators [Zimova-Herknerova and Plavka, 2006] used in neonatal clinical practice.

2.4.1.2 CO_2 sensors

The concentration of CO_2 measured in the expired air is approximately 4 % (V/V) [Guyton et al., 2012], a figure which is similar to concentration of this gas the alveolar volume. For reference, its athmospheric concentration is around 0.04 % [Dlugokencky and Tans, 2015]. The measurable difference in the end-tidal concentration of CO_2 with respect to atmospheric levels is the premise of this class of methods, which monitor respiration by measuring the concentration or partial pressure of this species at the opening of the nose or mouth. This is achieved through capnography.

Initial realisations of this technique used mass spectrometers [Lenz et al., 1991] to sense end-tidal tidal $[CO_2]$, but current monitors use IR spectroscopy based on absorption in a band near 4.3 μm (*e.g.* Kavanagh et al. [1992]). Light emitted through the gas sample at this wavelength is preferentially absorbed by CO_2 molecules with the light absorbance being proportional to the concentration of the absorbing molecules (Beer's Law). In practice, black-body radiators are usually used as emitters, with pyroelectric devices used as detectors [Johansson and Hök, 2008].

The large oscillation in $[CO_2]$ observed during the respiratory cycle explains the high sensitivity of this approach, whose only limitations are the relatively large size of the IR-based $[CO_2]$ sensors required to obtain adequate resolution and the need for an invasive nasal cannula. Additionally, in patients who alternate between mouth and nose breathing, the exclusive monitoring of a single channel can cause disturbances.

2.4.1.3 Chest wall motion sensors

Impedance Pneumography

With the advent of cardiography, early physiology studies have noted that respiratory-related fluctuations could be distinguished in electrical impedance signals recorded by electrodes placed on either side of the human chest [Allison et al., 1964; Goldensohn and Zablow, 1959].

Although in this context these fluctuations were considered to be a corruption of the cardiographic signal, and thus respiration a source of noise to be extracted or otherwise limited, it was soon noted that it could offer a viable measurement of the subject's respiratory effort [Sherwood et al., 1990]. In fact, transthoracic impedance pneumography

(IP) has become an ubiquitous technique to monitor RR in critical settings, particularly in intensive care scenarios where patients are already monitored by electrocardiography, as is the case in neonatal units [Monasterio et al., 2012].

IP is implemented using a set of two or four surface electrodes [Prutchi and Norris, 2004] placed across the chest (Figure 2.7). Under both configurations, a potential difference between the voltage sensing electrodes on the surface of the chest is generated by injecting a low-amplitude, high-frequency (50 to 500 kHz) alternating current (I_{AC}) through the drive electrodes. The corresponding voltage change can be then measured across the same electrodes (bipolar measurement), or across a different pair of electrodes (tetrapolar measurement). Both electrode arrangements are shown in Figure 2.8. The electrical impedance across the thorax (*i.e.* the transthoracic impedance) is found trivially as the ratio between the voltage difference of the two receiving electrodes to the current that flows between the two and through the tissue.

In the absence of respiratory motion, the baseline impedance across the chest at rest is kept approximately constant (typically around 500 Ω [Grenvik et al., 1972] in adults and 40 Ω [Sexson et al., 1991] in preterm infants). During inspiration, transthoracic impedance increases mainly due to two factors: (a) an increase in the gas volume in the chest in relation to (higher conductivity) fluids, and (b) an increase in the electrical path as a result of chest expansion.

Depending on the physiognomy of the monitored subject, these respiration-induced changes in impedance can range from 0.1 Ω to 1 Ω [Gupta, 2011]. The voltage difference between the two receiving electrodes, in turn, depends on the magnitude of the injected current.

Monitoring of respiration rate by IP comes with an array of advantages. It is minimally-invasive and very convenient if the patient is already monitored with ECG electrodes. However, this method is not without deficiencies, and a number of factors which lead to

(a)   (b)

Figure 2.7
Neonatal impedance pneumography (IP). The IP electrodes are positioned bilaterally across the infant's chest and below the insertion of the diaphragm [Di Fiore, 2004]. (2.7a) Whole body view. (2.7b) Front view.

inaccurate readings have been identified. Firstly, the surface electrodes record impedance changes across all tissues, including muscle, rendering this technique very susceptible to motion artefact. Additionally, the superimposition of cardiac-synchronous variations, usually referred to as the cardiogenic artefact, is a source of error inherent to impedance measurements. Lastly, similarly to other respiration monitoring techniques whose measurement principle is based on the respiratory effort, it cannot highlight episodes of obstructive apnoea, a severe limitation of this form of monitoring when used on its own.

Impedance Plethysmography

In parallel with the developments in electrical impedance pneumography, an alternative approach to monitoring chest excursions has been proposed which exploits the behaviour of inductive materials under strain. Proposed in 1977 by [Milledge and Stott, 1977], this technique was initially applied in paediatric and veterinary asthma research and later adapted for use in human polysomnography during the late 1990s [Primiano Jr, 2009].

In present-day configuration, respiratory inductance plethysmography (RIP) uses a set of two belts placed around the chest and abdomen (Figure 2.9). Each of these belts contains a sinusoidal coil sewn into their elastic mesh. The working principle can be understood from the Lenz's law of electromagnetic induction, according to which a change in the area enclosed by a conducting wire loop generates an opposing current inside the wire proportional to this change in area. Thus, if a small AC current is injected through this coil (thus generating a magnetic field by Faraday's law), changes in the cross-sectional area of the patient's body can be measured as changes in inductance (as the coil elongates or shortens).

The strengths of this technology are clear; by monitoring both rib cage and abdominal excursions, RIP allows for the detection of obstructive apnoea presented as paradoxic out-of-phase excursions in the abdominal and thoracic signals. In addition, studies on
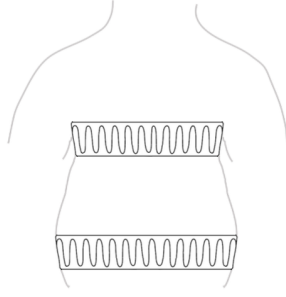
neonates have shown that, given appropriate calibration, this technique can provide *semi-quantitative* estimates of tidal volume in preterm infants that show good agreement with a pneumotachograph [Brooks et al., 1997]. These advantages come at the expense of increased invasiveness and a more delicate set-up. For this reason, RIP has been used more extensively in laboratory settings, in the the context of sleep studies and home monitoring.

2.4.1.4 Indirect methods

ECG-derived respiration

Electrocardiography(ECG) consists in the recording of the surface electric potentials generated by the activity of the heart. In critically-ill patients, ECG monitoring is primarily used to detect instantaneous changes in HR, which may or may not be a consequence of respiratory events.

To detect these potentials, electrodes are attached to the surface of the skin. Conventional heart rate monitoring in intensive care uses a six-lead set-up. In neonatal care, however, three smaller-sized electrodes are placed on the infant's body with the usual placement being right arm (RA), left arm (LA) and left leg (LL) or abdomen, as shown in Figure 2.10.

According to the mechanisms of cardiorespiratory interplay reviewed in Section 2.2.4, cardiac frequency is modulated by respiratory-induced changes in intra-thoracic pressure. Although the primary aim of standard electrocardiography is the surveillance of cardiac events, there is strong evidence to suggest that it can also provide an indirect measurement of the modulating respiratory signal.

In general, ECG-derived respiration can be obtained from two phenomena: changes in beat morphology and respiratory sinus arrhythmia.

Figure 2.10

Neonatal electrocardiography. Two ECG leads are attached to the chest (RA and LA) and one to the left leg or lower abdomen (LL); RA = right arm; LA = left arm; LL = left leg. (2.10a) Whole body view. (2.10b) Front view.

Beat Morphology

During the respiratory cycle, the filling of the lungs with inspired air causes a periodic shift in the projection of the electrical axis of the heart over the ECG leads [Moody and Mark, 1985]. This shift can be identified as an oscillating QRS amplitude in the electrocardiographic signal recorded by each lead (Figure 2.11).

During normal respiration, several factors are responsible for this phenomenon; (a) the thoracic expansion causes the heart to move and to rotate with regards to the electrode positions, and (b) as air fills the lungs, the impedance across the chest (and hence between the ECG electrodes) is increased. Although both factors create a modulating effect on ECG based on the respiratory cycle, the former is widely regarded as the main contributor of the observed amplitude modulation of the signal [Moody and Mark, 1985].

An advantage of this approach is that the performance of QRS detection algorithms (upon which it depends almost solely) have become relatively robust to muscle artefact
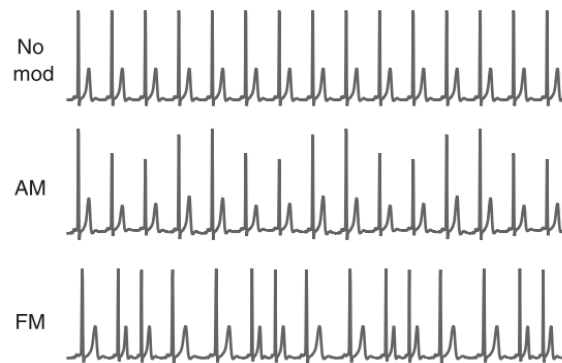
Figure 2.11

Modulation of the ECG waveform with the respiratory signal. From top: no modulation, amplitude modulation through changes in teh height of the QRS complex (AM), and frequency modulation by the RSA (FM). Adapted from Charlton et al. [2016] with permission.

[Helfenbein et al., 2014]. A drawback of this approach is that its measurand is strictly the respiratory effort. Therefore, it may not provide an accurate reproduction of the respiratory cycle in instances where chest movement occurs without airflow, a scenario which occurs during episodes of obstructive apnoea.

Respiratory sinus arrhythmia

The second approach relies on the modulation of heart rate with the respiratory-induced pressure changes in the main and peripheral arteries.

This phenomenon, which physiology books commonly refer to as the respiratory sinus arrhythmia, can be summarised as follows: during inspiration, the intra-thoracic pressure is decreased, thus increasing venous return to the right heart, causing the right atrium to stretch and an elevation of heart rate through the Bainbridge reflex described in Section 2.2.4. Conversely, during expiration, the reverse of this process results in a slowing of instantaneous HR (Figure 2.11), or equivalently, an elongation of inter-beat intervals (often measured between two consecutive R waves). One should note that respiratory activity is not the sole modulator of this time series. In reality, both sympathetic and parasympathetic activity of the nervous system make frequency-specific contributions to the heart rate power spectrum [Akselrod et al., 1981]. Both systems are considered to be involved in the short-term control of heart rate fluctuations (on the time scale of seconds to minutes). These systems act to maintain homeostasis by responding to perturbations sensed by a collection of chemo and baroreceptors and modulating beat-by-beat cardiac output accordingly.

Work by Selman et al. [1982] and Eckberg [1997] has revealed that parasympathetic activity largely accounts for high frequency modulations (0.15 to 0.4 Hz), while low frequency ones (0.04 to 0.15 Hz) are comprised of outputs of both parasympathetic and the sympathetic activity. The RSA (which is parasympathetically mediated) usually manifests itself as the peak in the high frequency band. A sample spectrogram of heart rate variability is given in Figure 2.12. Given the normal respiratory range (see Table 2.1) the respiratory component can at different times display some degree of overlap with both LF and HF spectral components, especially at very low respiratory rates.

There exist, however, limitations to the use of RSA for the extraction of respiration. There is a marked age dependence in the prominence of the RSA component in the heart rate variability (HRV) spectrum. While RSA is very pronounced in children, without sufficient cardiovascular exercise it is gradually attenuated as individuals approach adulthood. In addition to maturational factors, it is also affected by physiological ones. Whilst adults in excellent cardiovascular health (such as athletes in heavily aerobic sports) can typically maintain a strong vagal tone[1]), poor cardiovascular health, or the onset of neurological conditions affecting the parasympathetic system can also severely affect the

---

[1]often used as synonym of parasympathetic activity, since it is mediated by the *vagus* nerve.
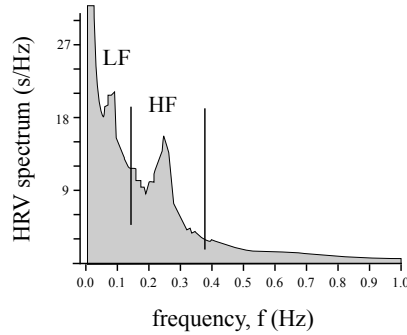
prominence of this frequency component [Ravenswaaij-Arts et al., 1993], thus rendering respiratory monitoring through tracking the frequency of RSA oscillations an inadequate technique in some cohorts, such as patients with diabetes or cardiovascular disease.

The existing methodology for extracting respiratory information from both amplitude modulation and frequency modulation of the ECG is reviewed in Chapter 4.

PPG-derived respiration

Photoplethysmography (PPG) is a well-known optical technique used to measure volume changes at capillary level and its variations due to blood flow. In its most basic configuration, PPG employs an infra-red (IR) light source in the vicinity of a photodetector unit, which integrates the light emitted by the former and transmitted through or reflected from a body part [Kamal et al., 1989].

The underlying signal of interest is the blood volume pulse (BVP) that propagates throughout the body in time with heart rate [Allen, 2007]. During the cardiac cycle, the BVP causes volumetric changes in the vascular bed of the skin dermis, which modify the path length of incident light such that changes in amount of detected light at this wavelength (800 - 960 nm) indicate the timing of cardiovascular events. More relevant to the scope of this thesis, the PPG signal contains components synchronous not only with the cardiac pulse, but also with ventilatory rhythms [Akselrod et al., 1981].

Similarly to the ECG, the PPG signal exhibits both frequency and amplitude modulations by the respiratory cycle (Figure 2.13). Whilst the frequency modulation of the PPG signal is also known to originate from the respiratory synus arrhythmia, the mechanisms underlying the respiratory-induced intensity variations remain to be fully elucidated (see the review paper by Meredith et al. [2012] and the references therein).

The body of literature on photoplethysmographic monitoring of respiration is extensive. Although most studies to date have predominantly focused on finger and forearm plethysmography, the search for a respiratory signal has prompted researchers to examine
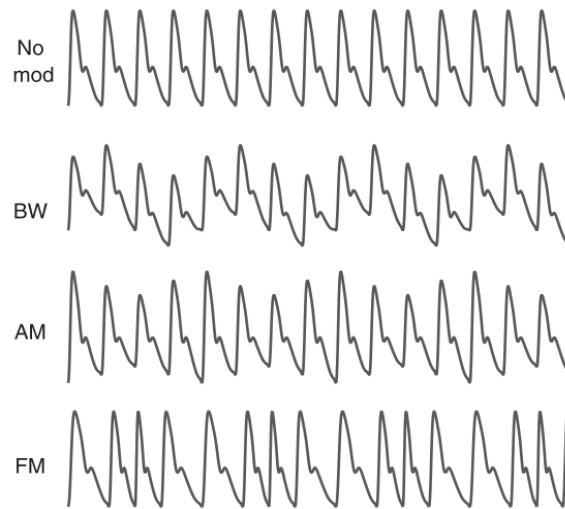
Modulation of the PPG waveform with the respiratory signal. From top: no modulation, baseline wander (BW), amplitude modulation (AM), and frequency modulation (FM). Adapted from Charlton et al. [2016] with permission.

new anatomical sites. Using spectral analysis, Shelley et al. [2006] observed that forehead and ears yield higher signal amplitude for ventilation rhythms when compared to other PPG sites, essentially due to their proximity to the chest and weaker sympathetic relay to the vasculature, thus reducing the attenuation effects of vasoconstriction on signal power.

Numerous groups have demonstrated that the extraction of the ventilatory-synchronous fluctuations from the PPG signal is technically attainable. The various methodologies to achieve this are reviewed in Chapter 4.

Photoplethysmography is often allied to pulse oximetry, *i.e.* the measurement of the relative concentration of oxygen-saturated haemoglobin species in arterial blood (spO_2). A neonatal oximeter consists of a sensor placed on the infant's hand or foot (see Figure 2.14). Since its introduction in neonatal monitoring, oximeters plays an integral part in documenting hypoxic episodes, which may occur in association with apnoea [Tin et al., 2001].
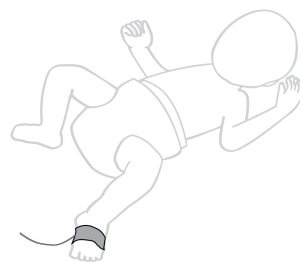
Neonatal pulse oximetry. In this scenario, the pulse oximetry is placed inside a pad conforming to the shape of the infant's foot.

2.4.2 The need for non-contact monitoring techniques

Respiration monitoring in neonatal units is generally performed with impedance pneumography. In comparison to the previously presented methods, IP is a convenient method if patients are already monitored for ECG but it presents clear technical and clinical shortcomings in the monitoring of neonates.

IP is notoriously prone to inaccurate readings due to a number of factors including poor ECG electrode placement, motion artefact and high sensitivity to physiological events which cause thoracic movements (such as coughing, or crying) [Brouillette et al., 1987; Warburton D, Stark AR, 1977]. In addition, the placement and removal of adhesive sensors and the attachment of wires can cause discomfort, stress, and even pain due to epidermal damage [Afsar, 2009; Baharestani, 2007]. The latter occurs particularly in infants of less than 27 weeks of gestational age, as the bond between the sensor and the dermis is stronger than the epidermal-dermal junction in these infants. In an infant of 26 weeks gestation the epidermis is only 2 - 3 cell layers thick and the keratinised stratum corneum barely developed [Evans and Rutter, 1986]. It is not until 34 weeks gestation that the stratum corneum is well defined and the skin functionally competent [Shwayder and Akland, 2005]. These early insults have been reported to adversely affect the infant's wellbeing and long-term developmental outcome [Chen et al., 2010]. Finally, the obtrusiveness of wire probes impairs parent-child bonding, particularly during kangaroo care[1]. A study by Anand and Scalzo [2000] has shown that pain, stress and maternal separation have a negative impact on the cognitive development of NICU patients.

In an effort to reduce patient discomfort and make respiratory measurements more ubiquitous, several methods have been proposed which do not require contact sensors to be attached to the patient, such as thermal imaging [Pavlidis et al., 2007], microwave-based [Baltag, 2008] and radar-based [Li et al., 2009] methods. Although technically these are non-contact approaches, they require specialist instrumentation and complex set-up procedures.

In the last decade, the topic of remote sensing has seen the emergence of numerous approaches to the measurement of vital signs using visible-light video cameras. This growing body of research is reviewed in Section 2.4.3. Although several studies in neonates have also been reported, most have been conducted to extract heart rate using video camera monitors, with few approaches aimed at estimating respiratory parameters. Section 2.5 addresses the use of visible-light cameras for the monitoring of premature and young infants specifically.

---

[1]Kangaroo care is a method of care, where the infant is held skin-to-skin with their parent or substitute caregiver.

2.4.3 Camera-based monitoring and prior art

2.4.3.1 Relevant literature

The aim of this section is to review previous work in the field of non-contact monitoring of respiration through visible (390 to 700 nm [Starr, 2005]) and near-infrared imaging (0.75 to 1.4 μm [Byrnes, 2009]). We cover both active and passive approaches but restrict the subject of our review to techniques based on camera devices with a spectral response in the 400 nm - 1 μm range. Most notably, this excludes the work on vital sign estimation via mid-infrared thermography by Pavlidis et al. [2007] among others [Abbas et al., 2011; Farah and Saatchi, 2012], or CO_2 imaging [Fei et al., 2005].

Several approaches have been designed to extract respiratory information from videos acquired in the 400 nm - 1 μm range and they can be categorised into either PPG-based or vision-based methods (Table 2.3). The distinction here is a subtle one. On the one hand, techniques based on video PPG or remote PPG (commonly shortened to rPPG or vPPG) regard the video signal acquired from a perfused skin region as a gateway to the underlying plethysmographic signal. The signal of interest is, therefore, the blood volume pulse. As seen in Section 2.4.1.4, according to conventional PPG theory, the BVP is modulated by respiratory events (or rather, their mechanical effects on the circulatory system) and thus the respiration waveform can be extracted from this signal using the state-of-the-art methods devised to perform this task on PPG signals acquired using standard pulse oximeters[1]. On the other hand, vision-based techniques present an alternative for extracting respiratory information by exploiting the fact that the output signal of camera devices is a sequence of images, and so carries information in features such as edges, ridges, shapes and a panoply of other visual descriptors. This approach encompasses an array of techniques which aim to perceive breathing information from a set of visual cues, in much the same way a human observer would approach this problem.

We adopt such a classification merely to provide some structure to this review. In practice (as the reader will realise over the course of this section) there is considerable overlap between the two approaches not only in aim but also in methodology. In practice, both approaches often coexist within the same paper as, strictly speaking, very stringent experimental conditions would be required to segregate the contributions of motion and the BVP and so obtain a signal that is truly plethysmographic in nature, *i.e.* arising exclusively from the blood pulsations in the microvasculature of the skin. Hence, the distinction between the two approaches should not be held too rigidly.

The remainder of this Section is thus split into the two main strands of research mentioned above (PPG-based approaches in Section 2.4.3.2 and vision-based approaches in Section 2.4.3.3). Over the course of both sections, we will review the keynote papers under these two categories and highlight the ways in which they differ.

---

[1]This topic is developed in Section 4.5.2.
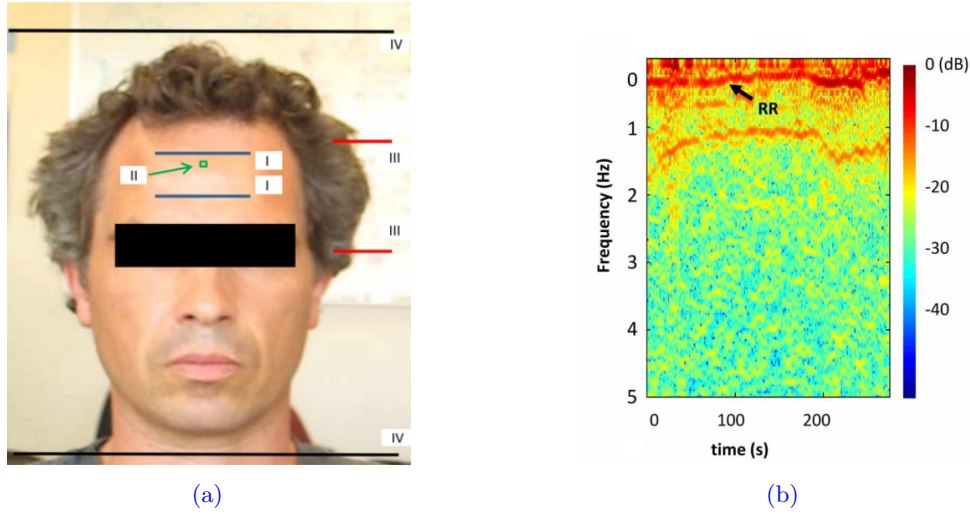
2.4.3.2 Photoplethysmographic approaches

Over the last 30 years, PPG monitoring using finger or ear pulse oximeters has become widespread in clinical practice. Thus far the optoelectronic setup for PPG has required a dedicated light source (*e.g.* at red and/or infra-red wavelengths), but recent work has shown that PPG signals can be acquired remotely [Wieringa et al., 2005], even using ambient light as the illumination source [Verkruysse et al., 2008]. In his pioneering study, Verkruysse et al. [2008] established that PPG signals could not only be acquired remotely (> 1.5 m away from the measurement site) but also from a region of interest over the human face. Additionally, signal-to-noise ratio (SNR) was such that up to four harmonics of the fundamental HR frequency were extracted along with the lower frequency RR modulation (Figure 2.15).

However, as PPG sensors are now at a distance from the measurement site, the variations in light intensity registered by them as result of subject or source motion are far more severe than the ones registered by ordinary PPG devices [Cennini et al., 2010]. Since the publication of these proof-of-principle studies, a mere ten years ago, numerous solutions have been proposed to address the low signal power and the deleterious effects of motion artefacts on remote measurements. Most of these studies suffer from deficiencies characteristic of proof-of-concept studies; most have been conducted in small cohorts of healthy cooperative volunteers and more emphasis has been placed on reporting methodologies than their performance statistics, with validation often performed on the same dataset as that used in their design and/or quantified using customised performance metrics computed for short periods of time. To summarise the vast literature available into a few paragraphs, we review the main options proposed so far to perform each of

the required processing steps along with the main lessons learned from them.

Measurement site

Regarding the site of PPG measurement, the face has so far been the preferred anatomical site [Bousefsaf et al., 2013; Poh et al., 2011; Schmitz, 2011; Sun et al., 2011; Takano and Ohta, 2007; Tarassenko et al., 2014; Verkruysse et al., 2008]. As it is generally uncovered, it accounts for a large locus of exposed skin pixels. Nevertheless, respiratory measurements from other body parts have also been documented, *e.g.* inner arm [Wieringa et al., 2005], upper chest [Zhao et al., 2013], torso [Villarroel et al., 2014] hands and extremities [Huelsbusch and Blazek, 2002; Sun et al., 2011].

Based on unpublished experiments, Verkruysse et al. [2008] claim that, when compared to other anatomical locations, they experienced stronger PPG modulation from videos of the human face, particularly over the forehead region. Their findings are in agreement with a previous study by Shelley et al. [2006], who observed that forehead and ear contact measurements yield a higher signal amplitude density for ventilation when compared to other common PPG sites. This was deemed to be a cumulative consequence of (a) the proximity of these structures to the chest and (b) the weaker sympathetic relay to the skin vasculature, and thus the reduced attenuating effects of vasoconstriction on signal power.

Selection of the region-of-interest

Given the short duration of experiments, the ROI has generally been selected manually at the start of the analysis. Automatic ROI selection has been attempted using the face detection algorithms by Everingham et al. [2006] ([Tarassenko et al., 2014]) or Viola-Jones [Viola and Jones, 2001] ([Bousefsaf et al., 2013; Poh et al., 2011; Zhao et al., 2013]) given that most of the protocols described had subjects sit facing the recording apparatus. Apart from Sun et al. [2011], who performed a trial whereby a subject underwent strenuous exercise, and Tarassenko et al. [2014] who monitored subjects in a clinical setting, no tracking or image stabilisation techniques were employed. In Sun et al. [2011], 2-D correlation between the first frame and each subsequent frame was applied for ROI registration while in Tarassenko et al. [2014] the Kanade-Lucas-Tomasi algorithm [Baker and Matthews, 2004] was used. No semantic approaches to image registration (of faces or other body structures) have been attempted.

Optoelectronic set-up

In terms of the experimental set-up, apart from one study which attempted a detection range of 5 m [Schmitz, 2011], most studies have required subjects to stand within a distance of 0.5 to 1 m from the recording device. As seen in the next Section 2.4.3.3, this
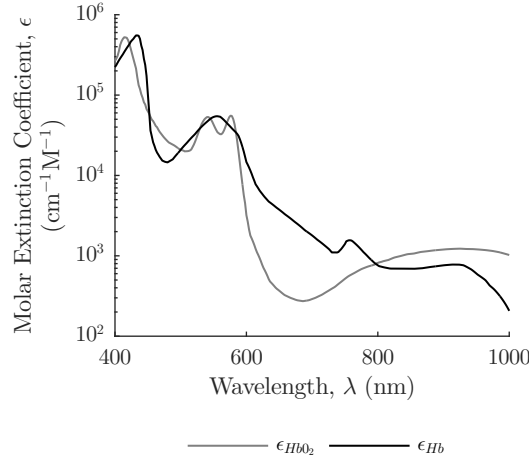
range is somewhat lower than that used in vision-based approaches although differences in set up (*e.g.* passive *versus* active imaging, or detected wavelength) mean that studies should be considered on an individual basis.

Concerning the acquisition of the optical signal, most approaches have relied on cameras sensitive to visible, *e.g.* [Aarts et al., 2013; Bousefsaf et al., 2013; Poh et al., 2011; Schmitz, 2011; Takano and Ohta, 2007; Tarassenko et al., 2014; Verkruysse et al., 2008; Villarroel et al., 2014; Wieringa et al., 2005; Zhao et al., 2013], or near-IR regions [Sun et al., 2011; Wieringa et al., 2005; Zhao et al., 2013] of the spectrum, in which case active illumination with an LED-ring light was provided.

The strongest PPG signals under visible light are found in the green channel [Verkruysse et al., 2008], in agreement with early spectroscopic studies which noted that haemoglobin has absorption peaks within the green/yellow (520 - 580 nm) region of the visible spectrum [Crowe and Damianou, 1992] (see Figure 2.16). While this may be surprising when considering that traditional PPG operates at much longer wavelengths, it is explained by the strong association of plethysmography with pulse oximetry, which has traditionally used red and IR wavelengths due to their spectral location on either side of the isobestic point[1] for oxy- and de-oxyhaemoglobin in the near-infrared (at 800 nm).

Using a data-driven approach, Lewandowska [2011]; Sahindrakar et al. [2011] published extensive studies on dimensionality reduction in colour spaces aimed at identifying the colour component with most power over the heart rate frequency band (as a proxy for PPG modulation due to the arterial pulse). A consensus was reached with both studies reporting that the R-G direction of the RGB (red, green and blue) colour space, or the *u**

---

[1]The isobestic point of a mixed solution refers to the wavelength (or frequency) at which the absorption of light by the solution remains constant as the equilibrium between the components in the solution changes. The requirement for an isobestic point to occur is, therefore, that the absorption spectra of two species, represented either as molar extinction (as in Figure 2.16) or absorbance coefficients constant molar concentration, cross each other for some wavelength.

chromaticity component in the International Commission on Illumination (CIE) Lu^*v^* system are preferable for PPG extraction.

As for sampling frequency, Sun et al. [2012] have compared the performance of a low-cost webcam and a high-sensitivity camera to conclude that, after interpolation, the signals provided by a 30 Hz webcam and by a 200 Hz camera yield estimates of HR and HRV of comparable accuracy.

Management of low signal-to-noise ratio

Remote PPG sensing is severely affected by low SNR. Huelsbusch and Blazek [2002] estimate that only about 0.1% of the backscattered signal originates from arterial blood volume, the only pulsatile component in the sampled tissue. This results in a very small AC oscillation on top of a large DC level, which consumes most of the dynamic range of recording devices. This inherent biological noise [Roald, 2013] and the flicker noise of the CCD camera can both easily be overridden by averaging the pixel intensity over enough pixels so that the subtle pulsation increases above the noise level. Motion artefact is by far the most troubling noise source, however.

Motion noise is dominant near the edges of objects in the image. Pixel values at the edges will change due to voluntary as well as involuntary movement, and both periodically (in time with RR and HR) or abruptly, as a result of sudden movements of the imaging device or in the field of view. These are normal occurrences in real-world settings, and so steep and sudden changes in DC levels must be expected. To aggravate this, since the frequency spectrum of motional noise overlaps the physiological range of the RR, they cannot be removed by linear filtering using fixed-frequency cut-offs.

To mitigate these artefacts, several strategies have been employed. Some authors have taken advantage of the multi-dimensional character of the signal input (RGB) to apply denoising tecniques based on blind source separation, such as independent component analysis (ICA) [Poh et al., 2010, 2011], principal component analysis (PCA) [Lewandowska, 2011] or constrained independent component analysis (cICA) [Tsouri et al., 2012]. Faced with a single IR channel, Sun et al. [2011] applied single-channel independent component analysis (SCICA). Zhao et al. [2013] used ICA to uncover the underlying vital signals from an embedding matrix constructed from a series of delayed versions of a single-channel video signal. Cennini et al. [2010] implemented a hybrid approach (combining hardware and software) based on the combination of multispectral illumination and least-mean squares adaptive filtering. Cennini assumed that the short penetration depth of the blue light made the intensity of the reflected signal at this wavelength a good substitute of motion-induced changes alone, which in turn would contribute to the signal reflected at other wavelengths in a linearly additive way. Whilst all these denoising approaches have shown good performance on the datasets they were tested (and developed) in, in the absence of studies which compare their performance on

the same dataset, their relative merits are difficult to assess.

In the frequency domain, in an attempt to filter out undesirable spectral components and isolate normal breathing, authors have resorted to digital filters in conjunction with frequency-based analysis, such as Fourier analysis [Poh et al., 2011; Schmitz, 2011; Takano and Ohta, 2007; Verkruysse et al., 2008; Wieringa et al., 2005], auto-regressive analysis via pole cancellation [Tarassenko et al., 2014], continuous wavelet transforms [Bousefsaf et al., 2013] or the Lomb periodogram [Poh et al., 2011].

2.4.3.3   Vision-based approaches

As alluded to earlier in this Chapter, manual methods remain the gold-standard for respiratory assessment in several areas of medical practice. Whilst these methods are reliable for the detection of individual breaths, they are impractical and ultimately unreliable for continuous monitoring over long periods [Cretikos et al., 2008].

With this in mind, several methods rooted in computer vision have been designed. Their main proposition is to utilise the same visual cues as used by human observers to carry out respiratory measurements.

The recognition and analysis of human motion is one of the most active topics in the field of computer vision. The ability of recent algorithms to quantify the motion of moving objects, their speed, and volumetric deformation means that, in theory, not only could continuous respiration monitoring be achieved, but one would also retain some of the advantages of human observations, such as a richer description of breathing patterns and tidal volumes.

The degree of complexity of the desired description (*i.e* the set of model features) is naturally proportional to the computational burden placed upon the algorithms used to extract the features. To date, we have seen three major approaches to respiratory monitoring from video data (Table 2.3). In this section, we will review them in ascending order of computational complexity.

Temporal differencing

The simplest methods are based on temporal differencing, *i.e.* the computation of the intensity difference between two consecutive frames in digital image data.

This principle has been applied successfully to determine respiratory rate in profile videos of a subject acquired using a CMOS (complementary metal-oxide-semiconductor) camera [Bai et al., 2010] (Figure 2.17). At the same time, Tan et al. [2010] filmed subjects engaged in normal computer usage, computed the intensity differences between frames acquired at half second intervals, and thresholded the resulting images to obtain a number of *hits*, or pixel locations where a considerable change in pixel intensity had taken place over the sampling time. When the distribution of the total number of hits
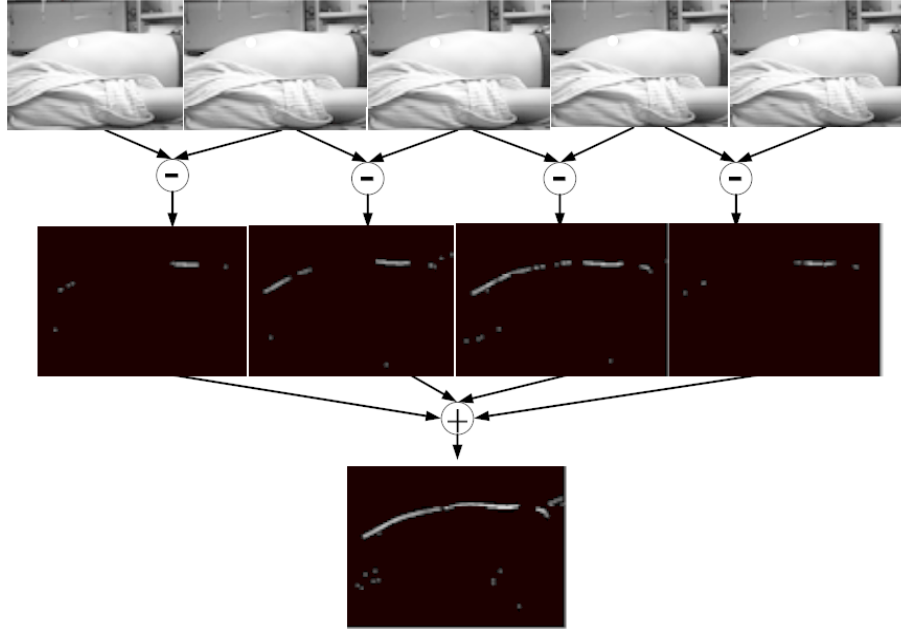
Chest displacement from frame differencing. A chest displacement image (bottom layer) is obtained from the cumulative sum of (absolute) differences between consecutive frames (top layer). Figure reproduced from Bai et al. [2010] with permission.

per difference image was plotted over a period corresponding to one breathing cycle, one could distinguish two clear lobes: one due to chest inflation and the other due to its deflation. The authors observed that the inspiratory lobe was generally *peakier* in shape and shorter in duration. This observation that is consistent not only with those of others [Yu et al., 2012] but also with the well-known mechanics of ventilation processes reviewed in Section 2.2.2; while inspiration is an active process initiated by muscular activation of the respiratory musculature in the torso, expiration is a slower process caused by the elastic recoil of the lungs and chest cage.

To reduce the dimensionality of this problem, Bartula et al. [2013] proposed a method based on classic projection-based motion estimators. Their method was based on the principle that image translation results in translation of image projections. A one-dimensional representation (profile) of each frame (or ROI) was obtained through a projection-like transformation onto a vertical axis. Breathing motion was then obtained by cross-correlation between this vector and that of earlier frames, with chest excursions defined as a function of the shift (in pixel dimensions) corresponding to the peak of the cross-correlation function.

In short, apart from the simplicity of these approaches, which makes them ideally suited for real-time implementations, frame differencing methods adapt well to sudden changes in the background, environment lighting or spurious motion. However, the small amplitude of abdominal and chest movements (typically of the order of several millimetres [De Groote et al., 1997]) means that subjects are often required to stand very close to

the camera (< 1 m) [Bai et al., 2010; Tan et al., 2010] or wear high-contrast patterns [Tan et al., 2010]. Furthermore, without incorporating the outline of the moving objects, it is our belief that the insights that can be gained by such a simplistic treatment of motion are also limited.

Depth sensing

A second approach that has recently been gaining momentum is driven by recent advances in depth-sensing technology. The last five years have seen the proliferation of low-cost devices that can capture human motion in three-dimensional space. The most iconic is perhaps the Microsoft Kinect^{TM} (Microsoft Corporation, Redmond, Washington, USA), released in 2010, and current holder of the record for fastest-selling consumer electronics device in history [BBCNews, 2011]. The Microsoft Kinect^{TM} offers an RGB sensor and a depth-sensing unit, comprising an IR projector and near-IR monochrome CMOS detector at 827 μm. This integrated system, often called RGB-D, outputs depth information at a frequency of 30 Hz, a sampling rate well above that required for RR detection.

The first attempt to use this device to detect subtle breathing motion can be attributed to Burba and Bolas [2012], who performed preliminary experiments by averaging the distance between the sensor and a rectangular region in the chest of a subject sat upright and facing the device at close range. Although their results were encouraging as proof of concept, some difficulties associated with the poor depth resolution of the device (up to 1 cm at 2 m distance [Kinect, 2011]) were evident. These were addressed in several follow-up studies. A study by Xia and Siochi [2012] addressed this issue by overlaying a small translation stage (9 in × 12 in) over the chest of four stable breathing subjects lying in the supine position. Although this protocol resulted in higher correlation between RGB-D estimates and a chest belt reference (0.96 - 0.98), it can no longer be deemed to be a non-contact setup. Martinez and Stiefelhagen [2012], on the other hand modified the Kinect^{TM} itself to project large IR dots and tracked the position of this fixed dot pattern along the epipolar line. After all trajectories were merged using PCA, auto-regressive spectral methods were employed to estimate RR as the dominant frequency in the principal component selected. In tests performed on an artificial chest, this active implementation showed a marked improvement over Burba and Bolas [2012] in accuracy over a range of breathing rate regimes (4 - 49.2 breaths/min) and measurement distances.

While previous approaches in this section have focused on measuring breathing rate alone, Yu et al. [2012] succeeded in extending extraction to other variables. In their first study, using an unmodified camera Kinect^{TM} placed 1.4 m away from a seated subject, the chest wall volume was estimated for three ROIs on the subject's chest placed over the abdomen and left and right thorax. The proposed image-based method was then tested in scenarios where asymmetries between regional breathing movements were expected. An

isovolume manoeuvre[1] revealed that symmetrical volume changes took place in the whole thorax and abdomen. Further tests on a patient that underwent a pulmonary lobectomy revealed a marked difference in the working volumes of the right *versus* the left thorax.

A second study by the same author saw the device mounted above a bed as part of a setup for sleep monitoring. At a distance of 1.25 m from the subject, both head and torso were in the field of view of the camera. These structures were used as ROIs and modelled geometrically. The average depth of the torso yielded RR estimates with an accuracy between 89.7% and 93.4% when compared with an IP reference. Although the estimates of tidal volume manifestly underestimated their true value (which we suspect may be due to the fact that back movements were not accounted for) these two studies have the merit of introducing an understanding of morphology and skeletal pose into the process of motion tracking. In other words, the authors could not only measure the periodicity of respiratory movements but also provide some, albeit limited, functional assessment of lung inflation. In the context of sleep monitoring, this study is to our knowledge the only study to have evaluated camera technology for long periods, with 42 hours of realistic overnight sleep monitoring, and whose error statistics therefore lend themselves to broader interpretation. However, we interpret the high accuracy reported in Yu's studies with reservations; it may be a consequence of the fact that the study was conducted during sleep, during which motion artefacts are minimal.

Optical flow theory and motion tracking

The last category consists of a set of methods built on optical flow theory and motion tracking. This approach has found proponents in Nakajim et al. [2001]; Nakajima et al. [1997a,b]. In a set-up similar to the one just described, the authors used a standard monocular CCD camera to monitor an elderly subject during sleep. Optical flow over the chest region was computed using the gradient-based method and breathing rate, postural changes and sleep apnoea events were extracted with minimal assessment of performance characteristics.

Building on optical flow, Wu et al. [2012] developed a framework whereby subtle periodic changes in pixel intensity are shown as high amplitude motion. In essence, their method works as follows. First, every frame in the video sequence is decomposed into different spatial frequency bands by computing its Laplacian pyramid [Burt and Adelson, 1983]. Then, all pyramid levels are temporally band-passed to isolate specific temporal frequencies relevant to a given application and to remove any temporal DC component.

The ability to arbitrarily amplify certain movements based on their periodicity means that this technique can be regarded as a microscope for motion at a given frequency. Although the authors envision applications primarily in computer visualisation and video

---

[1] An isovolume manoeuvre involves a breath-hold test whereby air is moved from the abdominal to the thoracic region and *vice versa*

(a)

(b)
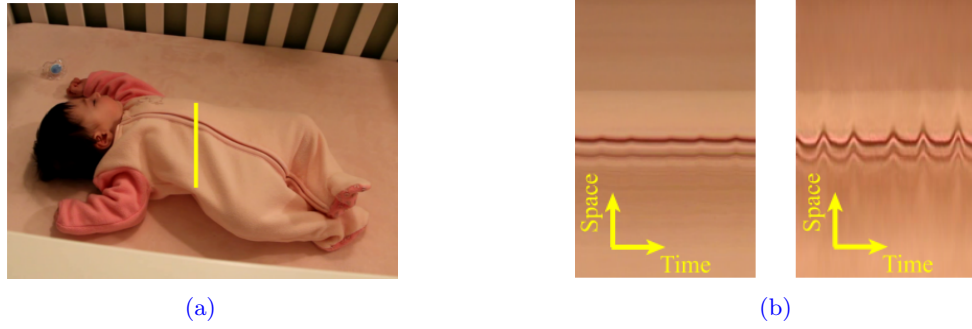
Figure 2.18

Magnifying phase. A sample frame and ROI are shown in (2.18a). On the right side (2.18b), the phase changes over a timeslice from an apparently static video are magnified ($\times$ 20) to reveal breathing motion. The processing amplifies small motions by modifying local phase variations in a complex steerable pyramid representation of the video. Figures reproduced from Wadhwa and Rubinstein [2013] with permission.

enhancement, they hint at how such a technique could be applied for the quantitative assessment of periodic biological phenomena. Several authors have since applied this framework as a pre-processing step to boost the SNR of pixel intensity (or local phase) fluctuations due to breathing or the BVP prior to the measurement of respiration [Al-Naji and Chahl, 2016; Alinovi et al., 2015; Koolen et al., 2015] or pulse rate [Klaessens et al., 2014] from videos. An alternative to their method using local phase variations to measure motion instead of explicit computation of optical flow has also been developed [Wadhwa and Rubinstein, 2013] using complex-valued steerable pyramids [Simoncelli et al., 1992] (Figure 2.18).

Alternative approaches focus on tracking a group of feature points. In Li et al. [2014]'s experiments, the authors measured RR in a sleep laboratory (under IR illumination) using a camera positioned 1.4 m above the head of the bed as a motion tracker. Specifically, each frame was subdivided into a grid of rectangular blocks. A Harris corner detector [Harris and Stephens, 1988] was used to detect feature points inside each block, while the Lucas-Kanade algorithm was used for their tracking. Two different methods of identifying respiratory-related motion from feature point trajectories were then used, including a simple averaging approach and one based on PCA. Each method condensed the feature point trajectories into a compact set of signals. The expectation was that the signal with the highest periodicity should be the most representative of the respiratory waveform. Model parameters such as the number of blocks, the number of feature points per block and criteria for signal periodicity were arbitrily set for this monitoring scenario. The system was evaluated on 5 participants in 4 simulated sleep scenarios, each lasting for five minutes. Their results indicate that for supine, lateral and obscured supine positions, the PCA-based method could detect respiratory rate within a 1 bpm margin in more than 97 % of time of simulated sleep.

A disadvantage of techniques based on optical flow is that its computation poses a challenge to processing power. This is a disadvantage in relation to the other approaches amenable to real-time computation. In the future, this burden on computing power could

be lessened if this approach was combined with efficient ROI detection and tracking to avoid the need to compute the flow over large image areas, thus limiting integration to the local neighbourhood around features of interest. The automatic localisation of valid ROIs for reliable respiratory signal extraction remains an open problem. Several approaches based the selection of regions with high SNR have been proposed with varying degrees of complexity; Tarassenko et al. [2014] have been able to construct accurate maps of the spatial distribution of respiratory information from AR-modelled area intensity while Janssen et al. [2016]; Lukáč et al. [2014] devised signal quality scores based on the temporal consistency and spatial redundancy of the optical flow fields.

2.5 Remote monitoring of neonates

The deployment of camera-based algorithms for non-contact monitoring of neonates or young infants (up to one year old) is comparatively recent and has so far focused mainly on the extraction of pulse rate.

In the first study conducted in a clinical setting, Scalise and Bernacchia [2012] established the feasibility of measuring pulse rate in 7 preterm infants (g.a. 30- 33 weeks) during supine sleep. Neonates were exposed to a green light source and HR measurements were taken for a maximum of 30 seconds with a webcam placed 20 centimetres away from the face.

Aarts et al. [2013] were the first to successfully attempt measurements under the ambient light of a neonatal unit. In their work, the authors monitored 19 infants (g.a. 25 - 42 weeks) with a digital camera placed on a tripod at approximately 1 metre from the infant. In 13 out of 19 infants, it was possible to derive heart rate estimates which matched those derived from standard ECG (within 5 beats·min^{-1}) for 90% of the time, even in recording scenarios with added low-frequency noise, such as kangaroo care, or high-frequency noise, *e.g.* during ventilation. Their method consisted in the analysis of the green channel of the RGB data over manually-selected ROIs over the infants' head, arms or thorax. Heart rate was estimated from frequency analysis with the Fast Fourier Transform (FFT).

The first investigation into respiratory monitoring of neonates using visible light cameras can be found in Villarroel et al. [2014]. Villarroel et al. [2014]'s is also the first account of continuous measurement of a vital sign in this population. In addition to the estimation of RR and spO$_2$ changes for a 20-min period including a severe desaturation, HR was estimated for over 20.1 hours (RMSE = 3.95 beats·min^{-1}, MAE = 2.83 beats·min^{-1}) in two infants (g.a. 27 and 31 weeks). Villarroel et al. [2014] positioned a 3-CCD camera through a hole cut in the top of the incubator canopy. They extracted HR and RR from an ROI using ICA and analysed the frequency content of the average signal over this region using digital filtering and auto-regressive spectral methods.

In another study published in the same year, Klaessens et al. [2014] reported on spot measurements of HR from six infants (g.a. 24 - 39 weeks) in an open incubator using a Sony camera and of respiratory rate using an infra-red thermal camera.

Eulerian Video Magnification (EVM) developed by the Freeman group [Wu et al., 2012] was used to amplify the colour variations in the digital camera images and visualise the pulse. Respiratory rate was also measured using an IR thermal camera by tracking the changes in temperature around the baby's nostrils during the breathing cycles as observed in the long wavelength infrared (LWIR) window (8 to 14 μm) of the electromagnetic spectrum. More recently, Koolen et al. [2015] would also use EVM to amplify the respiration-related intensity changes in videos acquired during polysomnographic recordings in 7 preterm infants (g.a. 33 - 40 weeks). The optical flow was computed between frames of the amplified videos and the Short-Term Fourier Transform (STFT) was used to represent the frequency components of the signal extracted from an ROI manually cropped to contain only the chest and abdomen region over time. Cross-correlation between respiratory rate extracted from a reference signal provided by a strain belt and the signal obtained by the optical flow algorithm was computed for different strategies for the construction of the respiratory signal from optical flow. Al-Naji and Chahl [2016] developed a motion magnification technique to magnify the respiratory chest movements of an 8-month infant in different positions using wavelet pyramid decomposition instead of Laplacian pyramid decomposition and an elliptic band pass filter instead of a Butterworth bandpass filter. The authors then used a threshold on frame differences between the magnified frames to binarise the images and quantify motion in the breathing frequency range. Their experimental results on short videos (15 s) of each position recorded with a DSLR video camera (at 30 fps[1]) indicate that RR was successfully measured with an accuracy of approximately 99 % and MRE of 1.314 %.

Alternative techniques for contactless HR measurement under ambient light based on similar principles (cardiac-synchronous colour variations) have also been proposed and tested in small cohorts of 8 (g.a. 37 -40 weeks) [Mestha et al., 2014] and 7 (median g.a. 32 ± 4 weeks) NICU patients [Fernando and Haan, 2015] for short periods of time (30 min and 1- 24 min per subject, respectively).

Finally, we note two other studies aimed at respiration monitoring in adults but which also included neonates among their study subjects. Based on the prior work of Lukáč et al. [2014] on estimating RR using motion vectors as features (*e.g.* optical flows) and dimensionality reduction (singular value decomposition or PCA), Janssen et al. [2016] extracted RR in two infants in 20 videos from different viewpoints with a high Pearson's *P* coefficient (0.65 to 1, depending on viewpoint). An alternative approach using EVM and pixel binarization prior to frequency estimation over the ROI was tested on two short videos of two newborn infants (of approximately 5 and 3 minutes duration) [Alinovi et al., 2015], where it produced RR estimates within 15 % confidence bounds of the IP reference.

---

[1]frames per second.

2.6 Chapter conclusions

In this chapter, we have reviewed the relevant background in respiratory physiology of the neonate and non-contact methods for respiratory monitoring using visible-light and near-infrared cameras.

Respiratory conditions are the primary cause of admittance of newborn infants to a neonatal unit. Respiratory conditions can arise for several aetiologies, many of which are unique to this age group. Therefore, the continuous and accurate monitoring of respiration and the timely detection of adverse events form integral parts of neonatal care. Current monitoring solutions based on contact sensors have clear deficiencies: they require adhesive sensors to be attached to the fragile skin of the infant and are not sufficiently robust to infant motion.

Under the paradigm of remote sensing, the latest decade has seen the emergence of numerous approaches to the measurement of vital signs using visible-light video cameras. With considerable overlap, previous approaches to extract respiratory parameters from visible light imaging can be classified into those based on either a photoplethysmographic or computer vision approach to signal extraction and analysis.

In light of these advances, some preliminary work to assess the feasibility of using camera-based methods in the clinical monitoring of vital signs has been reported. Few studies have addressed the technical challenges inherent to the monitoring of the neonatal population in the real-life settings of neonatal units.

Chapter 3

Datasets

Contents

3.1 Datasets and their relevance to this thesis

This chapter describes the two datasets used for validation of the signal and video processing techniques implemented in subsequent chapters of this thesis. The main features of these datasets are summarised in Table 3.1.

The Hypercapnoea dataset

The Hypercapnoea dataset (Section 3.2) was obtained in 50 young healthy male volunteers at rest. During a 45-minute period their respiratory rate was artificially increased via hypercapneic stimuli (*i.e.* an abnormally high partial pressure of carbon dioxide). In addition to video recording, subjects were monitored with standard ECG and PPG technology and respiration was double-monitored with two RIP belts and a nasal airflow pressure transducer. The aim of this dataset was to provide physiological data from healthy subjects during spontaneous breathing in which indirect methods for measuring respiration could be tested under controlled conditions. This work is described in Chapter 4.

Table 3.1
Summary of dataset contents.

	NICU	Hypercapnoea
Cohort		
	Preterm infants admitted to the Neonatal Intensive Care Unit at the John Radcliffe Hospital, Oxford.	18 to 40-year-old healthy males in 6 groups of different phototypes in the Fitzpatrick scale.
Size		
No. participants	30	50
Recording time	Up to four days each. ($\sim$ 5 hours per day)	2 hours each. First hour under hypoxia. Second hour under hypercapnoea.
Recorded data		
Video	1628$\times$1236 pixel frames of 24-bit RGB acquired at 20 Hz by a 3-CCD camera on top on incubator.	1628$\times$1236 pixel frames of 36-bit RGB acquired at 16 Hz by a 3-CCD camera located 1 m away from subject's face
Physiological signals	Foot PPG (125 Hz), single-lead ECG (250 Hz), IP (62.5 Hz) and derived vital signs (HR, RR and spO$_2$ at 1 Hz)	2 $\times$ finger PPG (256 Hz), single-lead ECG (256 Hz), thoracic and abdominal expansion belts (64 Hz), nasal airflow pressure (64 Hz) and derived vital signs (HR, RR and spO$_2$) at 1 Hz.
Other modalities	Audio inside incubator (44.1 kHz).	Ambient thermometer

The NICU dataset

The NICU dataset (Section 3.3) was acquired during an observational study in a neonatal intensive unit. This dataset is divided into 90 recording sessions from 30 neonatal patients. Each recording session consists of up to 5 hours of vital sign data collected as part of standard care, video footage, and patient case reports. The case reports contain a list of clinically observable events such as apnoeas, hypo- and hyperventilation, transient tachypnoea, changes to spontaneous breathing and changes in tidal volumes. The aim of this dataset was to assess the merits of our camera-based algorithms for respiratory surveillance (Chapter 7) and the detection of adverse events (Chapter 8) in this patient group.

3.2 The Hypercapnoea dataset

3.2.1 Overview

This section describes an exploratory study (CUREC: MSD-IDREC-C3-2014-003) carried out in collaboration with the Department of Physiology, Anatomy and Genetics (DPAG). The main goal of this investigation was to evaluate the performance of several physiological monitoring algorithms under controlled conditions of visible light illumination, subject movement, skin phototype, and ambient gas concentrations.

The study took place inside a dedicated normobaric chamber. This apparatus, which has been employed in multiple studies that emulate the human response to altitude

[Smith et al., 2008], allows for the composition of inspired gases to be adjusted in a safe and tightly-controlled manner. This design avoids the need for a mouthpiece or a mask to control the concentration of end-tidal gases, which in turn avoids the hyperventilation typically caused by these devices [Hirsch and Bishop, 1982].

The Hypercapnoea dataset was acquired as part of a two-part study. Over the course of the two sub-studies, the relative concentrations of atmospheric gases (N_2, O_2 and CO_2) inside the chamber were gradually modified so as to elicit changes in the vital signs of the participants. During the first sub-study, the fraction of inspired oxygen inside the chamber (FiO_2) was altered so as to induce a change in the seric levels of oxygen-carrying haemoglobin (the *hypoxia sub-study*). After this period, the subject was allowed to rest before re-entering the chamber for a second sub-study. For this study, (the *hypercapnoea sub-study*), FiO_2 was kept at normoxic[1] levels while increasing concentrations of CO_2 were delivered in a step-wise manner. The aim of this protocol was to induce an increase in minute ventilation, which in turn can be achieved through an increase in RR and/or V_T.

This section is concerned with the latter sub-study. Appendix A presents some considerations regarding how each sub-study affected the respiratory rhythm, and therefore, why the hypercapnoea sub-study was selected for analysis.

The study cohort consisted of young healthy male volunteers. All subjects were double-monitored with both protocols carried out under concurrent video and vital sign monitoring with pulse oximeters, ECG, nasal pressure cannulae and strain gauge belts. A summary of the monitoring equipment and protocols for data acquisition is given below.

3.2.2   Protocol

Each subject entered a normobaric hypoxia chamber with an FiO_2 set to 12%, where they sat comfortably on a standard office chair in a semi-reclined position. Prior to measurements, a 10-min period of rest was allowed. After this period, the protocol described below and shown in Figure 3.1 was adopted for the $FiCO_2$ level inside the chamber. The rationale behind this protocol and the physiological mechanisms at play during brief exposures to incremental hypoxia and hypercapnoea are discussed in Appendix A.

Phase 1   Normal breathing. After the chamber had been sealed, recordings were taken for a lead-in period of 3 minutes of normocapnoea and normoxia (*i.e.* under ambient gas concentrations). During this period, the subjects were allowed to breathe at their resting RR and V_T.

Phase 2   Hypercapnoeic breathing. For this part of the protocol (lasting approximately 32 minutes) ambient $FiCO_2$ inside the chamber was gradually increased in even steps while monitoring RR and end-tidal gases. We aimed for a small

---

[1]Normoxia is a term used to describe arterial oxygen tensions between 10 - 21%. Normocapnoea is the state of normal arterial carbon dioxide pressure, typically about 40 mmHg [Guyton et al., 2012].
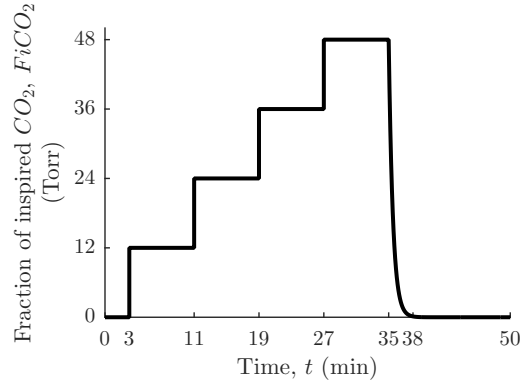
increase in $FiCO_2$ (~ 12 Torr) above its resting value every 8 minutes, holding the $FiCO_2$ at 4 predetermined levels under poikilo-oxic conditions (*i.e.* FiO_2 was allowed to vary naturally).

Phase 3 Return to eucapnoea. The chamber door was opened and recording resumed for approximately 3 minutes. This phase served to allow the subjects to return to their resting RR. When this value was attained, the recording stopped and the subject was allowed to exit the chamber.

During the three phases, subjects were instructed not to talk and to remain as still as possible. Audio entertainment was provided.

3.2.2.1 Sample size

50 healthy young male subjects were recruited for the study.

3.2.2.2 Eligibility criteria

Male subjects between 18 and 40 years of age who had capacity to consent were recruited. Participants with different phototypes on the Fitzpatrick scale [Fitzpatrick, 1988] (10 participants for Types I-IV and 5 participants for Types V-VI) were recruited[1]. Subjects had no upper or lower respiratory complaints or history of major medical problems. No conditions which contraindicate exposure to hypoxia or hypercapnoea, such as epilepsy, asthma, claustrophobia, or significant cardiorespiratory disease, were present. Subjects under medication with possible effects on the sympathetic nervous system were also

---

[1]Investigating the influence of skin pigmentation on the accuracy of the estimated vital signs was one of the goals of the Hypoxia protocol. Under the Hypoxia protocol, the volunteers were rendered hypoxic (*i.e.* subject to an abnormally low partial pressure of oxygen, pO_2) to assess the sensitivity of the camera monitor to changes in spO_2. Results under this protocol are not reported here.
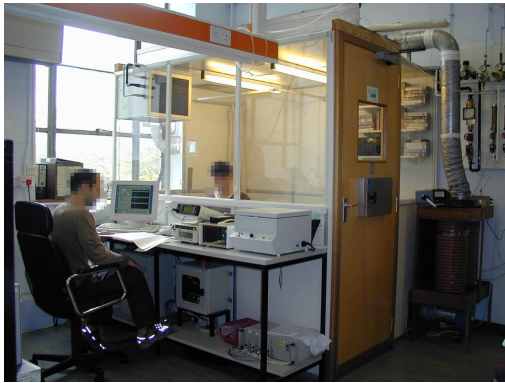
(a)   (b)

Figure 3.2

Experimental apparatus at the Physiology Laboratory, Department of Physiology, Anatomy and Genetics, Oxford, UK. (3.2a) Outside view. Customised chamber and equipment for experimental control. (3.2b) Inside view. Study participant and data acquisition apparatus.

excluded. Subjects were asked to shave 2 - 12 hours before the session so as to maximise the area of skin in field of view of the camera.

3.2.2.3   Volunteer recruitment protocol

Adverts for participation were posted in public sites including student halls, gyms, and University buildings. A small financial reward was offered to participants to compensate them for their time.

3.2.2.4   Time commitment and duration of study

Video and vital sign data was collected for a period of approximately one hour for individual participants. Study sessions took place between November 2014 and May 2015.

3.2.2.5   Ethics

The study protocol was approved by the applicable institutional ethics review boards, the University of Oxford Central University Research Committee (CUREC). The study complied with the Data Protection Act, which requires data to be anonymised as soon as it is practical to do so.

3.2.3   Data acquisition

3.2.3.1   Hardware for data acquisition

Room   The normobaric chamber used in this study is shown in Figure 3.2. Ambient room temperature inside the chamber was kept between 17 - 22 ^{o}C. To provide a spectrally

uniform lighting environment of homogeneous intensity, the chamber door, window and walls were lined with a layer of 5.3 mm-thick Tyvek® 1443R paper supplied by DuPont, Wilmington, USA. Due to its material properties, Tyvek has a reflectivity function with a predominant diffused specular component. In addition, its reflectance spectrum exhibits a flat response for wavelengths in the spectral range between 400 and 800 nm.

Illumination Illumination was delivered by two 5600 K LED mosaic panels (576 diodes over a 30×30 cm^2 area), each held by a stand approximately 2 m away from the subject and set at a height of 1.6 m as shown in Figure 3.2b. The LED panels were aimed at diffusive pads so that light could diffuse through the room, thus minimising specular reflection over smooth surfaces. A colour checker chart (BST-13®, X-RITE, Grand Rapids, USA) was placed on the wall behind the subject and in the field of view of the video recording equipment.

Subject monitoring and sensors The experimental setup for the acquisition of reference data is presented schematically in Figure 3.3. Respiration was monitored in two ways. The changes in thoracic and abdominal circumference were measured using two respiratory inductance plethysmography (RIP) belts (XactTrace®, Embla, Ontario, Canada). Subjects breathed through a tight-fitting nasal cannula (EMBLA, Embla, Ontario, Canada) fitted with a nasal pressure transducer. The nasal prongs were trimmed

as needed for participant comfort. All signals were sampled at 64 Hz.

A PPG sensor (Masimo M-LNCS DBI®, Masimo, Irvine, USA) was applied to the left index finger to acquire this signal at 256 Hz. Continuous 1-lead ECG signals were collected and digitized at 256 Hz.

All raw signals and reference vital signs (HR and spO$_2$ at 1 Hz) were recorded to a Visi BlackShadow® system (Stowood Scientific Instruments, Oxford, UK), as shown in Figure 3.3.

Sensing optics Video RGB data was collected by a 3 CCD (ICX274AL®, Sony, Tokyo, Japan) digital camera (JAI AT-200C®, Glostrup, Denmark) held approximately 1 m away from the subject face at face level by a camera arm (244RC®, Manfrotto, Cassola, Italy). This imaging device was set to acquire 36-bit true colour images (12-bit per colour) at a pre-set rate of 16 frames per second and at a resolution of 1628 × 1236 pixels[1].

Collection optics The RGB JAI camera was equipped with a C-mount positive lens GMHR3D26018C® (Imaco, Warsaw, Poland) with ∅ 30 × 62.7 mm, focal length $f = 6$ m and focal ratio set to 1.8.

Workstations The volume of video data collected over the course of each session posed challenges regarding bandwidth and storage requirements, which were addressed through the use of two separate machines.

The RGB camera was interfaced via Camera Link to a framegrabber card (Microenable IV AD4-CL®, Silicon Software, Mannheim, Germany) installed on a 1.90 GHz Intel Xeon processor machine with 4 × 8 GB RDIMM and 8 × 4 TB SATA (PowerEdge T420, Dell, Austin, USA). A Latitude E5540® laptop (Dell, Austin, USA) running a Windows 7 operating system (Microsoft, Redmond, USA) was used to (a) control the recording equipment via a LAN (Local Area Network) connection and (b) synchronise the clocks on the workstations with that of the *Visi BlackShadow* equipment at the start of each sub-study.

3.2.3.2 Software for data acquisition

The study machines ran in-house software to synchronisation of data aquisition between workstations and devices. This is described in detail in Section 3.3.3.2.

---

[1]In the context of this thesis, this dataset served to test the implementation the signal processing methods for the analysis of the physiological signals acquired using contact devices and to develop the evaluation procedure for evaluating the performance of these methods. With this in mind, the results of vital sign estimation from the video signals collected using this device are not reported here.

3.2.3.3 Software for chamber control

A desktop computer running dedicated software (Chamber v1.00, University Laboratory of Physiology, Oxford, UK) based on LabView® (National Instruments, Austin, Texas, USA) generated the partial pressures of O_2 and CO_2 specified by the protocol described. This computer received feedback from a pulse oximeter (3900 Pulse Oximeter®, Datex Ohmeda, Misissauga, Canada connected to a TruSignal TS-E4-H®, General Electric, Fairfield, USA or a OXY-F-UN®, Datex Ohmeda, Misissauga, Canada), a single-lead ECG monitor (Micromon 0740-000®, Kontron Augsburg, Germany) and a pnemotachograph (Normocap Oxy®, Datex Ohmeda, Misissauga, Canada) with a differential pressure transducer. These devices measured HR, RR, spO_2 and end-tidal gas concentrations to facilitate the monitoring of the physiological status of the volunteer by the on-site clinical staff.

3.2.4 Cohort description

A group of sessions in this dataset was randomly selected for further analysis in Chapter 4. This group consisted of 12 hypercapnoea sessions on 12 volunteers (2 of each phototype in the Fitzpatrick scale). The physiological response of this sub-group to the hypercapnoea protocol is quantified in Figure 3.4 in terms of changes in the three monitored vital signals with respect to their baseline values at the start of the study session.

3.3 The NICU dataset

3.3.1 Overview

This section presents an observational study of non-contact vital sign monitoring of premature infants in the Neonatal Intensive Care Unit (NICU) at the John Radcliffe Hospital (MONITOR Study - REC: 13/SC/0597).

The NICU is an intensive care and high-dependency unit offering specialist care for critically-ill or premature newborn infants. As the outputs of data collection in this unit, we obtained (a) RGB video footage recorded using a digital video camera positioned over the incubator inside which study infants were nursed, and (b) conventional vital sign data collected in parallel as part of routine care. This dataset is supplemented by (c) clinical annotations in patient case report forms (CRFs), where baseline demographic data, physiological events, and clinical interventions are recorded. In addition, (d) manual breath counts were obtained for 5 subjects during respiratory rate measurements by clinical staff, and (e) potential episodes of apnoea were retrospectively annotated through the analysis of the acquired vital sign data by clinical experts, as described in Chapter 8.

In comparison with previous investigations of non-contact monitoring techniques in
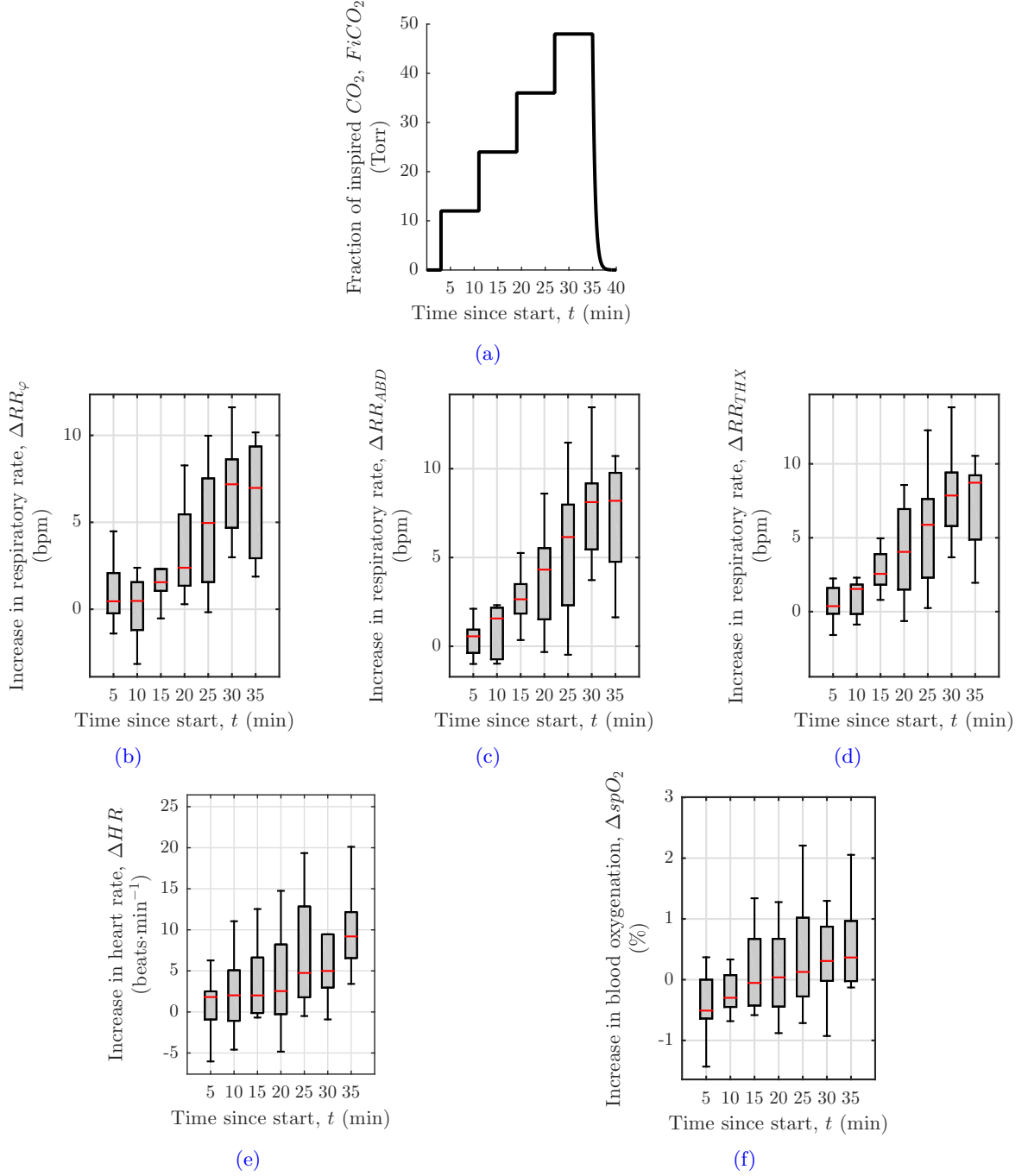
Figure 3.4

Physiological response to hypercapnoea delivered in steps of ascending $FiCO_2$ (3.4a). Boxplots are employed to represent the change in RR, HR (3.4e), and spO_2 (3.4f) sampled at 5-min intervals into the study session. The respiratory rates were derived from three different sources: the nasal airflow pressure transducer (3.4b) and the abdominal (3.4c) and thoracic (3.4d) RIP belts (using the methods described in Chapter 4). Statistics were computed for 12 volunteers selected randomly from the dataset.

the neonatal unit [Aarts et al., 2013; Abbas and Leonhardt, 2014; Abbas et al., 2011; Klaessens et al., 2014], we believe that our dataset is unique in a number of ways. First, the study infants were monitored for several hours and under continuous video recording. Understandably, this poses challenges on the protocol for data acquisition, primarily in terms of data storage and management. In this section, we describe how these issues were addressed. Furthermore, all prior studies were undertaken in small cohorts, characteristic of proof-of-concept studies, with much lower enrolment numbers than our study. Finally, as our enrolment criteria did not exclude subjects from non-white backgrounds, it allows for the assessment of non-contact vital sign monitoring in infants with a variety of skin colour tones.

The protocol for data acquisition is summarised in Section 3.3.2 and the experimental set-up is presented in Section 3.3.3. A brief description of the NICU cohort and some monitoring statistics are provided in Section 3.3.4.

3.3.2 Protocol

3.3.2.1 Sample size

A total of 30 preterm infants were enrolled under our protocol.

3.3.2.2 Patient eligibility

Informed consent was sought from parents of neonates nursed in the High Dependency Area (HDA) of the Neonatal Unit who satisfied the following inclusion criteria: (a) less than 37 weeks of corrected postmenstrual age, (b) requiring high dependency care, and (c) needing to be nursed exposed in an incubator.

Criterion (a) was adopted as it is in this age range that physiological instability, and consequently, episodes of apnoea, bradycardia, and desaturation more commonly manifest themselves in preterm infants. Criterion (c) was adopted so as to maximise the area of skin available for the extraction of a camera-derived PPG waveform.

The exclusion criteria were (a) the need for intensive care at time of data collection, or (b) the presence of a life-threatening congenital anomaly.

3.3.2.3 Patient recruitment protocol and process

The clinical staff in the NICU identified appropriate patients and approached their parents with a patient information sheet and a verbal explanation prior to obtaining written consent. A trained registered nurse carried out the data collection.

3.3.2.4 Ethics

The study received ethical approval from the South Central — Oxford A Research Ethics Committee and was sponsored by the Oxford University NHS Hospitals Trust.

3.3.2.5 Time commitment and duration of study

At nightime, ceiling lights on the unit are dimmed so we chose to concentrate data acquisition on daylight hours. When selecting the time periods during which to record, we noted that times of staff handover and open visiting hours were moments of frequent interaction with the infant. As part of routine care in the HDA, for a designated *quiet period* of approximately 6 hours every 24 hours, from 9 am to 3 pm, patients are allowed to rest and are not disturbed except in case of an emergency. During this period, a digital video camera was positioned over the incubator and used to film the study infant under normal lighting conditions in the HDA ward. Study sessions ran between February 2014 and May 2015.

3.3.2.6 Baseline demographics, physiological events and clinical procedures

Baseline patient characteristics including gender, ethnicity, gestational age at birth (weeks), corrected gestational age at start of study (weeks), birth weight (g), weight at start of study (g), medications administrated and requirement for respiratory support (*i.e.* none, high-flow therapy or low-flow nasal cannula), entereal feeds (ml/kg/day), feeding regime or continuous infusions (ml/kg/day) were collected for all subjects.

Any intervention during the study period, defined as a deliberate interaction with the baby or its immediate environment was also logged on the patient case report form (CRF) and later transcribed on to an anonymised eCRF platform (OpenClinica). Examples of procedures commonly associated with changes in physiology are the non-invasive measurement of blood pressure or capillary refill time, the administration of a feed or the withdrawal of blood by heel lancet for blood gas analysis.

For the manual respiratory rate measurements performed on the last five infants recruited, the clinical staff was instructed to press a set key ('Spacebar') on the keyboard of one of the study workstations (Workstation A) at the time each breath was registered. The specific moment during the patient's respiratory cycle corresponding to the key press was left at the discretion of the clinical staff so long as the same reference point on the respiration cycle (end of inhalation or end of exhalation) was used consistently for all breaths in each counting cycle. In the event of excessive movement by the infant or fading breathing motion rendering individual chest excursions visually indistinguishable over the course of an ongoing counting cycle, the current count was aborted and a new one was restarted once the infant settled and visible breathing movements resumed.

Data acquisition apparatus. Monitoring equipment imaging a mannequin inside the study incubator in the Oxford Neonatal Unit.

3.3.3 Data acquisition

3.3.3.1 Hardware for data acquisition

Issues regarding data collection, storage and transfer had to be carefully considered in our study design. We also received input on (a) the usability of the recording apparatus from nursing staff, who operated the equipment on site, (b) on health and safety from the Department of Medical Physics and Clinical Engineering at the John Radcliffe Hospital, and (c) practical considerations on the physical kit from the team of technicians at the mechanical workshop of the Institute for Biomedical Engineering, who modified the incubator and assembled the monitoring equipment.

The result was a standalone kit which groups most of the necessary equipment in a customised hospital roll stand. This not only facilitated access to the patient from all angles with minimal impact on standard patient care but also allowed for the entire kit to be quickly removed from the incubator area if required. Furthermore, all components were selected so that any parts in the vicinity of the incubator could be routinely sterilised or replaced.

The monitoring equipment is shown in Figure 3.5. Attached to the base of the stand are two workstations; one to store the vital sign data acquired by the patient monitor (used as part of standard care in the NICU) and one to manage data acquisition and video storage. Devices for user input/output are located at waist level on a flip-up keyboard and mouse tray (EMUL510-P15-KUC-MW®, Medical Mounts, Aldershot, UK) mounted on the same stand. A camera arm (244RC®, Manfrotto, Cassola, Italy) was clamped to the stand at one end, from which it extendes horizontally in order to hold the video camera securely above the incubator canopy. Lastly, a microphone and a colour checker chart (BST-13®, X-RITE, Grand Rapids, USA) were placed inside the incubator for the duration of each recording session and sterilised or replaced before the next study infant was nursed in the incubator.
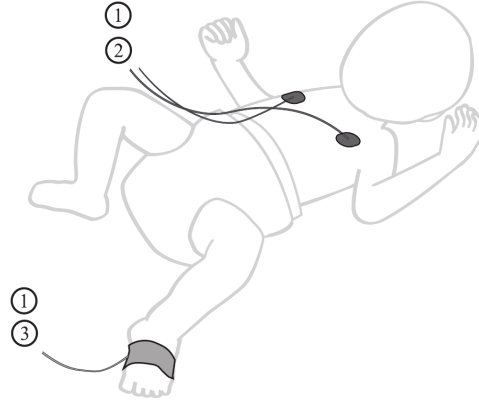
Monitoring equipment for the acquisition of reference data in the NICU dataset. Attached sensors are: (1) 2-lead ECG, (2) IP electrode and (3) foot-wrap pulse oximeter. The IP and the ECG signals from the RA and LA leads are acquired using the same set of electrodes on the chest.

The various components are now described in more detail below.

Workstations Data acquisition took place using a 1.90 GHz Intel® processor workstation with 4×8 GB RDIMM (registered memory module) and 8×4 TB SATA disks running in-house software under a Red Hat *Fedora* 20 Linux operating system (Workstation A). All conventionally-monitored signals were relayed to a separate workstation, a standard PC with a 4 TB disk and 8 GB RAM under Windows 7 (Workstation B) to meet the OS requirements of the software for data capture from patient monitors (ixTrend®, ixellence GmbH, Wildau, Germany).

Patient monitor and sensors An IntelliVue® MX800 patient monitor (Philips, Amsterdam, Netherlands) used as part of standard care in the Oxford NICU was used to record physiological data. Using proprietary software, reference values of HR, RR, and spO$_2$ were derived from the three available sources (Figure 3.6) at one-second increments. HR was derived from the single-lead ECG (sampled at 250 Hz), RR was derived from the bipolar IP signal collected (at 62.5 Hz) by the set of neonatal electrodes provided with the monitor, and both spO$_2$ and a secondary estimate of HR were derived from the PPG signal acquired by a SET LNCS Neo® probe (Masimo Corporation, Irvine, USA) and sampled at 125 Hz. All raw signals and derived vital signs relayed to Workstation B via the RS-232 serial communication protocol.

Incubator All infants were nursed inside a Giraffe Omnibed® incubator (General Electric, Fairfield, USA). A hole was drilled in the centre of the canopy to allow the camera lens to penetrate and tilt inside the chamber and thus obtain an unobstructed view of the study infant without signal attenuation or reflection from the perspex layer. The modified incubator was subject to a two-week period of humidity and temperature

tests to confirm that the same degree of control over these variables could be achieved in the study incubator as in an unmodified one.

Audio recorder   We measured noise levels within the incubator in order to track vital signs changes associated with different levels of noise. A TASCAM DR-07 MKII® portable digital recorder (TASCAM, Montebello, USA) was employed in conjunction with a microphone (SMX-10® directional stereo microphone, Azden Corporation, Arlington, USA later replaced by a disposable HD Ghost® microphone, Drift Innovation, Gerrards Cross, UK) attached to the inside of the posterior wall of the incubator to record the audio signals (at 44.1 Hz, in 16-bit WAV format) within the incubator. The audio signals acquired using these devices will be the subject of future work to assess whether they may serve as a proxy for nurse or parent interaction with the infant inside the incubator (which effectively acts as an echo chamber). This analysis falls outside the scope of this thesis.

Collection optics   This component (SV-0614H®,VS Technology, Tokyo, Japan) consists of two optical elements meant for delivering incident light to the photosensing unit. The first element is a positive lens with $f = 6$ mm and focal ratio ranging from 1.4 to 16 to collimate the light beam and focus it on the camera photodetectors. The latter provides full control of the lens aperture (ratio) in the 1 to 1.4 range. Both parameters were adjusted by the attending clinical staff at the beginning of each recording session and throughout the session so as to ensure that image brightness levels were adequate and that as much of the infant's skin as possible was imaged on the focal plane.

Sensing optics   We used a 3-CCD (Sony ICX274AL®, Sony, Tokyo, Japan) digital camera (JAI AT-200C®, JAI, Glostrup, Denmark) interfaced via Camera Link to a framegrabber card (Microenable IV AD4-CL®, Silicon Software, Mannheim, Germany).

Most single-CCD colour cameras use a Bayer filter, in order to detect a single colour per pixel. The other two-thirds of colour information for each pixel must then be interpolated with a demosaicing algorithm leading to a much lower effective resolution [Wootton, 2005]. By contrast, 3-CCD cameras use a beamsplitter and a chip arrangement to split the incoming light into red, green, and blue colour channels. The main advantage is that the 3-CCD sensor can operate at full resolution, with no loss of detail and all of the colours are spatially aligned.

The imaging device was set to acquire 24-bit true colour images (8-bit per colour) at a resolution of 1628×1236 pixels for an individual pixel size of 4.4 μm^2 at a nominal rate of 20 frames per second (fps). In practice, frames were acquired at an effective rate of 20.3 fps. Frame buffers were processed using a Spartan® FPGA (Field Programmable Gate Array) board (Xilinx, San Jose, USA) installed in Workstation A.

Central data storage and data transfer Given that the storage capacity in the study workstation was limited to 32 TB, a storage cluster hosted in the Oxford Institute for Biomedical Engineering Data Centre was acquired. The storage cluster is comprised of a Dell MD3200® head node (Dell, Austin, USA) and several MD1200® storage nodes (Dell, Austin, USA). Each MD1200® storage node contains 12×4 TB hard disks in a RAID 6 (Redundant Array of Inexpensive Disks) configuration. After lossless compression, video data was stored in *bzip2* file format.

Video and reference data for each recording session was transferred from the workstations to this server via an encryted 8 TB external storage device (2big Quadra USB 3.0®, Paris, France) configured in RAID 1.

User input To meet hygiene requirements within the NICU, a medical-grade keyboard (Accumed 105®, Accuratus, Alton, UK) and mouse (Accumed Mouse®, Accuratus, Alton, UK) were installed. Both peripherals are sealed so as to comply with water and dust-proofing requirements (IP67). They also comply with industrial standards (JIS Z 2801) for antibacterial activity due to a silver nanoparticle coating capable of deactivating bacteria commonly found in hospital settings.

3.3.3.2 Software for data acquisition

Software developed by our research group for routine tasks (*e.g.* to set up and query an NTP server, transfer patient data from workstations to external disks, encrypt patient data, compress video files, *etc.*) was adapted in accordance to the requirements of this study. Novel scripts were developed to perform project-specific tasks as described below.

Camera configuration Using the manufacturer's API, software was written to operate the camera on continuous trigger mode at a nominal rate of 20 frames per second and at full resolution. An internal buffer of 1200 frames was set, after which data was streamed to PC-memory. Each video file then consists of collated frame data collected over a period of 1 minute. Each pixel carries uncompressed true RGB colour data (8-bit per colour)[1]. In decoding the luminance values, a linear transfer function was adopted between the analogue signals and bit allocation, thus avoiding the non-linearities introduced either by *gamma* correction or by knee compensation[2]. The automatic gain control (AGC) as well as the gain for individual colour channels were also disabled. To reduce the spatial dependence between pixel intensity values, white-balance was adjusted only once in each acquisition cycle, *i.e.* in 'one push'.

---

[1] Note from the video acquisition parameters described, that each uncompressed video file (comprising 1 minute of footage) is approximately 6.7 GB in size.

[2] *Gamma* correction would replace the linear transfer function with a power series. Knee compensation would implement an exponential decay after a given analogue value to avoid output saturation at higher analogue inputs.

Logging manual breath counts   A custom kernel-based key logger was developed and installed on Workstation A, which recorded the time instant of each key press (time of button down) during video recording. As it operated on the time server from which the clock the camera software drew its timestamps, camera frames and key presses were automatically synchronised up to the nanosecond, a time interval much lower than human reaction times [HumanBenchmark, 2016]. The software developed underwent stress testing to ensure that no frames were lost during simultaneous operation of the camera and key logging routines.

Synchronisation between workstations   An NTP (Network Time Protocol) server was set up on Workstation A. A server's clock is likely to drift considerably between reboots. To address this, at the beginning of each recording session the client in the Windows workstation requested the current time from this server, and used it to set its own clock. As a result, at the beginning of each recording session, synchronicity between the timestamps gathered by the patient monitor and by the video camera was ensured to an accuracy of a millisecond.

Synchronisation between devices   Given that raw signals were collected using different devices before they are displayed by the patient monitor, in addition to synchronisation between the study workstation further processing was necessary to ensure synchronicity between these time series. This process is described in Appendix B.1.

3.3.4   Cohort description

Table 3.2

Selected population demographics of infants enrolled in the NICU study. Gestation age and weight are shown as mean(standard deviation).

Total subjects	30
Total sessions	90
Gender (no. participants)	
Male	18
Female	12
Ethnicity (no. participants)	
Asian or Asian British - Pakistani	2
Black British or Black African	1
Mixed - White British and Japanese	1
Mixed - White and Asian	2
Mixed - White and Black Caribbean	2
Mixed - Any Other Mixed	2
White British	18
White - Any Other Background	2
Gestation age (weeks)	31.1 (1.9)
Weight (g)	1240 (252)
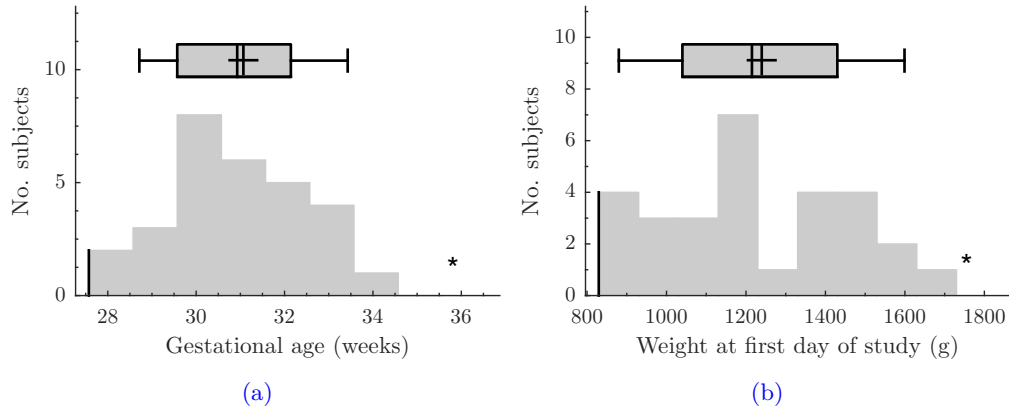
3.3.4.1 Baseline Population Characteristics

Between February 2004 and May 2005, 30 preterm infants (13 male; 17 female; g.a. 27.6 - 36.4 weeks) were screened for eligibility and enrolled in this study. Their demographic characteristics are detailed in Table 3.2. These infants exhibited a range of different ethnicities, with White British as the clear mode of this distribution. The distribution of study infants by g.a. and weight at the start of the study is shown in Figure 3.7.

3.3.4.2 Baseline vital signs

Baseline values for the monitored vital signs are shown in Table 3.3. Figures 3.8a through 3.8c contain the distribution of vital signs values registered during the study. These Figures were obtained from the analysis of the estimates produced by the patient monitor for each of the vital signs at 1-second increments over the entire duration of the study.

A first analysis of these variables shows that whilst the distribution of HR values appears as a classical bell-shaped distribution, the distribution of either RR or spO$_2$ display a degree of skewness, which explains the difference between their respective means and median shown in Table 3.3. Regarding RR, rates ranged from a lowest recorded value of 2.7 bpm to a maximum of 121.8 bpm. The mean rate was 51.1 bpm with a standard deviation of 18.5 bpm and a median of 49 bpm. Since the use of pulse oximetry became standard in the NICU, preterm infants are subject to tightly-controlled spO$_2$ targets. As a result, the distribution of this vital sign is mostly confined to the $]90;100]\%$ range with an evident mode on the $]99;100]\%$ histogram bin.

3.3.4.3 Monitoring statistics

Statistics regarding monitored times are reported in Table 3.4 and Figure 3.9.

(a)   (b)   (c)

Figure 3.8

Histograms of reference vital signs: Heart rate derived by the patient monitor from ECG (3.8a), respiratory rate derived by the patient monitor from IP (3.8b) and spO$_2$ derived by the patient monitor from the pulse oximeter (3.8c). + is the data mean, the boxplot bounds the 25% and 75% quartiles and the whiskers bound the 9% and 91% of the data. The black curve in (3.8b) shows a normal fit to the data.

3.4 Chapter summary

This chapter described the two datasets used to obtain the results reported in this thesis. Of all aspects of experiment design, a great emphasis was placed on the apparatus used in their acquisition, as addressing the challenges posed by the two experimental protocols was a non-trivial part of the work described here. The aim was not only to describe the hardware and software components involved, but also to explain the reasons why these particular devices or applications were selected.

The two resulting datasets were characterised in terms of their population characteristics of their cohorts and baseline vital sign data.

The Hypercapnoea dataset is analysed in Chapter 4, while the NICU dataset is analysed in Chapters and 7 and 8.

Table 3.4

Monitoring time over study subjects and sessions. Total monitoring time: 362.6 hours.

Monitoring time (hours)	Mean (SD)	Median (IQR)
Over subjects	14.0 (4.6)	15.0 (7.6)
Over sessions	4.7 (1.8)	4.7 (2.9)

SD = standard deviation; IQR = interquantile range
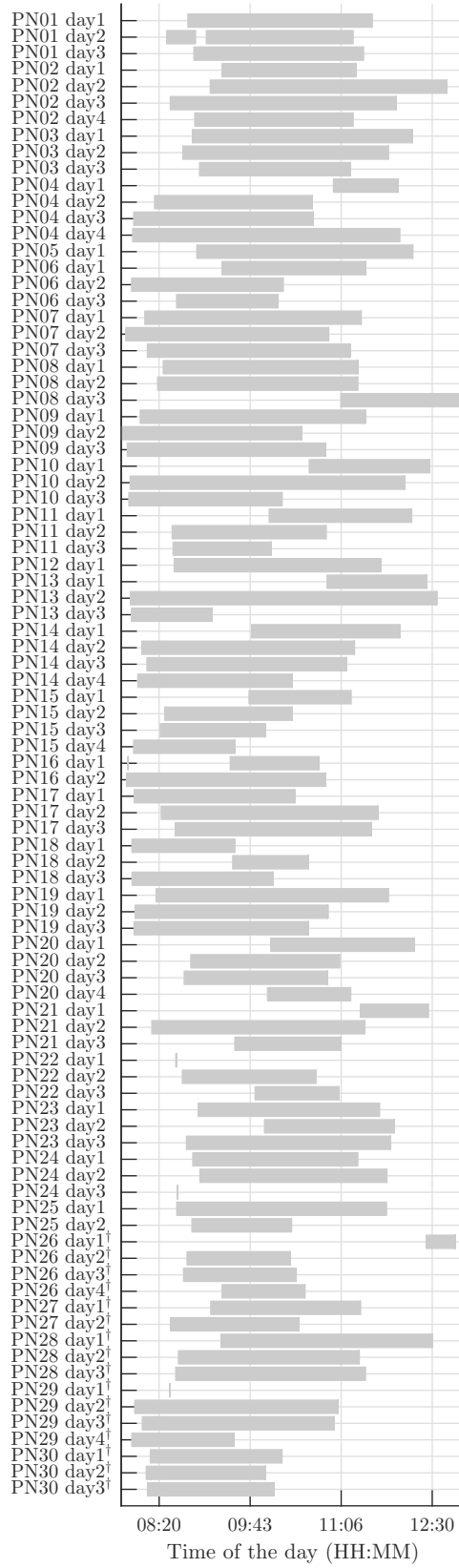
Recording times for the 90 study sessions in the NICU dataset. Manual breath data was available for the sessions marked with †).

61

Chapter 4

Signal processing methods

Contents

4.1 Chapter summary

This chapter presents the signal processing methods developed in this thesis for the analysis of physiological signals recorded using conventional methods.

First, we introduce methods for segmenting these signals into their repeating patterns. Secondly, we describe signal quality indices (SQIs) to identify segments corrupted by artefacts introduced during the recording of the signals. Thirdly, we examine how these signals are modulated by the respiratory cycle and, finally, we present signal processing

(a)

PPG
ECG
→ Beat Segmentation *(Section 4.2)* → PPG$_{peaks}$ PPG$_{onsets}$ ECG$_{QRS}$

→ SQI *(Sections 4.4.1, 4.4.2)* → SQI$_{PPG}$ SQI$_{ECG}$

→ Extraction of respiratory rhythms *(Sections 4.5.1, 4.5.2)* → RIAV RIIV RIFV RSA EDR

(b)

φ_{nasal} → Breath Segmentation *(Section 4.3.1)* → φ onsets → SQI *(Section 4.4.4)* → SQI φ

φ_{nasal}

(c)

IP → SQI *(Section 4.4.3)* → SQI$_{IP}$

IP
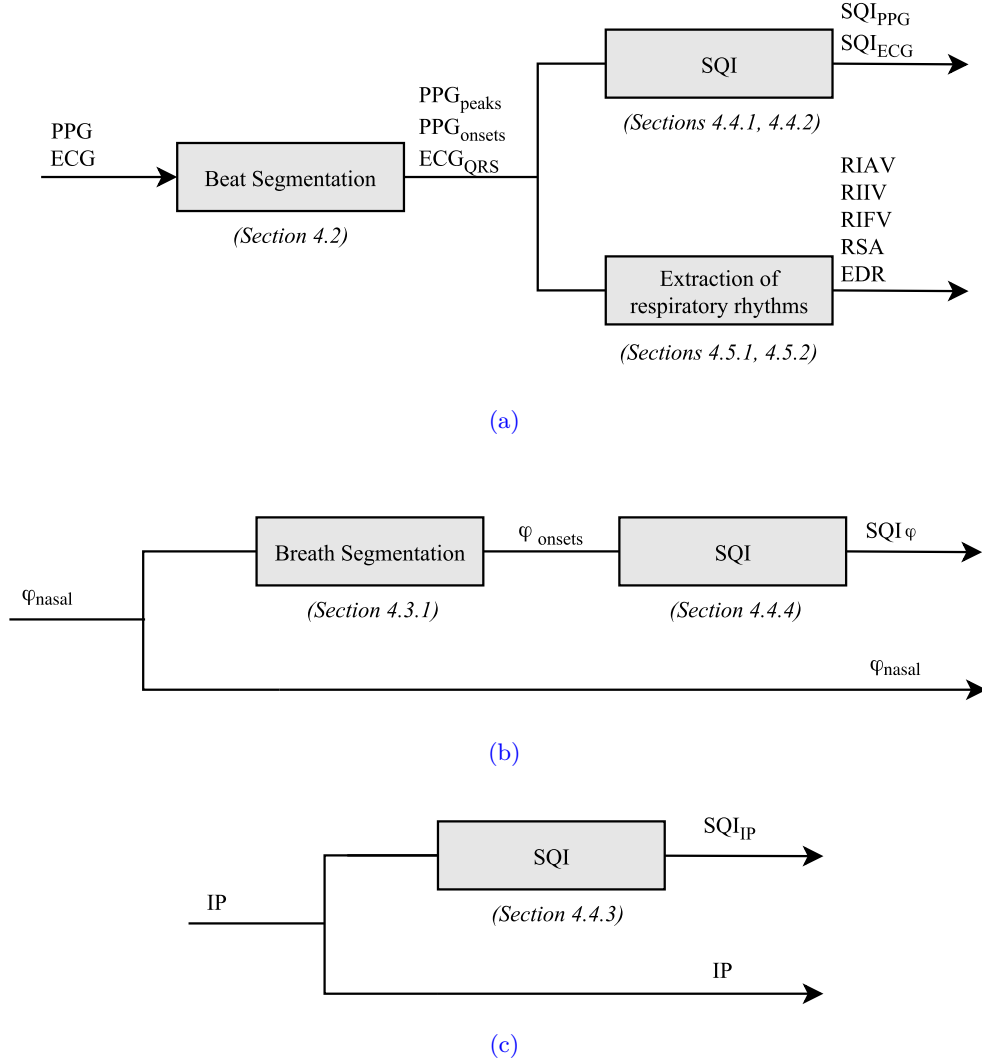
Figure 4.1

Computation of the respiratory waveform and signal quality indices. (4.1a) Cardiac beats were segmented from ECG and PPG waveforms; then, signal quality measures (SQI_{ECG}, SQI_{PPG}) and surrogate respiration signals were derived. (4.1b) Nasal airflow signals φ_{air} were segmented into breaths to permit the use of a morphological algorithm to compute SQI_{φ}. (4.1c) A signal quality index for the IP signal (SQI_{IP}) was also considered. Other quantities in the diagram are introduced later in this chapter.

techniques for extracting the respiratory rate (RR) from each source.

Our general approach to deriving these quantities is shown schematically in Figure 4.1. For each block (covered in more detail in Sections 4.2 through 4.5) we have selected methods whose performance has been tested on annotated datasets available in the public domain. Additional methodology is presented for tasks for which no generally accepted technique has been proposed. These are presented in Sections 4.3.1, 4.4.3, and 4.4.4.

The proposed methodology was validated against reference nasal airflow measurements from a subset of 12 healthy subjects (all spontaneously-breathing, all male, with an age range of 18 - 40 years old) whose minute ventilation was artificially elevated via a moderate hypercapnoeic stimulus supplied under the experimental protocol described in

Section 3.2. The typical physiological response to this protocol was quantified in Figure 3.4 in terms of changes in the three monitored vital signals with respect to their baseline values prior to the onset of the hypercapnoeic stimulus.

4.2   Segmentation of cardiac beats

The segmentation of cardiac beats in ECG or PPG signals relies on the occurrence of characteristic waveform patterns in these time series that are synchronous with the cardiac cycle (the QRS complex in the ECG and the BVP in the PPG). Thus, by segmenting the raw signals at fiducial points within their repeating units, one can obtain single-beat waveforms.

4.2.1   Segmentation from the electrocardiogram

ECG beat detection was performed using the well-known Hamilton and Tompkins algorithm [Pan and Tompkins, 1985]. The beat detector by Zong [2003] was also implemented, given its role (alongside the Tompkins detector) in the definition of ECG signal quality indices in Section 4.4.1.

Both algorithms implement a denoising step using digital filters, followed by a signal transform to enhance the sharp amplitude increase that characterises QRS complexes. While in the case of Pan and Tompkins [1985]'s detector the latter step is based on moving window integration of the squared (and filtered) ECG signal, in Zong [2003] it is handled by a length transform. Finally, in both cases, fiducial marks are identified in these transformed signals using adaptive thresholds. Local search strategies are then applied to find Q, R and S waves (here collectively referred to as ECG_{QRS}) in a neighbourhood of the threshold-crossing point.

Both algorithms have shown a high accuracy in previous studies using the single-channel ECG records in the MIT-BIH Arrhythmia database [Moody, 1990] (99.3% for Tompkins et al., and 99.7% for Zong et al.).

4.2.2   Segmentation from the photoplethysmogram

Much previous work exists concerning the segmentation of the PPG into cardiac beats from reference points in this waveform [Aboy et al., 2005; Karlen et al., 2012; Li et al., 2010; Zong et al., 2003]. Three methods available in the public literature for beat pulse segmentation were implemented and their results on the finger-PPG traces of our subset of 12 volunteers were analysed. The results of this comparative analysis are shown in Appendix D. As an outcome of this analysis, we opted for the use of the PPG peak and onset locations provided by Karlen et al. [2012] in the next sections of this thesis.

4.3 Segmentation of breaths

4.3.1 Segmentation from the nasal pressure flow signal

The segmentation of the subject nasal pressure flow signal (φ_{nasal}) into individual breaths has received surprisingly little attention in the literature. Nevertheless, as alluded to in Section 2.4.1.1, the morphology of pressure flow signals sensed at the opening of the nostrils is well-documented. A typical pressure airflow waveform is depicted in Figure 2.6 and labelled with the events that comprise one respiratory cycle. As evidenced in this figure, φ_{nasal} is a quasi-periodic signal with a characteristic morphology.

Building on this knowledge, we adapted one of the methods used for the segmentation of PPG pulses in Section 4.2.2 for breath segmentation on φ_{nasal} signals. To do this, we proceeded in the following way:

1. For each time window w (with a typical length of 60 seconds), we defined the time-reversed windowed φ_{nasal}, $\tilde{\varphi}_{nasal}$, as the signal given by[1]

$$\tilde{\varphi}_{nasal}(w, n) = \varphi_{nasal}(w, N_s^{(w)} - n), \quad n = 0, ..., N_s^w - 1 \tag{4.1}$$

   where n is the sample index and $N_s^{(w)}$ is the number of φ_{nasal} samples in window w.

2. To $\tilde{\varphi}_{nasal}$ we applied Zong et al. [2003]'s algorithm with an appropriate choice of parameters. We used an analysis window of 0.9 s to compute the values of the slope sum function (SSF) defined within this method. When the SSF signal crossed the detection threshold, a 0.7-second window around the crossing point was used to search for minimum and maximum SSF values. The initial threshold was set based on the first 15 seconds of each window.

4.4 Signal quality indices

Working with continuous measurements of biomedical signals one often encounters the need to distinguish between instances where the recorded signals have been acquired reliably from those where they are corrupted by movement artefacts, poor probe contact, electrical noise or other sources of measurement error. Under these circumstances, the collected signals are inadequate representations of the underlying physiological events.

The use of reliable signal quality metrics is, therefore, of fundamental importance as a pre-processing step to signal analysis so that the latter is only attempted in epochs for

---

[1]This step assumes no overlap in the windowing of the airflow signal although Equation 4.1 could trivially be adapted to accommodate for the general case.

which the recorded signals are artefact-free. In this section, we present the signal quality indices (SQIs) used throughout this thesis.

In this study, windows for SQI computation were set to have the same length as the windows over which the respiratory signals were derived (*e.g.* 60 s for the signals in the Hypercapnoea dataset).

4.4.1 SQI for the electrocardiogram

Prior work

Every measure of signal quality is designed to assess the fitness of a signal for a particular purpose. For example, an index designed to be applied to ECG waveforms prior to the extraction of respiratory rate (encoded in the low-frequency oscillations of the ECG) is manifestly different from one designed to be part of a system for the detection of myocardial events (which cause the elevation or depression of cardiac waves), or biometric identification (which relies on the variability in the high-frequency detail of QRS complexes). In essence, implicit to the definition of any SQI is the purpose for which the signal is to be processed and the subsequent analysis it will undergo.

In the field of clinical ECG analysis, a great deal of emphasis has been placed on the definition of SQI metrics to reduce false alarm rates in intensive care units (ICUs), a problem which leads to an increase in the workload of nursing staff and eventually their desensitisation to the alarms [McClure et al., 2016]. With this in mind, most definitions of ECG signal quality found in the literature have been developed for ICU patient data, particularly in the works of Behar et al. [2013]; Deshmane [2009]; Li et al. [2008]. The latter definition was adopted here given its widespread use in the literature.

Method

The acquired ECG signals were lineraly downsampled to 125 Hz and SQI of Li et al. [2008] was used by combining four quality measures; two based on beat detection:

SQI_{ECG}^b the comparison of multiple beat detection algorithms on a single lead;

SQI_{ECG}^i the comparison of the output of the same beat detection algorithm on different ECG leads.

and two binary indices based on:

SQI_{ECG}^k the evaluation of signal kurtosis;

SQI_{ECG}^s the ratio of the ECG spectral power found to be within a certain physiological frequency band on different ECG leads.

The quantities $SQI_{ECG}^{j}, j \in \{b, i, s, k\}$ are defined in Equations 4.2 to 4.7 and combined in the way shown in Equation 4.8. The rationale behind their definitions is covered in Sections 4.4.1.1 to 4.4.1.4. An in-depth discussion of these definitions and the parameters upon which they depend can be found in Li et al. [2008]. The parameters used by the authors (and adopted here) reflect domain knowledge in biomedical signal analysis.

4.4.1.1 Multiple beat detection algorithms on a single lead

Since the comprehensive study of the noise sensitivity of nine QRS detection algorithms by Friesen et al. [1990], it has been known that different QRS detection algorithms exhibit distinct failure modes, which in turn render them susceptible to different types of noise.

QRS detection in this thesis was performed on each channel individually using the two open-source detectors in Pan and Tompkins [1985] and Zong [2003]. As the former is less sensitive to noise [Friesen et al., 1990], discrepancies in their output can be regarded as an indicator of the presence of signal artefacts. A single-lead signal quality index for window w, $SQI_{ECG}^{b}(w)$, was then defined as the ratio of beats detected synchronously by both algorithms (*i.e.* within a tolerance interval, ξ) over the total number of beats detected (by either algorithm) within the window w (Equation 4.2).

$$SQI_{ECG}^{b}(w) = N_{b \times b\ matches}(w)/N_{QRS\ detected}(w) \tag{4.2}$$

where $N_{b \times b\ matches}(w)$ is the number of beats in window w simultaneously detected by both algorithms (subject to $\xi = 150$ ms) and $N_{QRS\ detected}(w)$ is the number of beats detected by one of the algorithms. As the least noise-sensitive of the two techniques, the Pan and Tompkins' detector was selected for this task.

4.4.1.2 Beat detection on different ECG leads

When several ECG leads are available, an inter-channel SQI (SQI_{ECG}^{i}) can be calculated. This is to safeguard against scenarios in which the quality of one ECG trace on one lead is such that one of the QRS detection algorithms misses one or more beats (due to low QRS amplitudes on that lead) or detects extra beats (due to artefact or high amplitude T waves). Whilst one can calculate SQI_{ECG}^{b} for each lead separately, only beat-by-beat comparison between different leads (using the same QRS detection algorithm), allows us to ascertain if the measurements from that lead are reliable.

For a given analysis window w, $SQI_{ECG}^{i}(w)$ was computed in the following way

$$SQI_{ECG}^{i}(w)_{i'} = \max_{\forall j', i' \neq j'} (N_{b \times b\ matches}(w)_{i',j'}/N_{QRS\ detected}(w)_{i'}) \tag{4.3}$$

where i' represents the current ECG lead and j' each of the other leads.

4.4.1.3 Kurtosis SQI

Kurtosis is defined analytically as the fourth standardised moment of a distribution. Given a signal sample x of size M with empirical mean $\hat{\mu}_x$ and standard deviation $\hat{\sigma}_x$, the empirical estimate of kurtosis, $\widehat{kurt(w)}$, is given by

$$\widehat{kurt(w)} = \frac{1}{M} \sum_{i=1}^{M-1} \left| \frac{x_i - \hat{\mu}_x}{\hat{\sigma}_x} \right|^4 \tag{4.4}$$

Graphically, kurtosis measures the relative *peakedness* of a distribution with respect to a Gaussian distribution. The kurtosis of a Gaussian distribution is equal to 3. He et al. [2006] found that the kurtosis of clean ECG traces normally exceeds 5. Common sources of electrocardiographic noise, such as muscle artefact or mains interference, are correlated over long time scales and thus tend to have lower kurtosis of around 5 and below. Based upon these findings, a single threshold rule was used to convert the value of $\widehat{kurt(w)}$ into a binary measure of signal quality, $SQI_{ECG}^k(w)$ (Equation 4.5).

$$SQI_{ECG}^k(w) = \begin{cases} 1, & \text{if } \widehat{kurt(w)} > 5 \\ 0, & \text{if } \widehat{kurt(w)} \leqslant 5 \end{cases} \tag{4.5}$$

4.4.1.4 Spectral SQI

In a typical QRS complex of normal duration, virtually all of the signal power is contained in frequencies below 30 Hz, with peak power occurring in the range of 4 to 12 Hz [Murthy et al., 1978]. Thus, a spectral-based measure of clean sinus rhythm ECG can be defined as the ratio of power over this band compared to the power in the overall signal (Equation 4.6).

$$SDR(w) = \int_{5Hz}^{14Hz} \mathcal{P}_w(f)df \Big/ \int_{5Hz}^{50Hz} \mathcal{P}_w(f)df \tag{4.6}$$

where $SDR(w)$ is the spectral distribution ratio for an ECG segment w with PSD given by $\mathcal{P}_w$.

It has been noted that when the SDR is low (< 0.5), high-frequency noise contamination in the ECG, usually by muscle artefact, is the likely cause [Li et al., 2008]. Conversely, when the SDR is high (> 0.8), artefacts with a QRS-like presentation, such as those caused by electrode motion, are likely to be present. Hence, moderate values of SDR indicate a reliable ECG trace. Based on these observations, the $SDR(w)$ for a given window w can be converted to $SQI_{ECG}^s(k)$ using the two thresholds set empirically by Li et al. [2008].

$$SQI^s_{ECG}(w) = \begin{cases} 1, & \text{if } SDR(w) \geqslant 0.5 \wedge SDR(w) \leqslant 0.8 \\ 0, & \text{otherwise} \end{cases} \tag{4.7}$$

The SQI quantities defined in the previous sections were then combined in the following way to obtain a global SQI for use in ECG signals SQI_{ECG}.

$$SQI_{ECG}(w) = \begin{cases} max(SQI^b_{ECG}(w), SQI^i_{ECG}(w)), & \text{if } SQI^k_{ECG}(w) = 1 \wedge SQI^s_{ECG}(w) = 1 \\ SQI^b_{ECG}(w), & \text{if } SQI^k_{ECG}(w) = 1 \wedge SQI^s_{ECG}(w) = 0 \\ \eta \, max(SQI^b_{ECG}(w), SQI^i_{ECG}(w)), & \text{if } SQI^k_{ECG}(w) = 0 \wedge SQI^s_{ECG}(w) = 1 \\ \eta SQI^b_{ECG}(w), & \text{if } SQI^k_{ECG}(w) = 0 \wedge SQI^s_{ECG}(w) = 0 \end{cases} \tag{4.8}$$

where η is arbitrarily set to 0.7 according to Li et al. [2008]. The rationale behind each branch of this function is as follows:

- if both SQI^k_{ECG} and SQI^s_{ECG} indicate good signal quality ($SQI^k_{ECG} = 1 \quad \wedge$ $SQI^s_{ECG} = 1$), SQI_{ECG} is set to SQI^i_{ECG} or SQI^b_{ECG} (whichever is higher) as both metrics should be reliable (first branch);

- if SQI^s_{ECG} indicates the presence of spectral noise ($SQI^s_{ECG} = 0$), SQI^i_{ECG} is ignored since artefact is often present on both channels and this leads to matched detections being registered on both channels (second and fourth branch);

- if SQI^k_{ECG} is low, SQI^b_{ECG} is multiplied by an attenuating factor η lower than 1 (third and fourth branch).

The following modifications were made to Li et al's method prior to its application in our datasets: a shorter window of 100 ms was applied for beat-by-beat comparisons to achieve an increase in sensitivity; where only single-lead recordings were available (*e.g.* in the Hypercapnoea dataset), SQI^i_{ECG} was set to 1.

In this study, SQI_{ECG} was evaluated for 10-second segments with 90 % overlap and its results filtered with a moving average filter with length equal to that of the analysis window used for the extraction of the respiration signal (60 seconds in the Hypercapnoea dataset and 30 seconds in the NICU dataset).

4.4.2   SQI for the photoplethysmogram

Prior work

PPG signals are highly prone to motion artefacts which pose a serious obstacle to the extraction of reliable cardiac [Gil and Bailon, 2010] and respiratory [Madhav et al., 2010]

information from this signal. Whilst a variety of algorithms have been designed to cope with the undesirable effects of such artefacts, only a few studies have produced explicit SQI metrics.

Statistical methods have been proposed [Gil and Bailon, 2010] that compute the relationship between the spectral moments of the PPG power spectrum (Hjorth parameters [Hjorth, 1970]) to assess signal quality. These have been successfully used for alarm suppression in arrhythmia [Deshmane, 2009] and apnoea [Monasterio et al., 2012] surveillance systems.

Approaches based on morphological analysis are also available. Using a simple decision tree and a set of arbitrary thresholds, Sukor et al. [2011] classified the signal quality of individual PPG pulses into one of three categories (good, poor or bad) based on five wave measurements (pulse amplitude, pulse width, trough depth difference, Euclidean distance and amplitude ratio) performed with respect to a running beat template. The performance of this method was tested on intervals of PPG records acquired with finger probes whose quality was artificially degraded with synthetic motion noise. In comparison to manually annotated beats, the mean sensitivity, specificity and accuracy for beat detection were respectively $89 \pm 11\%$, $77 \pm 19\%$ and $83 \pm 11\%$ on 104 fingertip PPG signals, acquired from 13 healthy volunteers.

More recent studies by Karlen et al. [2012] have also implemented morphology analysis by comparing each pulse with previously detected pulses superimposed at primary peak locations. Pulse detection was performed using a computationally-intensive combination of time-based methods (through differentiation) and frequency-based methods (through repeated Gaussian filtering on the Discrete Cosine transformed PPG signal). Whilst this algorithm achieved promising results for pulse detection on both the Complex Systems Laboratory (CSL), Portland, USA dataset [Goldberger et al., 2000] (with a sensitivity of 96.21% and a positive predictive value of 99.22%) and the CapnoBase, British Columbia, Canada [Karlen et al., 2010] (with a sensitivity of 96.44% and a positive predictive value of 99.80%), only a qualitative assessment of the SQI defined was offered. As a subsidiary result, the authors reported significant differences in the distributions of SQI values for pulses labelled as artefact and regular pulses on the combined CapnoBase and CSL benchmark datasets.

Li and Clifford [2012] provided a morphological formulation of the assessment of the signal quality of PPG pulses that has gathered ample support in the PPG literature. Their algorithm was trained and tested on 1055 segments of PPG, each with a duration of 6 s, recorded from 104 separate recordings on normal and verified arrhythmic events from a publicly available clinical database of patients in ICU. A sensitivity of 99.0% and a positive predictive value of 95.2% were found for quality estimation against manual annotations provided by experts.

Method

To assess the reliability of PPG signals we follow the steps described previously in Li and Clifford [2012]. In brief, Li et al's SQI metric, SQI^b_{PPG}, takes as input parameters the peak or onset locations that segment each beat (found using the methods in Section 4.2.2) in order to build a dynamic beat template. Such template was constructed by averaging the initial ten beats in a given time window w. The similarity between each subsequent PPG beat and the template was then computed using cross-correlation between the beat template and (a) the original beat (for SQI^o_{PPG}); (b) the interpolated and linearly detrended beat (for SQI^i_{PPG}), and (c) the beat obtained after dynamic time-warping[1] (for SQI^{dtw}_{PPG}). SQI^{dtw}_{PPG} was added to account for small morphological changes that a PPG pulse can undergo over the course of the analysis window (*e.g.* due to naturally-occurring physiological factors such as changes cardiac frequency or output). Additionally, a binary classifier to detect clipping of the PPG signal was added as a fourth SQI measure (SQI^c_{PPG}). The authors then fused these four measures using a set of heuristics (Equation 4.9) to obtain a beat-by-beat SQI, which stratifies beat quality into three categories (good, average or bad quality) with quality indices 1, 0.8 or 0.

$$SQI^\star_{PPG}(b) = \begin{cases} 1, & \text{if all the } SQI^{o,i,dtw,c}_{PPG}(b) \geqslant 0.9 \\[2mm] 0.8, & \begin{cases} \text{if 3 of the 4 } SQI^{o,i,dtw,c}_{PPG}(b) \geqslant 0.9 \\ \text{if all of the 4 } SQI^{o,i,dtw,c}_{PPG}(b) \geqslant 0.7 \\ \text{if } median(SQI^o_{PPG}(b), SQI^i_{PPG}(b), SQI^{dtw}_{PPG}(b)) \geqslant 0.8 \\ \quad \wedge SQI^o_{PPG}(b) \geqslant 0.5 \\ \quad \wedge SQI^c_{PPG}(b) \geqslant 0.7 \end{cases} \\[2mm] 0, & \text{otherwise} \end{cases}$$

(4.9)

In our implementation, we formed an interval-based SQI_{PPG} by adopting the mean $SQI^\star_{PPG}(w)$ score in our analysis window w as the $SQI_{PPG}(w)$ for that window[2].

4.4.3 SQI for the impedance pneumogram

The selected index adopted here for determining the quality of IP signals is spectral purity (Γ_s). Γ_s was derived from the EEG domain. Its use as a measure of signal periodicity was first proposed by Goncharova and Barlow [1990] in EEG studies and later extended to other physiological signals by Gil et al. [2008] and successfully applied to IP signals by Monasterio et al. [2012]; Nemati et al. [2010].

Γ_s is an heuristic related to the Hjorth descriptors [Hjorth, 1970] (namely the com-

---

[1]Time warping adjusts the time axis in order to minimise the cumulative distance between the beat and the template. A description of this procedure can be found in [Keogh and Ratanamahatana, 2005].

[2]Note that, unlike the previously defined $SQI_{ECG}(w)$, $SQI^\star_{PPG}(b)$ was defined on a beat basis instead of a window basis.

plexity descriptor $\mathcal{H}_3$), which describe the spectrum of a signal in terms of its moments. The nth order spectral moment $\bar{\Omega}_n$ of a signal are given by the integral

$$\bar{\Omega}_n = \int_{-\pi}^{\pi} \Omega^n \mathcal{P}_w(e^{j\Omega}) d\Omega \tag{4.10}$$

where $\mathcal{P}_w(e^{j\Omega})$ is the power spectrum of the windowed signal and Ω is the angular frequency ($\Omega = 2\pi f$, with f in physical frequency units). Note that, due to the symmetry of the power spectrum (*i.e.* $\mathcal{P}(e^{j\Omega}) = \mathcal{P}(e^{-j\Omega})$), the odd-numbered moments are all identical to zero. The spectral purity index is then defined as the ratio between the squared 2nd order moment, and the product of the 0th order moment (*i.e.* the total power) and 4th order moment [Sornmo and Laguna, 2014]:

$$\Gamma_s = \frac{\bar{\Omega}_2^2}{\bar{\Omega}_0 \bar{\Omega}_4} \tag{4.11}$$

Γ_s is designed to reflect the bandwidth of the signal[1], *i.e.* how well the signal may be described by a periodic signal with a single dominant frequency. Following Equation 4.11, Γ_s is a strictly positive quantity with a maximum value of unity for a sinusoidal signal.

In the presence of a regular breathing pattern, the IP signal should resemble a sinusoid with a frequency given by the respiratory rate and, hence, $\Gamma_s = 1$. It should approach zero in epochs of poor signal quality, during which the acquired signal is non-sinusoidal.

Finally, to obtain SQI_{IP} we have computed Γ_s over sliding windows (of 60 s with 5 s shift between windows, in the Hypercapnoea dataset). Γ_s was estimated using the temporal expressions of the spectral moments. When Equation 4.10 is rewritten in terms of signal derivatives, expressions for the second and fourth moments can be implemented as successive first-order differences on the signal without the need for the expensive computation of $\mathcal{P}_w(e^{j\Omega})$.

4.4.4 SQI for the nasal pressure airflow signal

Given the lack of a conventionally accepted signal quality index for the nasal flow signal, we have designed a simple SQI based on the agreement between the respiratory rate estimates (in bpm) computed by two different methods for the same window. The two estimates were derived from distinct measurement principles: a time-domain method and a frequency-domain method.

Time-domain estimates ($RR_{\varphi,onsets}$) were found by computing the median inter-breath interval between the inspiratory onsets within the analysis window (using the

---

[1]and for this reason, it is related to the Hjorth complexity descriptor $\mathcal{H}_3$, given by $\mathcal{H}_3 = \sqrt{\frac{\bar{\Omega}_4}{\bar{\Omega}_2} - \frac{\bar{\Omega}_2}{\bar{\Omega}_0}}$.
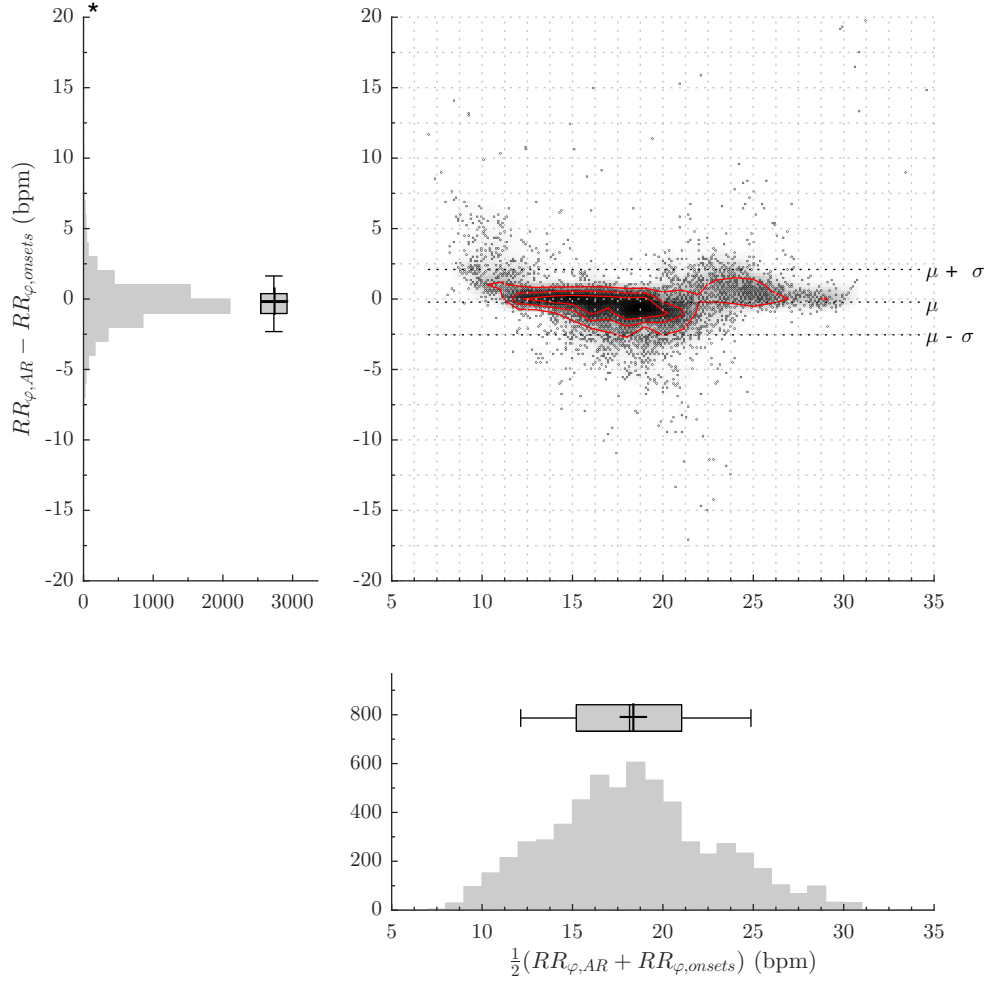
Bland Altman plot between time-domain estimates ($RR_{\varphi,onsets}$) and frequency-based estimates ($RR_{\varphi,AR}$) of respiratory rate from the nasal pressure airflow signal φ_{nasal}. RR estimates computed for all 60-second time windows of the study sessions of 12 young healthy volunteers. The red curves represent isolines of the local density of data points. μ and σ represent the mean and standard deviation of the $RR_{\varphi,AR} - RR_{\varphi,onsets}$, respectively.

methodology described Section 4.3). To compute the frequency-based estimate ($RR_{\varphi,AR}$) for the same window, we downsampled the φ_{nasal} signal to 4 Hz and modelled the resulting time series using a 9th order AR model. The respiratory estimate was given by the frequency corresponding to the highest peak in the Burg's power spectrum of this signal over the range of plausible frequencies for RR (5 to 50 bpm for adults at rest). The, SQI_φ for window w was given by Equation 4.12.

$$SQI_\varphi(w) = \begin{cases} 1, & \text{if } \mid RR_{\varphi,AR}(w) - RR_{\varphi,onsets}(w) \mid \leqslant \Omega_\varphi \text{ (in bpm)} \\ 0, & \text{otherwise} \end{cases} \tag{4.12}$$

To set the threshold Ω_φ, we analysed first how the respiratory rates were distributed for the 12 volunteer sessions studied (shown in Figure 4.2), and secondly, the effect, in

73
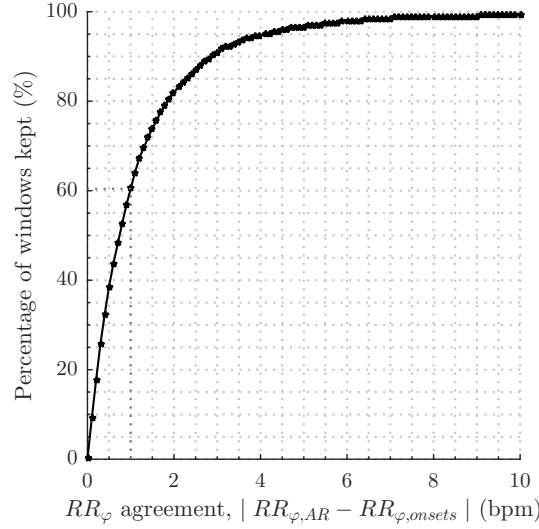
Percentage of φ_{nasal} windows retained when the threshold Ω_φ for agreement between $RR_{\varphi,AR}$ and $RR_{\varphi,onsets}$ is set to different values in the 0 - 10 bpm range.

terms of proportion of windows analysed[1], of setting Ω_φ to values in a range of 0 to 10 bpm (shown in Figure 4.3).

We note that no meaningful bias existed between RR estimates provided by the two methods. In fact, most RR estimates for the same window were within 1 bpm of each other throughout the range of RR values observed (from 5 to 35 bpm). Thus, this value (1 bpm) is a plausible choice for Ω_φ. It allowed us to exclude the most of the outliers in Figure 4.2, as it lies close to the inflection point of the curve in Figure 4.3, setting Ω_φ to this value means that most of the φ_{nasal} windows (60.4%) were kept for subsequent analysis.

4.5 Extracting the respiratory signal

We define the respiratory signal as a waveform which resembles a sinusoid whose fundamental frequency is the respiratory rate. The physiological signals introduced in Section 2.4.1 fall into two categories. Some measure respiration explicitly, either in the shape of tidal flow (*e.g.* capnography), respiratory effort (*e.g.* impedance pneumography) or other measurand from which the respiratory waveform can be directly extracted, whereas in others, such as the physiological signals monitored more routinely (*e.g.* ECG or PPG), respiration exerts only a modulatory effect on the original signal, and thus it is not so readily available.

In the latter case, strategies for extracting the respiratory signal from the original

---

[1]This is equivalent to what is often referred to as the coverage probability. The coverage probability is the probability that the difference between two independent measurements falls within pre-defined bounds (parametrised here by Ω_φ). It is usually expressed as a percentage calculated using the empirical cumulative distribution of the absolute differences [Barnhart et al., 2007].

signals are needed. In this section we describe how the respiratory signal was extracted from the signals collected. In the examples shown, raw signals were segmented into 60 s windows from which the respiratory waveform was extracted. After each computation, the window was shifted by 5 seconds and another minute of raw signal data was investigated.

4.5.1 Respiratory signal from the electrocardiogram

In Figure 4.4 we show a segment from a single-lead ECG record during a period of quiet breathing for a subject in the Hypercapnoea dataset. The locations of the R-peaks were found using Pan and Tompkins [1985]. We notice the presence of some low-amplitude oscillation in this waveform. This is the well-known amplitude modulation of ECG by the respiratory cycle discussed in Section 2.4.1.4. To extract this modulation, as well as the frequency modulation signal, the following to approaches may be considered: using the changes in the beat morphology caused by respiration (see Section 4.5.1.1) or the modulations of the cardiac frequency (see Section 4.5.1.2).
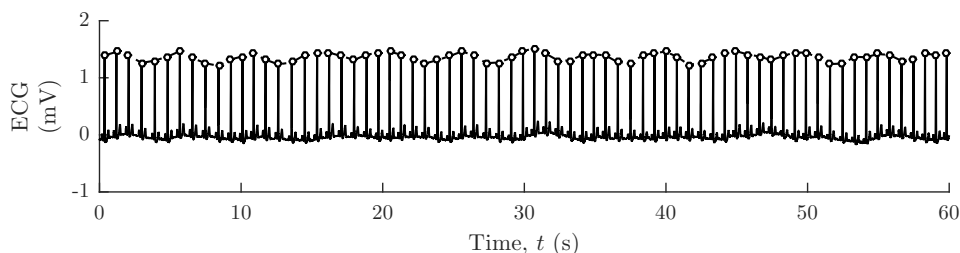
Figure 4.4

Excerpt from a single-lead ECG record of a healthy subject over a 60-s window. The R-peaks have been calculated (circles) using the Pan and Tompkins [1985]'s algorithm. The fluctuations in the height of the QRS complex are clearly visible. The area under the QRS complexes and R-R intervals in this record are analysed in Figure 4.6.

4.5.1.1 Beat morphology

Prior work

Common ways of extracting the respiratory-synchronous modulation of beat morphology include the measurement of the amplitude of the R-wave (*e.g.* Khaled et al. [1992]), the QRS height[1] (*e.g.* Dobrev and Daskalov [1999]), or the area under the QRS complex in relation to a running baseline (*e.g.* Moody and Mark [1985]).

The methodology for extracting respiratory information from respiratory-induced variations in beat morphology was first explored in a seminal paper by Moody and Mark [1985], which exploited the oscillations in the ratio of the QRS areas measured in two ECG leads as a measure of respiratory activity. There has since been a surge of interest

---

[1]The QRS height is the difference in amplitude between the R and S waves.

in this topic resulting in a multitude of signal processing algorithms to derive respiratory-induced modulations from both multi-lead [Travaglini et al., 1998; Zhao et al., 1994] and single-lead [Cysarz et al., 2008; De Chazal et al., 2003; Dobrev and Daskalov, 1999; Khaled et al., 1992; Mason and Tarassenko, 2001; Park et al., 2008; Yi and Park, 2002] ECG signals.

In recent research, it has been established that single-lead methods generally yield more robust estimates than methods based on multi-lead tracking of the mean electrical axis [O'Brien and Heneghan, 2007], and so we adopt the approach pioneered by Moody and Mark [1985] to derive a respiration effort from a single ECG lead.

Method

After subtracting the signal baseline (given by a 25-ms moving-average of the ECG signal), the area under each of the QRS complexes in an ECG segment w was measured over a fixed window centred on each R-peak. The length of this window was determined during the learning phase of the algorithm which extends for ten QRS complexes. During this phase, the mean duration of the complexes was approximated as average interval from the PQ junction to the J-point of the QRS complexes in the learning window)[1]. This process is illustrated in Figure 4.5a. We defined N_{QRS}^w as the total number of QRS complexes detected in window w.

The time series obtained as a result of this process, $\text{EDR}^\star(w; t)$, was then linearly interpolated between the time instants corresponding to the first and last R-peaks in each window (t_{R_1} and $t_{R_{N_{QRS}^w}}$, respectively) to obtain an evenly sampled time-series $\text{EDR}(w; t)$ at a sampling rate of 4 Hz (Figure 4.6a).

4.5.1.2 Respiratory sinus arrhythmia

Prior work

The second approach relies on the well-known modulation of heart rate by the respiratory-induced pressure changes in the main and peripheral arteries (Section 2.2.4).

Although the respiratory sinus arrhythmia (RSA) may also be extracted by means of time-domain methods, spectral analysis of the time series of beat-to-beat R-R intervals (the RR tachogram), which searches for RSA as the peak in the heart rate variability (HRV) spectrum over the expected HF range [Akselrod et al., 1981], remains the preferred method to extract the variations in the instantaneous heart rate governed by RSA [Cysarz et al., 2008; Mason and Tarassenko, 2001; Park et al., 2008; Schäfer and Kratky, 2008; Thayer et al., 2002].

---

[1]Since the window length is fixed, this area is proportional to the mean amplitude of the signal, or equivalently, the projection of the mean cardiac electrical vector on that electrode lead axis.

(a)

(b)
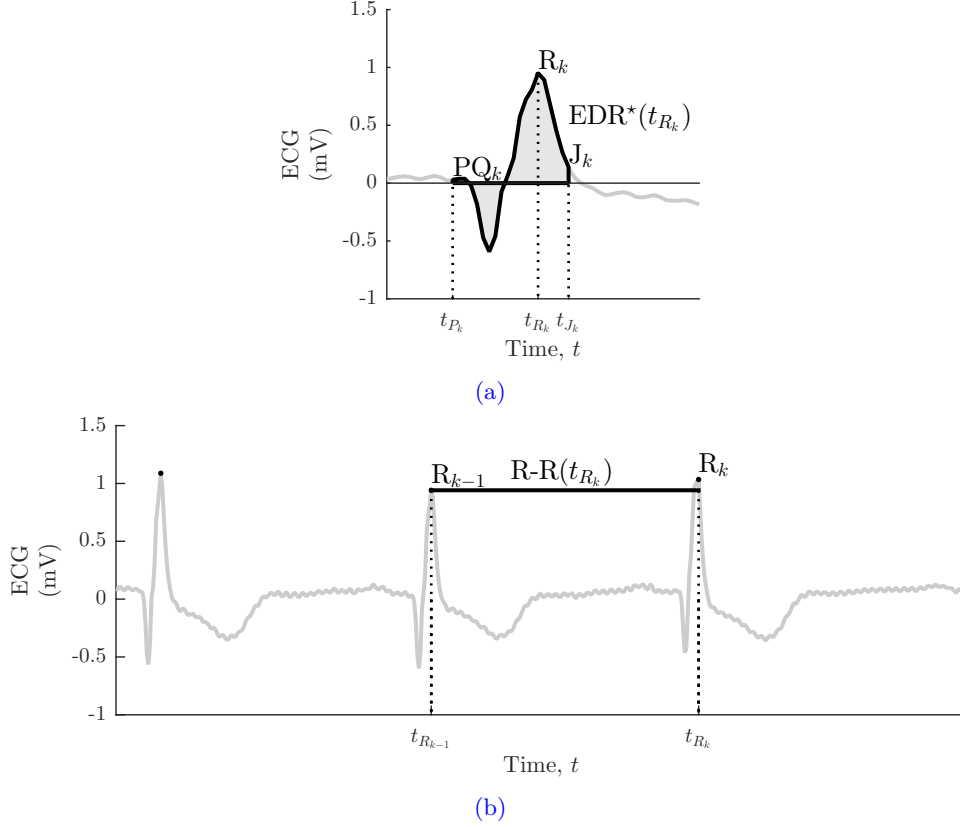
Figure 4.5

Excerpts from a single-lead ECG record of a healthy subject illustrating the measurement of the ECG-derived respiration from QRS morphology ($\text{EDR}^{\star}(t_{R_k})$) and R-R interval for one sample beat k. (4.5a) A typical QRS complex illustrating the measurement of the time-voltage area of each QRS complex used to derive $\text{EDR}(t_{R_k})$. Moody and Mark [1985]'s algorithm measures the area inscribed by each QRS complex from the PQ-joint as denoted by the shaded area. (4.5b) The R-$R(t_{R_k})$ intervals between consecutive R-waves. The time of the end R-peak was taken as the time of R-$R(t_{R_k})$.

Method

The N_{R-R}^{w} R-R intervals in the analysis window w were calculated as the difference between the times of successive R-peaks (found using Pan and Tompkins [1985] and shown in Figure 4.4) according to Equation 4.13.

$$R\text{-}R(w;t_{R_k}) = t_{R_k} - t_{R_k-1}, k = 2, ..., N_{R-R}^{w} \tag{4.13}$$

This process is shown schematically in Figure 4.5b. The R-$R(w;t)$ series obtained was then linearly interpolated between the time instants corresponding to the first and last R-peaks in each window (t_{R_1} and $t_{R_{N_{QRS}^w}}$, respectively) to obtain an evenly sampled time series $\text{RSA}(w;t)$ at a sampling rate of 4 Hz (Figure 4.6a).
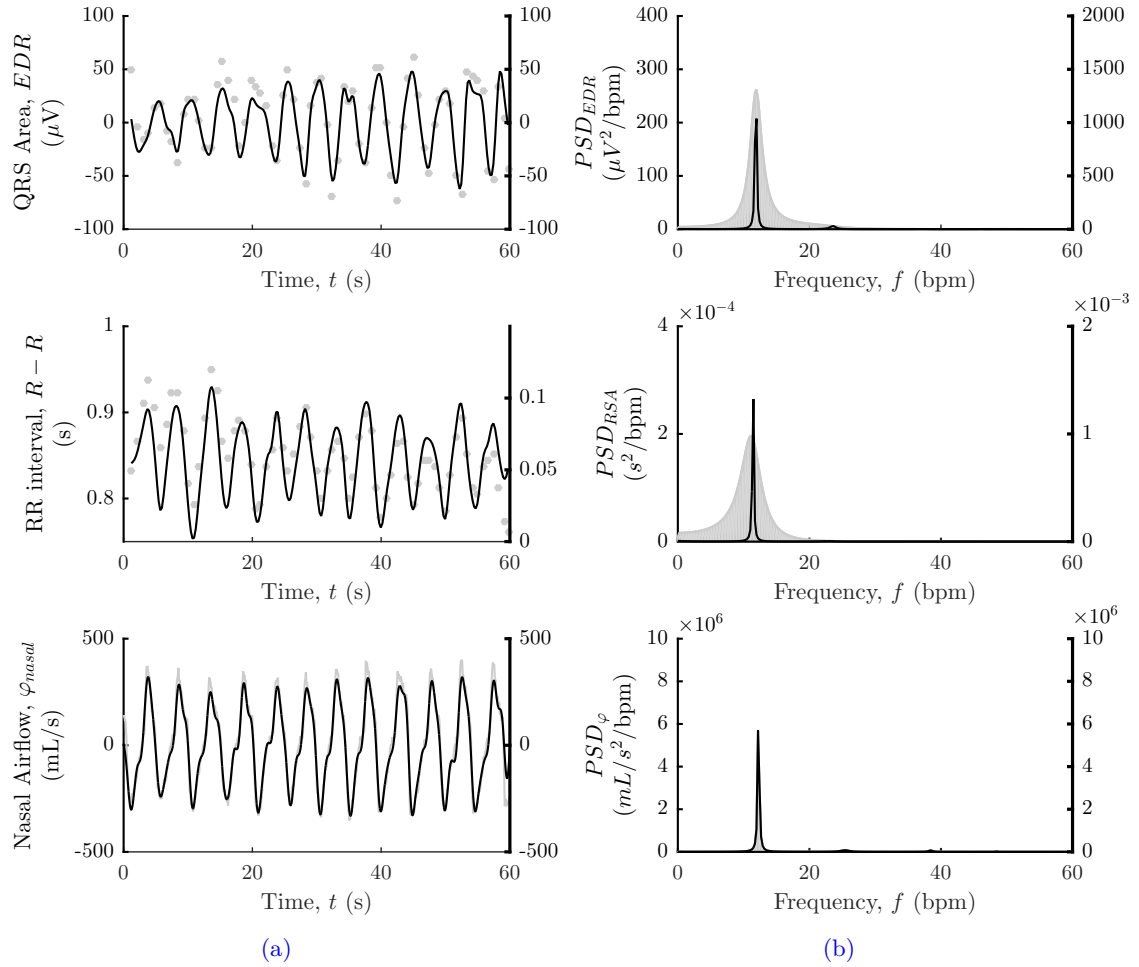
Respiratory waveforms derived from a 60-second interval of ECG and accompanying airflow trace. (4.6a) From top to bottom: QRS area under a single ECG lead at the times of the corresponding R-peaks (EDR⋆, dots) and same time series resampled at 4 Hz and bandpass-filtered (EDR, solid line); Duration of R-R intervals at the times of the corresponding R-peaks (R-R, dots) and the series resampled at 4 Hz and bandpass-filtered (RSA, solid line); Accompanying airflow trace. It may be seen from the power spectral density of these signals on the right-hand side (4.6b) that the respiratory signal is present in all three of these waveforms for this particular window. Power spectral density estimates for the original (shaded area) and the filtered time series (solid line) were computed using a 9th order AR model. The units on the vertical axes on the left side of the panels pertain to the curves plotted as solid lines.

4.5.2 Respiratory signal from the photoplethysmogram

The physiological mechanisms whereby the respiratory waveform modulates the PPG were discussed in Section 2.4.1.4. In Figure 4.7 we show an excerpt from a PPG record during a period of quiet breathing for a subject in the Hypercapnoea dataset. The locations of the waveform peaks and onsets were found using the methods in Section 4.2.2. As this figure suggests, PPG displays some periodic behaviour at the fundamental respiratory frequency. It should, therefore, be possible to extract a PPG-derived respiration. Albeit less extensive than for eletrocardiography, the body of literature covering photoplethysmography as a tool for monitoring respiration offers a range of methodologically distinct approaches.

Prior work

Much previous work exists concerning the extraction of the respiratory signal from the PPG using a combination of digital filtering and frequency-based spectral analysis [Foo and Wilson, 2005; Johansson et al., 1999; Johnston and Mendelson, 2004; Lindberg et al., 1992; Nilsson et al., 2005; Olsson et al., 2000; Wertheim et al., 2009]. As an alternative to Fourier methods, parametric techniques have been proposed to analyse the PPG spectra [Fleming et al., 2008; Fleming and Tarassenko, 2006; Lee and Chon, 2010] yielding mean absolute errors as low as 0.04 bpm in respiratory rate estimation from a dataset consisting of 45-minute sections from seven individual records in the MIMIC database [Fleming and Tarassenko, 2006].

In recent years, some groups have noted that the subtle transient components of PPG modulation by ventilation can be obscured by the averaging inherent to frequency-only methods. A number of time-frequency methods have since been applied. These include the short time Fourier transform (STFT) [Shelley et al., 2006], complex demodulation (CMD) [Chon et al., 2009; Dash et al., 2010] and the continuous wavelet transform (CWT) [Addison et al., 2014; Clifton et al., 2007; Leonard et al., 2004, 2006], which has become
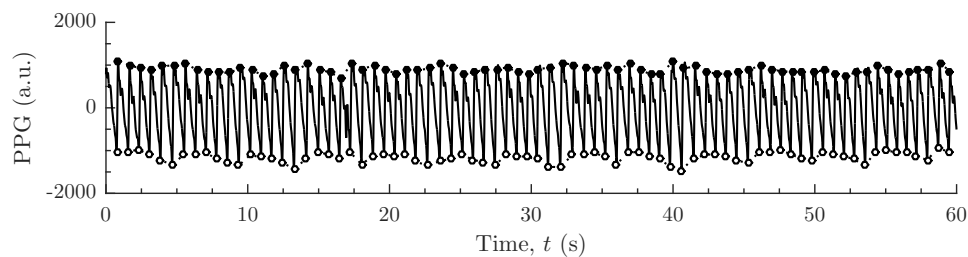
Figure 4.7

Excerpt from the PPG record of a healthy subject over a 60-second window. The peaks (dots) and onsets (circles) of the pulse wave peaks have been found using Karlen et al. [2012]. Some amplitude and frequency fluctuations can be distinguished: baseline modulation (cardiac pulses on top of a baseline varying over the respiratory cycle), amplitude modulation (cardiac pulse amplitudes varying over the respiratory cycle) and RSA respiratory sinus arrhythmia (pulse period varying over the respiratory cycle). The amplitude, intensity and frequency modulations for this segment are analysed in Figure 4.9.

the most favoured tool of researchers among time-frequency methods, mainly due to its frequency-dependent windowing.

Other approaches have been suggested to remove frequency components associated with non-breathing oscillations from the PPG waveform spectra. For instance, Johansson [2003] used neural networks, while an ICA-based approach has been successfully applied by Zhou et al. [2006]. More sophisticated spectral analysis, has also been employed; Garde et al. [2014] examined respiration and heart rate extraction using the correntropy spectral density of the photoplethysmogram.

Finally, respiratory-led fluctuations can be decomposed into three fundamental waveform modulations [Meredith et al., 2012]. Often the preferred methods consist in a combination of these modulations and that is the approach we adopt here. Karlen et al. [2013] has taken a fusion approach which combined respiratory rate estimates from time series extracted from the PPG waveform measurements. Lázaro and Gil [2011] extracted respiration as a combination of three parameters present in the PPG signal: pulse rate variability, pulse amplitude variability and pulse width variability.
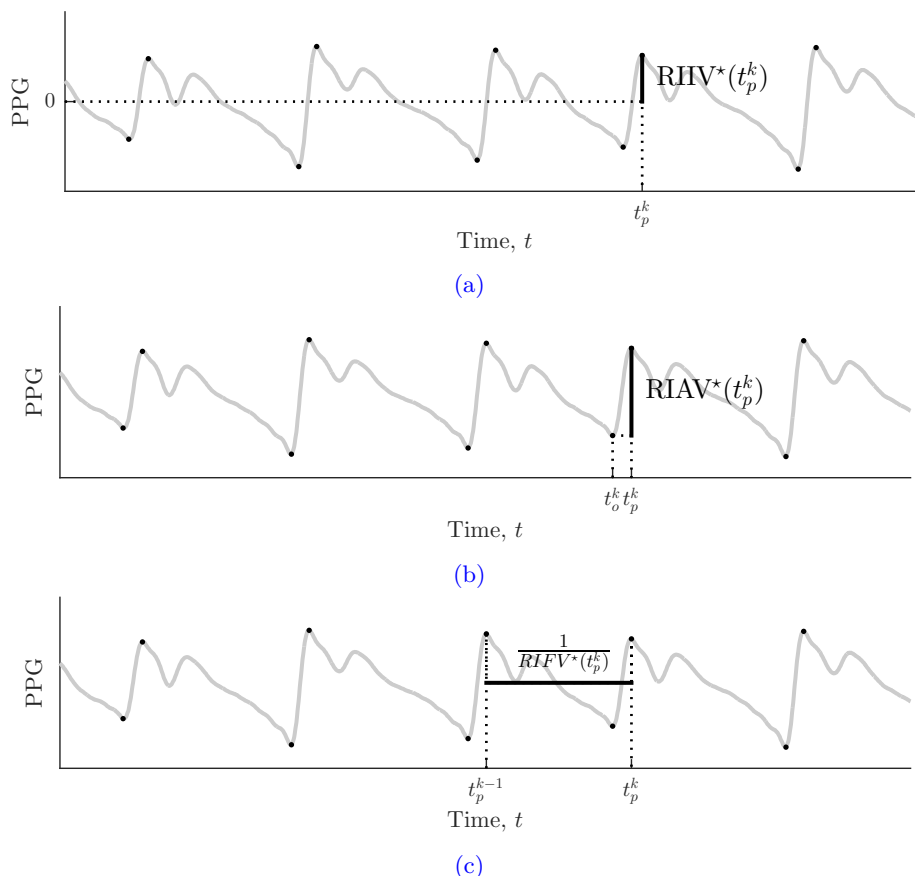
Figure 4.8

Excerpt of a PPG record (8 seconds) from a healthy subject illustrating the measurement of the PPG-derived respiration rhythms: (4.8a) RIIV*, (4.8b) RIAV*, and (4.8c) RIFV* time series for one sample pulse k.

Method

Given a PPG record segmented into a series of peaks and troughs (*e.g.* by one of the algorithms in Section 4.2.2), the following rhythms are usually computed as surrogates of the respiratory signal [Nemati et al., 2010]:

Respiratory-induced intensity variations (RIIV*) consisting of the time series of PPG peak amplitudes (Figure 4.8a).

Respiratory-induced amplitude variations (RIAV*) consisting of the time series of amplitude differences between the peak and the onset of a PPG pulse (Figure 4.8b).

Respiratory-induced frequency variations (RIFV*) consisting of the time series of intervals between consecutive PPG peaks[1] (Figure 4.8c).

A segment of PPG exhibiting the three modulations is shown in Figure 4.7. The RIIV*, RIAV* and RIFV* time series were then linearly interpolated at a sampling rate of 4 Hz to obtain the evenly-sampled time series RIIV, RIAV and RIFV (Figure 4.9a).

4.6 Performance analysis

4.6.1 Spectral coherence analysis

We used coherence analysis (a review of which can be found in Appendix C.1) to assess the spectral similarity between the nasal airflow reference signal and the extracted respiratory signals.

Spectral coherence is analogous to time correlation, but assesses the similarity between two signals in the frequency domain, rather than in the time domain. As time correlation is sensitive to even small-scale amplitude differences between signals, the correlation coefficient degrades quickly in the presence of noise regardless of the spectral properties of the noise source. Thus, to estimate the amount of overlap in the structure of two signals over a range of frequencies (the respiratory range), we used spectral coherence analysis.

The auto- and cross-spectra were estimated numerically via Welch's averaged modified periodogram [Welch, 1967]. Signals were split into $K = 8$ segments of 256 points[2] (64 s long) overlapping by one half segment. A Hamming window was applied to each segment to reduce spectral leakage.

---

[1] This quantity was inverted to reflect frequency variations.

[2] Note that for an analysis window of 60 s and a sampling frequency of 4 Hz, this is the nearest power of two of the window length.
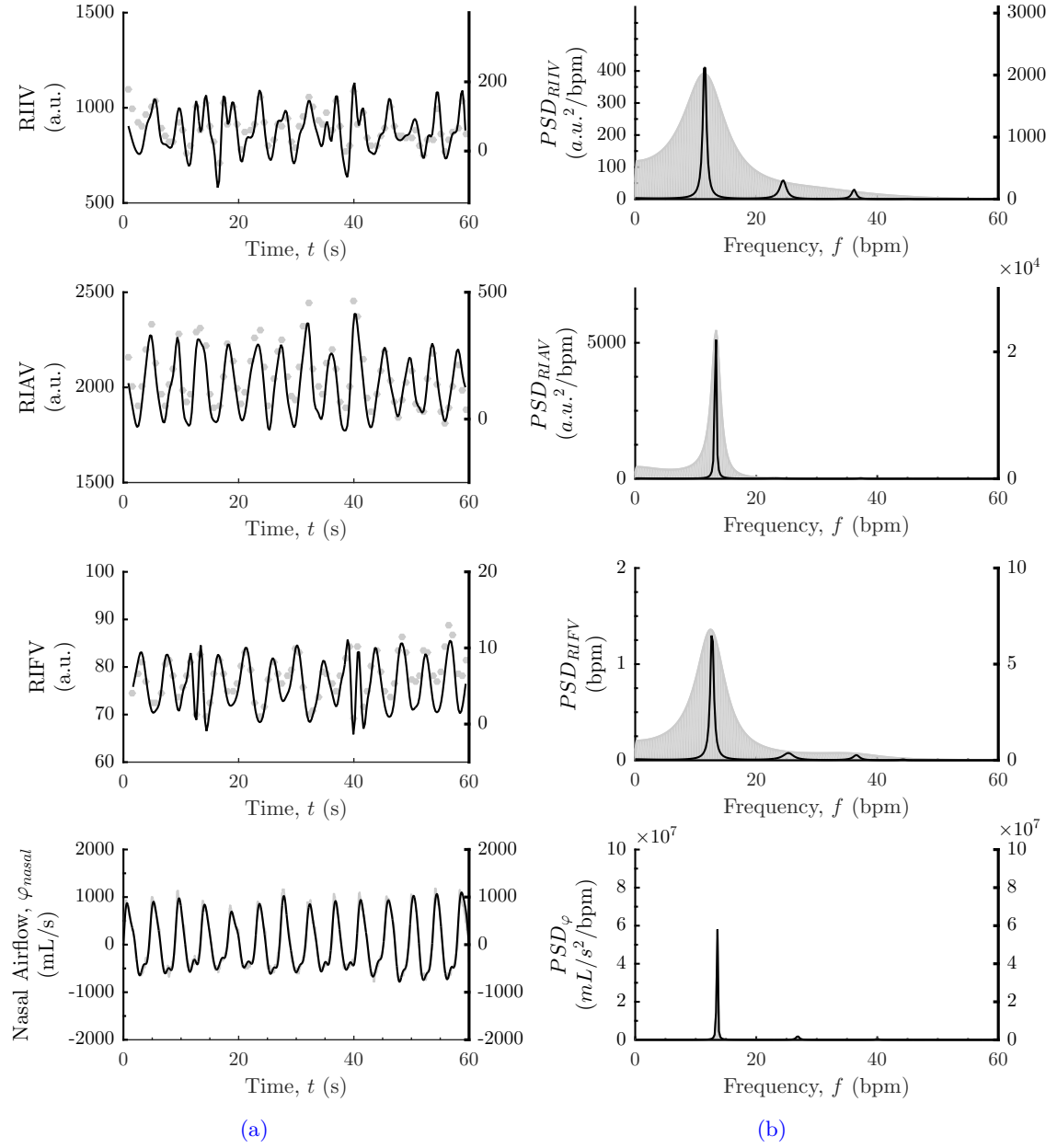
Figure 4.9

Respiratory waveforms derived from a 60-second interval of PPG and accompanying airflow trace. (4.9a) From top to bottom: Amplitude of pulse peaks (RIIV*, dots) and the series resampled at 4 Hz and bandpass-filtered (RIIV, solid line); Peak-to-onset height at the times of the corresponding peaks (RIAV*, dots) and the series resampled at 4 Hz and bandpass-filtered (RIAV, solid line); Pulse frequency variations (RIFV*, dots) and the series resampled at 4 Hz and bandpass-filtered (RIFV, solid line); Accompanying airflow trace. It may be seen from the power spectral density of these signals on the right-hand side (4.9b) that the respiratory signal is present in all three of these waveforms for this particular window. Power spectral density estimates for the original (shaded area) and the filtered time series (solid line) were computed using a 9$^{\text{th}}$ order AR model. The units on the vertical axes on the left side of the panels pertain to the curves plotted as solid lines.

Statistical analysis

For assessing the reliability of the cross-spectral estimates, we compared the coherence values against a significance threshold $\gamma^2_{th,\alpha}(f)$, such that a frequency bin k with coherence greater than $\gamma^2_{th,\alpha(f)}$ means that the null hypothesis that the φ_{nasal} and the derived respiratory signal are not related at that frequency k can be rejected at a significance level α (for $p < 0.05$).

Assuming independent (Hamming-windowed) segments, for $K = 8$, we have $\tilde{K} = 7.31$, which gives a threshold value $\gamma^2_{th,\alpha} = 0.38$ (see Equation C.5). As we ran multiple tests (one for each frequency bin in the respiratory range), p was set to a lower value according to the Bonferroni correction [Dunn, 1961]. The new threshold value became $\gamma^2_{th,\alpha} = 0.65$.

4.6.2 Respiratory rate analysis

The respiratory rate was derived through analysis of the frequency content of the extracted respiratory signals. Many techniques are available to do this. In this chapter, we use autoregressive modelling. A review of autoregressive models and their use within parametric spectral analysis is covered in Appendix C.2.

Prior to model-fitting, it is useful to remove the baseline wander and high-frequency noise. This is usually done through the use of digital filters. The respiratory time series were filtered using a 40th order bandpass, linear phase, finite impulse response (FIR) filter, using a Hamming window and cut-off frequencies which encompass the range of plausible respiratory rate values for the population under study (5-50 bpm in young healthy adults).

To estimate a parametric PSD, a 9th order AR model was fitted to the filtered waveforms using the Burg's algorithm [Burg, 1975]. The frequency corresponding to the peak of the autoregressive PSD in the aforementioned range of frequencies was taken as the respiratory rate estimate and denoted

$$RR_i, i \in \{\varphi_{nasal}, EDR, RSA, RIIV, RIAV, RIFV\} \qquad (4.14)$$

where the underscript index i refers to the original respiratory signal from which this rate was computed. This process is illustrated in Figures 4.6 and 4.9 for a 60-second analysis window of φ_{nasal} and ECG- or PPG-derived respiratory waveforms, respectively.

Error analysis

For a quantitative comparison between the performance of the methods implemented, the mean absolute error (MAE), the root-mean-square error (RMSE), the error means (bias) and standard deviation (SD) were computed for RR estimates.

4.7 Results

The methodology described in this chapter was validated against reference φ_{nasal} measurements from a subset of 12 healthy subjects.

In Section 4.7.1, spectral coherence analysis was used to assess the similarity between the φ_{nasal} reference signal and the derived respiratory signals over the range of respiratory frequencies. The respiratory rate, *i.e.* the dominant frequency within this range, was obtained for each of the derived respiratory signals and compared against that derived from φ_{nasal} in Section 4.7.2.

4.7.1 Spectral coherence analysis

Examples of coherence spectral maps for the ECG and PPG-derived respiratory waveforms are shown in Figures 4.10b to 4.14b during a typical study session of the Hypercapnoea study.

Regarding the phase of the derived respiratory signals, we have measured the phase lags between the derived rhythms and φ_{nasal}, given by the phase response of the cross-spectra $\mathcal{G}_{j\varphi}, j \in \{$EDR, RSA, RIIV, RIAV RIFV$\}$. These are shown above their respective coherence maps, in Figures 4.10a to 4.14a.

The respiratory rates estimated from the ECG- and PPG-derived respiratory waveforms during one study session are shown in Figures 4.15 and 4.16, respectively. This is the same study session shown in Figures 4.10b to 4.14b.

4.7.2 Respiratory rate analysis

For a meaningful comparison between the ECG- or PPG-derived rates and the ones computed from the nasal airflow reference, only RR estimates from windows of good φ_{nasal}, PPG and ECG signal quality were kept for comparison. We refer to these as valid windows.

Valid φ_{nasal} windows

Only φ_{nasal} windows for which the agreement between the frequency-domain and the time-domain estimates of RR (using extrema detection) was better than $\Omega_\varphi = 1$ bpm were kept as valid windows (with $SQI_\varphi = 1$). These corresponded to approximately 60.4% of the total number of analysis windows (Figure 4.3 and Table 4.1). The relationship between this choice of Ω_φ and the reduction of the mean absolute error of RR estimates from ECG- and PPG-derived rythms is shown in Figure 4.17.
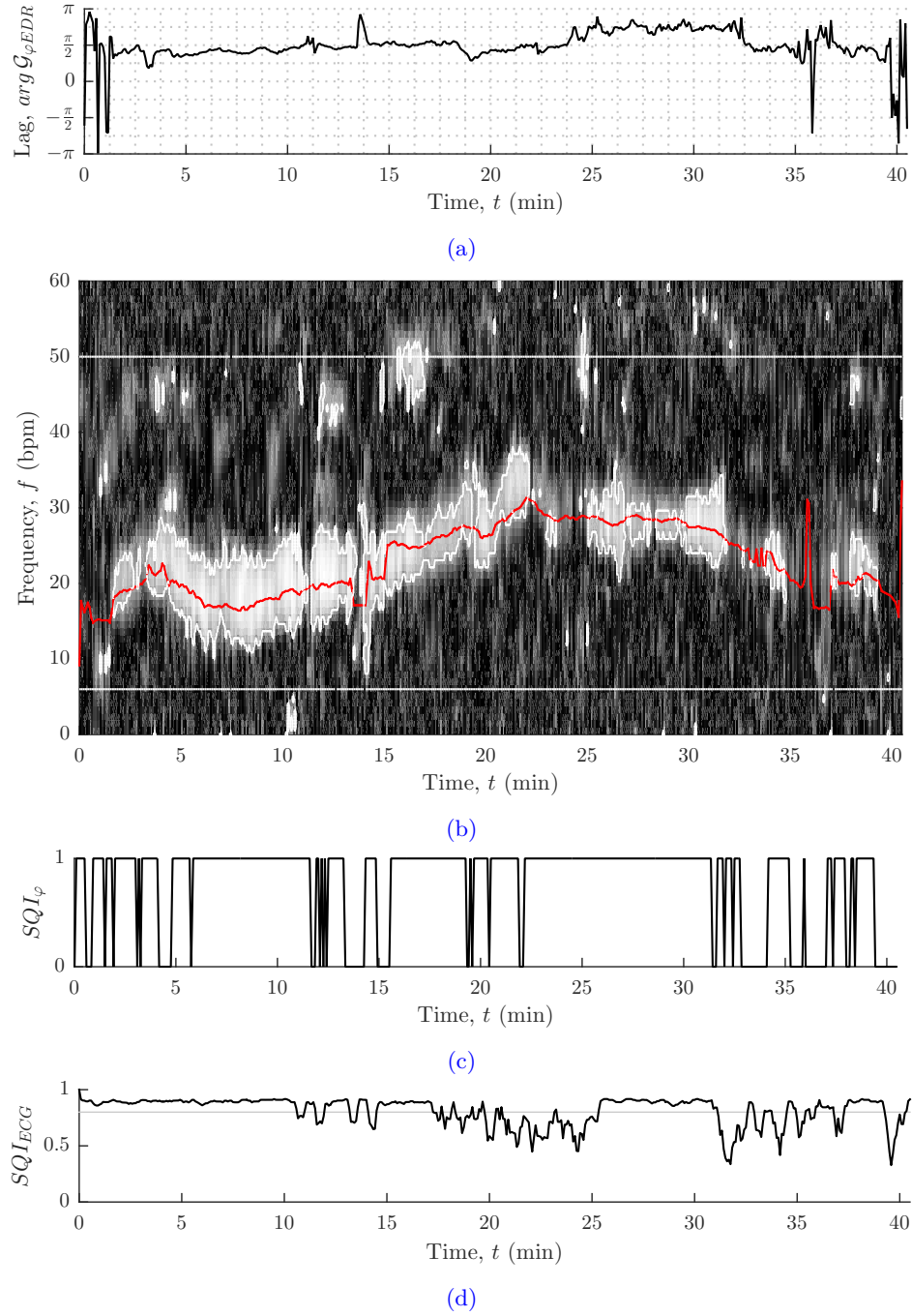
Figure 4.10
Example of coherence spectral analysis between EDR and φ_{nasal} for one healthy male volunteer. From top to bottom: (4.10a) Phase lag between EDR and φ_{nasal} at the respiratory frequency RR_φ; (4.10b) Squared spectral coherence map between EDR and φ_{nasal}. Bright areas indicate high coherence. The red continuous line indicates the respiratory frequency derived from φ_{nasal}, RR_φ. White horizontal lines indicate the frequency limits for hypothesis testing. White-delimited areas correspond to frequency bins whose coherence exceeds the significance threshold at 95% significance level; (4.10c) Nasal airflow signal quality; (4.10d) ECG signal quality. The SQI threshold for this signal is shown as an horizontal grey line.

(a)

(b)

(c)

(d)
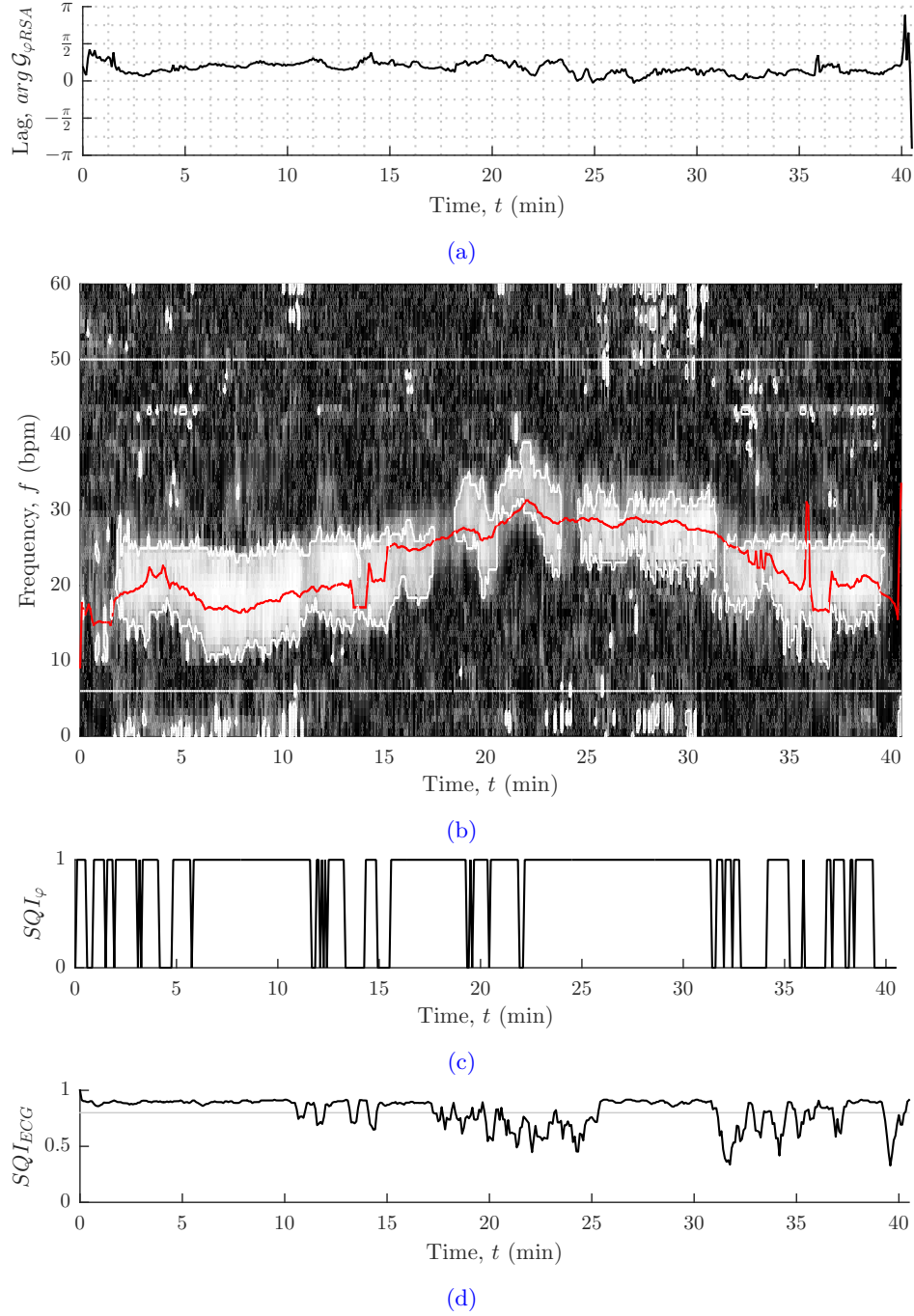
Figure 4.11

Example of coherence spectral analysis between RSA and φ_{nasal} for one healthy male volunteer. From top to bottom: (4.11a) Phase lag between RSA and φ_{nasal} at the respiratory frequency RR_φ; (4.11b) Squared spectral coherence map between RSA and φ_{nasal}. Bright areas indicate high coherence. The red continuous line indicates the respiratory frequency derived from φ_{nasal}, RR_φ. White horizontal lines indicate the frequency limits for hypothesis testing. White-delimited areas correspond to frequency bins whose coherence exceeds the significance threshold at 95% significance level; (4.11c) Nasal airflow signal quality; (4.11d) ECG signal quality. The SQI threshold for this signal is shown as an horizontal grey line.

(a)

(b)

(c)

(d)
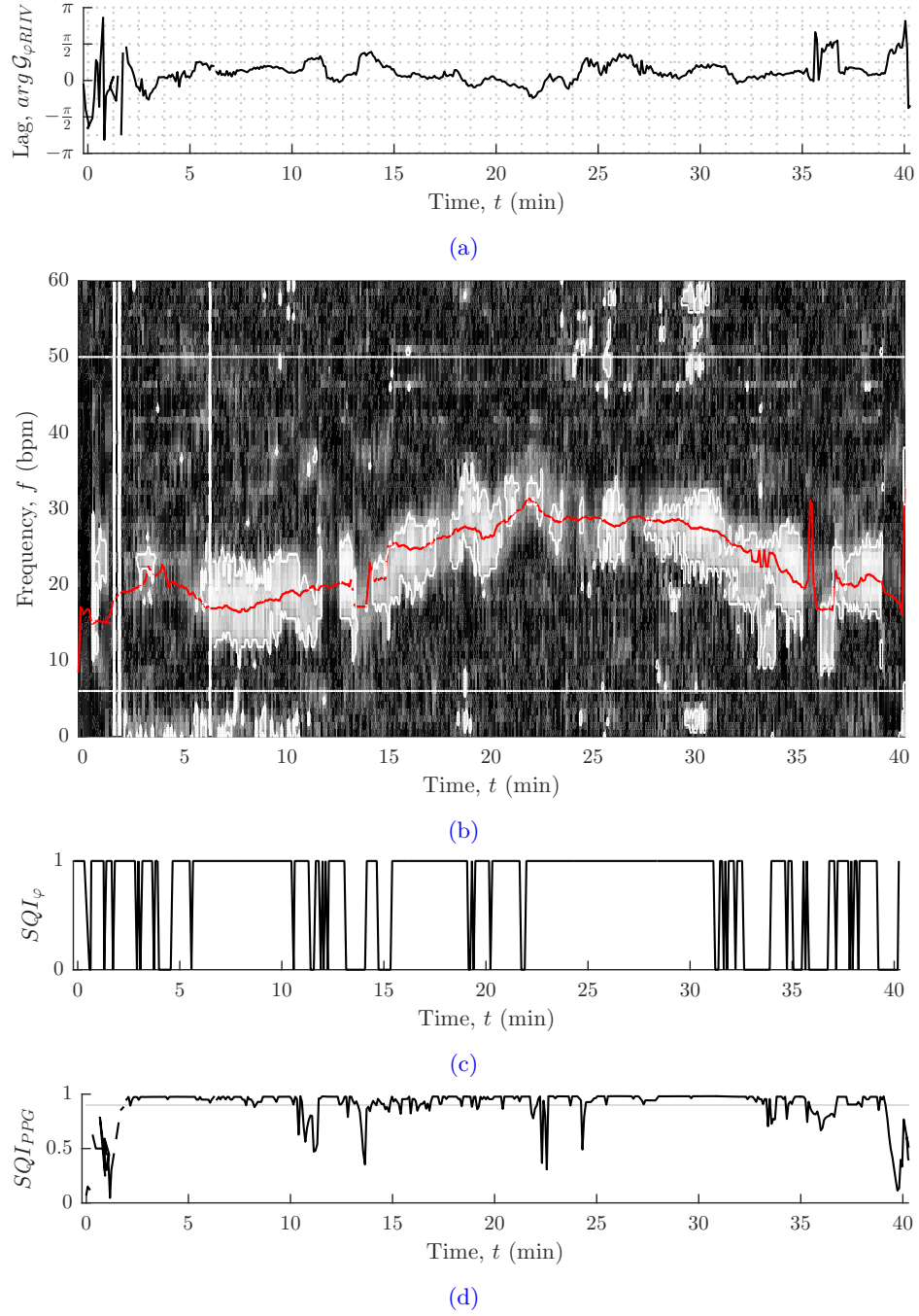
Figure 4.12

Example of coherence spectral analysis between RIIV and φ_{nasal} for one healthy male volunteer. From top to bottom: (4.12a) Phase lag between RIIV and φ_{nasal} at the respiratory frequency RR_φ; (4.12b) Squared spectral coherence map between RIIV and φ_{nasal}. Bright areas indicate high coherence. The red continuous line indicates the respiratory frequency derived from φ_{nasal}, RR_φ. White horizontal lines indicate the frequency limits for hypothesis testing. White-delimited areas correspond to frequency bins whose coherence exceeds the significance threshold at 95% significance level; (4.12c) Nasal airflow signal quality; (4.12d) PPG signal quality. The SQI threshold for this signal is shown as an horizontal grey line.

(a)

(b)

(c)

(d)
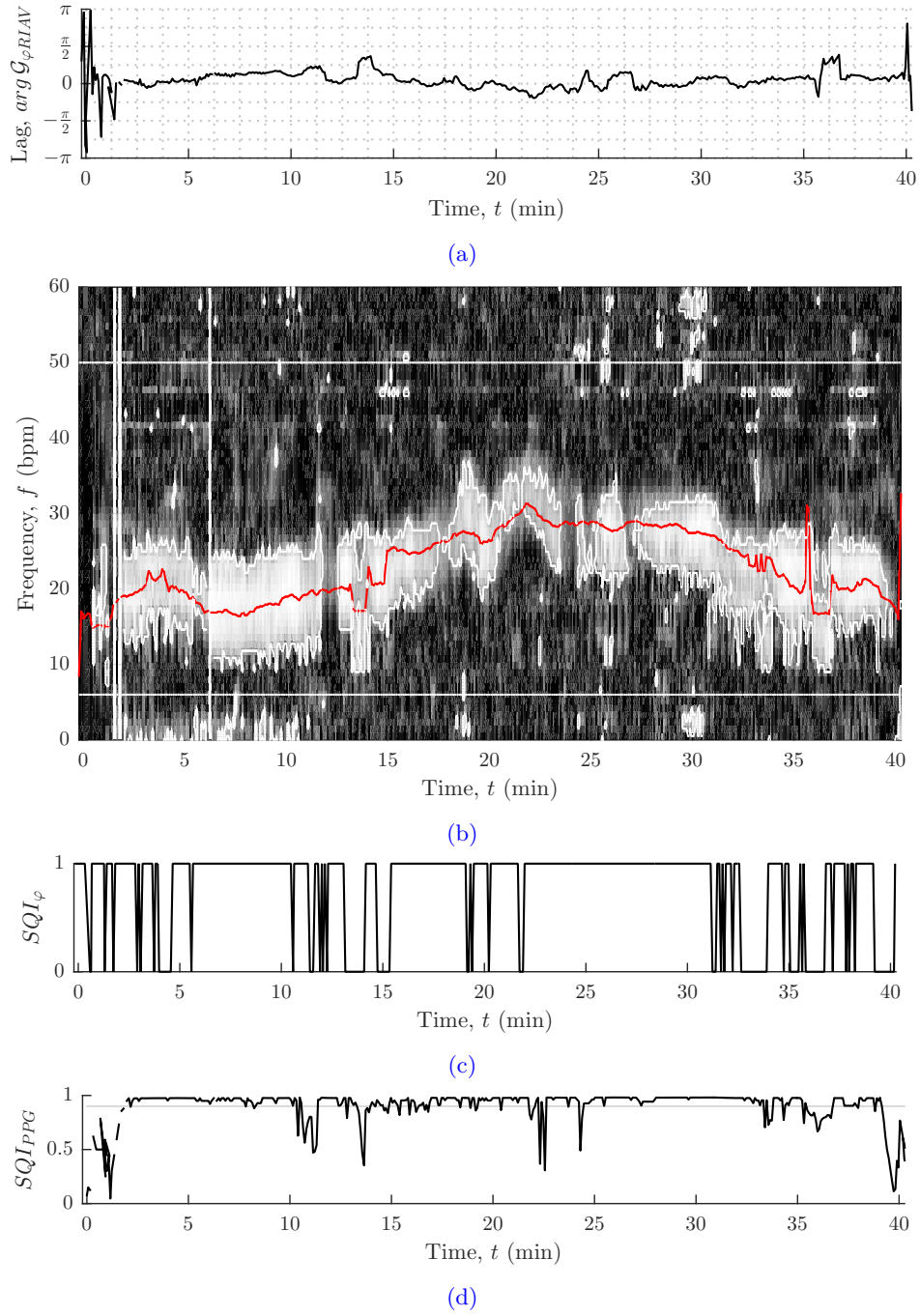
Figure 4.13

Example of coherence spectral analysis between RIAV and φ_{nasal} for one healthy male volunteer. From top to bottom: (4.13a) Phase lag between RIAV and φ_{nasal} at the respiratory frequency RR_φ; (4.13b) Squared spectral coherence map between RIAV and φ_{nasal}. Bright areas indicate high coherence. The red continuous line indicates the respiratory frequency derived from φ_{nasal}, RR_φ. White horizontal lines indicate the frequency limits for hypothesis testing. White-delimited areas correspond to frequency bins whose coherence exceeds the significance threshold at 95% significance level; (4.13c) Nasal airflow signal quality; (4.13d) PPG signal quality. The SQI threshold for this signal is shown as an horizontal grey line.

(a)

(b)

(c)

(d)
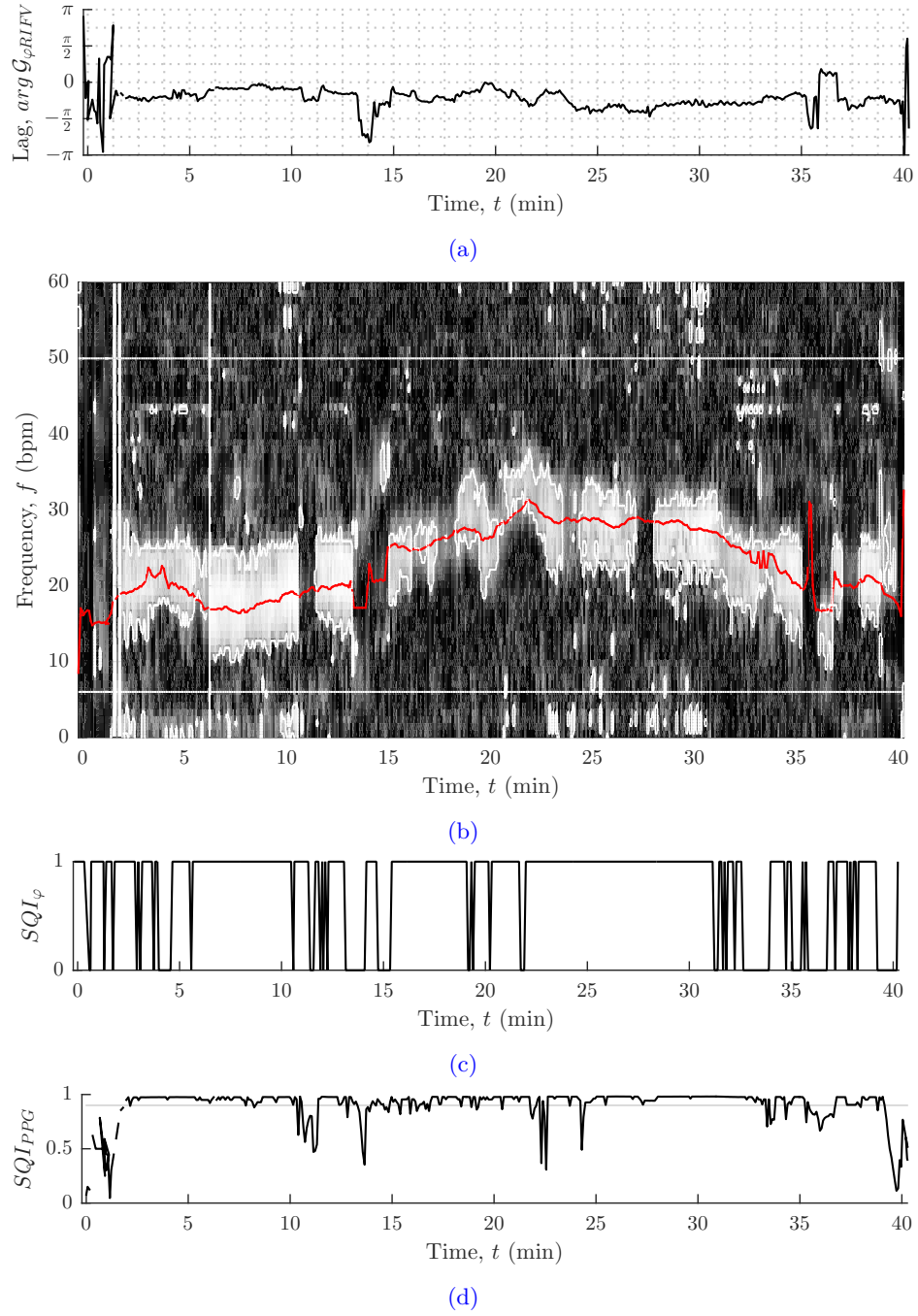
Figure 4.14

Example of coherence spectral analysis between RIFV and φ_{nasal} for one healthy male volunteer. From top to bottom: (4.14a) Phase lag between RIFV and φ_{nasal} at the respiratory frequency RR_φ; (4.14b) Squared spectral coherence map between RIFV and φ_{nasal}. Bright areas indicate high coherence. The red continuous line indicates the respiratory frequency derived from φ_{nasal}, RR_φ. White horizontal lines indicate the frequency limits for hypothesis testing. White-delimited areas correspond to frequency bins whose coherence exceeds the significance threshold at 95% significance level; (4.14c) Nasal airflow signal quality; (4.14d) PPG signal quality. The SQI threshold for this signal is shown as an horizontal grey line.

(a) RR_{EDR} (solid line) *versus* RR_φ (dots).

(b) RR_{RSA} (solid line) *versus* RR_φ (dots).
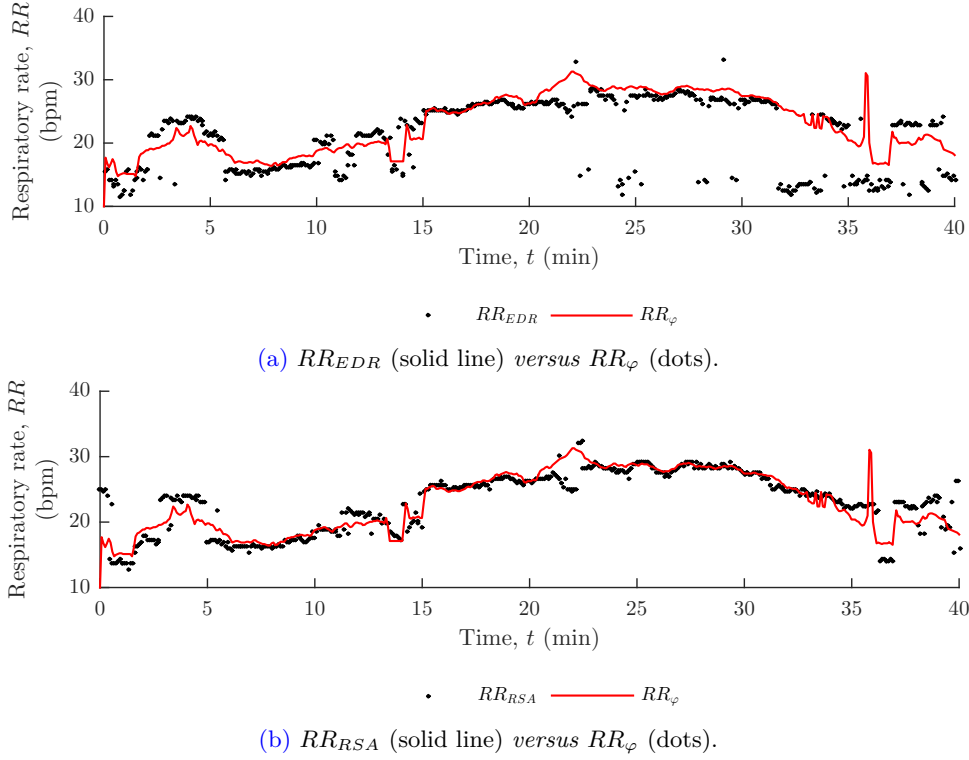
Figure 4.15

Example of respiratory rates derived from the nasal airflow reference and calculated using the (4.15a) EDR technique and (4.15b) RSA technique for one healthy volunteer session. A progressive increase in respiratory rate is seen as the session progresses as a result of the increasingly hypnocapnoeic environment.

Valid ECG and PPG windows

Only ECG windows with $SQI_{ECG}(w) > 0.8$ (77.3% of the data, from Table 4.1) and PPG windows with $SQI_{PPG}(w) > 0.9$ (63.7% of the data, from Table 4.2) were retained for the calculation of error metrics. These thresholds were selected from the qualitative analysis of Figures 4.18 and 4.19 to achieve high accuracy while retaining over 60 % of estimates.

As a result of thresholding the applicable SQI metrics, roughly 48.8% of the ECG data (approximately 4.1 hours) were used to validate the ECG-derived estimates (Table 4.2). This figure decreases to 43.5% of windows, or approximately 3.6 hours, in the case of PPG data (Table 4.2).

In these segments, we quantified the accuracy of our methods in terms of mean bias, and standard deviation of the errors, mean absolute error (MAE) and root-mean square error rates (RMSE). These quantities are shown in Table 4.3 for the 12 study sessions. An analysis of individual MAE for each study session is given in Figure 4.20.

90
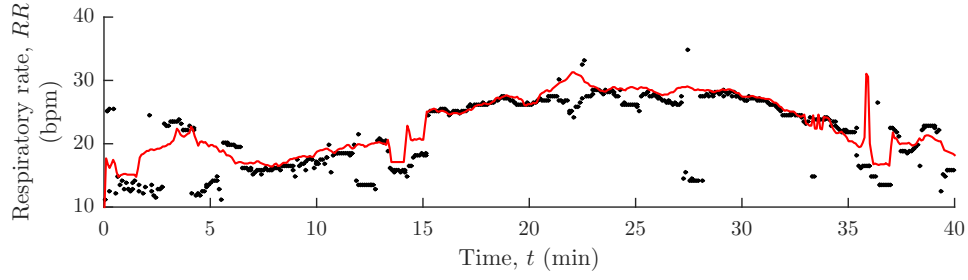
(a) RR_{RIIV} (dots) *versus* RR_{φ} (solid line).
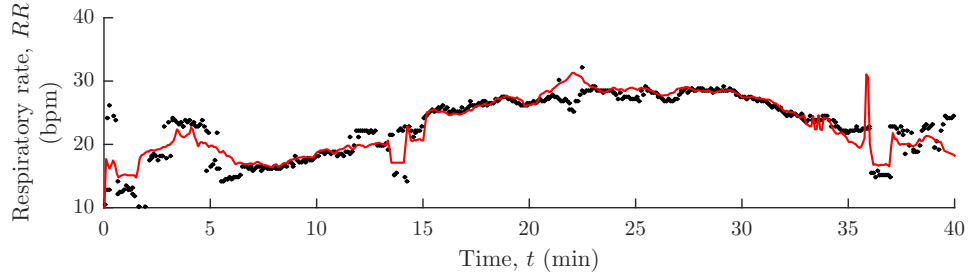
(b) RR_{RIAV} (dots) *versus* RR_{φ} (solid line).
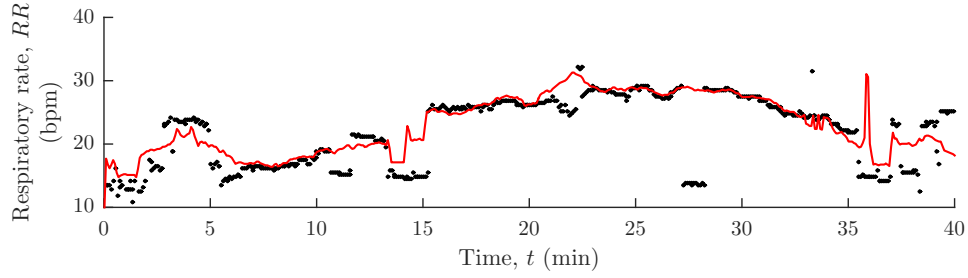
(c) RR_{RIFV} (dots) *versus* RR_{φ} (solid line).

Figure 4.16

Example of respiratory rates derived from the nasal airflow reference and calculated using the respiratory-induced PPG variations: (4.16a) RIIV (4.16b) RIAV, and (4.16c) RIFV for one healthy volunteer session. A progressive increase in respiratory rate is seen as the session progresses as a result of the increasingly hypnocapnoeic environment.
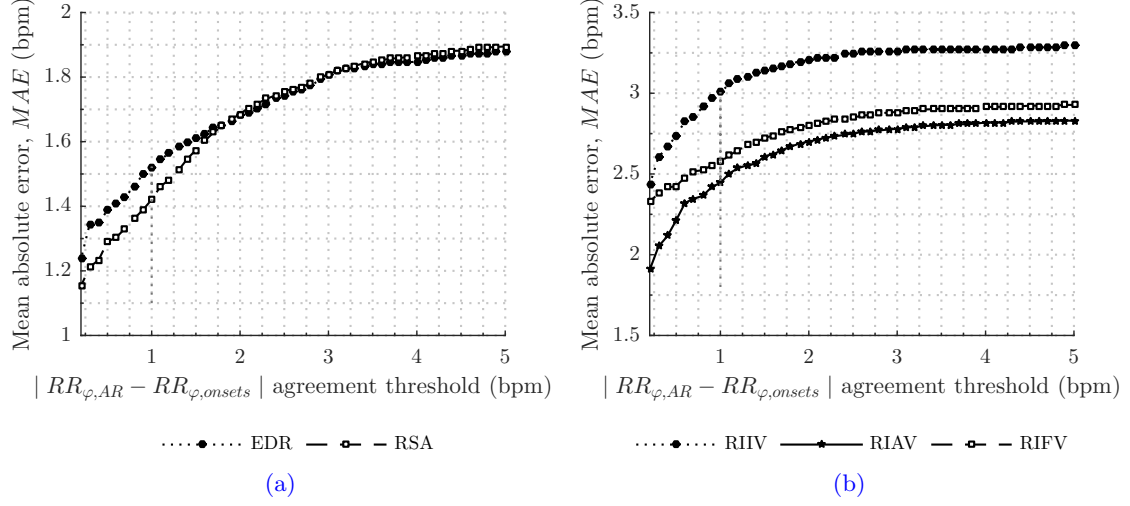
Figure 4.17

Mean absolute error (MAE) between φ_{nasal}-derived and (4.17a) ECG-derived or (4.17b) PPG-derived respiratory rates, when Ω_φ is set to different values for $RR_{\varphi,AR}$ and $RR_{\varphi,onsets}$ agreement.
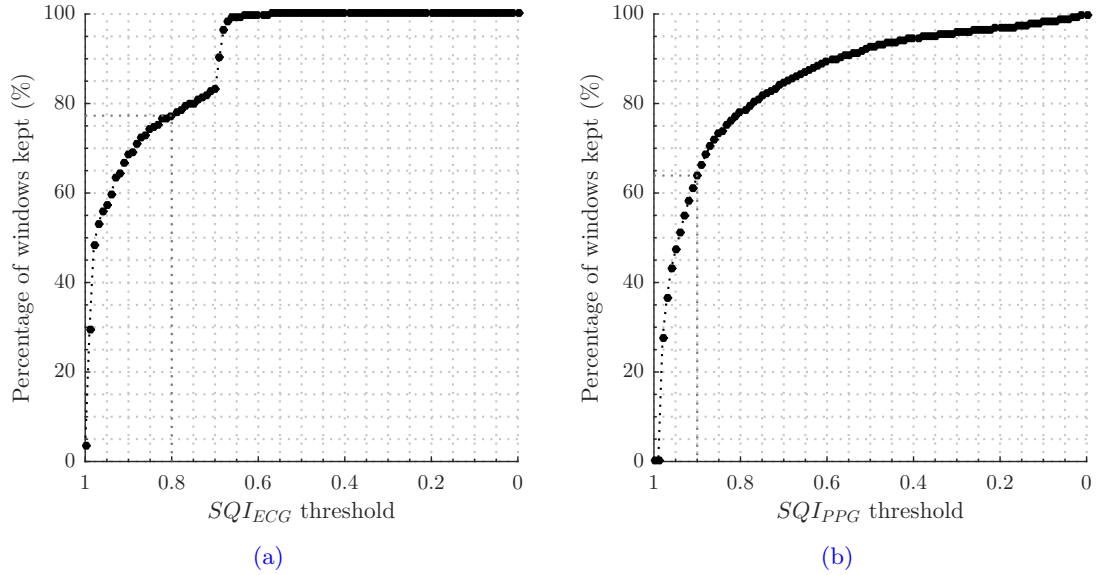
Figure 4.18

Relationship between the percentage of windows kept and the thresholds on signal quality measures. (4.18a) Percentage of ECG windows kept when the threshold on SQI_{ECG} is set to different values. (4.18b) Percentage of PPG windows kept when the threshold on SQI_{PPG} is set to different values.
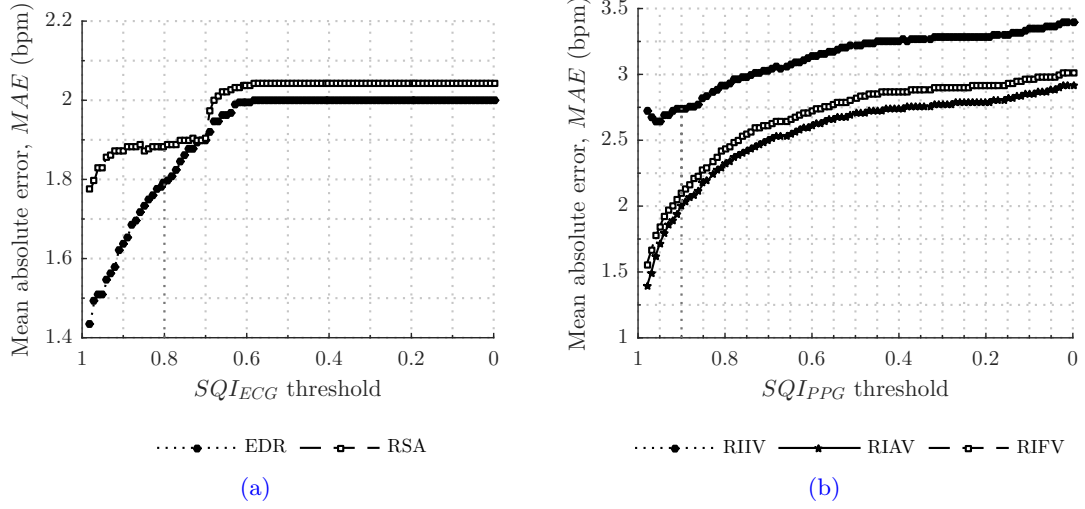
Relationship between mean absolute error (MAE) in RR estimation and the thresholds on signal quality measures. (4.19a) MAE between ECG-derived and φ_{nasal}-derived respiratory rates when the threshold on SQI_{ECG} is set to different values. (4.19b) MAE between PPG-derived and φ_{nasal}-derived respiratory rates when the threshold on SQI_{PPG} is set to different values.

4.8 Discussion

Spectral coherence analysis

A visual inspection of the individual coherence spectra confirmed the presence of statistically significant coherence values between the φ_{nasal} reference and each of the ECG- and PPG-derived surrogates of the respiratory signal. For illustration, the squared coherence spectra for one study session was shown for the ECG-derived respiratory signals (Figure 4.10 to 4.11) and PPG-derived signals (Figure 4.12 to 4.14). In all the cases presented, we confirmed that for most of the duration of the study session ($\sim$ 40 min) both signals oscillate at similar frequencies over the respiratory range (5 - 50 bpm) with significant coherence values at frequencies surrounding the instantaneous respiratory rate measured by the airflow reference (RR_φ, red trace).

The degree of agreement with φ_{nasal} is generally higher for ECG signals, where, for each time window (corresponding to the columns of the coherence maps in Figures 4.10b to 4.14b), the number of frequency bins with γ^2 values above the $\gamma^2_{th,\alpha}$ (inside the white-delimited areas) is higher, showing a consistent frequency coupling between the two signals over a greater range of frequencies.

Although γ^2 values reached significance for most windows, instances occurred when γ^2 fell below the 95% significance level threshold. These instances were at least partly explained by poor signal quality. Consider the two signals in the bottom panels of Figure 4.10 to 4.14 showing signal quality measures for φ_{nasal} and either ECG (in Figures 4.10 to 4.11) or PPG (in Figures 4.12 to 4.14). There appears to be a close temporal relationship

Monitoring time (in minutes) and proportion of windows with RR estimates after thresholds on SQI_φ, SQI_{ECG} and $SQI_{ECG} \wedge SQI_\varphi$ are applied.

		Proportion of time with RR estimates					
	Total time	$SQI_\varphi = 1$		$SQI_{ECG} > 0.8$		$SQI_{ECG} > 0.8 \wedge SQI_\varphi = 1$	
Session	(min)	(%)	(min)	(%)	(min)	(%)	(min)
1	40.3	67.1	27.0	97.1	39.0	66.3	26.67
2	42.8	59.6	25.5	94.5	40.4	54.2	23.1
3	40.6	36.9	15.0	100	40.6	40.0	15.0
4	41.2	47.6	19.6	100.0	41.2	47.6	19.6
5	40.1	71.5	28.7	100.0	40.1	71.5	28.7
6	43.3	32.9	14.3	46.2	20.0	22.9	9.9
7	40.8	91.6	37.3	100.0	40.8	91.6	37.
8	40.6	77.6	31.5	32.0	13.0	25.3	10.3
9	41.9	63.3	26.5	99.4	41.6	63.3	26.5
10	42.8	50.1	21.4	63.0	26.9	37.6	16.1
11	43.1	78.4	33.8	72.0	31.0	54.2	23.3
12	43.9	50.5	22.2	28.7	12.6	18.4	8.1
Total	501	60.4	302.6	77.3	387.2	48.8	244.6

Monitoring time (in minutes) and proportion of windows with RR estimates after thresholds on SQI_φ, SQI_{PPG} and $SQI_{PPG} \wedge SQI_\varphi$ are applied.

		Proportion of time with RR estimates					
	Total time	$SQI_\varphi = 1$		$SQI_{PPG} > 0.9$		$SQI_{PPG} > 0.9 \wedge SQI_\varphi = 1$	
Session	(min)	(%)	(min)	(%)	(min)	(%)	(min)
1	40.3	67.1	27.0	90.7	36.5	65.2	26.3
2	42.8	59.6	25.5	81.5	34.8	51.7	22.1
3	40.6	37.0	15.0	13.6	5.5	7.6	3.01
4	41.2	47.6	19.6	7.7	3.2	5.7	2.3
5	40.1	71.5	28.7	50.5	20.3	37.4	15.0
6	43.3	32.9	14.3	75.1	32.5	27.7	12.0
7	40.8	91.6	37.3	95.7	39.0	89.2	36.3
8	40.6	77.6	31.5	77.6	31.5	65.7	26.7
9	41.8	63.3	26.5	58.8	24.6	42.0	17.6
10	42.8	50.1	21.4	46.4	19.8	20.1	8.6
11	43.1	78.3	33.8	94.8	40.8	75.4	32.5
12	43.9	50.5	22.2	69.6	30.6	34.9	15.3
Total	501	60.4	302.7	63.7	319.0	43.5	217.8

Summary of error analysis for the algorithm for respiratory rate extraction from the ECG and PPG-derived respiratory signals. The errors were computed against the reference signal RR_φ. All the figures are shown in breaths·min^{-1}.

	Sources				
Error metrics	ECG-derived		PPG-derived		
	RR_{EDR}	RR_{RSA}	RR_{RIIV}	RR_{RIAV}	RR_{RIFV}
ME($\pm$SD)	-0.31(2.57)	0.64(1.73)	0.20(4.61)	-0.58(3.32)	-0.81(2.83)
MAE	1.41	1.33	2.46	1.61	1.74
RMSE	2.59	1.84	4.62	3.37	2.94
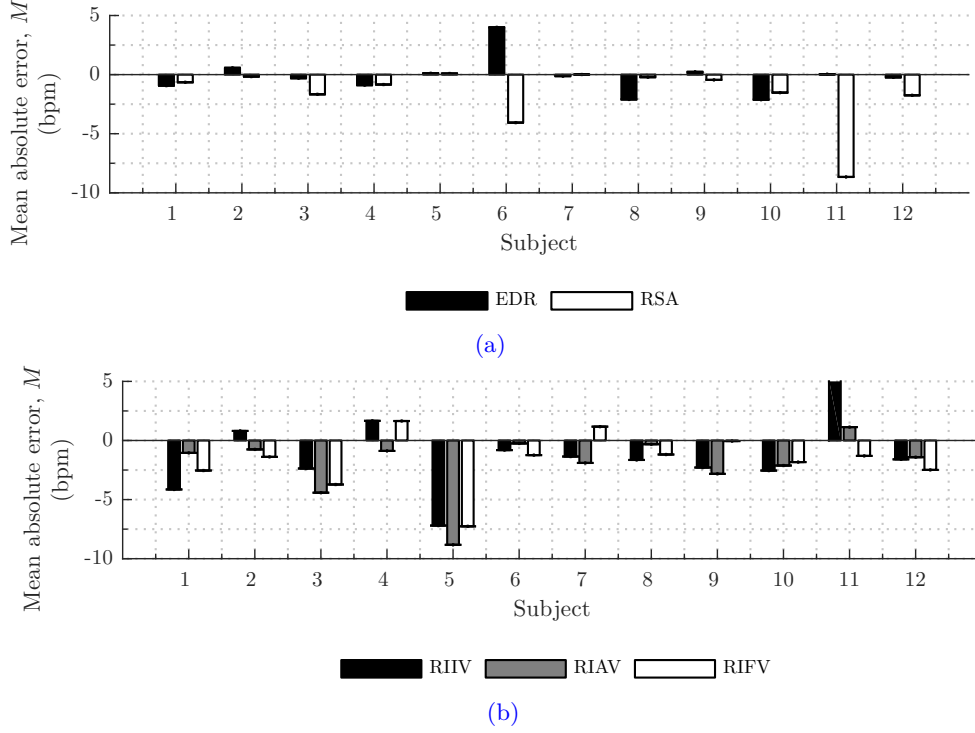
Mean absolute errors (MAE) between φ_{nasal}-derived and ECG-derived (4.20a) or PPG-derived (4.20a) respiratory rates for each study subject.

between low γ^2 levels and poor ECG or PPG signal quality, as quantified in SQI_{ECG} and SQI_{PPG}.

It is also clear that low SQI_φ has an adverse effect on γ^2 levels and is usually accompanied by low SQI_{ECG} and SQI_{PPG}, particularly during instances of sustained poor signal quality (over the time-scale of minutes) as those seen from $t = 10$ min to $t = 15$ min or $t = 33$ min to $t = 37$ min. Hence, the same artefactual motion that degrades SQI_φ tends to also affect both SQI_{ECG} and SQI_{PPG}.

Nevertheless, whilst significant γ^2 values are contingent on the good quality of the raw signals, the presence of artefacts alone was not sufficient to explain poor coherence (see *e.g.* $t = [27; 28]$ min in Figures 4.12 to 4.14). It is known that frequency-modulated rhythms (RSA and RIFV) are attenuated when humans are asked to breathe at a specific pace, thus overruling the autonomous nervous system by cortical inputs [Larsen et al., 2010]. As each hypercapneic step was accompanied by increased fan noise inside the study chamber, we cannot exclude the possibility that for certain periods the subjects consciously adjusted their minute ventilation in anticipation of an impending increase in the hypercapnoeic challenge, thus momentarily cancelling this source of ECG modulation.

Regarding the phase of the derived respiratory signals, during normal respiration, a relatively constant phase shift $(0 - \frac{\pi}{4})$ was measured between RSA and φ_{nasal} (Figure 4.11a). The same phenomenon was also observed for the respiratory signals derived from

RSA in the other volunteer sessions. However, this phase coupling was not observed between φ_{nasal} and the (mechanically-mediated) EDR modulation. Unlike for frequency modulation, there is no physiological dependence on the phase between the these two signals; approaches based on shifts in QRS morphology have respiratory effort as their measurand, and therefore, a peak in EDR may not correspond to a peak in the tidal flow rate measured by the nasal pressure transducer. For instance, in Figure 4.10a we observed that a relatively constant phase difference exists between the two signals at the start of the recording session. This shift is maintained until subject motion from $t = 17$ min to $t = 25$ min resets this phase difference to $\frac{3}{4}\pi$. We hypothesize that the noisier phase trace observed for the following 7 minutes (from $t = 25$ min to $t = 32$ min) is a result of larger electrode displacement with the exaggerated chest excursions required to circulate larger V_T as the $[CO_2]$ inside the study chamber was increased.

Respiratory rate analysis

To ensure that we had the highest possible confidence in the RR derived from the extracted respiratory signal ($RR_j, j \in \{$EDR, RSA, RIIV, RIAV RIFV$\}$) and the airflow reference (RR_φ), we took steps to eliminate regions of poor quality in the φ_{nasal}, ECG and PPG waveforms, which might typically be included in normal device operation.

We ignored segments with $SQI_\varphi < 1$, and proceeded similarly for ECG and PPG signals, where ECG windows with $SQI_{ECG} < 0.8$ and PPG windows with $SQI_{PPG} < 0.9$ were excluded from the analysis. These thresholds were set arbitrarily from analysis of Figures 4.3 and 4.17 to 4.18 and could be altered so as to achieve lower MAE rates or a higher percentage of windows included in the analysis, and thus effective RR monitoring time.

A summary of the error analysis of the RR extracted from the derived respiratory signals is given in Table 4.3. We observed that superior performance was obtained for ECG-derived methods, with $\text{MAE}_{EDR} = 1.41$ bpm; $\text{RMSE}_{EDR} = 2.59$ bpm and $\text{MAE}_{RSA} = 1.33$ bpm; $\text{RMSE}_{RSA} = 1.84$ bpm, while retaining 48.8% of the analysis windows. Both techniques slightly outperformed PPG-derived RR estimates, for which $\text{MAE}_{RIIV} = 2.46$ bpm; $\text{RMSE}_{RIIV} = 4.62$ bpm, $\text{MAE}_{RIAV} = 1.61$ bpm; $\text{RMSE}_{RIAV} = 3.37$ bpm and $\text{MAE}_{RIFV} = 1.74$ bpm; $\text{RMSE}_{RIFV} = 2.94$ bpm, with 56.5% of all analysis windows rejected by the SQI methods. The higher accuracy of ECG methods in estimating spontaneous respiratory rates in healthy adults has previously been reported [Charlton et al., 2016].

Both EDR and RSA methods provided good approximations of RR_φ, with marginally better respiratory estimates from RSA ($\text{RMSE}_{RSA} = 1.84$ bpm *versus* $\text{RMSE}_{EDR} = 2.59$ bpm). These results are hardly surprising. ECG-derived respiration algorithms based on beat morphology have been shown to be appropriate even when only single-lead ECG traces are available [Khaled et al., 1992]. As presence of motion artefact proved to be the

primary cause of failure of this technique [Helfenbein et al., 2014], reliable EDR rates are expected when monitored subjects are at rest and the respiration effort is not masked by subject muscle activity.

Regarding RSA-derived respiration, it is known that this rhythm is very pronounced in children and that it is gradually attenuated as individuals approach adulthood [van Ravenswaaij-Arts, 1993]. This renders the RSA-approach unviable in elderly individuals [Cysarz et al., 2008]. By design, our study included young adults in good cardiorespiratory health exclusively, and thus the RR estimates based on RSA are very reliable. A exception is seen for example for subject 11, whose RR was not captured by our methods based on RSA (Figure 4.20a).

The following aspects are also noteworthy. Our RR estimates based on amplitude modulation from PPG-derived rhythms performed slightly worse than ECG-derived ones on all error metrics. This observation is in line with previous results [Nemati et al., 2010]. With the exception of RR_{RSA} and RR_{RIIV}, all the RR extracted from the derived respiratory signals exhibited a relevant negative bias. This bias was within the standard deviation of error for all RR measures.

We briefly compare the figures in Table 4.3 with those of previous work in this area. As the quantities shown can only be compared in a meaningful way when different methods are applied to datasets acquired on similar populations and under comparable conditions, our comparison of error metrics is merely a qualitative one.

Regarding ECG-derived respiration, Cysarz et al. [2008] compared approaches based on RSA and RPA (*i.e.* the respiratory-synchronous variation given by the peaks of the ECG R-wave) using ECG data from 52 healthy adults (25 to 85 years in age) during night-time and wake time. In young adults (< 50 years) during wake time, the mean bias and standard deviation of errors in RR estimation was 0.29(1.02) bpm for RPA methods and 0.07(0.95) bpm for RSA methods. All rates registered were, however, in a very limited range of 13 to 19 bpm. Schäfer and Kratky [2008] measured RR from RSA in 40 supine subjects. The MAE found for 1-minute segments was 1.14 bpm for 20 young subjects (21-34 years old). Using AR spectral methods, Orphanidou et al. [2013] reported an MAE of 1.31 and 0.81 bpm from RPA and RSA-based methods, respectively, on the same dataset of 20 young adults.

Regarding PPG-derived respiration, Karlen et al. [2013] calculated RR from PPG data obtained from 29 children and 13 adults collected during elective surgery and routine anaesthesia and found RMSE rates of 5.8, 6.2, and 3.9 bpm for RIFV, RIAV and RIIV-derived rates, respectively.

In conclusion, the error rates reported here are similar to those reported in previous studies in the adult population.

4.9   Chapter conclusions

Our aim in this chapter was to establish the methodology for (a) the extraction of respiratory signals and (b) the assessment of signal quality for the signals analysed in this thesis.

We demonstrated this methodology on a population of 12 healthy young adults in a controlled hypercapnoeic environment. First, surrogate respiratory signals were derived from ECG and PPG waveforms; secondly, we established the presence of respiratory modulation in the derived signals by detecting significant values of spectral coherence between these signals and φ_{nasal} measured concurrently; then, respiration rates were estimated from the respiratory signals using an AR model; finally, we demonstrated that by applying the SQI measures described earlier, we were able to reduce the error in the RR estimates from each signal source.

In this study, ECG-based methods provided the closest approximation to the true respiratory rate derived from a nasal pressure airflow reference, an observation that is consistent with previous findings in the scientific literature on the topic. The errors reported are within their expected range for the population studied and the experimental protocol used for acquiring these signals.

Chapter 5

Image processing methods

Contents

5.1 Chapter summary

This chapter comprises the image analysis algorithms applied towards the detection of the region-of-interest from which the camera-based respiration is extracted.

The detection of the subject in the video frame and the selection of a region-of-interest are important first steps in non-contact vital sign monitoring. To perform the first task, we used a simple skin detector based on colour attributes and parametric mixture models. Whilst it is plausible that recent classification methods based on convolutional neural networks [LeCun and Bengio, 1995] could outperform this approach, the use of classical methods for skin segmentation was effective for the analysis presented in this thesis. Such methods achieve reasonable accuracy while avoiding the need for large training sets or

complex training procedures (*e.g.* data augmentation) to reduce overfitting and ensure good generalisation performance of the learned network.

This chapter is organised as follows. Section 5.2 introduces the problem of skin segmentation in video data. Sections 5.2.1 and 5.2.2 provide an overview of the state-of-the-art in pixel-based skin detection and motivate our choice of skin model based on prior art[1]. The practical questions related to our methodology are answered in Section 5.3. In Section 5.5 we evaluate our methods against ground-truth labels by clinical annotators and discuss these results in Section 5.6. The main conclusions are drawn in Section 5.7.

5.2 Skin Detection

Skin is a commonly used primitive in human computer vision systems, with algorithms for skin detection being routinely employed for subject localisation in the preliminary steps of many high-level vision systems.

In essence, skin detection is the process of finding the pixels or regions in an image that exhibit skin-like appearance as dictated by visual attributes such as tone or texture.

Without delving into the complex mechanisms of light interaction with this organ, the visual appearance of skin is ultimately determined by the spectral and spatial distributions of propagated light. These are influenced by both the structural and the biophysical properties of skin, which Baranoski and Krishnaswamy [2010] grouped into those of an endogenous or an exogenous origin[2].

Let us first consider the distinction in the first clause in our problem definition. Skin detection methodologies can be categorised into either *pixel-based* or *region-based* approaches. While pixel-based approaches classify each pixel as skin or non-skin individually, region-based approaches exploit texture, shape or other spatial features extracted from the local arrangement of skin pixels in an image [Jedynak et al., 2002; Poudel et al., 2012; Yang and Ahuja, 1998]. Most of the research in this field, however, falls under the former approach, whereby each pixel is classified as skin or non-skin individually and independently from its neighbours. It is also the approach we adopt here.

We formulate the problem of skin detection as a pixel-wise classification one, *i.e.* that a certain colour-pixel[3] $c = [c_1, c_2, c_3]^T$ is mapped to an output label $l \in \{skin, \overline{skin}\}$, where each c_i is a colour component in a triplet of a given a colour space and $\{skin, \overline{skin}\}$ is a

---

[1]An in-depth review of this field lies outside the scope of this thesis, hence we refer the interested reader to the review works by Vezhnevets et al. [2003] and Kakumanu et al. [2007]. We believe that jointly their work provides the most comprehensive survey of classical skin colour models for classification published to date.

[2]While endogenous agents account for skin constituents (*e.g.* pigments, fibres and cellular arrangements), exogenous agents account for surrounding environmental conditions (*e.g.* temperature and prevailing illumination) and the presence of external bodies (*e.g.* hair, sweat, oils). Changes in these agents across skin tissue result in different spectral and spatial light responses, and therefore, visual attributes, a reality that poses numerous challenges to skin recognition algorithms.

[3]The operator $[.]^T$ denotes the transpose.

set of binary labels representing skin and non-skin classes[1].

Solving this problem involves addressing the following sub-problems sequentially [Vezhnevets et al., 2003]:

1. Representing colours using a *suitable* colour space (Section 5.2.1);

2. Describing the distribution of skin pixel intensities using a *suitable* model (Section 5.2.2);

3. Defining classification rules to discriminate between *skin* and $\overline{skin}$ pixels (Section 5.2.3).

The last decades have seen considerable of research on these areas. On the one hand, progress in colorimetry and the establishment of image data transmission and printing standards have fostered a better understanding of colour properties, specification systems and the conversions between them. On the other hand, recent advances in CCD and CMOS technology have driven down the cost and size of imaging sensors, with the result that colour cameras are now packaged into most mobile communication devices as a standard feature. This has contributed to the proliferation of online digital images in recent years, which in turn has provided the computer vision community with access to vast training and test data for skin modelling and classification. The emergence of free access datasets labelled for skin/$\overline{skin}$, such as the COMPAQ [Jones and Rehg, 2002] and the ECU [Phung et al., 2005] datasets, has been particularly helpful in providing a better assessment of the relative merits of skin recognition algorithms by allowing for their validation over larger and richer datasets.

Sections 5.2.1 to 5.2.3 provide a brief overview of the state-of-the-art in pixel-based skin detection and justify our choice of colour space, skin model and classification rules based on previous findings (Figure 5.1).

5.2.1 Colour Representation

Colour is a low-level image feature and is computationally inexpensive. For this reason, it is an attractive starting point in object detection. In addition, it is robust to partial occlusions (such as those typically seen in the field of view of a camera placed above a NICU incubator, as a result of blankets, diapers, ECG leads or other medical equipment) and invariant to rotation and scale transformations [Ohta et al., 1980].

Nonetheless, colour-based methods are known to suffer from poor performance under changing illumination conditions which alter the spectral or spatial energy distribution of propagated light. In their seminal book of 1987, Hunter and Harold [1987] asserted

---

[1]It follows from this problem statement that all pixel-based methods are inherently colour-based. In fact, colour (or, simply put, pixel intensity) can be regarded as the 0th order spatial model in which the neighbourhood is restricted to a single pixel.
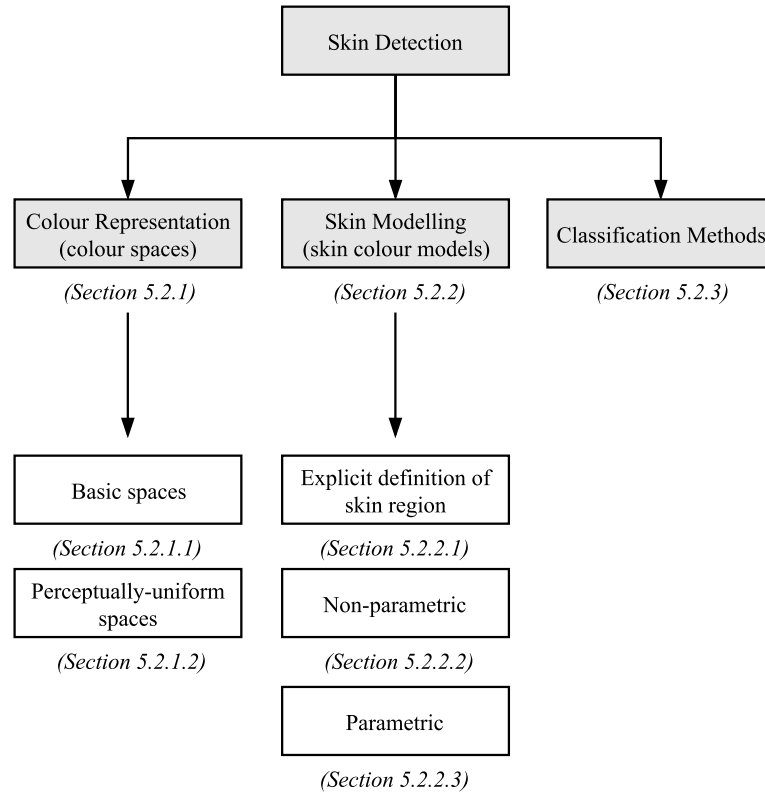
Figure 5.1
The three basic steps of skin detection. Approaches to colour representation, skin colour modelling, and classification methods discussed in this chapter.

that variations in the spectrum of the propagated light affect colour lightness[1], hue and saturation[2] of an imaged object, and therefore, its *perceived* colour. This is known as the colour constancy problem. In contrast, changes in the spatial distribution of propagated light are deemed to affect appearance characteristics such as glossiness, reflection haze, transmission haze and translucency.

Apart from changes in prevailing illumination, the presence of complex backgrounds, which may include surfaces with skin-like appearance (such as some types of wood), and the individual variability of skin tone between individuals are equally known to pose challenges to skin detectors.

Using a biophysical model, Storring et al. [1999] studied the problem of colour constancy in human skin. As the main conclusions of this study, the authors verified the

---

[1]not to be confused with brightness. Brightness is an absolute term that refers to the degree at which a source appears to be radiating or reflecting light, whereas lightness is a relative (perceptual) term based on the judgement of an object's brightness relative to the brightness of a similarly illuminated area [Hunt and Pointer, 2011].

[2]Following the definitions set in Fairchild [2005], hue is the attribute of colour perception by means of which two objects are judged to be similar in red, yellow, green, and blue colour content. The saturation of a colour is an attribute determined by a combination of light intensity and the extent to which it is distributed across the visible spectrum or, in other words, the colourfulness of a colour relative to its own brightness.

existence of the so-called *skin locus*, a well-defined region in the colour spaces normally occupied by skin colour tones under controlled illumination. Although this region was confined to a restricted area of the colour plane, its coordinates were dependent on acquisition parameters. Therefore, the location of *skin locus* must be determined either empirically or theoretically, if prior knowledge of illuminants, skin spectral response and camera specifications is available. However, this knowledge is not commonly accessible *a priori*, as is the case for the videos captured in the Oxford NICU under ambient light supplied by a wide variety of both artificial and natural light sources.

The selection of an appropriate colour system for a skin segmentation task is therefore an important one. Over the years, computer graphics and standards for video data transmission have originated a wealth of colour spaces for different use cases. A wide sample of spaces have been used for skin modelling [Terrillon and Akamatsu, 1999; Vezhnevets et al., 2003]. A reasonable choice of colour space should take into account several factors:

- **Colour stability** The stability of the colour space at extreme values of each component is desirable so as to avoid colour clipping. In addition, in the event that parametric methods should be used to model the colour distribution of skin regions, colour spaces which offer a greater degree of separability between the desired classes (in our case, *skin versus $\overline{skin}$*) minimise the overlap between them and so achieve higher recognition rates [Terrillon and Akamatsu, 1999].

- **Perceptual uniformity** The term *skin colour* is not a physical property of an object, but a perceptual property which relies on human perception of colour information. In this regard, colour representation models similar to the colour sensitivity of the human vision system should in theory aid the performance skin detection algorithm. Perceptually-uniform colour spaces have been proposed whose dimensions relate more closely those of the human vision system. This subject will be discussed in the next section.

- **Inclusion of luminance** Regarding the treatment of luminance, many works on skin recognition have suggested that the exclusion of the luminance-related components from skin colour models can increase the robustness of the cluster description to changes in illumination conditions, an hypothesis dismissed empirically by Shin et al. [2002] on a dataset of 805 images with different skin tones and illumination. Shin's results on this modest dataset were corroborated by Khan et al. [2012], who compared the performance of nine non-parametric methods and reported that in almost all combinations of colour space and classifier, three-dimensional colour models outperform two-dimensional ones (chrominance models). An alternative argument for the exclusion of luminance stems from the fact that skin hue is largely invariant across different ethnic groups [Elgammal et al., 2009]. In other words,

the concentration of colour pigments in the dermal layers accounts primarily for changes in luminance, and only to a lesser extent for changes in skin hue.

In conclusion, authors have argued that the exclusion of luminance from the classification process is not beneficial but can be advantageous in some instances, where it can help generalise sparse training sets (*i.e.* with few images under the same illumination conditions) [Vezhnevets et al., 2003]. Notwithstanding, in the absence of conclusive evidence for its inclusion, we exclude this component in our analysis of skin colour.

In this chapter, we adopt the CIE *L*a*b** space in order to exploit the perceptual uniformity warranted by this colour space. In the following subsections we briefly introduce this colour space and compare it to the widely-known RGB space. We review both spaces with regards to the factors discussed above (colour stability, perceptual uniformity, and their treatment of the luminance component), and their relative advantages in the context of skin segmentation problems.

5.2.1.1 Basic colour spaces: RGB and normalised RGB

The RGB colour space originated from CRT (Cathode Ray Tube) technology. CRT screens produce colour displays by additive combination of the emission spectra of three cathode rays (red, green and blue). Since then, this space has seen widespread adoption for digital colour representation. RGB is the native space for hardware devices such as displays, scanners and digital cameras with its 24-bit format being the most common image format on the World Wide Web [Gonzalez and Woods, 2002].

Graphically, the RGB gamut is represented by a 3D cube (Figure 5.2) where the R, G and B components occupy the three orthogonal axes originating in one of the vertices in the base of this cube corresponding to pure black. This colour space is sometimes normalised by $R + G + B$ to give the so-called normalised RGB space. Note that, unlike RGB, this space is two-dimensional.

This colour space is perceptually non-uniform, a consequence of the well-known fact that the sensitivity of the human visual system exhibits a varying response over the
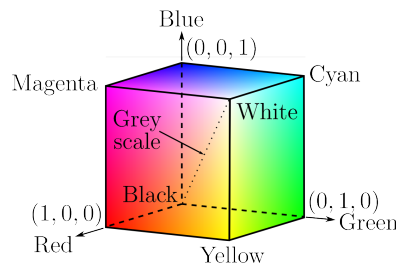
Figure 5.2
The RGB colour space, reproduced from Cattin [2010] with permission.
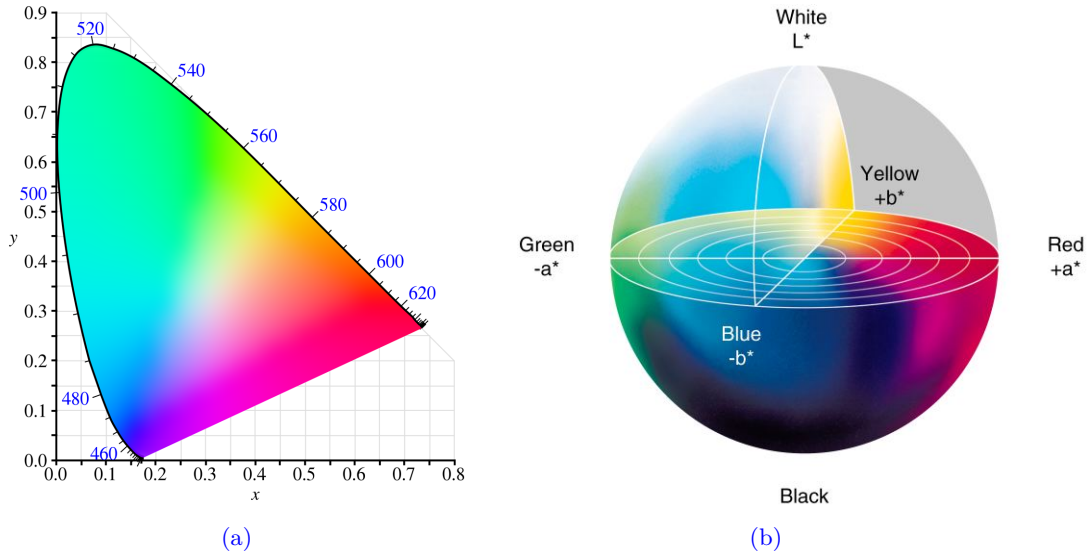
Figure 5.3
Colour gamuts for CIE 1931 *xy* and CIE *L*a*b** perceptually-uniform colour spaces. (5.3a) The CIE 1931 *xy* chromaticity diagram. The *xy* components represent CIE XY normalised by $X + Y + Z$. The outer curved boundary shows the spectral (or monochromatic) wavelengths, in nanometers. Reproduced from color reference system [2016] with permission. (5.3b) CIE *L*a*b** colour space. The CIE *L*a*b** space has components *L**, to describe lightness of colour ($L* = 0$ yields black and $L* = 100$ indicates diffuse white), and *a** and *b**, for the colour-opponent dimensions. Reproduced from color reference system [2016] with permission.

visible range of the electromagnetic spectrum. Therefore, changing the luminance over a skin patch affects all three colour components of this space.

Despite this drawback, RGB has been extensively used for skin segmentation given its simplicity and straightforward implementation. As an example, it was applied by Brand and Mason [2000] and Jones and Rehg [2002]. Both studies implemented non-parametric (histogram-based) skin models, which are known to be independent of colour space up to an invertible coordinate transformation (more on this on Section 5.2.2).

5.2.1.2 Perceptually-uniform colour spaces: CIE family

Colour spaces in the CIE family were standardised by CIE (Comission International de l'Éclairage) following extensive psychophysical experiments by Poynton [2012] aimed at reproducing the colorimetric response of a standard human observer.

The CIE 1931 XYZ space (a projection of which is shown in Figure 5.3a) was one of the first colour spaces defined mathematically so as to establish quantitative links between monochromatic colours in the electromagnetic visible spectrum (*i.e.* wavelengths), and perceptual colours, as perceived by human vision system. It serves as a basis to other colorimetric spaces, such as the CIE 1976 *L*a*b** space (Figure 5.3b) used in this thesis. Although CIE XYZ is related to the RGB colour space through a linear transformation, the increased perceptual-uniformity afforded by the CIE *L*a*b** colour space comes at the expense of additional computational load as colour conversions involving this space

demand non-linear operations (cubic roots) intended to replicate the non-linear response of the human perception of colour [Gonzalez and Woods, 2002].

The CIE *L*a*b** and other spaces in the perceptually-uniform family have been employed for skin colour segmentation by Cai and Goshtasby [1999]; Kawato and Ohya [2000]; Lai and Li [1998]; Yang and Ahuja [1998]; Zarit et al. [1999] among others.

5.2.2 Skin modelling

The end goal of skin segmentation is to obtain a decision rule which assigns pixel membership to the *skin* or $\overline{skin}$ classes. This is typically achieved through the definition of a distance metric for 'skin-likeliness' that is inherent to a particular choice of skin colour model. Several approaches to skin modelling have been proposed. The focus of this chapter is on parametric models and on Gaussian mixtures models in particular. Alternative research streams based on an explicit definition of the skin locus and histogram-based non-parametric models are mentioned only to provide a general perspective of how the proposed methodology fits into the field[1] (Figure 5.1).

5.2.2.1 Explicit definition of the skin region

One method to build a skin classifier is to design a skin cluster filter through a set of static rules that specify its boundaries in a given colour space. This method was first attempted by Jordão et al. [1999] and has since gathered many supporters attracted by the intuitiveness of its approach. The appeal of such a straightforward model is that these hard fast rules lead to the construction of very rapid classifiers. The accuracy of these methods relies on finding both a good colour space and adequate decision rules empirically. Recently, methods that automate this process using machine learning algorithms have been proposed in order to achieve a more accurate labelling of skin pixels.

5.2.2.2 Non-parametric models

Non-parametric models are a class of statistical methods centred on the idea of using skin colour histograms as a surrogate for the skin probability density function over a colour space. Based on their distributions one can then draw look-up tables that assign a skin probability value to each point of a discretised colour space (histogram bin).

Popular non-parametric approaches include the Skin Probability Maps (SPMs) of Brand and Mason [2000]'s, the naive Bayes classifier of Jones and Rehg [2002] and the face detection and tracking algorithms by Zarit et al. [1999] and Soriano et al. [2000].

The main advantages of non-parametric methods are that they are relative fast to train and use, and that their performance has been shown to be theoretically [Albiol

---

[1]The more recent dynamic skin distribution models are excluded. A review of this family of methods is available in Section 3.4 of Vezhnevets et al. [2003].

et al., 2001] and empirically [Shin et al., 2002] independent of the shape of skin *locus* (and therefore, colour space).

The main drawback is the considerably higher storage space required for storing skin probability maps, as opposed to a modest set of model parameters provided by parametric models (*e.g.* the probability map of an 8-bit image in a 3D colour space is an array of $2^{8\times3} = 2^{24}$ elements). The inability to interpolate results and generalise sparse training data [Sigal et al., 2004] is also a major concern.

In summary, in general terms non-parametric approaches perform well when they are presented with sufficient and representative training data and the histograms are quantised properly.

5.2.2.3 Parametric models

Parametric statistical approaches characterise the skin-colour distribution using a finite number of parameters. Although normal (Gaussian) distributions are certainly the most popular choice of density functions to model the distribution of skin tones, an alternative has been suggested by Lee and Yoo [2002], who have observed that, in several colour spaces, the skin cluster is elliptic in shape and hence not well described by a Gaussian model due to the asymmetry of this function with regards to its density means.

In any case, whenever an explicit distribution model is used, the question of model suitability arises. Few authors have validated their model choice. Yang et al. [1997] have shown that the distribution of skin colour for single subject under fixed lighting conditions in normalized RGB space does obey a Gaussian distribution. Yang and Ahuja [1998] have validated their hypotheses of skin colour gaussianity in the CIE Luv space as well as the suitability of a two-class Gaussian Mixture Model (GMM) with statistical tests. However, most authors have relied simply on the normal *appearance* of the chromatic distributions of the skin cluster and held their classifier's performance as the criterium for model validity.

As the goodness of fit for these parametric models is more dependent on the shape of the input distribution than for the non-parametric models, the colour space used has a profound effect on the ability to separate the mixture classes.

The main advantages of parametric models are the relatively small size of training sets required for generalisation and the compactness of the model representation. The main disadvantages are the need for computationally intensive algorithms required to estimate model parameters, particularly in the case of Gaussian mixtures, and the *a priori* selection of model order.

The choice of model order is an important one. On the one hand, an overly simple classifier may not be able to learn the meaningful features of each class appropriately and yield classification results of poor accuracy. On the other hand, a model that is

unnecessarily complex may lead to over-fitting. The number of Gaussian components used by different authors has ranged from a single one [Hsu et al., 2002; Yang and Ahuja, 1998] to a mixture of 2 [Yang and Ahuja, 1998], 8 [Terrillon and Akamatsu, 1999] or even 16 [Jones and Rehg, 2002] components. In practice, the order is generally determined via heuristics. In constrained environments, however, the model order can be predefined based on prior knowledge of environmental conditions. GMMs have also been applied by Cai and Goshtasby [1999]; McKenna et al. [1998]; Oliver et al. [1997].

We adopted a parametric approach using Gaussian mixture models, however, we base our approach on a different set of assumptions with respect to the studies cited thus far. It is generally accepted that parameters of the skin-color distribution can vary significantly with background complexity, individual variations in skin tone and lighting conditions [Storring et al., 1999; Yang et al., 1997; Yang and Ahuja, 1998]. Thus, to build a system that is sufficiently general to model different skin tones or robust enough to handle even modest variations in illumination conditions, one must employ an algorithm that adjusts the parameters of the distribution accordingly.

Given the wide range of ethnicities of the subjects recruited (Table 3.2) we applied GMMs on an individual basis. In addition, we exploited the fact that the background of all the frames to be classified (*i.e.* the incubators in the Oxford HDA) exhibit similar colour properties, and therefore, use a constant GMM model order. In summary, instead of modelling the variety of subject skin tones we opted instead to model entire frames by relying on the fact that the number of components in the Gaussian mixture can be predefined based on known environmental conditions. We hypothesize that this approach is more robust to changes in lighting (which inevitably occur inside the incubator as part of the daily routine in the neonatal ward) as these have a minimal effect on model order.

Gaussian Mixture Models

GMMs the most commonly used examples of mixture distributions. A Kth order GMM describes a probability density $0 \leqslant p \leqslant 1$ formed by a weighted sum of K Gaussian densities in a D-dimensional (colour) space. For a random variable x, p is defined as:

$$p(x|\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{k=1}^{K} \alpha_k p_k(x|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{5.1}$$

with identically distributed $p_k(x|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$:

$$p_k(x|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^D \det \boldsymbol{\Sigma}_k}} e^{-\frac{1}{2}(x-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(x-\boldsymbol{\mu}_k)} \tag{5.2}$$

where $\alpha_k \in \mathbb{R}^+$ are mixing coefficients (subject to the constraint $\sum_{k=1}^{K} \alpha_k = 1$), $\boldsymbol{\mu}_k \in \mathbb{R}^D$

the mean and $\mathbf{\Sigma}_k \in \mathbb{R}^{D \times D}$ is the covariance function of the Gaussian component k.

The mixture of K Gaussians can then be fully characterised by the parameter set $\boldsymbol{\theta}$, whose elements $\boldsymbol{\theta}_k$ are given by

$$\boldsymbol{\theta}_k = \{\alpha_k, \boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}_{k=1,\ldots,K} \tag{5.3}$$

Thus, given a sample $\boldsymbol{x} = [x_1, x_2, \ldots, x_N], x_i \in \mathbb{R}^D$, the problem of finding the Gaussian mixture model which suitably describes the N observations consists simply in estimating $\boldsymbol{\theta}$ subject to K. Under a Maximum Likelihood (ML) formulation, this problem becomes that of finding $\boldsymbol{\theta}$ which maximises the log-likelihood function $\log \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})$:

$$\boldsymbol{\theta}^{ML} = \arg \max_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) \tag{5.4}$$

with $\log \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})$ defined as follows (under i.i.d.[1] assumptions),

$$\log \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta}) = \log \prod_{i=1}^{N} p(x_i|\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} \alpha_k p_k(x_i|\boldsymbol{\theta}) \tag{5.5}$$

While for a single Gaussian component ($K = 1$) there is a closed-form solution to this maximisation problem (see *e.g.* Bishop [2007]), for $K > 1$, iterative algorithms are needed to find a numerical solution to Equation 5.4.

5.2.3 Classification rules

Regarding the classification criteria employed for skin labelling in statistical methods, the posterior probability is often used a measure on skin-likeliness. For binary classification, this quantity is often thresholded using the receiver operating characteristic (ROC) analysis. Otherwise, in mixture models one can also compare the value of a metric (*e.g.* the Malahanobis distance) between the pixel $\boldsymbol{c}$ we wish to classify and the different cluster centres of the estimated distribution, with class membership being assigned to the closest one. Here, we resort to the former approach.

5.3 Method

A flowchart illustrating the main steps of our skin detector is shown in Figure 5.4. The details regarding the implementation of each process block are provided in Sections 5.3.1 through 5.3.4.

---

[1] independent and identically distributed random variables.
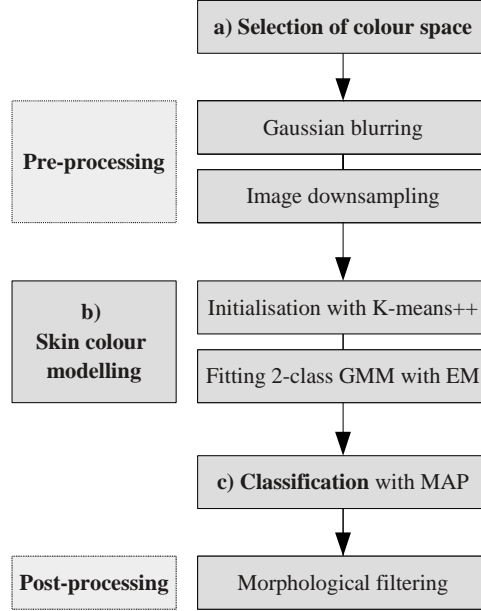
Figure 5.4
Flow diagram illustrating the image processing algorithm for skin segmentation in the NICU dataset.

5.3.1 Image pre-processing

Our starting point is the construction of colour histograms in the CIE *L*a*b** space. In order to render the global structure of the distribution more apparent, we applied a downsampling step. This step also reduces the computational load by reducing the size of the input data (pixel intensities) to the time-consuming process of fitting model parameters. To do this, we applied to each frame an image scale factor SF of 0.1 to the input frame, which produced reduced frames with one tenth of the size of the original video frames. SF is defined trivially as

$$SF = \frac{k_s\prime}{k_s} \tag{5.6}$$

where k_s and k_s' are the spatial sampling frequencies of the original image and the one obtained after resampling, respectively.

Following Nyquist, prior to downsampling we applied an anti-aliasing Gaussian filter with a cut-off frequency of $k_c = \frac{1}{2}k_s'$. Our choice of k_c is determined by the standard deviation (approximately 3 pixels) and size (10×10 pixels^2) of the filter kernel. For brevity, an in-depth discussion of our choice of parameters in this paragraph is provided in Appendix C.3.

5.3.2   Estimating Gaussian mixture densities

For rapid skin segmentation, we implemented a simple classifier system of two classes (skin *versus* non-skin pixels) for one-dimensional data (*a**) and Gaussian components.

Model fitting was performed using Expectation-Maximisation (EM) [Dempster et al., 1977], a well-established ML framework for estimating the underlying parameters of mixture distributions. The details of fitting GMMs within an EM framework can be found in Bilmes [1998]. Starting with some initial mixture model, the EM algorithm alternates between two steps: an Expectation step (*E*-step) for computing a lower bound of the log-likelihood $\log \mathcal{L}(\boldsymbol{x}|\boldsymbol{\theta})$ and a Maximisation step (*M*-step) for improving the current model with respect to this lower bound.

It has been demonstrated that EM is guaranteed to converge to a stationary point of the log-likelihood function under mild conditions [Dempster et al., 1977; Redner and Walker, 1984; Wu, 1983]. Unfortunately, as this function is generally non-convex, the convergence point of the EM algorithm bears a strong dependence on its initial conditions. Hence, one should ensure both the number of K components to be incorporated into the model and the initialisation of the mixture parameters are suited to the problem at hand. We initialised the EM algorithm parameters with the parameter array $\boldsymbol{\theta}^0$:

$$\boldsymbol{\theta}_k^0 = \{\alpha^0, \boldsymbol{\mu}_k^0, \boldsymbol{\Sigma}_k^0\}_{k=1,..,K} \tag{5.7}$$

where each sub-parameter is computed as follows:

- **K** It is a pre-requisite of GMM estimation that the number of mixture components be known prior to model fitting. As we wish to implement a simple classifier system with two classes (skin *versus* non-skin pixels), the choice of $K = 2$ is a trivial one[1].

- **$\boldsymbol{\mu}^0$** A reasonable choice of $\boldsymbol{\mu}^0$ is the set of K *a** chroma values corresponding to the peak coordinates in the *a** histogram over the reduced frames. As the ideal bin width, *i.e.* that which would provide a smooth histogram profile along *a**, is not known *a priori*, the histogram bin counts were first smoothed with a 3-bin median filter before peaks were detected. Peak detection returned several sets of closely located peaks. We determined their prominence[2], kept the two peaks for which it was highest and assigned $\boldsymbol{\mu}^0$ their *abcissa* values.

- **$\boldsymbol{\Sigma}^0$** After the *a** channel has been normalised, we assumed $\boldsymbol{x}$ is drawn from a

---

[1]Note that this use of GMM for segmentation is not the one normally used in the literature reviewed. Instead of constructing the mixture model using exclusively ground truth skin pixel intensities, we do so over the entire frame. See the discussion of our use of mixture models in Section 5.2.2.3.

[2]concept borrowed from topographic prominence. Prominence characterises the height of a mountain or hill's summit and is a measure of its *independence*. The prominence of a peak is given by the minimum height of climb to the summit on any route from a higher peak (or from sea level if there is no higher peak). The mathematical analogue is formally defined in MathWorks [2015] function reference for *findpeaks*.

collection of K independent Gaussian random variables with unit variance (*i.e.* $\sigma_1^0 = \sigma_2^0 = 1$).

- $\boldsymbol{\alpha}_k^0$ One popular heuristic used to initialise the mixture parameters is to apply k-Means++ [Arthur and Vassilvitskii, 2007], a variant of the K-means algorithm, to obtain initial class memberships. Assuming the centroids were located in the K $\boldsymbol{\mu}_k^0$ values and uniform initial mixing proportions, we let this algorithm provide us with binary class memberships for each x_k and computed each α_k^0 as the proportion of observations in our sample assigned to class k.

The ML estimates of the mixture model have been computed as

$$\boldsymbol{\theta}^{ML} = \{\boldsymbol{\alpha}^{ML}, \boldsymbol{\mu}^{ML}, \boldsymbol{\Sigma}^{ML}\} \tag{5.8}$$

we identified the skin class k_{skin} as

$$k_{skin} = \arg \min_{k \in \{1,2\}} |\mu_k^{ML} - a^*_{threshold}| \tag{5.9}$$

where $|.|$ denotes the $\mathcal{L}_1$-norm and $a^*_{threshold} = 10$.

5.3.3 Classification criteria

The posterior probability $p(skin|a^*)$ was used as a direct measure of skin-colour likeness. Classification was obtained by comparing these pixel-wise posterior with a threshold value estimated from the training data. We empirically set this threshold to 0.75. Pixels were classified as skin-pixels according to the criterium:

$$p(skin|\boldsymbol{c}) >= p^*_{skin}, p^*_{skin} = 0.75 \tag{5.10}$$

5.3.4 Post-processing

No image segmentation is perfect and often further processing is required before a useful segmentation mask is obtained. In order to separate pixels assigned the same label into regions without connecting components, the binary mask obtained after classification was upsampled to the original image size and then processed with a morphological operator, MATLAB's (MathWorks, Natick, USA) *imerode* algorithm.

The most common use of this operator is to remove unnecessary morphological detail from an image region. In a binary image, this operation sets the intensity of an output pixel to 0 if and only if any of the pixels in its neighbourhood are equal to 0 [Gonzalez and Woods, 2002]. Naturally, the number of pixels removed from the objects in an image

depends on both the size and the shape of the structuring element used to process the image. We adopted a 5×5 pixel^2 'cross' as this proved a good compromise; sufficiently large to break the pixel connectivity between skin pixels and ECG and IP wires and yet sufficiently small to preserve the edge integrity around the segmented skin areas.

Finally, the resulting mask was spatially filtered for connected components. The largest region was assumed to contain skin pixels. As the final step, the smallest (rectangular) bounding box that can contain said region was drawn and kept as the region of interest for further processing.

5.4 Performance analysis

To evaluate the success of a particular approach to object detection, it is common to compute performance statistics, such as measures of image overlap or receiver operating characteristics. Regardless of our choice of performance statistic, the merit of an image analysis algorithm is difficult to quantify in the absence of a ground truth or reference standard. The lack of an accepted reference standard is often a challenge in the analysis of clinical image data. A common approach to this problem is to compare algorithmic detections to those generated by domain experts. In Section 5.4.1, we describe how a dataset was constructed to assess the performance of the skin detector. In Section 5.4.2, we show how a reference was created by combining non-expert labels and demonstrate its use.

5.4.1 Construction of a dataset

A collection of 165 frames was selected. This set of frames was constituted by the first frame acquired at the start of each of the valid 165 manual breath counts performed by the clinical staff (see Appendix B.2 for details). In terms of infant activity, these frames were extracted from periods of relative quiescence, when the infants were resting in supine (63), right lateral (51), left lateral (43) and prone (3) positions[1] (Table 5.1). Manual breath counts were taken throughout the day thus resulting in frames being exposed to varying levels of ambient light. Therefore, this dataset ensures a good reproduction of the full range of lighting and postural characteristics observed in neonatal data.

---

[1]The infant position could not be classified in 5 out of the 165 frames.

Table 5.1

Infant poses at the start of manual breath counts.

Supine	Right lateral	Left lateral	Prone	NC	**Total**
63	51	43	3	5	165

113
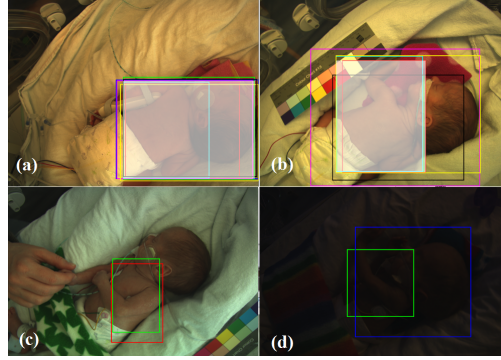
Figure 5.5

Four voting scenarios. Labels by individual annotators are represented by their bounding boxes. The filled rectangle represents the fused reference label according to a majority vote.

5.4.2 Construction of a reference standard

We evaluated our algorithm by assessing its performance against an estimate of the true labelling from observations of n annotators.

First, each frame was manually labelled by all annotators, who performed this task independently. Each annotator was instructed to draw the smallest bounding box that could enclose the largest contiguous skin region visible in the frame. Annotators were allowed to skip any frames for which they felt the skin boundaries could not be distinguished from the background. Frames were presented in a randomised order to minimise observer exhaustion.

While several methods have been proposed for estimating the hidden true label from a combination of human labels, the best strategy to fuse these pixel annotations remains unclear [Warfield et al., 2004]. In its simplest form, the method of estimating the true label can be reduced to a simple voting process. Here, we apply an approach based on a majority voting rule, whereby each pixel is deemed to be part of the true label if it is included in the labelled domains of at least half of the annotators who have seen the frame (including annotators who did not label it). In instances where none of the pixels in a given frame could secure a majority vote, no voting-based estimate was produced and this frame was excluded from the validation set. This occurred for 12 frames.

Figure 5.5 illustrates four possible voting outcomes. Figure 5.5a shows a frame whose labelling reached a consensual vote among annotators, whereas Figure 5.5b shows an example where inter-observer agreement is much lower. In the frames shown in Figures 5.5c and 5.5d no set of pixels secured enough votes for one label to be assigned.

5.4.3 Performance metrics

There is a wide variety of comparison criteria used in medical image object detection. The *best* way to measure the accuracy of detection depends largely on the consequences

that any error in the object detection might have.

In light of this, some authors have defined measures of accuracy based on volumetric estimates (*e.g.* for assessing the progression of brain lesions from MR imagery, where volume under- and over-estimations are undesirable). Alternative measures have been sought based on surface distance measures, such as the Hausdorff metric [Rockafellar and Wets, 2010], commonly applied in radiation therapy planning, where shape fidelity of the segmentation of the true lesion is essential to avoid any overinclusion (which would result in an delivery of dose to healthy tissue) or any underinclusion (which would leave a part of the tumour untreated).

We adopted a measure based on overlap ratios. Unlike the volumetric or contour measures just described, overlap measures are sensitive to displacement of the object label by design. In general terms, they are defined by ratios of the following quantities: true positives (TP = intersection between detections and ground truth), false positives (FP = detected parts not overlapping the ground truth), false negatives (FN = missed parts of the ground truth) and true negatives (TN = part of the image beyond the union of the detections and the ground truth).

To assess the performance of the skin segmentation method defined in this section, we computed its sensitivity, defined by the ratio $TP/(TP + FN)$, specificity, defined by the ratio $TN/(TN + FP)$ and the Dice similarity index [Dice, 1945], an overlap ratio measure defined as twice the ratio of the intersection of the automatically detected and ground truth sets over the sum of their sizes. Hence, it ranges from 0 (no overlap) to 1 (complete congruence). This index has recently become very popular in medical image analysis as it offers a compact representation of the compromise between sensitivity and specificity (its counterpart).

5.5 Results

As a first approach to colour-based skin segmentation, we explored different colour spaces in search for chrominance components with low chromatic spread over skin areas and high contrast between skin and background pixels. In Figure 5.6 we project a sample frame onto four colour spaces and analyse the the distribution of pixel intensities over the twelve colour channels. These spaces were selected so as to span a variety of colorimetric families: the standard RGB, the perceptual YCbCr, the orthogonal HSV and the perceptually-uniform CIE *L*a*b**. By examining the histograms in Figure 5.6, we found that the *a** chroma component in this colour space is ideally suited to our needs; its histogram exhibits two well-separated modes with limited overlap. In contrast, this degree of separability is not seen for any of the chrominance components in the other spaces for the frame shown. We note also that (as expected) the luminance-related components in all colour spaces (Y in YCbCr, *L** in CIE *L*a*b** and V in HSV) exhibit predominantly flat bin profiles.
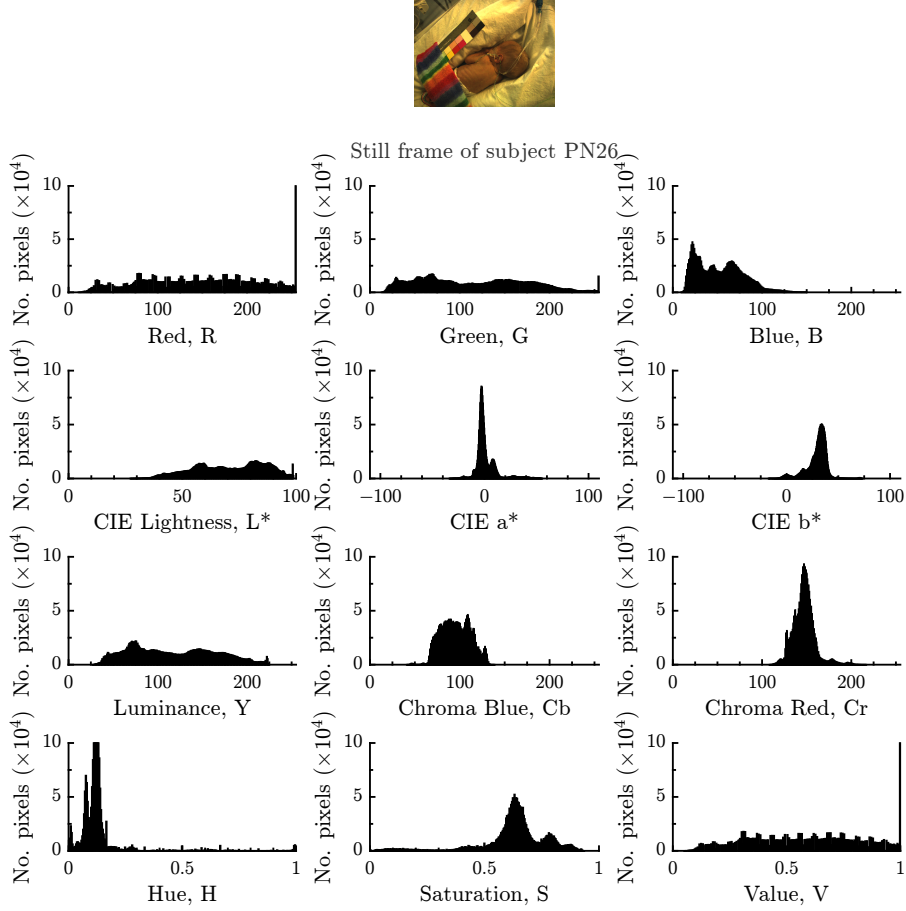
From top to bottom: Sample frame of subject PN 26 before a manual count, and its distribution of pixel intensities in RGB, CIE Lab, YCbCr and HSV colour spaces. The horizontal axes are set to the entire range of their respective component in the colour space gamut.

Based on these observations, pixels were clustered into classes based on their a^* chroma value. To do this, we constructed a 2-class Gaussian mixture model on the downsampled a^* image. The histogram of the reduced frame is shown in Figure 5.7a. The resulting mixture model is shown in Figure 5.7b, which presents the estimated mixture and likelihood probability density functions for *skin* and $\overline{skin}$ classes. As expected, the chromatically neutral ($a^* = 0$) class was given the highest mixing weight, a direct result of the fact that the pixels that make up the bedsheets occupy the largest area in the input frame. A smaller cluster of higher chroma and lower probability mass under the mixture density function is also clearly defined. The strict separation between these two classes is the basis of the good performance of our skin detector.

Figures 5.8a and 5.8b show detection results of the implemented skin classifier. Directly overlayed on each frame is the smallest bounding box containing the labelled skin pixels. Figure 5.8a shows results for different subject poses. Our results reinforce the fact

(a) (b)
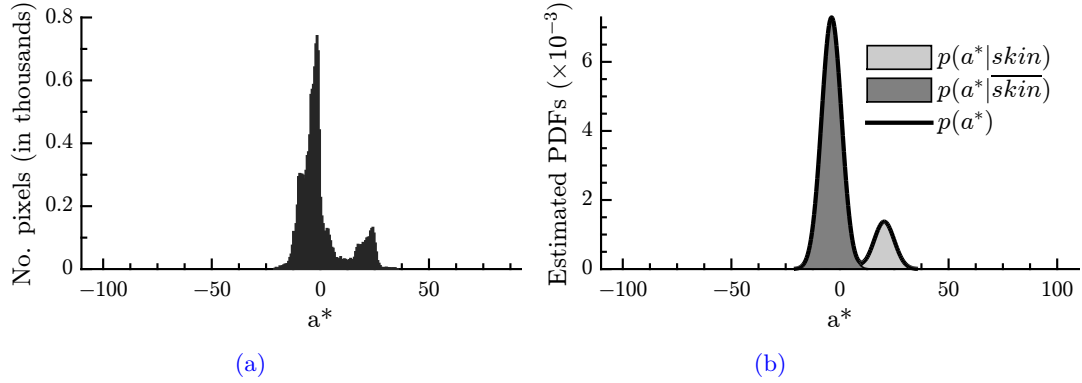
Figure 5.7

Histogram of pixel intensities and estimated Gaussian mixture model for a sample (reduced) frame of subject PN 26. (5.7a) Histogram of pixel intensities in the *a** channel. (5.7b) Estimated mixture and likelihood probability density function for *skin* and $\overline{skin}$ classes.

Table 5.2

Median values of Dice coefficient, sensitivity and specificity scores of the 2-class GMM method for each of the annotators ($n = 7$).

Annotator	Dice	Sensitivity	Specificity
1	0.76	0.63	1.00
2	0.85	0.78	1.00
3	0.95	0.99	0.95
4	0.88	0.80	1.00
5	0.86	1.00	0.89
6	0.90	1.00	0.92
7	0.92	0.93	0.99
Algorithm			
2−class GMM	0.78	0.97	0.83

that there is a distinct advantage in using colour-based methods for skin segmentation, especially concerning their robustness to changes in object scale and orientation. Figure 5.8b shows detection results under varying light conditions, a well-known challenge for colour-based segmentation algorithms.

Dice coefficients, sensitivity and specificity were calculated for each frame using a majority-voted manual label as ground-truth (from $n = 7$ annotators), where a voting-based reference was available (Figure 5.9). The same quantities were computed for each annotator and for our 2-class GMM classifier, with median values presented in Table 5.2.

5.6 Discussion

Based on our observation that skin and non-skin pixels exhibit a notable degree of separability on the *a** channel, the CIE *L*a*b** colour space (Figure 5.6), a clustering method based on a 2-class Gaussian mixture model was applied that exploited this property (Figure 5.7). Our results show that given a suitable choice of colour space, simple analysis of colour information allowed for effective skin segmentation in the NICU dataset.

117

(a)

(b)

Figure 5.8

Performance for different (5.8a) body poses and (5.8b) illumination levels. The images were recorded at the time of the first key press in each the manual breath count. These 1236×1635 images were segmented using a 2-class GMM classifier. The smallest bounding box containing the segmented *skin* region is shown in white.

For model fitting, we have used EM to compute the parameters of the Gaussian mixture distribution. EM is an iterative method which enjoys monotonic convergence and a linear rate of convergence [Dempster et al., 1977; Redner and Walker, 1984; Wu, 1983]. Although for some applications this rate may be unacceptably slow, fast convergence is observed for cases such as ours, where the Gaussian mixtures are well separated [Xu and Jordan, 1996]. In addition, given that their underlying components are multivariate Gaussian distributions, there is a closed-form solution to the expression to be maximised during the E-step of this method.

Regarding our choice of label fusion, it should be noted that while our approach may seem overly simple, it presents several advantages; it is unique in the case of binary labelling by an odd number of voters and its outcome reflects the overall preferred choice

average Dice = 0.78

average Sensitivity = 0.97

average Specificity = 0.83

(a) Dice

(b) Sensitivity
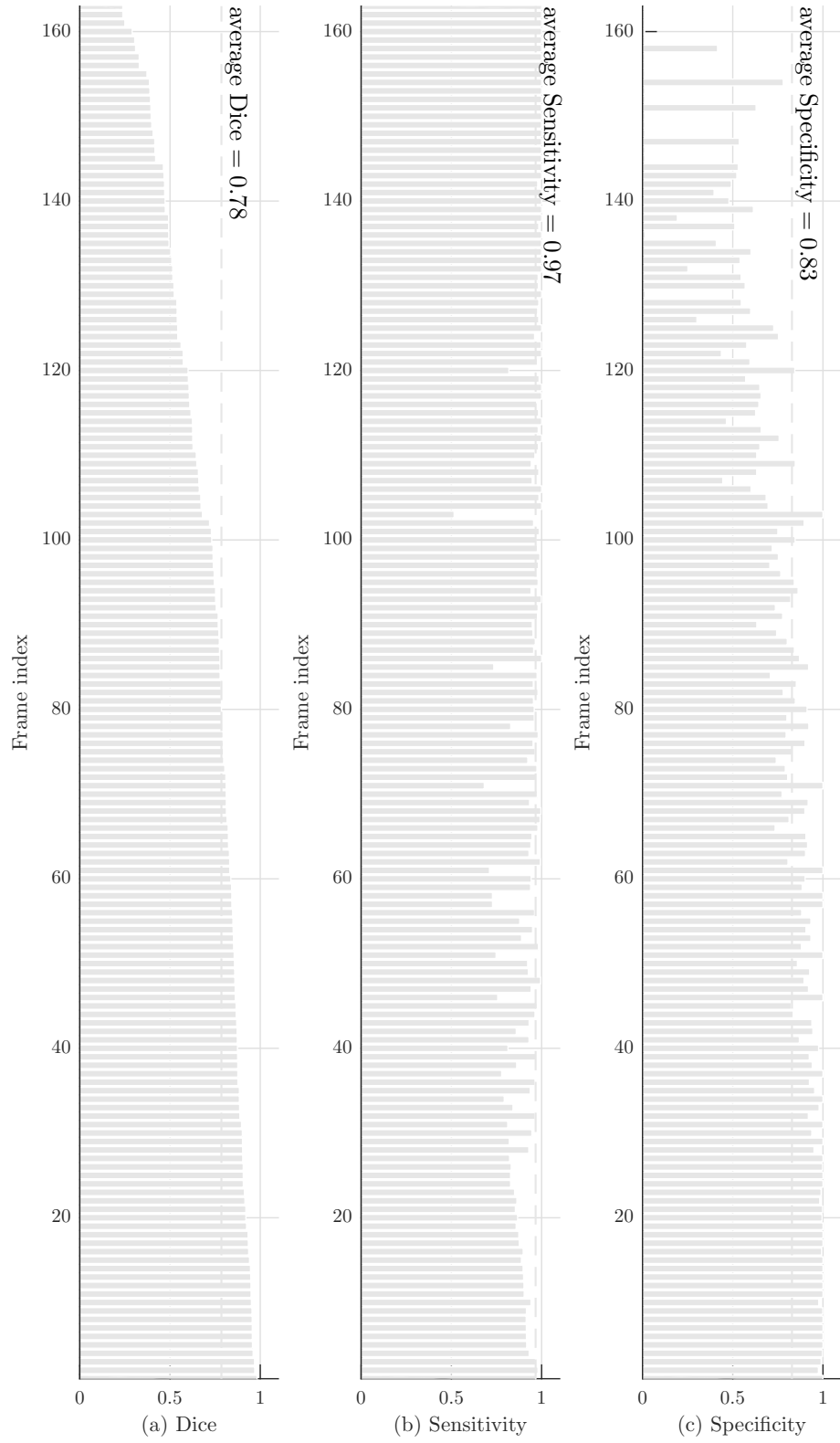
(c) Specificity

Figure 5.9
Dice coefficient (a), sensitivity (b) and specificity (c) scores of our method for each frame in the dataset.
Frames ordered by descending value of Dice score.

of the voters. However, it is a disadvantage of this strategy that it does not provide guidance as to how many annotators should agree before the structure is considered to be present, nor does it account for potential variability in quality or performance amongst the voters. Even though it does not allow for structural information regarding the object being segmented to be incorporated, our manual labelling process using bounding boxes reintroduces this constraint in practice.

In this chapter we have used heuristically-selected values for $a^*_{threshold}$ and for p^*_{skin}, the classification threshold on the posterior probability $p(skin|a^*)$ upon which our classifier relies. We acknowledge that this approach is not optimal and that in future work, potential performance gains could potentially be achieved through the adoption of a more data-driven approach to set these parameters, *e.g.* ROC analysis.

From the observation of the results shown in Table 5.2, it is evident that our automated method underperforms most human annotators. An exception is seen in the case of annotator 1, whose sensitivity and Dice measures score below their respective values for the GMM classifier. This behaviour is to be expected when the following factors are taken into account.

First, it is rare for classical computer vision algorithms to excel human annotators in visual tasks. When asked to identify an object from a still image, humans are able to draw from their experience. They can discern objects from the background by comparing the input against all of the objects they have seen and everything they have learned about those objects over the course of their life. This level of visual cognition represents a clear advantage over computational approaches. With the exception of recent methods based on deep neural networks, whose improvement in performance with the increase in the size of training data is maintained even for large volumes of training data, rarely do automated methods achieve near-human performance on image recognition tasks.

Secondly, we have opted for a simple colour-based approach to segmentation. In addition, in an effort to reduce the computational load on the GMM parameter estimation routine, we reduced the complexity of the colour model further to a single dimension by collapse the 3D colour model into a single channel (a^*). A predicted side effect of this oversimplification is that objects (or pixels) with skin-like tones become increasingly hard to resolve in the colour space. While this is not a problem when the infant and these objects are spatially disconnected, the performance of our automated algorithm can be affected negatively when this is not the case and the connectivity between them cannot be broken in the post-processing steps that follow mixture fitting. This error mode accounts for the higher false positive rate, or lower specificity in the algorithmic segmentations presented. It also impacts Dice scores adversely, although this effect is less pronounced.

Finally, the performance our methods is severely affected by poor illumination. It is unreasonable to expect that colour alone would be sufficient to give a robust solution in

these dark conditions, where it can merely serve as an important cue in a more integrated approach.

5.7   Chapter conclusions

In this Chapter, we have briefly described the problem of skin segmentation from a complex scene, designed an approach based on colour space segmentation and validated it on a subset of still images from our NICU dataset.

Overall, the method presented showed convincing accuracy on the dataset of 165 frames in which it was tested. The performance of the implemented skin classifier is very satisfactory (median Dice score = 0.78, median sensitivity = 0.97 and median specificity = 0.83), especially given that a single dimension (a^*) is used to solve what clearly is a multi-dimensional problem.

Our results suggest that this procedure can be used to solve the skin segmentation problem as part of the subject localisation step which that precedes vital sign estimation from NICU video footage.

Chapter 6

Video processing methods

Contents

6.1 Chapter summary

The video processing pipeline for the analysis of neonatal video data, which is presented schematically in Figure 1.1, yields two outputs: (a) a respiratory signal ψ and (b) a global activity signal α. In this chapter, we develop the video processing methods responsible for the derivation of these two signals.

The respiratory signal ψ was extracted from video data using a novel approach for the separation of a periodic signal from a set of noisy multivariate observations. The proposed framework has several advantages over conventional source separation techniques; it is more robust to narrowband noise and insensitive to phase differences across channels. While it is applicable to any problem where one seeks to extract a periodic source from a multivariate (and possibly degenerate) mixture of signal and noise sources, in this thesis, we used this method for the extraction of the subtle respiratory signal from neonatal video recordings. As we have seen previously, these measurements are typically affected by noise sources, such as *e.g.* subject movement, cardiac interference, uncontrolled illumination.

In Section 6.2, we introduce this approach and evaluate it on simulated and real data from a subset of neonatal video data.

In order to detect artefactual motion inside the incubator, we performed activity analysis by monitoring the presence of high-frequency noise in the absolute (pixel-wise) differences between consecutive frames. This is presented in Section 6.3 and assessed on an hour-long segment of neonatal video data manually annotated for motion events.

As presented in Figure 1.1, in both cases, prior image processing was necessary to restrict analysis to the frame area occupied by the monitored infant. This step, described in Chapter 5, precedes both activity and respiratory analysis described in this chapter. In Chapter 7, the respiratory signal extracted using the methodology described here is compared to that derived from both direct (IP) and indirect (ECG and PPG) sources during periods of manual breath counts. In Chapter 8, both the respiratory and the global motion signal are used as inputs in a classification algorithm for periods of cessation of breathing.

6.2 Respiratory signal from video data

6.2.1 Motivation

Consider the output of region processing shown in Figure 6.1, in which we show the average pixel intensity over the 25 image regions (5×5 non-overlapping boxes) during a 90-second period of quiet breathing and manual breath counting. The regions were obtained as subdivisions of the rectangular box which enclosed the largest skin region in the first frame in this time segment (B_{skin}) computed using the skin detection algorithm described in Chapter 5. The average pixel intensity over each subregion (ROI) was computed as the mean value of the blue colour channel over the region and for each frame. In an attempt to relate the resulting time series to identifiable respiratory events, the time instants of individual breaths (provided by the manual breath count) are shown as vertical lines between $t = 30$ s and $t = 63$ s.

As Figure 6.1b suggests, the average pixel intensity over the ROIs yields a set of signals with different degrees of periodic modulation at the fundamental respiratory frequency. It should, therefore, be possible to combine them in a way which enhances the periodic signature of the respiratory signal. Ideally, such combination would boost the components at the respiratory frequency (evident in the bottom traces shown in Figure 6.1b) while cancelling aperiodic noise. In this section, we present a linear source separation method to address this problem.

Our approach is partially based on the methodology of periodic component analysis (πCA) introduced in Saul et al. [2001] for the analysis of the periodic structure in speech. It uses singular eigenvalue decomposition [Parra and Sajda, 2003] but is optimised for

(b)
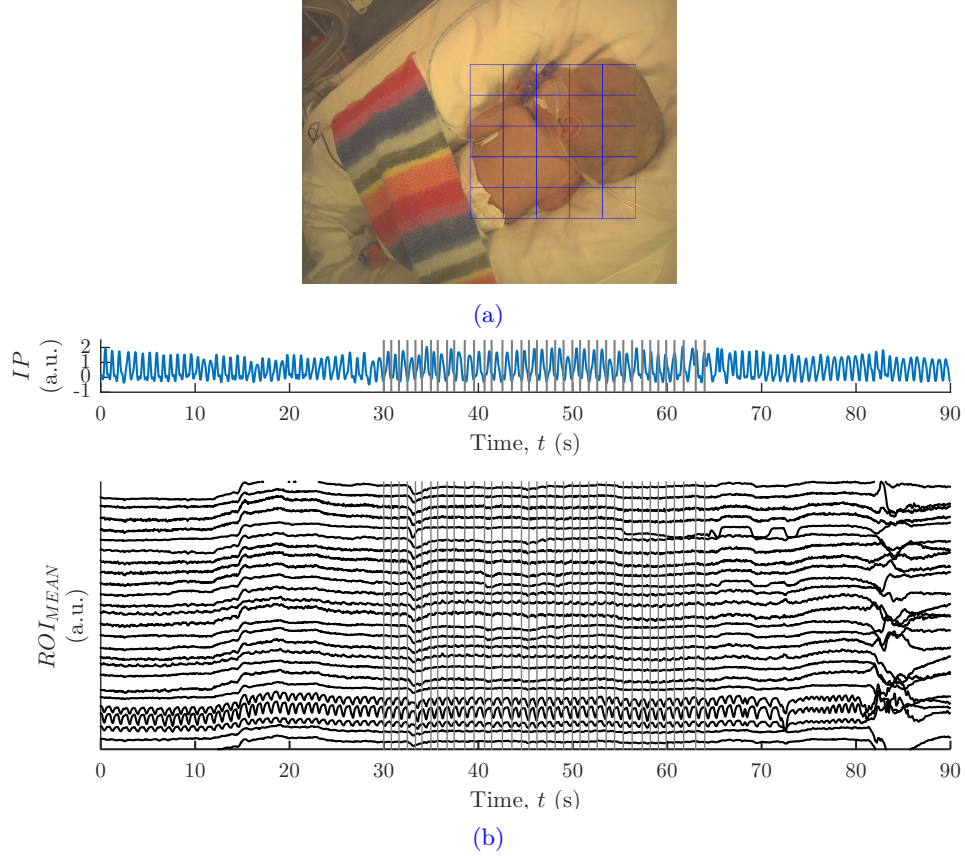
Figure 6.1
A typical 90-second segment of IP signal and average pixel intensity over skin regions during a manual breath count. (6.1a) Still image at the time of the first key press in a manual breath count. The bounding box containing the segmented skin region is divided into a grid of 5×5 rectangular subregions shown in blue. (6.1b) Segment of IP signal and average pixel intensity over the skin subregions. Each black trace represents the average pixel intensity (of the blue channel) over one of the ROIs in Figure 6.1a. Pixel intensities were normalised for display. The time instants of manually-annotated breaths are shown as vertical grey lines from $t = 30$ s to $t = 63$. During normal breathing, the IP and pixel intensity signals are modulated by the respiration cycle, with different regions exhibiting different degrees of modulation.

applications where the sources have a periodic structure.

6.2.2   Respiratory component analysis

We develop the method in three steps. These steps are designed to convey the main assumptions in our approach: (i) the use of an eigenvalue method for enhancing weakly periodic signals; (ii) the use of an analytic representation of the signals to compensate for phase differences across channels; and (iii) the extraction of the fundamental frequency using joint matrix diagonalisation.

Linear source separation

We begin with a mathematical simplification of the problem at hand. Given a set of m observed real-valued signals $\boldsymbol{x}(t) = \left[x_1(t), \ldots, x_m(t)\right]^T$, let us assume these are the

product of a linear combination of a set of n source signals $\boldsymbol{s}(t) = [s_1(t), \ldots, s_n(t)]^T$. Thus, the problem of source separation becomes that of finding the linear embedding of the observed signals $\boldsymbol{x}(t)$:

$$\boldsymbol{s}(t) = A^T \boldsymbol{x}(t) \tag{6.1}$$

where A is the matrix whose elements are the set of mixing weights $\{a_{i,j}\}_{i,j=1}^{m \times n}$ which maximise a given property of the source signals $\boldsymbol{s}(t)$.

In the case shown in Figure 6.1, where the observed signals $\{x_i(t)\}_{i=1}^m$ are average pixel intensities registered across the ROIs, the source signals driving the pixel intensity changes over the ROIs may be a combination of lighting changes, and local motion (*e.g.* limb movement, breathing excursions, ballistocardiographic motions of the head) or global motion (*e.g.* camera displacement, large infant movements), among other factors.[1].

Periodicity

The source of interest is the respiratory signal. For this reason, it is essential that our approach to source separation can return this signal as a single latent component. As seen in Chapter 2, the morphology of the respiratory waveform is highly dependent on the underlying physiological signal. For instance, while the electrical impedance signal collected at the chest wall will typically have a sinusoidal profile, the nasal pressure flow signal measured by a cannula is a quasi-periodic waveform with a rising edge (during the inspiration phase) followed by a sharp fall and a prolonged tail (during the expiration phase). Nevertheless, under resting conditions and at a constant rate, there is an assumption that the respiratory signal is periodic, or, in the very least quasi-periodic, with a period which equals the respiratory rate.

Conventional methods for solving linear source separation, exemplified by principal (PCA) and independent component analysis (ICA), seek source signals which are minimally correlated (PCA) or maximally independent in a probabilistic or information theory sense (ICA). These criteria are ill-suited for the problem at hand. When applied to a mixture of signals such as that shown in Figure 6.1b, PCA can often extract highly periodic components which explain very little variance in the original data. The component which explains the most variance is usually determined by voluntary subject motion, which introduces large intensity changes in the signal when compared to the subtle respiratory motion. The selection of high-variance components as the respiratory signal can, therefore, lead to inaccurate estimates of this signal from noisy windows. For quasi-periodic signals, a measure of periodicity of the extracted signals may be a more appropriate criterion. A simple modification of these methods can accommodate for this.

---

[1]In this scenario, the assumption of scalar mixing between the n sources is not a trivial one. We will discuss its implications later in this chapter.

We rephrase the problem of PCA, as formulated by [Pearson, 1901], using a measure of periodicity (instead of variance)[1]. Given an m-dimensional observation vector $\boldsymbol{x}(t)$, we seek to find the vector $s(t)$ given by the linear mixture $s(t) = \sum_{i=1}^{m} w_i x_i(t)$, or equivalently, $s(t) = \boldsymbol{w}^T \boldsymbol{x}(t)$, that minimises the prediction error Q at a given time lag τ:

$$Q(\boldsymbol{w}, \tau) = \frac{\mathbf{E}_t\{|s(t+\tau) - s(t))|^2\}}{\mathbf{E}_t\{|s(t)|^2\}} \tag{6.2}$$

where $E_t\{\cdot\}$ represents the average over the time index t. It follows from this definition that $Q \geqslant 0$, where $Q = 0$ indicates $s(t)$ is a fully periodic signal with period that equals τ. For simplicity, we have assumed that the signals are discretely sampled and that the period τ is an integer multiple of the sampling interval. It can be shown that, by expanding the right hand side of this equation in terms of w_i, Equation 6.2 may be rearranged as follows:

$$Q(\boldsymbol{w}, \tau) = \frac{\mathbf{E}_{t,ij}\{w_i w_j B_{ij}(\tau)\}}{\mathbf{E}_{t,ij}\{w_i w_j C_{xx,ij}(0)\}} \tag{6.3}$$

with the matrices $C_{xx}(\tau)$ and $B(\tau)$ given by:

$$C_{xx}(\tau) = \mathbf{E}_t\{\boldsymbol{x}(t+\tau)\boldsymbol{x}(t)^T\}$$
$$B(\tau) = \mathbf{E}_t\{[\boldsymbol{x}(t+\tau) - \boldsymbol{x}(t)][\boldsymbol{x}(t+\tau) - \boldsymbol{x}(t)]^T\} \tag{6.4}$$

Equation 6.3 can be written in a more compact form

$$Q(\boldsymbol{w}, \tau) = \frac{\boldsymbol{w}^T B(\tau)\boldsymbol{w}}{\boldsymbol{w}^T C_{xx}(0)\boldsymbol{w}} \tag{6.5}$$

with $\boldsymbol{w} = [w_1(t), \ldots, w_n(t)]^T$. We note also that, given that $C_{xx}(\tau)$ is symmetric and positive definite (for $C_{xx}(\tau)$ is the auto-covariance matrix of $\boldsymbol{x}$), $B(\tau)$ can be rearranged as [Sameni, 2008]

$$B(\tau) = 2\Big(C_{xx}(0) - C_{xx}(\tau)\Big) \tag{6.6}$$

and thus the right hand side of Equation 6.5 may be rewritten as

$$Q(\boldsymbol{w}, \tau) = 2\Big(1 - \frac{\boldsymbol{w}^T C_{xx}(\tau)\boldsymbol{w}}{\boldsymbol{w}^T C_{xx}(0)\boldsymbol{w}}\Big) \tag{6.7}$$

Hence, finding the set of weights $w \in \mathbb{R}^{m \times 1}$ which minimises the prediction error

---

[1]This is a restatement of the problem such as that seen in πCA [Saul et al., 2001].

$Q(\boldsymbol{w}, \tau)$ is equivalent to finding the set of weights which maximises the ratio between the quadratic forms $\boldsymbol{w}^T C_{xx}(\tau)\boldsymbol{w}$ and $\boldsymbol{w}^T C_{xx}(0)\boldsymbol{w}$.

According to the Rayleigh theorem in Strang [1988], the solution this problem is given by eigenvalue decomposition. Given a matrix W which can simultaneously diagonalise the symmetric matrices $C_{xx}(\tau), C_{xx}(0) \in \mathbb{R}^{m \times m}$, such that:

$$W^T C_{xx}(\tau)W = D \quad \text{and} \quad W^T C_{xx}(0)W = I, \tag{6.8}$$

where D is a diagonal matrix with respect to $W = [w_1, \ldots, w_n]$ with eigenvalues sorted by descending order of magnitude and I is the identity matrix, the first eigenvector w_1 corresponding to the largest eigenvalue, also maximises the ratio $\frac{\boldsymbol{w}^T C_{xx}(\tau)\boldsymbol{w}}{\boldsymbol{w}^T C_{xx}(0)\boldsymbol{w}}$ [Strang, 1988].

In other words, the weight vector $\boldsymbol{w}$ which minimises Equation 6.5 is given by the eigenvector corresponding to the largest eigenvalue of the matrix pair $C_{xx}(\tau), C_{xx}(0)$. That is, the transformation $W^T \boldsymbol{x}(t)$ returns the estimates of the periodic sources along the rows in descending order of periodicity. The first source (*i.e.* the most periodic), $\hat{s}_1(t)$[1], is assumed to contain the respiratory signal $\psi(t)$:

$$\psi(t) \equiv \hat{s}_1(t) \tag{6.9}$$

Phase invariance

The eigenvalue method in the previous section has one serious deficiency; it cannot account for phase differences across the input channels, *i.e.* any phase lags between the signals in the $]-\pi, \pi[$ range will cause the signals to interfere destructively when linearly combined.

To overcome this shortcoming, we replaced the real-valued inputs $\{x_i(t)\}_{i=1}^m$ by their analytic representations $\tilde{x}_i(t)\}_{i=1}^m$. The analytic representation of the signal $\boldsymbol{x}(t)$ was derived using the discrete-time Hilbert transform. We then restate the eigenvalue problem in Equation 6.3 for the general case where the input signals are allowed to be complex[2]

$$Q(\boldsymbol{w}, \tau) = \frac{\mathbf{E}_{ij}\{w_i^* w_j B_{ij}(\tau)\}}{\mathbf{E}_{ij}\{w_i^* w_j C_{xx,ij}\}} \tag{6.10}$$

---

[1] The superscript ˆ denotes an estimated variable.
[2] The superscript * denotes the complex conjugate.

where the matrix elements $\{B_{ij}\}$ and $\{C_{xx,ij}\}$ are determined by

$$C_{xx,ij} = \mathbf{E}_t\{\tilde{x}_i^*(t+\tau)\tilde{x}_j(t)\}$$

$$B_{ij}(\tau) = \mathbf{E}_t\{\tilde{x}_i^*(t)\tilde{x}_j(t) + \tilde{x}_i^*(t+\tau)\tilde{x}_j(t+\tau) - \tilde{x}_i^*(t)\tilde{x}_j(t+\tau) - \tilde{x}_i^*(t+\tau)\tilde{x}_j(t)\}$$

(6.11)

In this formulation, $\{w_i\}_{i=1}^m \in \mathbb{C}$ so that they can compensate for phase differences across channels. The optimal weights W are again given by the eigenvector corresponding to the smallest eigenvalue of the matrix pair $C_{xx}(\tau), C_{xx}(0)$. As $C_{xx}(\tau)$ and $C_{xx}(0)$ are Hermitian matrices[1], these eigenvalues remain real.

Extracting the periodic component

Using the properties above, a straightforward iterative algorithm can be devised to extract the most periodic component from a mixture of real signals $x(t)$. First, the analytic signals, $\tilde{\boldsymbol{x}}(t)$, are computed from $\boldsymbol{x}(t)$ (or over sliding windows of this signal) using the Hilbert transform. The next step is to diagonalise the matrix pair $C_{xx}(\tau), C_{xx}(0)$ of $\tilde{\boldsymbol{x}}(t)$ over a range of periods, $\tau \in [\tau_{min}, \tau_{max}]$, associated with the range of plausible respiratory rate values. The fundamental period is selected as the value of $\hat{\tau}_0$ which minimises the cost function in Equation 6.10. The final step is to use the transformation matrix found during the $C_{xx}(\hat{\tau}_0), C_{xx}(0)$ diagonalisation step to compute $\hat{s}(t) = W^T \boldsymbol{x}(t)$ and find $\psi(t)$ as the first component, $\hat{s}_1(t)$.

An important limitation of this approach needs to be addressed. The need to estimate the period of the source signal places a considerable burden on the algorithm: each iteration step requires (1) the computation of $m \times m$ cross-covariance $C_{xx}(\tau)$ and (2) the diagonalisation of the matrix pair $C_{xx}(\tau), C_{xx}(0)$. In order to reduce the computational load, we substitute the second step by the joint diagonalisation of the matrix set $C_{xx}(0), C_{xx}(\tau_{min}), \ldots, C_{xx}(\tau_{max})$ obtained after (1). The process of joint diagonalisation of eigenmatrices consists in finding the change of basis which makes these matrices *approximately* diagonal according to a given joint diagonality criterion. Intuitively, this set of basis vectors corresponds to the *average* eigenvectors of the matrix set, *i.e.* that retain important invariance properties. The joint diagonalisation procedure used here was developed by Cardoso and Souloumiac [1996] based on the joint approximate diagonalization of eigenmatrices (JADE) algorithm. This method of extracting the periodic component embedded in a multi-channel time series through a combination of generalised eigenvalue decomposition and joint digonalisation will be referred to as respiratory component analysis, or τCA.

_____

[1] $H^\dagger = H$, where † denotes the conjugate transpose.

6.2.3 Performance analysis

6.2.3.1 Synthetic data

In this chapter, we illustrate some features of τCA through its performance on simulated data. Two experiments were performed on two sets of synthetically generated data: a set of phase shifted sinusoids and a simplified breathing model.

The first experiment is intended to show the performance of the proposed method with respect to the conventional methods of blind source separation (Experiment A). The synthetic time series were normalised and then decomposed into their source signals using PCA and ICA. The two techniques were selected as they have found widespread application in biomedical signal processing both as a means of denoising or reducing the dimensionality of multivariate sets. Here, PCA was performed using singular value decomposition [Parra and Sajda, 2003]. For ICA, the algorithm based on JADE developed by Cardoso [1999] was used. In a second experiment, we sought to show the advantages of using analytic representations *versus* the original (real-valued) signals as the inputs to τCA (Experiment B).

Phase-shifted sinusoids

A reference signal $s_1(t)$ consisting of a sine wave with a frequency of $f_d = 2$ Hz sampled at a frequency of $f_s = 100$ Hz for a duration of 5 s was generated. A set of $N = 6$ signals $\{\check{s}_k(t)\}_{k=1}^N$ with the same analytic description but for a phase difference of $\frac{\pi}{N}$ radians between each other (Equation 6.12, Figure 6.2a) were linearly combined to obtain the set of observed signals $\boldsymbol{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ with $\{x_i(t) = \sum_{k=1}^N c_{ik}\check{s}_k(t)\}_{i=1}^N$ shown in Figure 6.2b.

$$\check{s}_k(t) = sin\Big(2\pi f_d t + \frac{\pi}{N}(k-1)\Big), k = 1, \ldots, N$$
$$\check{s}_1(t) = s_1(t) \tag{6.12}$$

The N^2 mixing coefficients $c_{ik} \in [0, 1]$ were drawn from a uniform distribution in

(a) Synthetic signals.   (b) Mixture of synthetic signals.
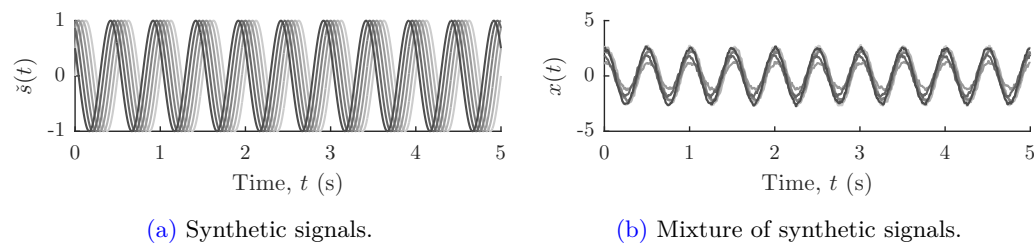
Figure 6.2
Synthetic data. A phase difference of $\frac{\pi}{6}$ exists between the signals shown in 6.2a. In 6.2b, the signals have been linearly combined using normally distributed weights and random Gaussian noise with a standard deviation of 0.1 was added to each component.

this interval (Equation 6.13). Zero-mean Gaussian noise with a standard deviation of 1 was added to the linear mixture.

$$c_{ik} \sim \mathcal{U}(0,\ 1), i = 1, \ldots, N, k = 1, \ldots, N \qquad (6.13)$$

Simplified breathing model

(a)

(b)

(c)

(d)
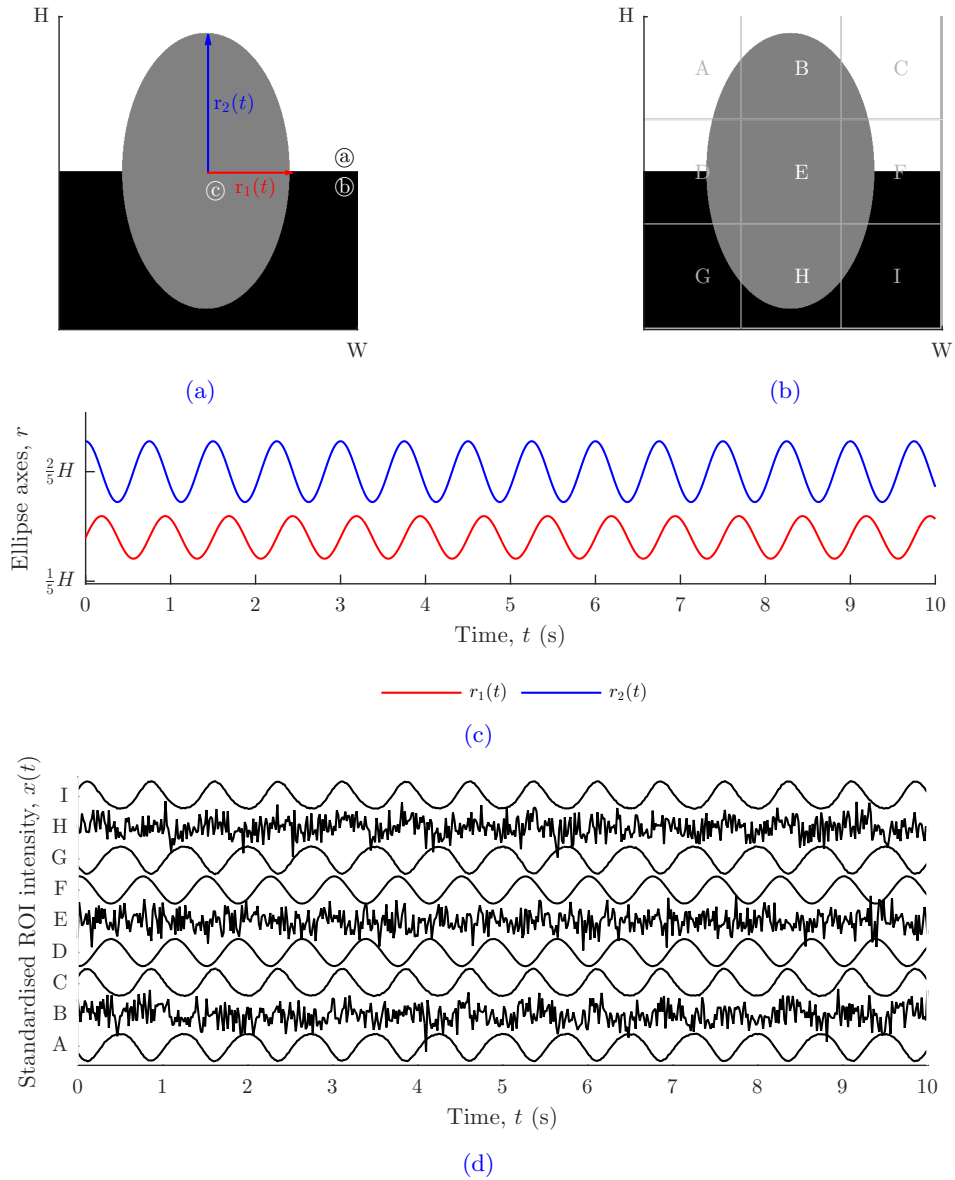
Simplified breathing model. (6.3a) Synthetic model. (6.3b) Regions of interest. (6.3c) Synthetic signals. (6.3d) Mixture of synthetic signals. The variables shown are defined in the main text of Section 6.2.4

The primary application of our method is the derivation of respiration from an RGB video signal. Let us consider a simplification of this problem shown in Figure 6.3. In this

model, each frame ($H \times W$) is composed of three sub-regions. The top half of the frame (ⓐ) is populated by pixels with intensity $I = 1$ while pixels in the bottom half (ⓑ) have $I = 0$. The centre of the frame (ⓒ) is occupied by an ellipse with $I = \frac{1}{2}$, whose axes oscillate over time t according to the function:

$$
\begin{aligned}
r_1(t) &= \frac{1}{R} r_2(t + \phi) \\
r_2(t) &= r_{min} + r_\delta sin(2\pi f_0 t) + \varepsilon(t)
\end{aligned}
\tag{6.14}
$$

where r_1 is the horizontal axis and r_2 is the vertical axis. The purpose of these three elements is to model a simple scene. ⓐ and ⓑ are bright and dark background elements (*e.g.* the white bedding and a colourful blanket), while ⓒ represents an infant breathing at a rate of f_0 breaths per minute.

In this example, the time vector t is sampled at regular intervals of $\frac{1}{f_s}$ in the interval $[0; 10]$ s. The remaining model parameters are defined below:

$$
\begin{aligned}
H = W &= 512 \text{ pixels} \\
r_{min} &= \frac{H}{2.5} \\
r_\delta &= \frac{H}{18} \\
R &= 7 \\
f_s &= 60 \text{ Hz} \\
f_0 &= 1.3 \text{ Hz} \\
\phi &= \frac{1}{2\pi f_0} \frac{\pi}{4}
\end{aligned}
\tag{6.15}
$$

In addition, normally distributed (i.i.d) noise with a variance of 0.1^2 was added to each pixel. To produce a multi-channel input, the average intensity was computed over 9 image blocks (A)$-$(I) arranged as shown in Figure 6.3b.

The average intensity over these regions over time is shown in Figure 6.3d. The intensity fluctuations with f_0 are clearly visible, particularly for the lateral regions. There are observable phase differences among these channels which arise as a result of ϕ.

6.2.3.2 Neonatal data

To illustrate the performance of τCA on natural video data, we generated a segment of multi-channel data by pixel averaging over visible skin regions in a segment of the NICU dataset in the manner shown previously in Section 6.2.1. This segment, shown in Figure 6.4, was obtained as an 8-second section of the signals shown in Figure 6.1a from $t = 15$ s to $t = 23$ s.

6.2.4 Results

6.2.4.1 Synthetic data

Phase-shifted sinusoids

In Figure 6.5, we show the results of decomposition using the aforementioned separation techniques, with the mixed signals $\boldsymbol{x}$ separated into their original components $\hat{\boldsymbol{s}}$ (Experiment A). We observe that PCA and ICA have each separated the sinusoidal source into two principal or independent components. Note also the similarity of these component pairs regarding their main frequency of oscillation. In contrast, τCA returned a single periodic component with the artificially-added Gaussian noise divided across the remaining τCs.

Shown along the left column of Figure 6.6 are two decompositions of the simulated signals above, when either real or analytic versions are fed to the τCA routine (Experiment B). In the column on the right-hand side of this figure, the single-sided amplitude spectrum of each component computed with the FFT is shown. When real-valued inputs are used, we notice that several τCs exhibit periodic oscillations with a period of $\frac{1}{f_d}$ and distinct initial phases. As an example, components 1 and 2 in Figure 6.6a are approximately in anti-phase. Conversely, this is not seen in the case where analytic signals are fed to the τCA method, with a single τC exhibiting a 2 Hz amplitude peak in the Fourier spectrum.

Simplified breathing model

The τCA algorithm was used to estimate the dominating periodicity and subsequently infer the source signal. Its results are shown in Figure 6.7 alongside the results of PCA and ICA for the same task (Experiment A). As in the previous example, we observe
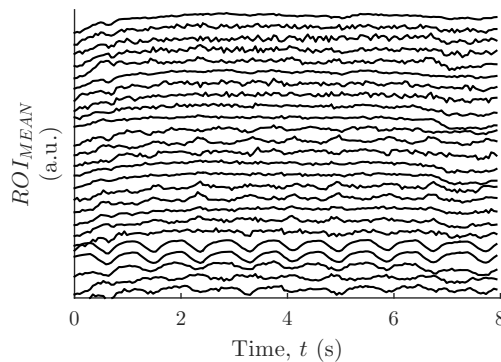
Figure 6.4
Segment of average pixel intensity over the 5×5 skin subregions in Figure 6.1b from $t = 15$ s to $t = 23$ s. Each black trace represents the average pixel intensity over one ROI. The 25 traces have been normalised for display.
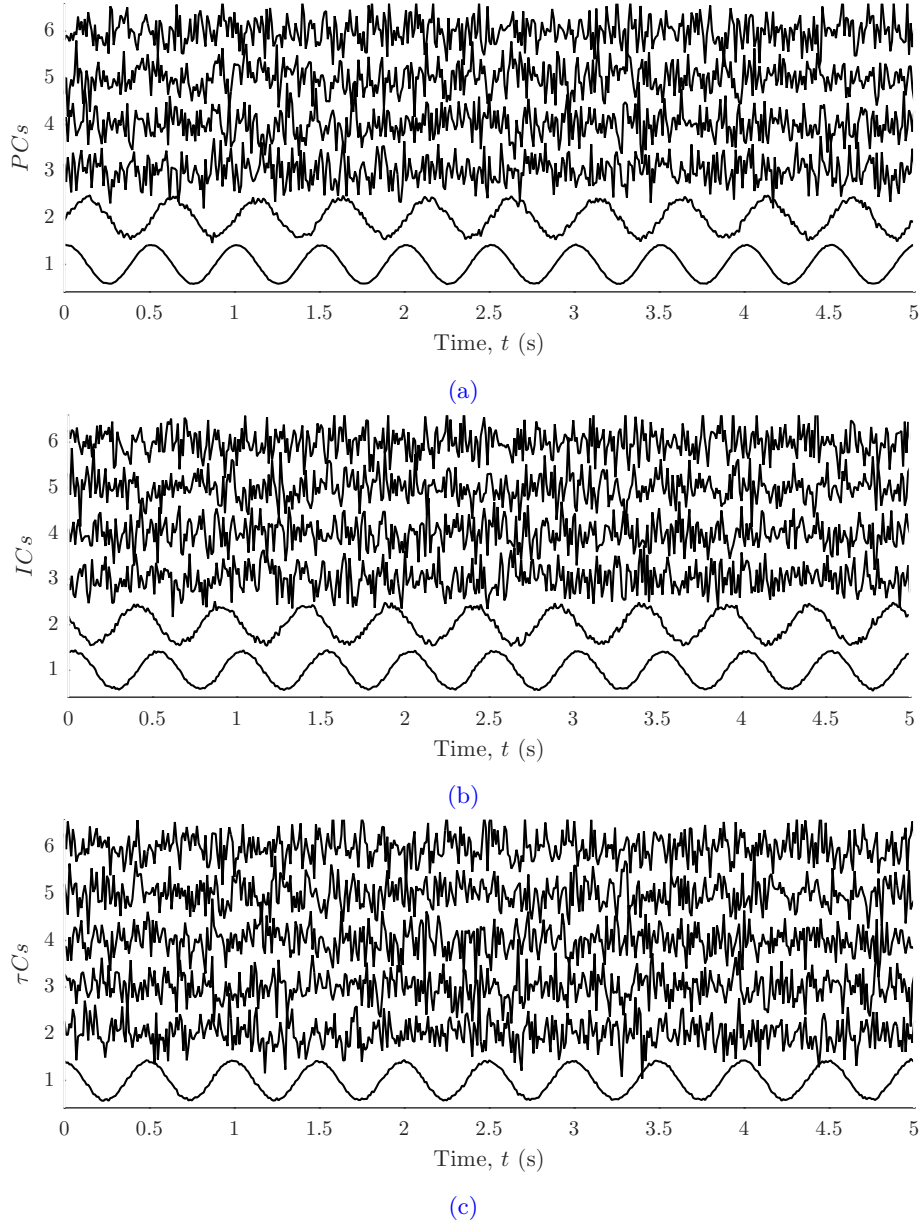
132

Components returned after the three source separation routines were applied to the synthetic signals consisting of a linear mixture of sinusoids with a phase-lag of $\frac{\pi}{6}$: (6.5a) PCs, (6.5b) ICs, and (6.5c) τCs.

that only τCA returned a single periodic component. Both principal and independent components returned a set of periodic components with varying amounts of Gaussian noise. Once again, the advantage of using complex inputs over real-valued ones was demonstrated (Figure 6.8) (Experiment B).

6.2.4.2 Neonatal data

In Figure 6.9, we observe that the intensity changes over the 25 image regions was decomposed into several components. From top to bottom, these include a respiratory
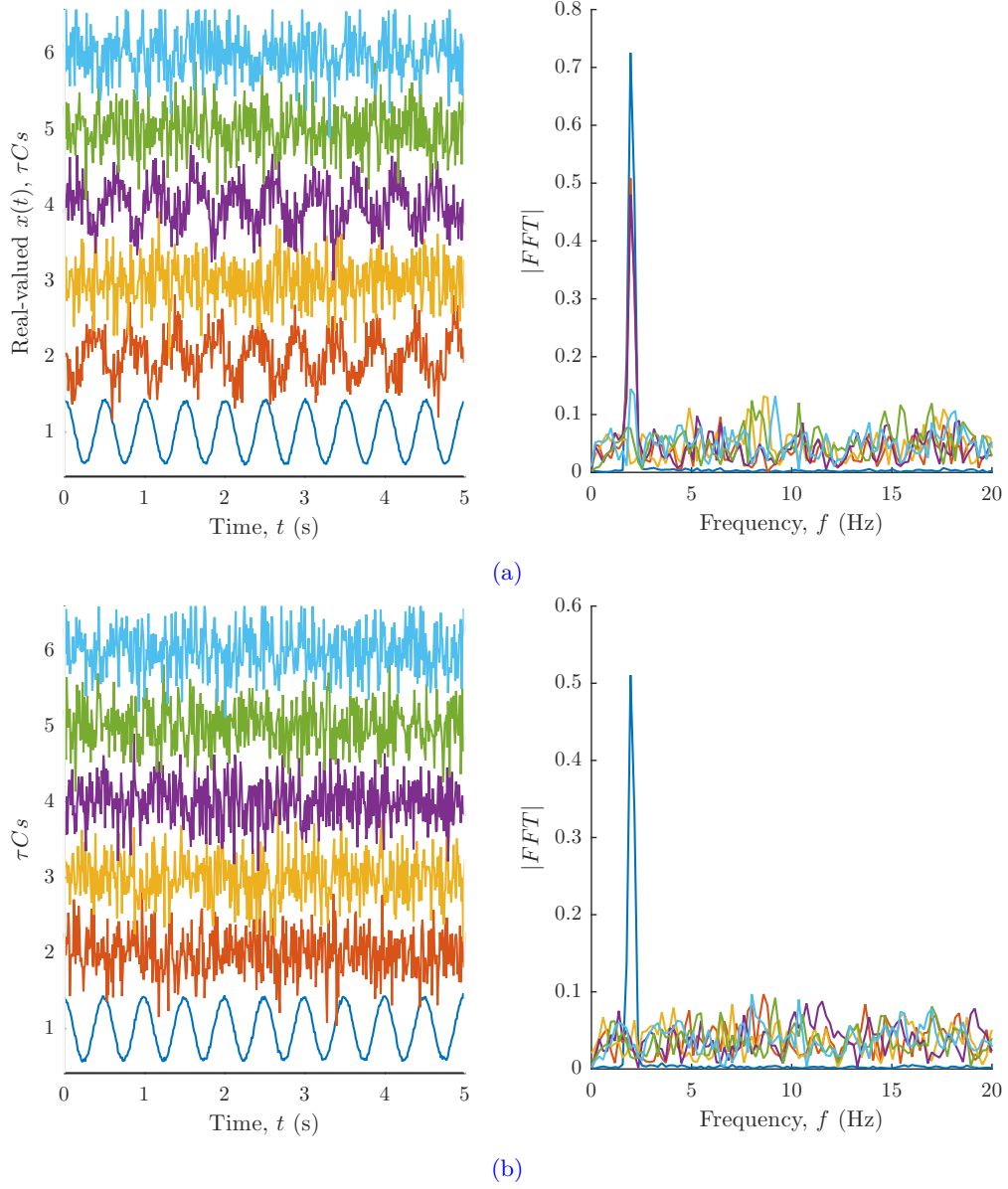
Figure 6.6

Components returned after τCA was applied to the synthetic signals consisting of a linear mixture of sinusoids with a phase-lag of $\frac{\pi}{6}$. Results are shown for both real-valued inputs (6.6a) and their analytic representations (6.6b). The frequency spectrum of the returned τCs is shown in the FFT periodograms on the right-hand side.

component (ψ, reproduced in the plot below), high-frequency noise, and low-frequency noise (the bottom two components, likely related to infant motion). A segment of the IP waveform acquired simultaneously is also shown. The advantages of using τCA are clearly observed; the latent respiratory signal seen across several channels of ROI_{MEAN} is returned as a single component.

Figure 6.10 shows the extracted τC ($\hat{s}_1$) from input signals extracted for a given window of interest using a range of ROI sizes. We observed that as smaller ROI sizes are used, the respiratory modulation of $\hat{s}_1$ is more pronounced as evidenced by the increased
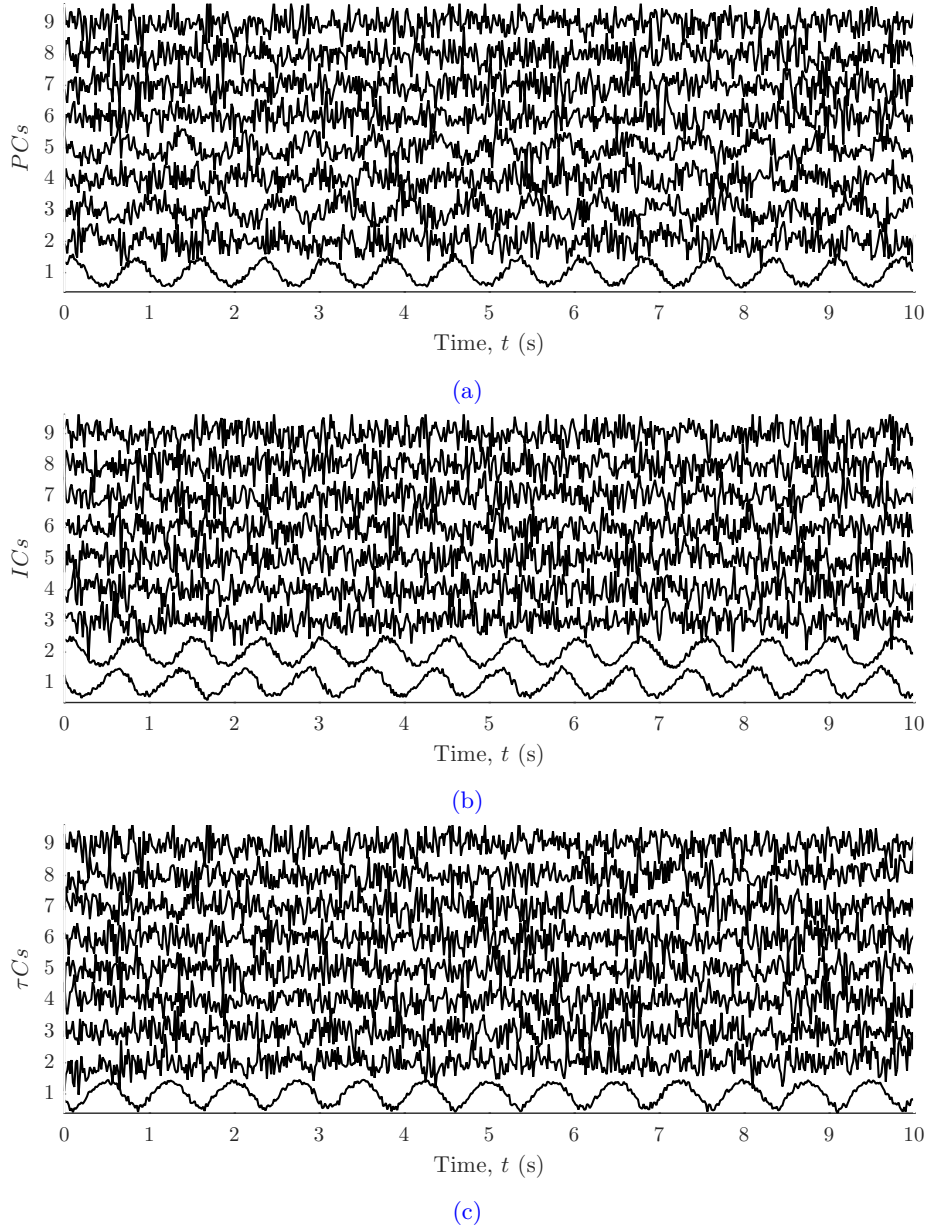
Components returned after the three source separation routines were applied to the synthetic signals generated using the simplified breathing model of Figure 6.3: (6.7a) PCs, (6.7b) ICs, and (6.7c) τCs.

magnitude of the PSD spectra at the respiratory rate frequency (note the change of scale of the vertical axes of Figure 6.10b). The size of the ROI to be selected is dependent on the physiological information we wish to extract. This is discussed further in Section 6.2.5.

In Figure 6.11, we present respiratory rate values (RR_ψ) derived from the ψ signal returned from τCA. RR_ψ was computed using the methodology for RR estimation developed in Chapter 4. In these examples, the length of the analysis window was adjusted to 8-seconds (with 1-second shifts). The pass band of the digital filters in Section 4.6.2 was shifted to 20 - 120 bpm in order to better capture the high respiratory rates observed

(a)

(b)
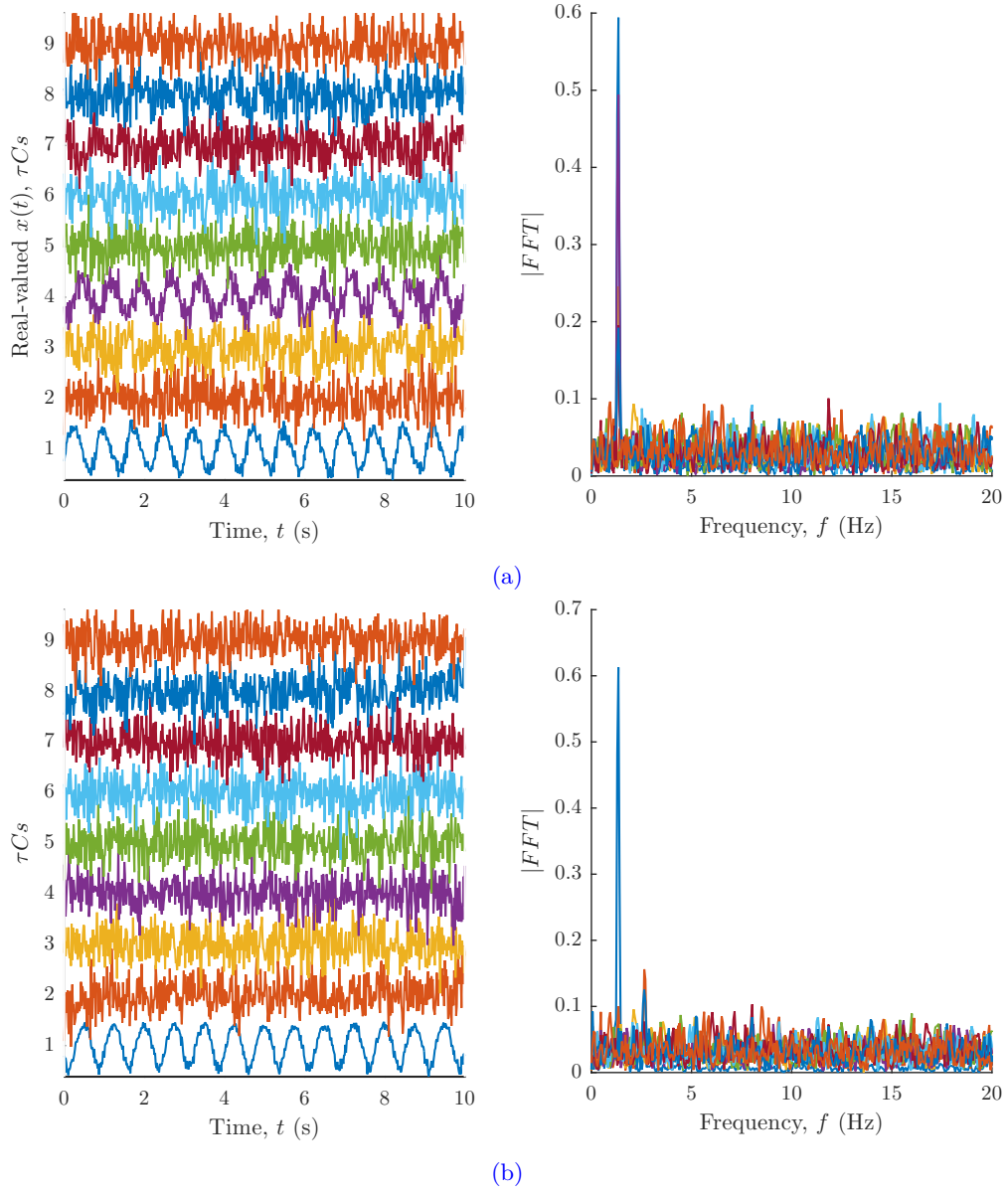
Figure 6.8
Video processing pipeline. Components returned after τCA was applied to the synthetic signals generated using the simplified breathing model of Figure 6.3. Results are shown for both real-valued inputs (6.8a) and their analytic representations (6.8b). The frequency spectrum of the returned τCs is shown in the periodograms on the left-hand side.

in neonates. The selection includes periods of stable breathing at low (Figure 6.11a) and high (Figure 6.11d) respiratory rates, as well as periods with changing respiratory dynamics (Figures 6.11b and 6.11c).

6.2.5 Discussion

The temporal structure of quasi-periodic signals, such as the respiratory signal, is rich in information. We proposed a novel source separation method τCA to recover these signals which incorporates this feature (periodicity) into the definition of the cost function Q. In essence, similarly to many algebraic methods for PCA or ICA, τCA consists in the diagonalisation of a set of matrices containing second or higher-order statistics of the dataset. We further adapted this method so as to make it robust to phase differences between the input signals and amenable to computation. We tested this method on artificial and real data.

Synthetic data

In the examples shown, the denoising properties of τCA were confirmed. The advantages of using τCA over conventional methods with regards to dimensionality reduction were also demonstrated: the latent periodical signal seen across several input channels was
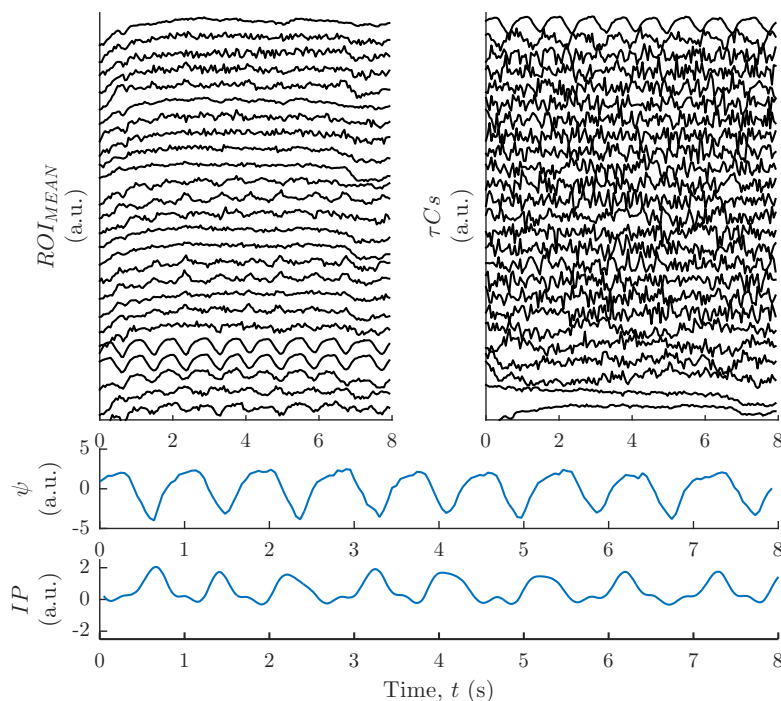
Figure 6.9

Components returned after τCA was applied to a 8 s window of neonatal video data. The extracted respiratory rhythm ψ for this segment and the IP waveform acquired using the chest electrodes are also shown.

137

(a) $s_1(t)$ (b) $PSD_{s_1}(f)$
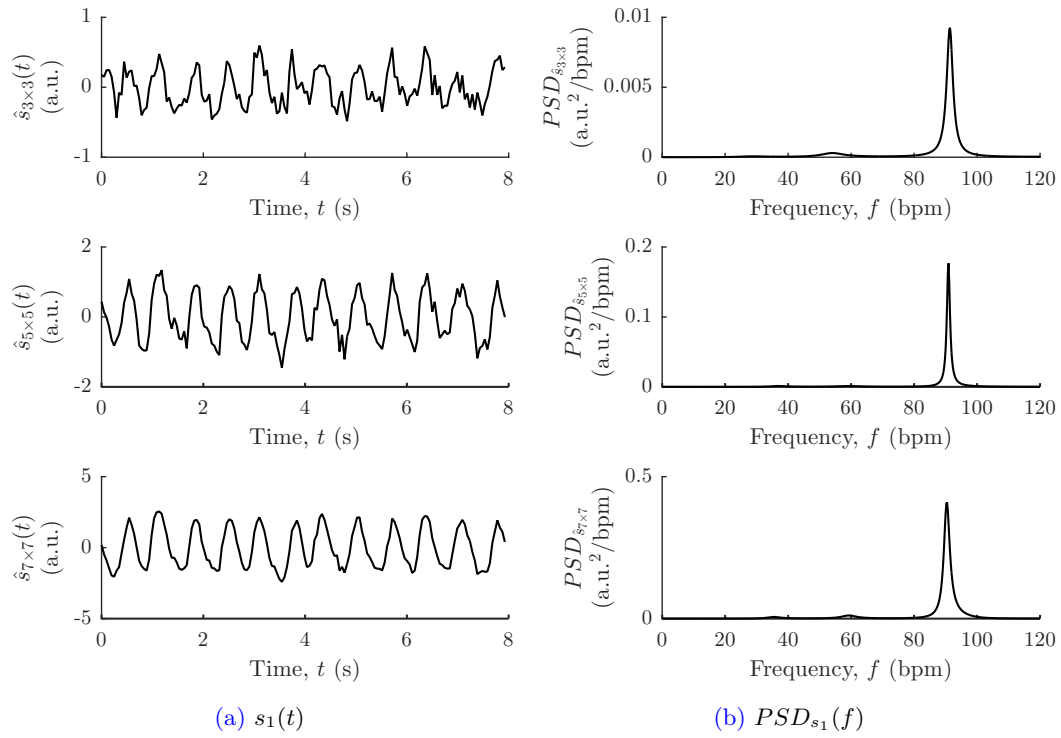
Figure 6.10
;

Estimated source s_1 returned after τCA was applied to a 8 s window of neonatal video data when skin area is divided into subregions with decreasing pixel sizes (3×3, 5×5 or 7×7 ROI divisions). Both (6.10a) time and (6.10b) frequency domain representations are shown. The power spectral density plots were obtained using the Burg method to estimate the parameters of an 9th order auto-regressive model.
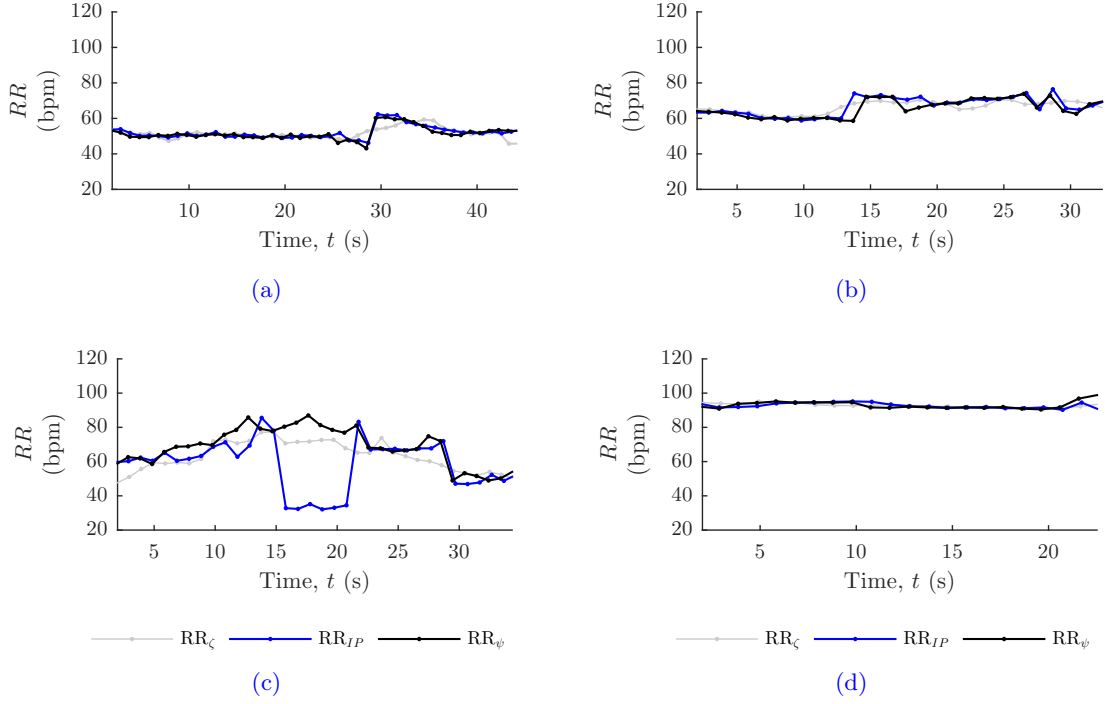
Figure 6.11
Respiratory rates estimates from manual(RR_ζ), IP (RR_{IP}) and camera signals (RR_ψ) during four sample periods of manual breath counts in the NICU dataset.

returned as a single component (Figures 6.5 and 6.7). In addition, as this component arises by design, we did not need to select it from the set of output signals using criteria defined *a posteriori* (*e.g.* to rank the components by SNR, periodicity or spectral power) as is often the case when BSS methods are used to extract physiological rates from video data, as ICA [Poh et al., 2010, 2011] or PCA [Lewandowska, 2011].

As evidenced in Figures 6.2 and 6.3, there were noticeable phase differences between the multivariate input signals produced by either the phase-shifted sinusoids or the simplified breathing model, which have been introduced in our design of these two sets. As illustrated in Figures 6.6 and 6.8, the use of analytic signals as inputs accounts for the presence of phase lags in the cases shown. These results provide a good basis for the use of analytic inputs to τCA (or to linear source separation techniques) when phase lags are suspected in the observed signals.

Neonatal data

A key aspect of remote photoplethysmography is the size of the measurement area. Tarassenko et al. [2014] have shown that heart rate can be extracted from larger regions of interest (150×150 pixels when the camera is placed approximately at 1 m distance from the subject face), whereas changes in the reflected light intensity as a result of breathing are more prominent in smaller regions of interest (with pixel sizes as low as 10×10 pixels). Thus, the size of the ROI determines the relative influence of pulse

and respiratory modulations in the signal. Similarly, in Figure 6.10b we observed that as smaller ROI sizes are used, the respiratory modulation of $\hat{s}_1$ is more pronounced as evidenced by the increased magnitude of the PSD spectra at the respiratory rate frequency.

The results of frequency analysis of the extracted respiratory component ψ was shown in Figure 6.11 for four data segments. These examples reveal that, through the use of a simple method for RR derivation, we obtained camera-derived RR values which follow the manually-derived rates with considerable accuracy, even without resorting to quality metrics, not only at low respiratory rates (Figure 6.11a or Figure 6.11b) but also at high respiratory rates (Figure 6.11d). In Figure 6.11c, we highlight an episode where the RR_{IP} values returned using the conventional methods to monitor this vital sign falsely report a decrease in respiratory rate, whilst RR_ψ and RR_ζ are in agreement ($\pm$ 10 bpm). In the next chapter we compare the performance of camera- and IP-based methods, as well as of previously developed methods to derive RR from the PPG and ECG waveforms, in a more systematic manner.

6.3 Activity signal from video data

We define the movement artefact in the video sequence, such as that caused by spurious patient movement, nursing interventions, or maternal touch, as the activity signal $\alpha(t)$.

6.3.1 Method

A straightforward way to detect the presence of movement artefact is to monitor the high frequency content in raw pixel intensities. We started by obtaining the value of the blue colour channel $I_B(t)$ over the bounding box B_{skin} containing the frame area occupied by the monitored infant at time t. B_{skin} was obtained using the methods described in Chapter 5 and updated at 1-minute intervals.

The blue channel was selected due to the low penetration depth of light in skin over this wavelength range [Cui et al., 1990]. As blue light is more affected by specular reflection than other colour channels, the intensity signal collected at this wavelength is much less modulated by the BVP, thus rendering it a good proxy for subject motion.

This signal was then processed using a high-pass filter designed to attenuate breathing motion (3rd order Butterworth; frequency cut-off at 1.6 Hz); this cut-off frequency approximates the maximum of the neonatal range of respiratory frequencies (100 bpm). A frame-wise measure of activity $\alpha^*(t)$ was obtained as the sum of the squared filtered pixel intensities I'_B as

$$\alpha^*(t) = \sum_{x,y \in B_{skin}} I'_B(x,y,t)^2 \tag{6.16}$$

which takes on a value for every frame t. Finally, this series was filtered using a moving 5-point average filter to achieve a smoother description of activity $\alpha(t)$. In the computation of $\alpha(t)$, 8-bit precision was used in all processing steps ($\alpha \in \{0, 2^8 - 1\}$).

6.3.2 Results

Figure 6.12 shows a manually-annotated 55-minute segment of neonatal video footage. This segment was annotated by a human observer, who was asked to identify periods of (a) upper limb, (b) lower limb or (c) torso movement, as well as (d) nursing interventions. This was achieved using a graphical interface where the annotator could indicate the time instant (to the nearest minute) of the onset and termination of each motion event identified. During the annotation process, the video was played at three times its original speed, at around 60 Hz. Whilst necessary to sustain attention levels, this approach greatly reduces the time resolution of the label timestamps. The labels (a) - (d) were merged and are shown in Figure 6.12c.

We used this segment to validate our algorithm for the detection of global motion in the video data. Figure 6.12a shows camera-derived measures of activity α(Figure 6.12a). We observe that motion events manifest as perturbations in this signal. When compared to the labels produced by thresholding α at $\alpha > 0$ (Figure 6.12b), a good agreement is seen between manual and camera-derived labels although, as expected, the time resolution of the labels obtained manually is comparatively lower.

(a)

(b)

(c)
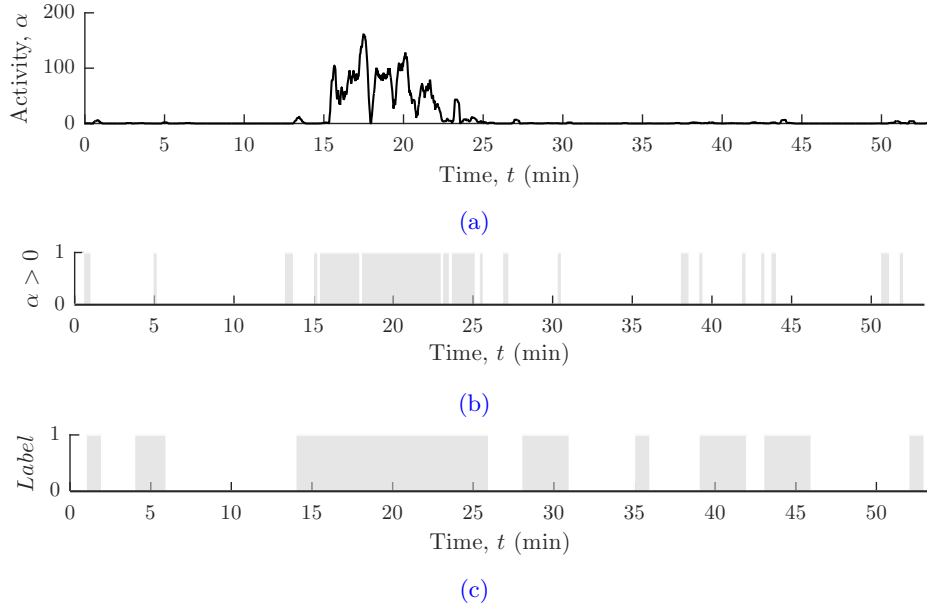
Figure 6.12

Segment of camera-derived infant activity α (6.12a) and manual labels. Binary labels for instants where $\alpha > 0$ are also shown (6.12b) to facilitate comparison with the manual labels in (6.12c).

6.3.3 Discussion

We believe the aforementioned method presents a simple measure of activity. Of particular interest is its low computational cost; as the computation of α relies on the computation of raw frame differences and the use of a low order digital filter, the method can be executed in real-time. Furthermore, the use of an high-pass filter ensures its robustness to slow-varying changes in the illumination environment.

The lack of available reference activity data poses a clear limitation to the development of novel methods for activity detection. Since patient activity data is only relevant on short time-scales, labelled data needs to be obtained with high time resolution (under a second) for it to be useable for the training and testing of these methods. This renders the labelling process extremely strenuous for human annotators. It is for this reason that only one hour-long segment was available for analysis.

As infant activity is expected to impact the quality of the collected physiological signals, improvements on the activity scores reported in this chapter could potentially be achieved if the signal quality metrics defined previously for ECG, PPG and IP were combined into our definition of activity, *e.g.* by using the set of SQI thresholds defined previously for the exclusion of noisy windows from our analysis. The reason this has not been performed in this study is the following. Since this measure of activity will be used in the camera-based detection of adverse events in Chapter 8, we required this parameter to remain independent from other signal sources. This approach guarantees that α could be computed from the video output exclusively, and therefore, the performance of the camera monitor for that classification task could be assessed independently of the input signals to the patient monitor.

6.4 Chapter conclusions

In this chapter, a general framework was presented for the separation of periodic sources from multivariate mixtures. We applied a novel approach, τCA, in which prior knowledge about the periodicity of the source signal is incorporated into a linear model. In simulations on synthetic data, we showed that this source is inferred more accurately as a result of the inclusion of this prior knowledge when compared to commonly applied source separation routines.

While the method is general and may be adapted to different applications, we focused on the extraction of the respiratory waveform from multi-area video recordings. Its behaviour was illustrated on a segment of video data from the neonatal dataset. The effect of increasing the number of ROI subdivisions (and, therefore, the dimensionality of the input) was explored.

A simple method for estimating the presence of movement artefact (activity) is also

presented and demonstrated on a manually annotated segment of video data.

Both algorithms were applied in later chapters for the estimation of the respiratory waveform (Chapter 7) and the detection of cessations of breathing (Chapter 8) from video data of premature infants in the NICU dataset.

Chapter 7

Analysis of respiratory rhythms

Contents

7.1   Chapter summary

The immaturity of the brain mechanisms governing the respiratory rhythm in preterm infants can result in irregular breathing patterns. In neonatal clinical practice, understanding the pathological instabilities of breathing can provide valuable insights into the cardiorespiratory status as well as the maturation level of critically-ill infants.

The respiratory assessment of the neonate involves primarily the inspection of respiratory rate. In addition to respiratory rate, the duration of inter-breath intervals (IBIs), *i.e.* the duration of the time intervals between successive inspiratory (or expiratory) efforts, is also an important measure of respiratory instability and plays an important part in understanding the irregularity in the breathing pattern. As this pattern is highly non-stationary [Frey et al., 1998], with rapid changes in breathing regimes over short periods of times, to date there is no algorithm which can reliably capture the instantaneous

changes in IBI duration [Indic et al., 2013].

This chapter is concerned with the measurement of both these variables in the respiratory rhythms of preterm infants. We compare RR and IBI traces estimated from (a) the chest impedance pneumogram, (b) the ECG-derived respiratory rhythms (EDR and RSA), (c) the PPG-derived respiratory rhythms (RRIV, RIAV and RIFV) and (d) the camera-derived respiration rhythm (ψ) against those measured on the gold standard of breath detection provided by attending clinical staff during manual breath counts (Figure 7.1).

7.2 Extraction of inter-breath intervals

7.2.1 Extraction of IBIs from manual breath counts

As a reference for IBI duration, we computed this quantity from the instantaneous breath information obtained during manual respiratory measurements (Figure 7.1a).

Manual measurements of respiration data were available for study subjects PN 26 - 30 (spread over a total of 14 sessions) in the NICU dataset in the form of logged key presses. The timestamps of key presses during these sessions were parsed into separate counting cycles in the manner described in Appendix B.2. In summary, a total of 209 independent counting cycles were found. Out of these, 39 breath counts were aborted by the clinical staff due large infant movements, wriggling motion, or other causes. Out of the remaining measurements, 5 were not recorded in the camera monitoring time. This left a total of 165 manual breath counts, or 6924 events if individual breaths are considered.

We modelled each counting cycle c as series of n_c breaths defined by a train of spikes $\zeta_c(t)$ at instants, $t_0^c, ..., t_{n_c-1}^c$ which mark the timing of individual breaths[1] in that counting cycle (Equation 7.1).

$$\zeta_c(t) = \begin{cases} 1, & \text{if } 1 \in t_0^c, ..., t_{n_c-1}^c \\ 0, & \text{otherwise} \end{cases} \tag{7.1}$$

The $n_c - 1$ inter-breath intervals, $IBI_\zeta(t)$, were then computed trivially as the time differences between consecutive breaths and assigned the timestamp of the later breath (Equation 7.2).

$$IBI_\zeta(t_n^c) = t_n^c - t_{n-1}^c, n \in \{1, ..., n_c - 1\} \tag{7.2}$$

---

[1] These time instants may correspond to either expiration or inspiration end times. The choice between selecting the end-inspiratory or the end-expiratory pause as the onset of each respiratory cycle was left at the discretion of the nursing staff performing the manual count. Therefore, this choice was consistent within the same manual count but not across counts.
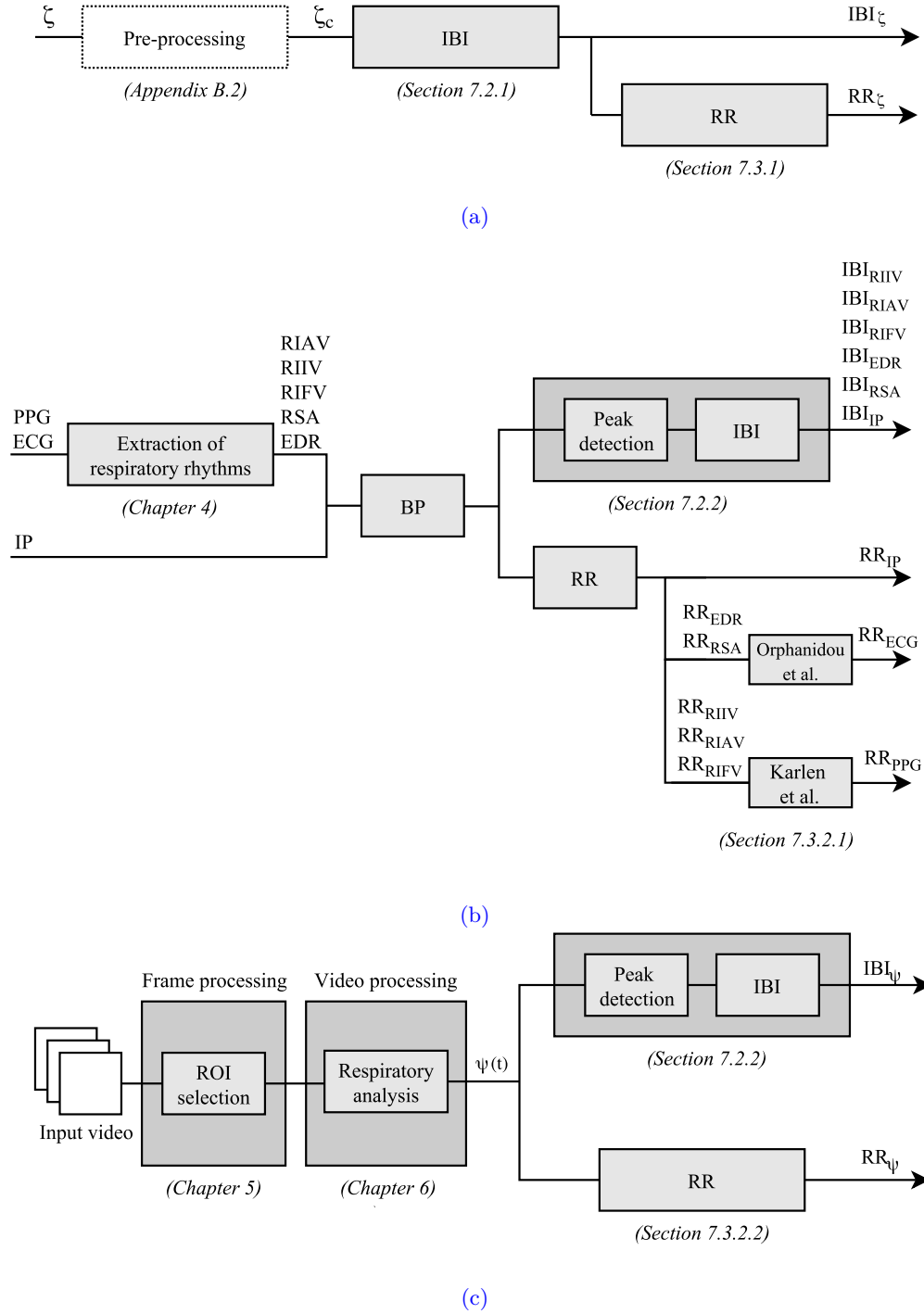
Figure 7.1
Computation of IBI and RR. (7.1a) A manual IBI signal was computed from the manual breath signals ζ_c directly, and these signals were used to derive RR estimates. (7.1b) Surrogate respiratory signals were extracted from ECG and PPG waveforms. The extracted signals and IP waveform were bandpass filtered before IBI or RR were derived. (7.1c) In deriving these variables from video data, the methods presented in earlier Chapters were necessary to produce a respiratory signal ψ from which IBI and RR measurements could be extracted. BP = Bandpass filter.

7.2.2 Extraction of IBIs from derived respiratory rhythms

IBI traces were also computed from respiratory rhythms extracted from video signals and the physiological signals collected routinely in neonatal units as part of standard patient care.

Respiratory signals were derived from IP, ECG and PPG waveforms using the methodology developed in Chapter 4 and from video data using the methodology of Chapter 6. After obtaining respiratory signals from these sources, the time instants of individual breaths were found as the time coordinates of the local maxima in these waveforms as returned by MathWorks [2015]'s *findpeaks*[1]. IBIs were then computed as the time differences between consecutive peaks and assigned the timestamps of the later breath.

7.3 Extraction of respiratory rate

Respiratory rates were calculated over 8-second windows with 1-second shifts. To facilitate comparison with reference manual rates, these windows were defined so that they overlapped with those used in the computation of manual respiratory rates RR_ζ.

7.3.1 Extraction of RR from from manual breath counts

For each counting cycle c, the start of the first analysis window was set to t_0^c. The last window for this cycle was the last one whose right limit was below $t_{n_c-1}^c$. For each time window w, $RR_\zeta(w)$ was computed as the inverse of $\overline{IBI}_\zeta(w)$, the mean value of the IBI_ζ data points within the window and assigned the timestamp corresponding to the centre time point of the window.

7.3.2 Extraction of RR from derived respiratory rhythms

7.3.2.1 from standard monitoring

Respiratory signals were derived from IP, ECG and PPG waveforms using the methodology developed in Section 4.5 of Chapter 4. The pass band of the digital filters was shifted to the frequencies corresponding to 20 and 140 bpm in order to capture the high respiratory rates observed in this cohort. The remaining parameters were used as defined in Chapter 4.

The challenges placed on indirect methods of extracting the respiratory signal are aggravated in real-life monitoring scenarios. For this reason, a post-processing stage,

---

[1] In 1-D data series, *findpeaks* assigns a peak to every data point whose amplitude is greater than its two neighbouring samples. Non-infinite endpoints are excluded. If a peak is flat, *findpeaks* returns the point with the lowest index only [MathWorks, 2015].

whereby simultaneous RR estimates corresponding to different modulations of the same source are fused, has received particular attention (Figure 7.1b).

A fused PPG-derived respiratory rate $RR_{PPG}(w)$ was estimated using the approach in Karlen et al. [2013] as the mean of the three respiratory-induced variations derived from PPG $\{RR_{RIAV}(w), RR_{RIIV}(w), RR_{RIFV}(w)\}$. In compliance with the authors, for time points for which the standard deviation of the three estimates exceeds 4 bpm, no RR estimate was provided.

The ECG-derived respiratory rate $RR_{ECG}(w)$ reported for window w was selected among $\{RR_{EDR}(w), RR_{RSA}(w)\}$, using the pole magnitude criterion [Orphanidou et al., 2013]. The RR estimate was chosen as the frequency of the highest magnitude pole of the two 9th order AR models obtained from auto-regressive spectral analysis of the resampled EDR and RSA respiratory signals.

7.3.2.2 from video data

The camera-derived respiratory signal ψ was obtained over the same time windows using the τCA method described in Chapter 6.2 (Figure 7.1c). Similarly to the example given in Section 6.2.1, we used a collection of signals derived from pixel averaging over image subregions as the input to the τCA routine. Here we define this process more rigorously and introduce a multi-scale approach.

Consider a set of d-bit RGB images acquired during the analysis window w at the sampling rate of the imaging device. Each digital image (*i.e.* a frame) is a two-dimensional function of the form $I(x, y) : \mathbb{N}^2 \mapsto \mathbb{N}^3$, where (x, y) are spatial coordinates. Its output is an array of three $R \times C$ matrices containing the pixel intensity for each colour channel, where R and C are the dimensions of the frame in number of rows and columns and the intensity value is a positive integer of the set $\{0, 1, \ldots, 2^d - 1\}$.

An RGB video is then a function described as

$$I(x, y, t) = [I_R(x, y, t) \quad I_G(x, y, t) \quad I_B(x, y, t)] \tag{7.3}$$

where I_R, I_G, and I_B are the matrices of the three colour channels and t is the discrete time index .

As shown in Figure 7.1c, the first step in our processing is to decompose the first frame of window w, $I(x, y, t)$, into its components in the *L*a*b** colour space and obtain the bounding box B_{skin} containing the largest skin cluster segmented from the initial frame at t_0 (*i.e.* at the start of the manual counting cycle) according to the methods described in Chapter 5.

This rectangular area is then partitioned into $n \times n$ non-overlapping blocks. Let $B_1, B_2, \ldots, B_{n \times n}$ be a partition of B_{skin} into a grid of $n \times n$ blocks such that

$$B_1 \cap B_2 \cap \cdots \cap B_{n \times n} = \emptyset,$$
$$B_1 \cup B_2 \cup \cdots \cup B_{n \times n} = B_{skin}. \tag{7.4}$$

This process is repeated for $n \in \mathcal{N}$ levels, so that at each iteration all n_w blue frames in the window w, $I_B(x, y, w; t), t \in \{t_1^{(w)}, \ldots, t_{n_w}^{(w)}\}$, are decomposed into a set of $n \times n$ images each representing B_{skin} at a different resolution. This is a multi-scale approach; with ascending n, we obtain a representation of B_{skin} at higher resolution and with finer detail. The blue channel was selected for the same reasons presented previously to justify its use in the activity detection algorithm of Section 6.3, *i.e* of the available wavelengths, blue light has the shortest penetration depth in human skin, and thus the signal reflected at this wavelength is a good indicator of subject motion.

Each level n, contributes to the formation of the input signal $\boldsymbol{a}(t)$ to the τCA routine in the following manner:

$$a_i^{(n)}(w; t) = \sum_{x,y \in B_i} I_B(x, y, w; t), \quad i = 1, \ldots, n \times n \tag{7.5}$$

For each n, the τCA algorithm was then applied to

$$\boldsymbol{a}^{(n)}(w; t) = \left[a_1^{(n)}(w; t), \ldots, a_{n \times n}^{(n)}(w; t)\right]^T, n \in \mathcal{N} \tag{7.6}$$

to yield a candidate respiratory signal $\psi^{(n)}(w; t)$ for window w and level n.

Candidate respiratory rates were then calculated from the resulting $\psi^{(n)}(w; t), n \in \mathcal{N}$ signals using the standard methods for RR extraction from respiratory signals described in Chapter 4 (resampling at 4 Hz; 9th order AR model). The camera-derived rate for that window was then selected as:

$$RR_\psi(w) = median\left(RR_\psi^{(n)}(w)\right), \forall n \in \mathcal{N} \tag{7.7}$$

and the true respiratory signal $\psi(w; t)$ given by

$$\psi(w; t) = \psi^{(r)}(w; t),$$
$$r \in \mathcal{N} : \forall_{c \in \mathcal{N}} \|RR_\psi^{(c)}(w) - RR_\psi(w)\| \geqslant \|RR_\psi^{(r)}(w) - RR_\psi(w)\| \tag{7.8}$$

Based on the analysis of Figure 6.10 of Chapter 6, we have used $\mathcal{N} = \{3, 5, 7, 9\}$ for the estimation of $\psi(w; t)$ from neonatal video data.

7.4 Signal quality indices

7.4.1 SQI for physiological signals

The methods used for the assessment of signal quality have been reviewed in Section 4.4. Here, they are applied as implemented with the exception of the following parameters, which were adjusted to address the unique physiology of the neonatal population.

- In the computation of SQI_{PPG}^b (Section 4.4.2; Equation 4.9), the minimum and maximum beat-to-beat interval used by Li and Clifford [2012] in the segmentation of this waveform were shifted to 0.5 and 1.5 s, that is, the inverse of the limits of the neonatal respiratory range ($[40; 120]$ bpm);

- In the computation of the location of the R-peaks using the *wqrs* open-source QRS detector, we shortened the refractory period used by Zong [2003] to half this value (125 ms);

- When assessing the quality of the single-lead ECG signal SQI_{ECG}^b (Section 4.4.1.1, Equation 4.2), a shorter tolerance interval $\xi = 50$ ms was adopted for computing the ratio of beats detected synchronously by both QRS detection algorithms.

7.4.2 SQI for camera-derived signals

We obtained a spectral-based measure of signal quality for the camera-based respiratory signal $\psi(w; t)$ derived for window w as

$$SQI_\psi(w) = \int_{RR_{min}}^{RR_{max}} \mathcal{P}_w^\psi(f) df \Big/ \int \mathcal{P}_w^\psi(f) df \qquad (7.9)$$

where $\mathcal{P}_w^\psi$ is the power spectral density of $\psi(w; t)$ estimated using the auto-regressive methods described in Section 4.6.2. The intervals of integration are the neonatal respiratory band ($[RR_{min}, RR_{max}] = [40, 120]$ bpm) for the numerator, and the range of frequencies for which $\mathcal{P}_w^\psi > 0$ for the denominator. In practice, this range is $[0, f_s/2]$, where f_s is the sampling frequency of the signal for which $\mathcal{P}_w^\psi$ is computed. In our case, this upper limit is 2 Hz = 120 bpm due to the 4 Hz resampling prior to the auto-regressive model fit.

Intuitively, $SQI_\psi(w)$ is given by the proportion signal power contained in the respiratory frequencies. Given our definition of respiratory as a sinusoidal waveform whose fundamental frequency is the respiratory rate (Section 4.5), a window with a clean ψ rhythm should have $SQI_\psi(w)$ close to 1.
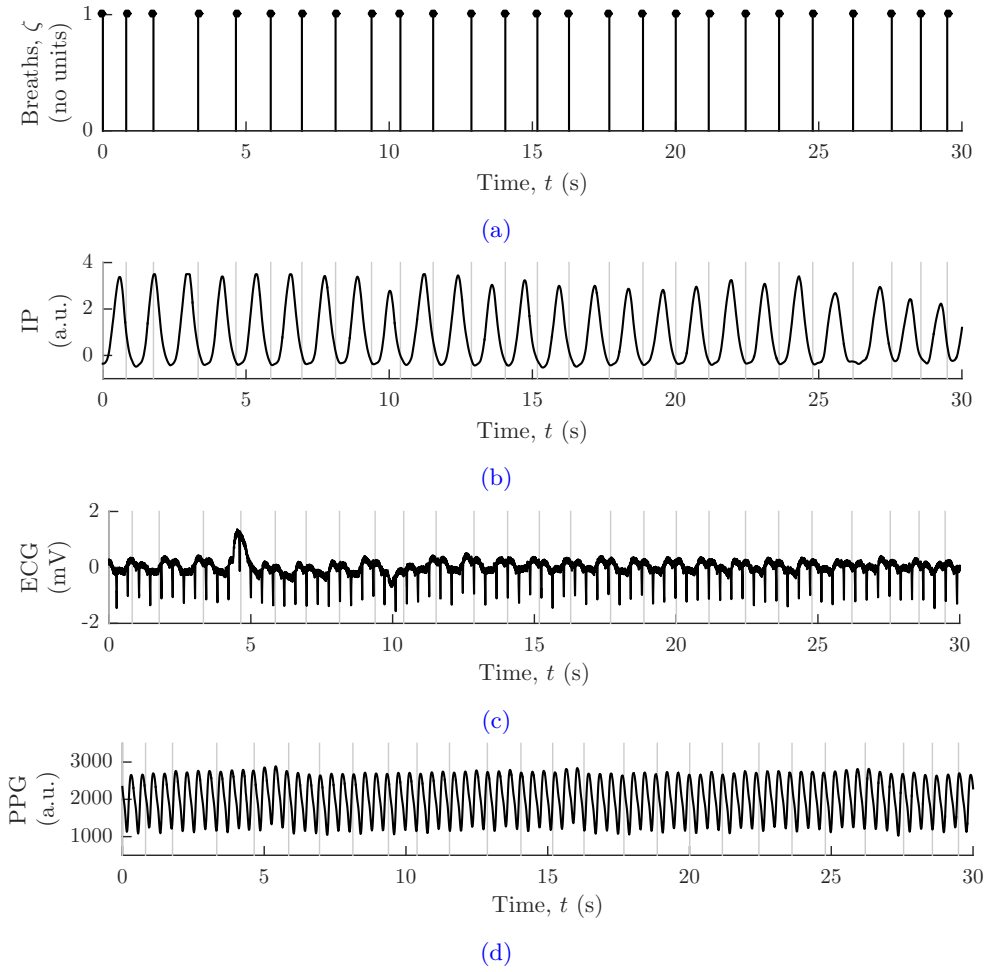
An example of the modulation of IP, ECG and PPG signals with the respiratory cycle during a manual counting cycle on PN 28. Modulation through 25 complete respiratory cycles is shown. The manually-counted breaths on the upper trace are reproduced as grey vertical lines on the subsequent plots.

7.5 Results

Figure 7.2 shows the respiratory-induced modulation of IP, ECG and PPG signals through a selection of 25 complete respiratory cycles. The modulation of the R-peak and the changes in IP with ventilatory effort are clearly seen. Less evident in this segment is the modulation of PPG with respiration.

Figure 7.3 shows the distribution of respiratory rates and inter-breath intervals in the manual counting cycles. The respiratory rates presented in Figure 7.3b were computed for each manual count as the inverse of the median $\mathrm{IBI}_\zeta(t_n^c) = t_n^c - t_{n-1}^c$ during that counting cycle c. In the remainder of this section we compare manual measurements of IBIs (Section 7.5.1) and RR (Section 7.5.2) against estimates of these variables derived from camera and physiological signals.
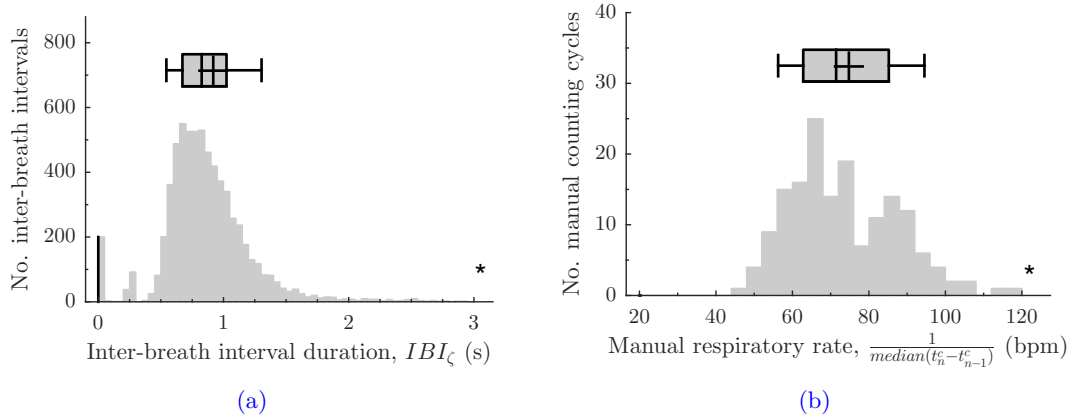
Figure 7.3

Histograms of (7.3a) inter-breath intervals IBI$_\zeta$ and (7.3b) respiratory rates RR$_\zeta$ for the manual counting cycles in PN 26 - 30. + is the data mean, the boxplot bounds the 25% and 75% quartiles and the whiskers bound the 9% and 91% of the data.

7.5.1 Analysis of inter-breath intervals

7.5.1.1 Spectral coherence analysis

We used coherence analysis (a review of which can be found in Appendix C.1) to assess the similarity between the IBIs computed from the manual reference signal and from the extracted respiratory signals.

As it is a requirement for coherence analysis that the two time series be regularly sampled and at the same time instants, the series of manual and derived IBIs were linearly interpolated at a frequency of 4 Hz (or at evenly spaced intervals of 250 ms) between the time instants corresponding to the first and last manual IBI points of each manual count (t_1^c and $t_{n_c-1}^c$).

For assessing the reliability of the cross-spectral estimates, we proceeded in a similar manner to that presented in Chapter 4; we compared the coherence values for each manual count against a significance threshold $\gamma_{th,\alpha}^2(f)$, such that for a frequency bin k, a coherence greater than $\gamma_{th,\alpha(f)}^2$ means that the null hypothesis that the reference IBI and the IBI derived from a given respiratory signal (*i.e.* IP, RIIV, RIAV, EDR, RSA) are not related at that frequency can be rejected at a significance level α (for $p < 0.05$).

In Figure 7.4, the results of IBI computation are shown for the 30-second signal segment of data in Figure 7.2. To estimate the amount of overlap in the spectral structure of the IBIs derived from the extracted respiratory signals and the manual IBIs, we have computed their spectral coherence on a window-by-window basis over the 165 manual counting cycles. For each window, we report spectral coherence as the integral of the statistically significant magnitude-squared coherence values over the range of breathing frequencies (40 - 120 bpm). The results are quantified in Table 7.1 in terms of mean and standard deviation.
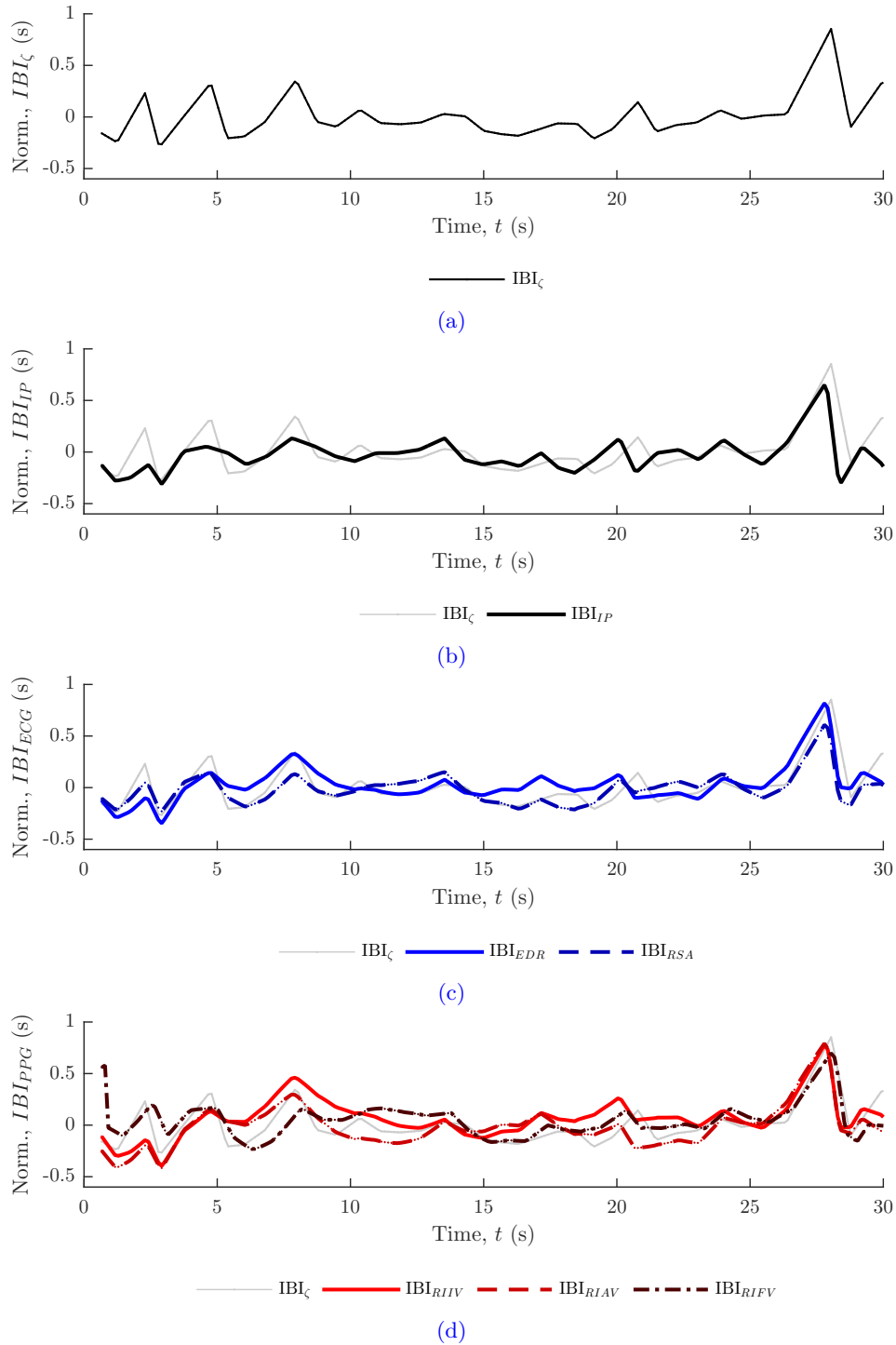
Figure 7.4
An example of detrended IBI measurements from IP, ECG and PPG signals. 30-second excerpt from a manual counting cycle on PN 28. All IBI traces have been linearly interpolated at 4 Hz so as to obtain simultaneous estimates of IBI.

Sum of significant magnitude-squared coherence (SMSC) values between derived IBIs and IBI_ζ. The mean($\pm$ standard deviation) shown was computed over the 165 manual counting cycles.

	Sources					
	ECG		PPG			IP
	IBI_{EDR}	IBI_{RSA}	IBI_{RIIV}	IBI_{RIAV}	IBI_{RIFV}	IBI_{IP}
SMSC (a.u.)	4.9($\pm$1.6)	2.9($\pm$4.4)	2.2($\pm$2.7)	4.0($\pm$2.8)	2.1($\pm$4.4)	8.3($\pm$1.4)

7.5.2 Analysis of respiratory rates

The distribution of respiratory rates as estimated from all signal sources for the periods of manual breath counts are shown in Figure 7.5. The aliasing effect caused by the sampling of the respiratory signal at the heartbeat frequency is evidenced by the failure of techniques for RR estimation from PPG or ECG in detecting accurate RR values above 80 bpm.

7.5.2.1 Respiratory rate from IP

In Figure 7.6, we compare the accuracy of re-derived RR_{IP} and the RR estimates provided by the patient monitor from the same source (RR_{MON}). By default, the monitor estimates are given at 1-second intervals. For comparison, these were interpolated at the time instants of the RR_{IP} data points.

We observe that the MAE of both RR_{IP} and RR_{MON} estimates decrease monotonically when only estimates from windows with progressively higher thresholds in SQI_{IP} are considered. In spite of the increase in accuracy achieved through SQI thresholding a considerable difference in the error performance between the two methods persists, with MAE(RR_{IP}) = 6.3 bpm and MAE(RR_{MON}) = 11.9 bpm at $SQI_{IP} \geqslant 0.75$. At this SQI threshold, respiratory rates were computed for 67.3% of the IP windows.

7.5.2.2 Respiratory rate from indirect methods

Figures 7.7 and 7.8 illustrate the reduction in the MAE error for the indirect methods of RR estimation from the ECG and the PPG respectively. For reasons which will be explained later in the chapter, this analysis is limited to estimates with $RR_\zeta \leqslant 80$ bpm.

RSA-derived RR was shown to be less accurate than EDR-derived estimates. This relationship was maintained at all SQI_{ECG} levels. RR_{EDR} and RR_{ECG}, selected using the pole magnitude criterion, show similar performance.

With respect to the PPG-derived rates, a substantial reduction in MAE was seen across all modulations of this waveform for stricter thresholds on SQI_{PPG}. Of the three sources of modulation of the PPG waveform, the lowest error rates were consistently seen in RIIV estimates (e.g. MAE(RR_{RIIV}) = 8.3 bpm at $SQI_{PPG} \geqslant 0.85$) followed
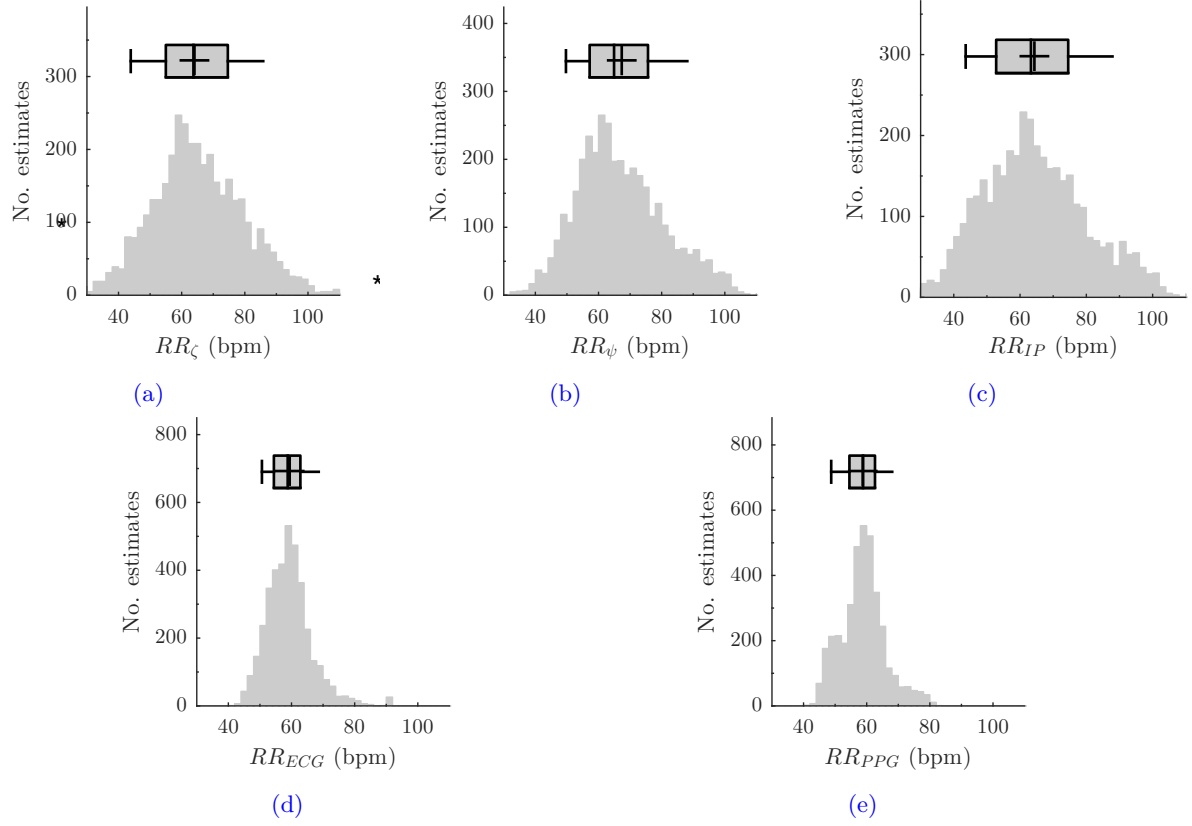
Figure 7.5

Distribution of respiratory rate estimates from periods of manual breath counts. Histograms are shown for (7.5a) manual, (7.5b) camera, (7.5c) IP-, (7.5d) ECG-, and (7.5e) PPG-derived rates. Asterisks represent outliers.
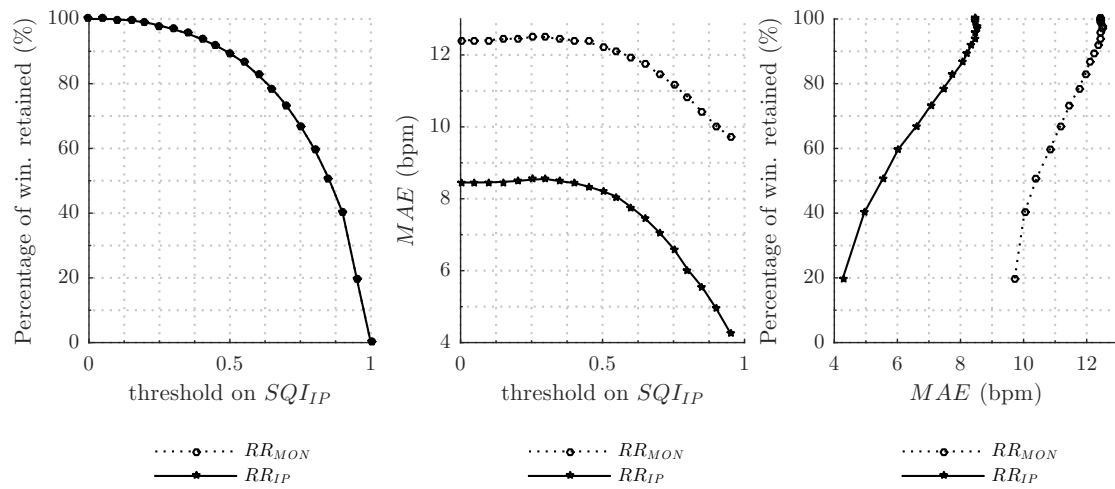
Figure 7.6

Relationship between the threshold on SQI_{IP}, the percentage of windows retained and the mean absolute error (MAE) of retained RR_{IP} and RR_{MON} estimates against RR_ζ during manual breath counts. RR_{MON} are the RR estimates provided by the patient monitor.
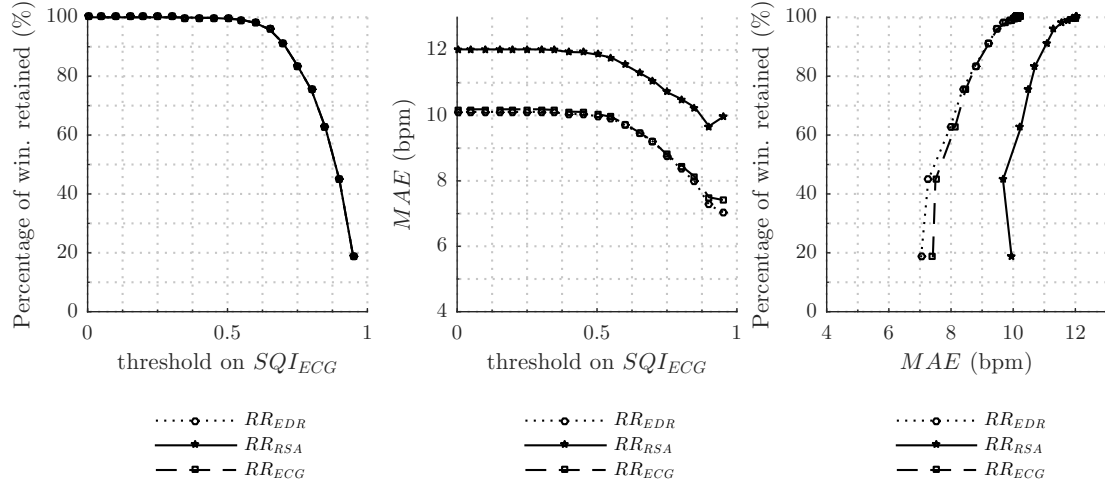
Figure 7.7
Relationship between the threshold on SQI_{ECG}, the percentage of windows retained and the mean absolute error (MAE) of retained RR_{EDR}, RR_{RSA} and RR_{ECG} estimates against RR_ζ for $RR_\zeta \leqslant 80$ bpm during manual breath counts.
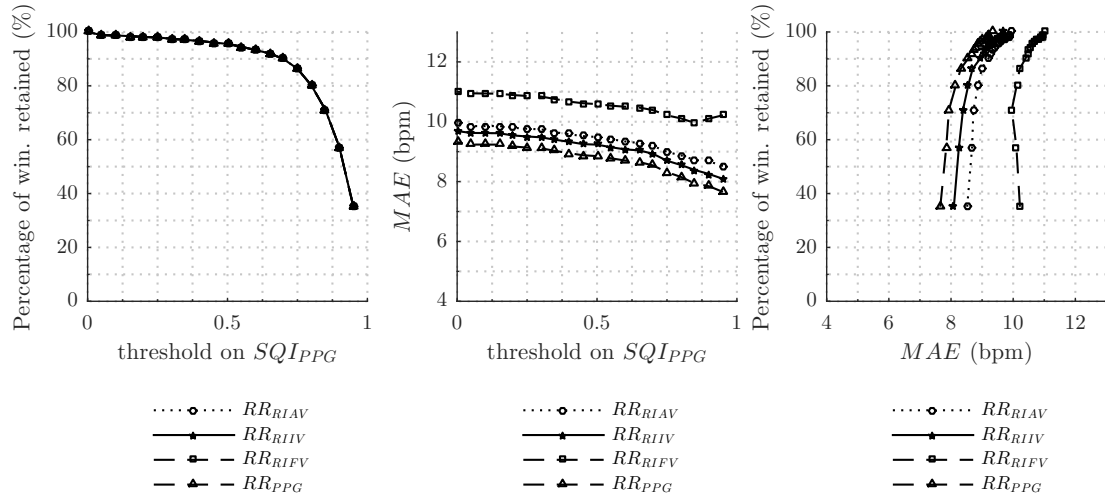
Figure 7.8
Relationship between the threshold on SQI_{PPG}, the percentage of windows retained and the mean absolute error (MAE) of retained RR_{RIIV}, RR_{RIAV}, RR_{RIFV} and RR_{PPG} estimates against RR_ζ for $RR_\zeta \leqslant 80$ bpm during manual breath counts.

closely by RIAV estimates (*e.g.* MAE(RR_{RIAV}) = 8.7 bpm at $SQI_{PPG} \geqslant 0.85$). Both are appreciably more accurate than frequency-modulated estimates (*e.g.* MAE(RR_{RIFV}) = 9.8 bpm at $SQI_{PPG} \geqslant 0.85$). As seen in RR_{ECG}, fusion techniques enabled a reduction in MAE. Using Karlen et al. [2013]'s criteria, a reduction in MAE of approximately 0.5 bpm in relation to RR extracted from the best-performing source of modulation (RR_{RIIV}) is seen at all SQI thresholds.
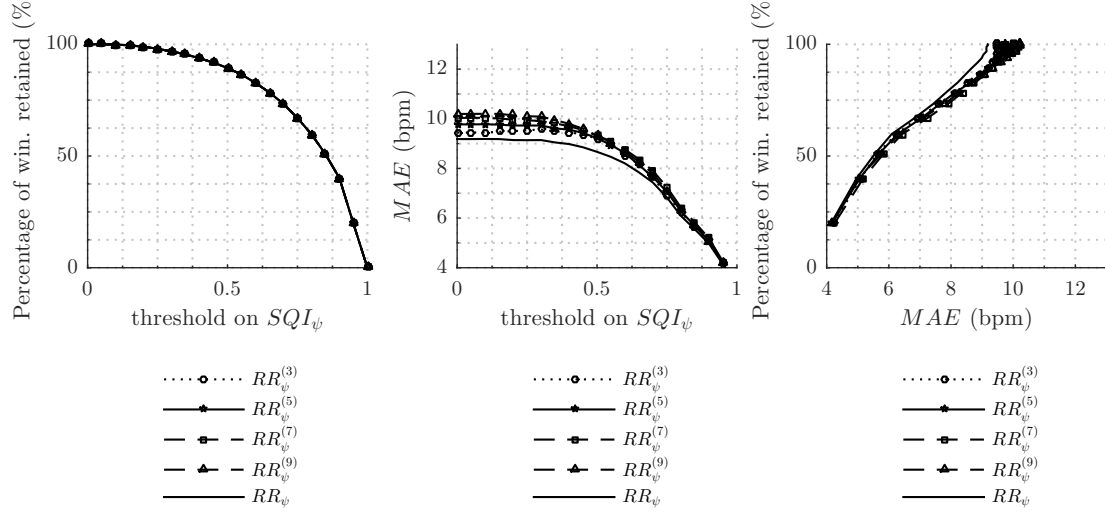
Figure 7.9

Relationship between the threshold on SQI_ψ, the percentage of windows retained and the mean absolute error (MAE) of retained $RR_\psi^{(3)}$, $RR_\psi^{(5)}$, $RR_\psi^{(7)}$, $RR_\psi^{(9)}$ and RR_ψ estimates against RR_ζ during manual breath counts. $RR_\psi(w) = median\left(RR_\psi^{(n)}(w)\right), \forall n \in \{3, 5, 7, 9\}$.

7.5.2.3   Respiratory rate from ψ

In Figure 7.9 we show the relationship between the accuracy of derived RR and the percentage of windows retained at different image scale levels $RR_\psi^i, i \in \{3, 5, 7, 9\}$. From analysis of these plots, it is clear that the SQI_ψ defined is successful in reducing the MAE of camera-derived RR estimates at all resolution levels. As expected, this improvement in performance is achieved at the expense of a reduction in coverage probability. In addition, a considerable reduction in MAE is achieved when the median RR for each window w is taken across the scale space (RR_ψ). This behaviour is maintained even when SQI_ψ is thresholded at progressively higher values.

7.5.2.4   Error analysis

The time series of RR estimates was linearly interpolated at the time instants of RR estimates from the reference measurements and mean (ME), standard deviation (SE), and mean absolute (MAE) error values were calculated over all manual counts.

Previous papers on respiratory rate estimation from the ECG have reported suboptimal performance at high respiratory rates [Orphanidou et al., 2013]. Thus, the performance of the RR estimation methods was assessed by calculating error values for two different frequency ranges: (a) $RR_\zeta < 80$ bpm (84.7% of windows) and $RR_\zeta \geqslant 80$ bpm (15.4% of windows). The results are summarised in Table 7.2.

Only those windows for which the raw signal was reliable (valid windows) were retained for RR extraction from their respective waveform. This was achieved by setting the following conditions: $SQI_{ECG} \geqslant 0.9$, $SQI_{PPG} \geqslant 0.9$, $SQI_\psi \geqslant 0.9$, $SQI_{IP} \geqslant 0.9$,. As a

Summary of error analysis for the algorithms for respiratory rate extraction from ψ, IP, ECG and PPG-derived respiratory signals. The errors were computed with respect to the reference signal RR_ζ. All the figures are shown in breaths·min^{-1}.

Error metrics	Sources				
		Direct methods		Indirect methods	
	RR_ψ	RR_{IP}	RR_{MON}	RR_{ECG}	RR_{PPG}
All RR_ζ range					
$ME(\pm SD)$	-0.0($\pm$7.8)	-0.8($\pm$6.4)	-3.1($\pm$12.7)	-7.2($\pm$12.5)	-5.5($\pm$16.9)
MAE	5.0	4.0	8.7	10.2	11.8
$RR_\zeta < 80$ bpm					
$ME(\pm SD)$	0.5($\pm$7.7)	-0.3($\pm$5.6)	-2.2($\pm$12.3)	-4.0($\pm$11.9)	-0.3($\pm$11.9)
MAE	4.7	3.6	8.2	7.5	7.8
$RR_\zeta \geqslant 80$ bpm					
$ME(\pm SD)$	-4.9($\pm$15.2)	-6.0($\pm$18.0)	-14.5($\pm$19.1)	-31.7($\pm$12.6)	-31.2($\pm$12.3)
MAE	9.3	10.7	16.9	31.7	31.2

result, 50.1% of the data was used in validating ECG-derived estimates, 54.7% for PPG-derived estimates, 54.7% for IP-derived estimates, and 73.2% for ψ-derived estimates.

The SQI thresholds set above for retaining windows of good signal quality could be altered so as to achieve lower MAE rates or a higher percentage of windows included in the analysis, and thus effective RR monitoring time. Figures 7.6 through 7.9 provide the relationship between these variables for each of the signal sources.

7.6 Discussion

The fast-changing respiratory rates observed in the preterm infant mean that this vital sign can change considerably over the course of the time interval conventionally allocated to manual breath counts (*i.e.* 60 seconds). We felt that a comparison of the RR values computed at the end of each counting cycle would fail to capture the full dynamic range of this vital sign and convey the false impression that this signal remains approximately constant over such a period. Thus, in this analysis, we opted instead to segment these periods into shorter time windows of 8 seconds and perform a comparison of the respiratory rates computed over these 8-second windows. To allow for this analysis, the timestamps of individual breaths were recorded. This constitutes a small change to the standard procedure for recording manual counts and thus, respiratory measurements recorded in this way were only available for the last five infants recruited.

The median respiratory rate (over manual counts) was found to be 61 bpm (Figure 7.3b), with a range extending from 44 to 120 bpm. These figures validate our choice of pass band for the digital filters used for RR derivation (20 - 140 bpm). When compare this distribution against the distribution of rates reported by the patient monitor for the entirety of the study period (shown in Figure 3.8b and parametrised in Table 3.3), we

observe that (a) the median rate obtained during the manual measurements is moderately higher than the median RR_{MON} reported by the patient monitor, and (b) breathing rates in the 20 - 40 bpm range have not been sampled during the manual breath counts. This was anticipated as this range comprises abnormally low rates that should prompt clinical inspection, and therefore, were not recorded under the stable conditions under which manual counts were scheduled.

We note also that the range of resting respiratory rates detected is much wider than that found in adults (12 - 20 bpm; Table 2.1). The immaturity of the respiratory centres in the premature infant leads to frequent periods of fast shallow breathing, and hence to fast-changing respiratory rates in these infants. This renders the problem of accurate RR estimation much more challenging in this population.

Errors associated with direct methods

We have re-derived RR from the raw IP signal using standard algorithms for RR extraction found in the literature. It was surprising to note the lower error rates of the RR estimates derived in this manner in relation to those provided by the patient monitor during manual breath counts (Figure 7.6); with $\text{MAE}(RR_{IP}) = 3.6$ bpm *versus* $\text{MAE}(RR_{MON}) = 8.2$ bpm for $RR_\zeta < 80$ bpm and $\text{MAE}(RR_{IP}) = 10.7$ bpm *versus* $\text{MAE}(RR_{MON}) = 16.9$ bpm for $RR_\zeta \geqslant 80$ bpm. This result suggests that there is room for improvement for the manufacturer's built-in software for monitoring respiratory rates in premature infants.

Errors associated with indirect methods

In this chapter, two data fusion methods have been used, both of which estimate the respiratory rate from waveforms derived from amplitude and frequency modulations of the ECG and PPG waveforms. In both cases, the mean absolute error was reduced as a result of the use of the data fusion approach (Figures 7.7 and 7.8).

Errors associated with camera-derived respiratory rates

Comparison of the ψ-derived RR estimates shows that performance was comparable across several image scales, as can be observed in Figure 7.9. There is an inverse relationship between the number of ROIs used n and their size; with ascending n each of the $n \times n$ ROIs will be smaller in pixel size. Only a modest decrease in performance was seen for smaller ROIs, which is presumably caused by the reduced number of pixels per subregion, leading to a lower SNR. The best performing ψ-derived method is given by the median RR_ψ. Compared to IP-derived RR_{IP}, this method showed similar performance, with overall $\text{MAE}(RR_\psi) = 5.0$ bpm *versus* $\text{MAE}(RR_{IP}) = 4.0$ bpm; the fused RR_ψ method showed marginally worse performance than RR_{IP} for $RR_\zeta < 80$ bpm, with $\text{MAE}(RR_\psi) = 4.7$ bpm *versus* $\text{MAE}(RR_{IP}) = 3.6$ bpm, yet slightly better performance for $RR_\zeta \geqslant 80$

bpm, with $\mathrm{MAE}(RR_\psi) = 9.3$ bpm *versus* $\mathrm{MAE}(RR_{IP}) = 10.7$ bpm. This behaviour was expected for chest motion-artefacts which corrupt the IP waveform are naturally expected to disrupt the visual cues upon which the video signal relies to detect breathing movement.

Errors associated with different respiratory ranges

We calculated the RR estimation errors for two frequency ranges of respiratory rates. The lowest errors were found for $RR_\zeta < 80$ bpm. It is clear from the analysis of Table 7.2 that none of the algorithms evaluated based on indirect methods could detect rates above this value (*i.e.* for $RR_\zeta \geqslant 80$ bpm, $\mathrm{ME} \simeq \mathrm{MAE}$ values are seen for both ECG and PPG-derived methods).

Both indirect methods suffer from one caveat. As the respiratory wave is only sampled once per cardiac beat, in order for these methods to reveal the fundamental frequency of the respiratory signal, this frequency cannot exceed half of the instantaneous heart rate during the same time epoch. As we have seen in Figure 3.8a, the HR of the study cohort has a bell-shaped distribution with a median of 157 bpm (Table 3.3). Thus, in instances where the respiratory frequency exceeds half this frequency (approximately 80 bpm), the extraction of RR through these methods is likely to be affected by aliasing. As shown in Figure 7.10, this was a common occurrence during the periods of manual counts.

Errors associated with amplitude-based respiration sources

We found that MAE errors were lower for RR_{IP} when compared to amplitude-based signals, *e.g.* RR_{RIAV}, RR_{RIIV}, and RR_{EDR} (Figures 7.6 to 7.8). In a clear contrast to the protocol in Chapter 4, where subjects were compliant and asked to keep movements to a minimum, subject motion during this observational study was largely unrestricted. Motion unrelated to respiration but within the range of breathing frequencies causes intensity variations that contaminate the amplitude modulation of the detected signals, which ultimately has a deleterious effect on the reliability of amplitude-based estimates of RR and IBI. This may help explain the poor performance of this class of methods in this study.

Given the arrangement of IP and ECG electrodes, a temporal association in the presence of motion noise between these sources was expected. This was confirmed by the comparable percentages of ECG and IP windows retained (50.1% and 54.7%, respectively) after thresholds were set on the applicable SQIs.

Errors associated with frequency-based respiration sources

The highest errors were found for methods based on frequency-based sources (RR_{RIFV} and RR_{RSA}), which rely on the RSA as a pathway to respiration.
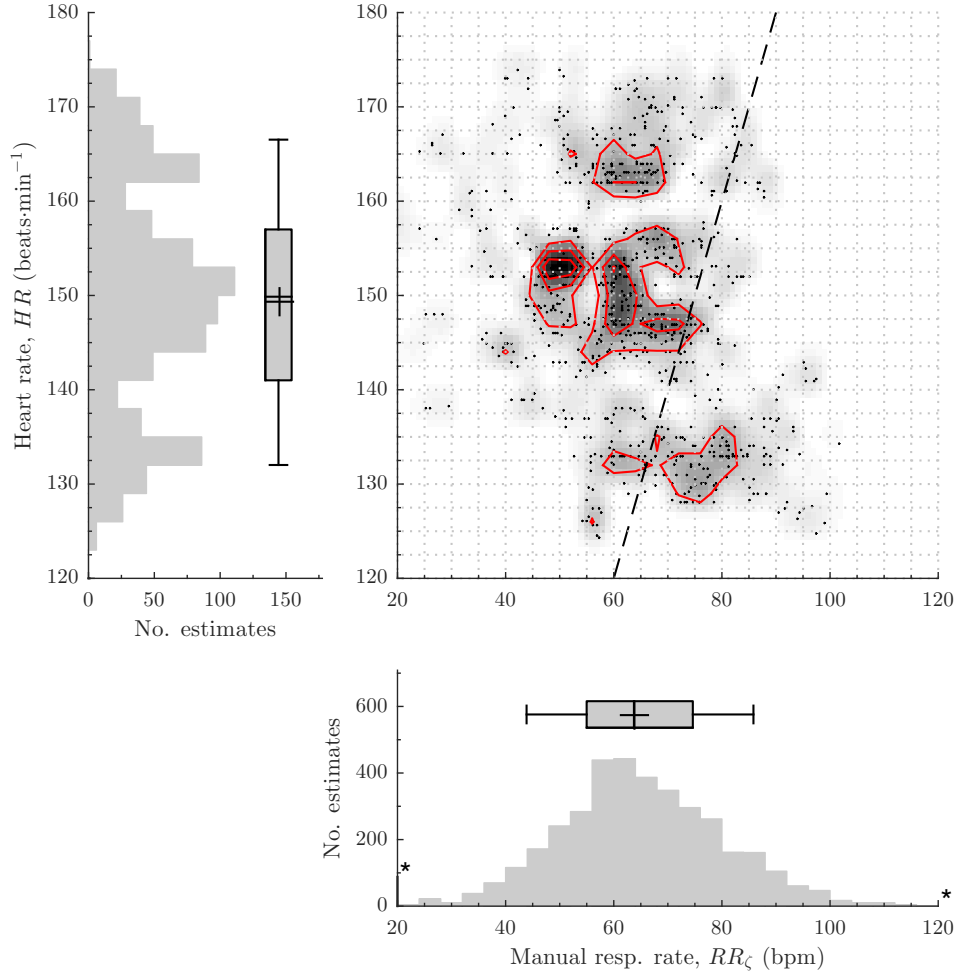
Heatmap of the distribution of (HR, RR_ζ) estimates. Marginal distributions of both variables are also shown. The HR values shown were obtained after the cardiac frequency estimates derived by the study monitor from the ECG trace were interpolated at the time instants of RR_ζ computation. The red curves represent isolines of the local density of data points. The dark dotted line represents the Nyquist limit for respiratory frequency estimation. For estimates below this line, $RR_\zeta > \frac{1}{2}HR$, so the respiratory signal is insufficiently sampled from cardiac beats for accurate detection of the true RR.

There are conflicting reports in the literature regarding the presence of RSA in preterm infants. Several studies which have estimated RSA in neonates [Indic et al., 2013] suggest that although there is a interaction between heart rate variability and respiration, these cardiorespiratory interactions are not as continuous as in adults, and may only develop later in some infants. In fact, Clark et al. [2012] suggested that RSA may serve to gauge the developmental maturity in neonatal intensive care patients.

The evidence for the use of RSA for continuous monitoring of RR in neonates is scarce [Johansson et al., 1999; Wertheim et al., 2009]. Monasterio et al. [2012] included both pulse rate and heart rate variability as features in a classifier for apnoea-related desaturations in preterm infants, nevertheless the reliability of these two signals was not assessed against a respiration standard.

7.7 Chapter conclusions

In this chapter, we have estimated respiratory rates and inter-breath intervals from video data and the three different signal sources available to the patient monitors used in neonatal practice.

By applying the techniques developed previously to a database of 165 manual breath counts, we have demonstrated that accurate RR estimates can be produced for neonates from all sources for respiratory rates in the 40 - 80 bpm range. On the basis of the results from this study, we have concluded that the measurement of respiratory rate in this population using indirect methods poses a fundamental challenge at rapid breathing regimes (upwards of 80 breaths per min). Camera-derived respiratory signals do not suffer from this limitation. The performance of proposed methods for the extraction of the respiratory signal from video data has been assessed in terms of accuracy in the RR estimates derived from this signal. Comparable error rates were found between these estimates and those derived from analysis of the IP signal acquired at the same time as manual breath counts.

Chapter 8

Detection of adverse physiological events

Contents

8.1 Chapter summary

In the previous chapter, we demonstrated the feasibility of camera-based measurements of respiration in neonates during periods of spontaneous breathing at rates varying from 40 to 120 bpm. In this chapter, we extend our analysis to periods of breathing cessation, as encountered during adverse physiological events.

This chapter is therefore concerned with the detection of respiratory pauses induced by apnoea of prematurity (AOP), a condition which affects a significant proportion of infants born prematurely in the first few weeks of postnatal life [Eichenwald et al., 1997]. As we have seen in Section 2.3.3, AOP is a major concern in patient care. Recurrent episodes of AOP can impair neurodevelopment and adversely affect patient outcomes. In spite of their clinical significance, the detection of AOP events remains a persistent challenge for patient monitoring systems based on contact sensors [Brouillette et al., 1987;

163

Warburton D, Stark AR, 1977].

In this chapter, we present an automated method for the detection of cessation of breathing events (COBEs) and assess its performance in segments of video camera signals from our NICU dataset. It begins by extracting the respiratory signal from these signals using the methodology established in Chapter 6. Then, two quantities were computed to characterise the respiration signal: its mean rate and signal quality. Finally, potential cessations of breathing were identified and reviewed by neonatologists and the positive predictive value of the cessation detection algorithm was assessed.

This chapter is organised as follows: Section 8.2 introduces previous work in the field of neonatal apnoea detection and Section 8.3 the dataset and proposed methodology. The experimental results are reported in Section 8.4 and discussed in Section 8.5. The main conclusions are presented in Section 8.6.

8.2   Related work

The automatic detection of apnoeic episodes in preterm infants is a technically challeging problem in pattern recognition. An inherent difficulty is the poor quality of the physiological signals available. This may explain why this problem has deserved scarce attention in the literature on signal processing techniques for apnoea detection, where most studies have focused on the diagnosis of obstructive sleep apnoea (OSA) either in adults or children.

The poor quality of the physiological signals typically available can be attributed to two main sources of signal artefact. On the one hand, voluntary and involuntary movements of the neonate, such as kicking, stretching, crying, and nursing interventions have been identified as the causes of motion artefacts which lower the signal-to-noise ratio of the oxygen saturation estimates and produce inaccurate detections [Tobin et al., 2002]. At the same time, pulse oximeters also suffer from bad connections and poor sensor contact [Chambrin, 2001]. As a result, the unimodal detection of apnoeic events through the associated desaturation events exhibits poor specificity. To mitigate alarm fatigue, McClure et al. [2016] experimented with different spO_2 averaging times and alarm delays in a study including 10 preterm infants of very low birth weight (< 1500 g). Their results indicated that incorporating an alarm delay with shorter spO_2 averaging times (*e.g.* 2 or 8 seconds) can reduce both alarm number and duration, and allow for a more accurate assessment of oxygenation.

On the other hand, Lee et al. [2012] identified a phenomenon whereby cardiac blood flow acts as a potential confounder in the IP signal during central apnoeic episodes. As heart rate slows during severe apnoea and intracardiac volumes are increased, chest impedance changes due to cardiac activity are exacerbated. In the absence of breathing movements, these increasingly large impedance fluctuations due to cardiac filling can reach

thresholds for breath detection thus foiling the apnoea alarm. The authors developed a filter design to tackle this error mode in the operation of the apnoea alarm. The analysis of this new filtered IP agreed with clinical experts in 212 cases out of 234 episodes (91% accuracy) in very low birth weight patients admitted to the University of Virginia NICU.

To address these problems, Monasterio et al. [2012] used machine learning techniques to detect the occurrence of apnoea-related desaturations in preterm infants and thus reduce the false alarm rate in NICU stays. The authors used PPG, ECG and IP derived features from the Physionet's MIMIC II dataset and applied a feature selection method based on mutual information combined with a support vector machine (SVM) to produce a classifier with a sensitivity of 86.2% and a specificity of 91.4% on the test data. Using the same framework, Daly et al. [2012] developed a monitor for use in resource-poor environments which relied solely on PPG-derived parameters, thus producing a classifier with a sensitivity of 76.9% and a specificity of 52.0% and a balanced accuracy of 64.4% on a test set of 663 events from 27 NICU stays in the same dataset. Both authors emphasized the need for high quality reference data to boost these figures.

In a later study, [Williamson et al., 2011] explored the predictive utility of features derived from respiratory rate and heart rate. Specifically, the multiscale correlation structure of IBI and heartbeat intervals was used to train a patient-specific apnoea prediction algorithm. The algorithm's prediction results were significantly better than chance only for three of the six patients on whom it was evaluated.

In conclusion, these preliminary studies suggest that although features in conventionally monitored cardiopulmonary signals can detect the occurrence of clinically significant apnoeas in preterm infants, the automatic detection of these events has been impaired by the poor quality of these vital signs, which are often corrupted artefacts introduced during their acquisition with traditional contact sensors.

8.3 Methodology

In this chapter, we present an automated method to detect cessations of breathing events (COBEs) in short time epochs of 300 s of camera signals gathered from the NICU dataset.

We started by extracting relevant information from these signals through the methodology established in previous chapters. The purpose of this analysis was two-fold; we sought to (1) detect breathing movements and estimate their frequency and (2) produce a binary signal quality index (SQI) for these estimates based on the level of activity on the video sequence. Finally, the information provided by these two signals was combined into a COBE classifier (as shown in Figure 8.1). A set of reference annotations were obtained from clinical experts as a gold standard to train the detection algorithm and evaluate its performance.
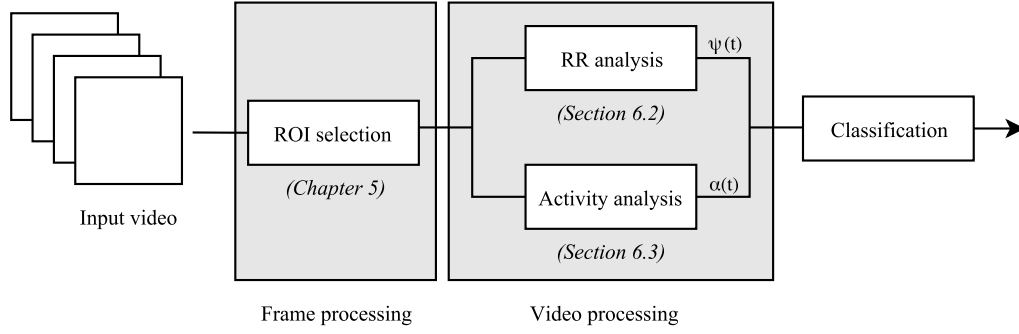
Flow diagram illustrating the algorithm for COBE detection in the NICU dataset. The quantities in the diagram are introduced in the main text.

8.3.1 Dataset

For the creation of training and validation datasets, we gathered a set of 107 time epochs of 300 s comprising episodes of cessations of breathing (positive events, **P**) and instances of normal breathing (negative events, **N**). These were extracted from the NICU dataset (Chapter 3.3) in two steps as described below. To facilitate the identification of positive events, we first processed the respiratory rate derived from the bipolar IP signal by the patient monitor (RR_{MON}) with a set of heuristics (Section 8.3.1.1). The purpose of this step was to flag potential apnoeic episodes (even if with low specificity). For a detailed discription of this problem, see Appendix E. Prior to their inclusion in the dataset, these episodes were then manually reviewed by clinical experts (Section 8.3.1.2).

8.3.1.1 Identification of potential cessation of breathing events

According to the definition of apnoea introduced in Chapter 2, apnoeic episodes are characterised by (a) a cessation of breathing for a minimum of 20 seconds or (b) a shorter pause accompanied by bradycardia (HR < 100 beats per minute), cyanosis, or pallor. [Committee on Fetus and Newborn, 2003]. As a first approach to the problem, in this work we focus primarily on the detection of instances where breathing movements cease, regardless of whether this event was accompanied by any of the medical signs described above or not. Thus, the following was considered to be a necessary condition for COBEs:

$$\text{RR}_{MON} < 20 \text{ bpm for a period of at least 20 seconds}[1].$$

The rationale behind the criterion placed on RR_{MON} is the following: the respiratory rate estimation routines implemented in the patient monitors attempt to estimate respi-

---

[1]as derived by the patient monitor.

ratory rate from the IP waveform even when this signal is not periodic. Thus, during a respiratory pause, it will output non-zero values of respiratory rate even in the absence of a respiratory effort during the analysis window. Hence, we have defined a potential cessation of breathing event as any period for which the respiratory rate produced by the patient monitor is lower than 20 bpm, a value well below the physiological range for this vital sign in neonates, as previously indicated by the distribution of RR values in the dataset shown in Figure 3.8b.

8.3.1.2 Manual validation

The set Ω_c of events detected by applying the criterion above to the dataset was annotated by a team of three clinical reviewers (including a consultant neonatologist, a clinical research fellow and a neonatal nurse) to ascertain whether each occurrence was artefactual or an actual COBE, and (if so) if this cessation was apnoea-related.

To do so, the reviewers were firstly given access to (i) the raw signals collected by the monitoring devices (single-lead ECG, IP, foot PPG signals), and (ii) autonomic responses (heart rate, oxygen saturation, and respiratory rate) derived by the monitor from these traces for a 300s period around the event trigger (*i.e.* the time instant at which RR_{MON} falls below 20 bpm). After this stage, the reviewers were shown (iii) the video footage collected during these events (when available) and were allowed to keep their intial label or assign a different one in light of contrasting visual evidence.

For each event, reviewers decided among three options: (a) the event was caused by noise or artefacts in the raw signals (b) the event was caused by a cessation of breathing (positive event), or (c) the event was not caused by a cessation of breathing nor by artefacts in the raw signals (*i.e.* RR_{MON} accurately reported a low respiratory rate value). A block diagram which illustrates this decision process is shown in Figure 8.2.

The (a) set was further divided into (a_2) motion artefacts and (a_1) IP artefacts (*i.e.* related to suboptimal probe placement or other acquisition artefacts which cause the IP signal to deviate from true respiratory effort). The (b) set was subdivided into cessations of breathing with (b_1) and without (b_2) an associated apnoea (*i.e.* with desaturation and bradycardia). Examples of signal segments for scenarios (a) - (c) are given in Figures 8.3 to 8.6.

At each stage, each reviewer selected one of the labels above as the cause for the observed decrease in RR_{MON} and a collective label determined after group discussion was assigned for further analysis. When no consensus could be reached, the event was excluded from further analysis.
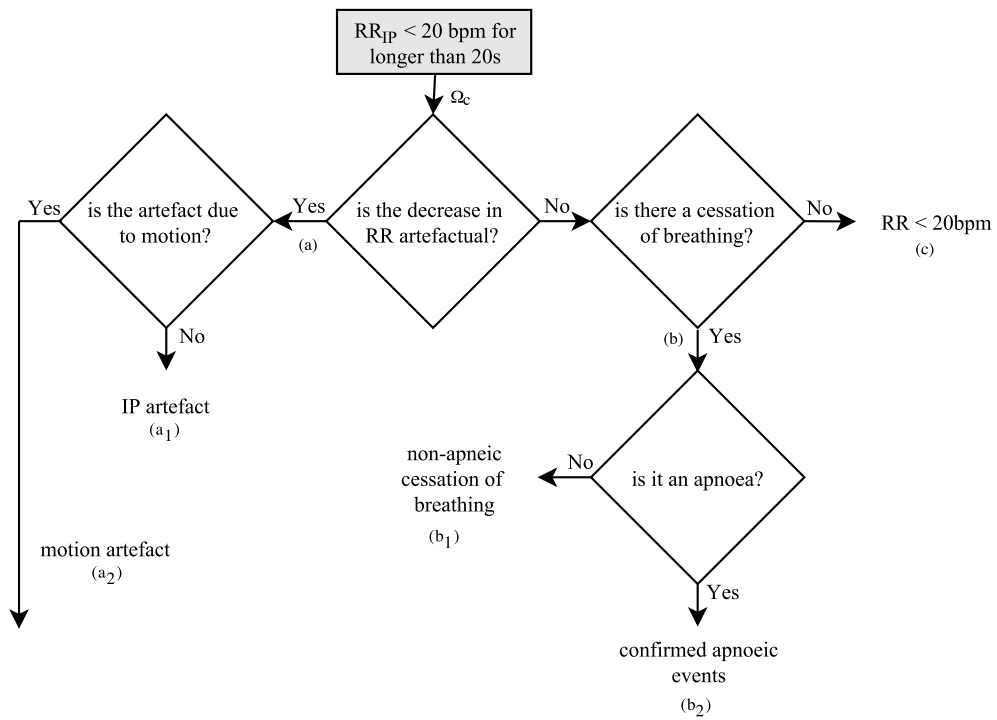
Decision tree for manual validation of the set of potential COBEs (Ω_c). In a first stage, the reviewers analysed (i) the raw signals collected by the monitoring devices and the vital signs derived from these traces for a 300s period around each event trigger. In a second stage, the reviewers analysed the video footage collected over the same period.
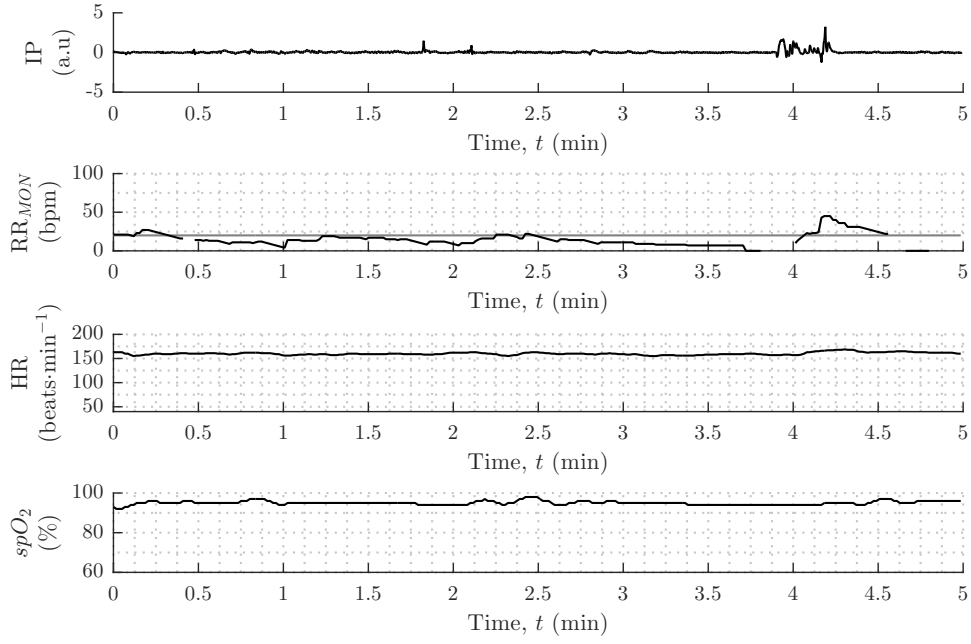
Segment of IP, RR, HR and spO$_2$ tracings during a period of recording of artefactual IP (a_1). The RR$_{MON}$ derived from the artefactual IP trace is abnormally low (< 20 bpm). Visual inspection of the video recording acquired during this period revealed that the infant was breathing at a rate of approximately 80 bpm throughout this period. The horizontal line at $RR = 20$ bpm illustrates the criterion used in Section 8.3.1 for the detection of potential COBEs.
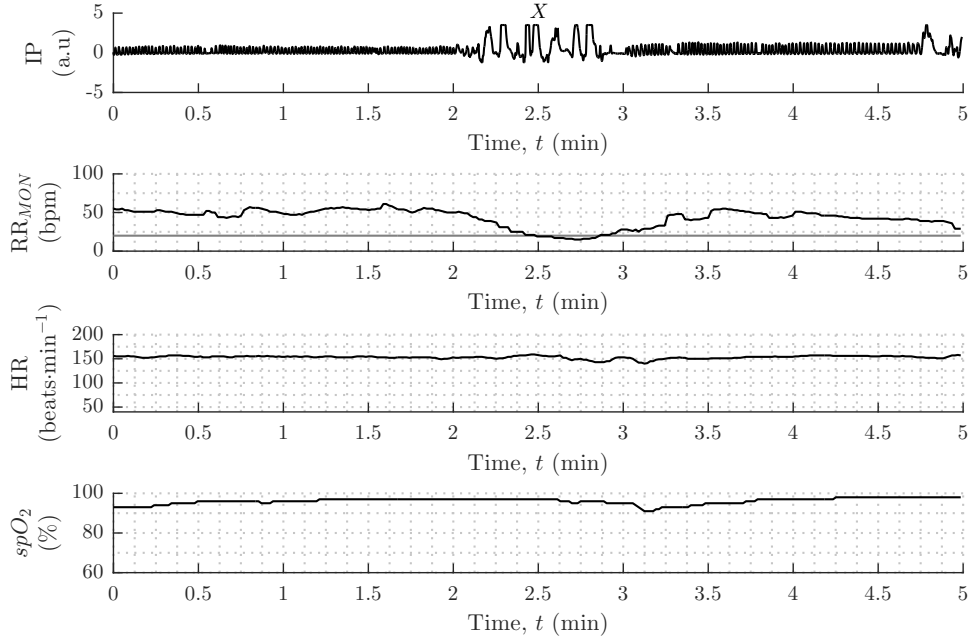
Segment of IP, RR, HR and spO$_2$ tracings during infant motion (a_2). The spO$_2$ tracing shows a small noise-related desaturation at t = 3 min. The HR tracing shows no signs of bradycardia. High amounts of noise are visible on the IP signal before the desaturation (X; from t = 2 min to t = 3 min). The horizontal line at $RR = 20$ bpm illustrates the criterion used in Section 8.3.1 for the detection of potential COBEs.
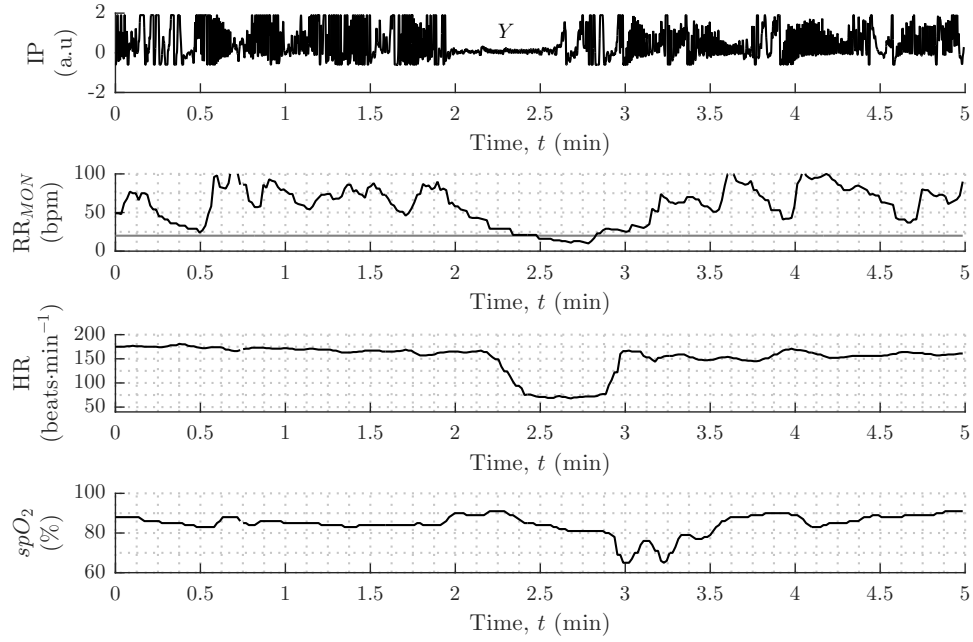
Segment of IP, RR, HR and spO₂ tracings during an apnoeic event in (b_1). A cessation of breathing lasting 40 s can be observed in the IP signal from t = 2 min (Y), followed by bradycardia around 30 s later; the peripheral arterial oxygen saturation (spO₂) falls to values near 60% at t = 3 min. The horizontal line at $RR = 20$ bpm illustrates the criterion used in Section 8.3.1 for the detection of potential COBEs.
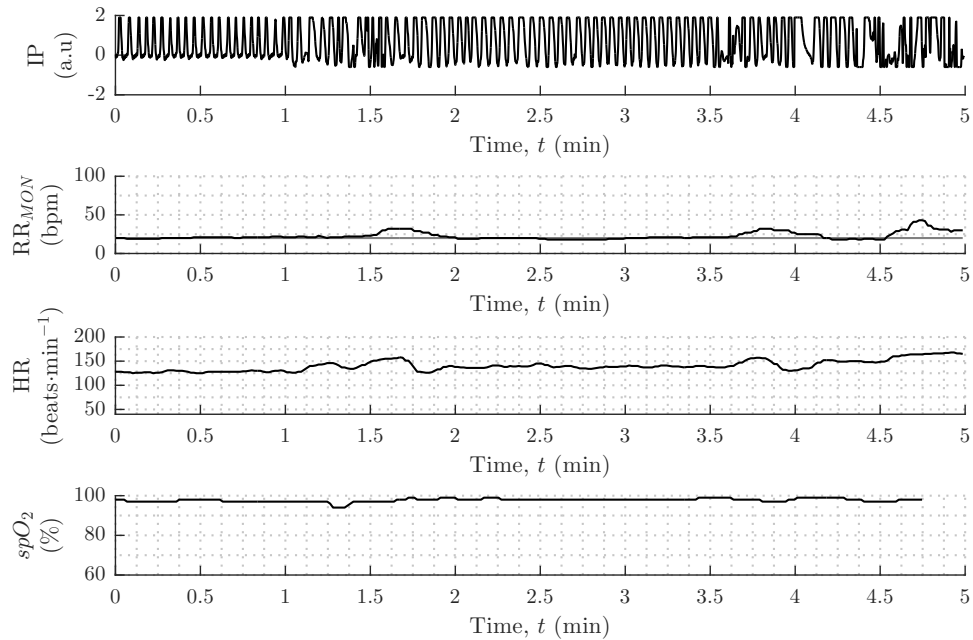
Segment of IP, RR, HR and spO₂ tracings during a period of slow breathing (c). This segment was acquired during the monitoring of a term infant (with a gestation age of 36 weeks and 3 days at the time of recording). The horizontal line at $RR = 20$ bpm illustrates the criterion used in Section 8.3.1 for the detection of potential COBEs.

8.3.2 Feature selection

When detecting cessations of breathing, a natural vital sign to enquire is RR. We derived camera-based RR_ψ estimates over the 300-second segments with the same parameters adopted in Section 7.3.2.2. However, as we observe in Figure 8.7, this feature is not effective in separating the elements in the (a) and (b) classes. As a consequence, we adopted SQI_ψ defined previously in Chapter 7. This poses the problem of finding cessations of COBEs as that of finding those time segments (*i.e.* set of consecutive window-derived estimates) for which the estimated ψ signals have low SNR at breathing frequencies.

(a) RR_ψ distribution during (a_1) events.

(b) RR_ψ distribution during (a_2) events.

(c) RR_ψ distribution during (b_1) events.

(d) RR_ψ distribution during (b_2) events.

(e) RR_ψ distribution during (c) events.
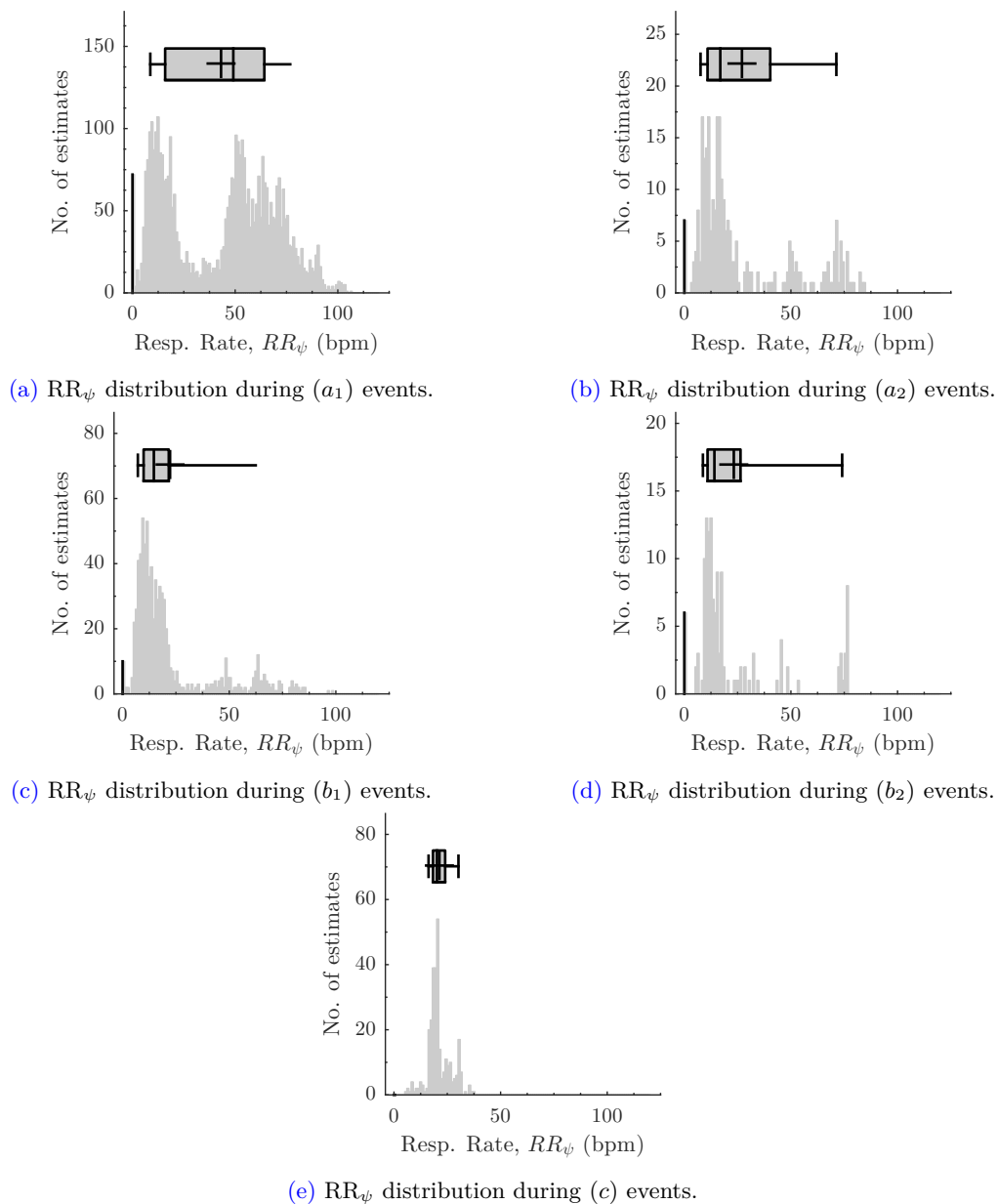
Figure 8.7
Distribution of RR_ψ estimates during 107 episodes of potential COBEs.

Naturally, this condition is also met in windows affected by movement artefact. In

these windows, SQI_ψ is low as a result not of the lack of breathing motion but of the presence of infant motion.

A second signal quality index (SQI_α) was introduced, based on the level of activity in the video sequence $\alpha(t)$:

$$SQI_\alpha(t) = \begin{cases} 1, & \text{if } \alpha(t) = 0 \\ 0, & \text{if } \alpha(t) > 0 \end{cases} \tag{8.1}$$

which takes on values for every frame. This signal was interpolated at the times of RR_ψ estimates. Prior to their use as a feature $\text{SQI}_\psi(w)$ values for windows w with $SQI_\alpha(w) = 0$ were discarded due to the presence of non-breathing motion in the video stream. After these SQI_ψ values were discarded, a further signal quality index $\text{SQI}_{\widetilde{\psi}}$ was obtained by linearly interpolating SQI_ψ between the remaining segments for which these were less than Δt^* apart. Δt^* is the detection threshold, a parameter of the classifier.

8.3.3 Classification

For each event, a COBE was deemed to have occurred if

$$SQI_{\widetilde{\psi}} < 0.9 \text{ for a period of at least 20 seconds.}$$

In other words, a non-existing respiratory rhythm was assumed for windows $\text{SQI}_\psi < 0.9$. This threshold was adopted for consistency with the one used previously in Chapter 7. We used ROC analysis to determine the optimal value for Δt^* on the training set.

8.3.4 Training

The manually classified events were distributed at random between test and training sets in a manner which preserved the balance between the number of elements of each subclass in each set (Table 8.1). An approach to obtaining model performance with respect to a learning parameter is to use cross-validation on two independent folds. The training set was divided into two sets, D_1 and D_2, such that the assignment to the two sets was based on the number of positive events. For each set, positive events were taken from the positive set P, and negative events were randomly sampled without replacement from the negative set N so that the number of positive and negative episodes were equal. First, a model was trained on D_1 and validated on D_2. Then, another model was trained on D_2 and validated on D_1. The results from both models were combined to produce an estimate of parameter $\Delta t^\star$. The parameter space was explored for $1 \leqslant \Delta t^\star < 8$ in

Distribution of annotated events in training and test sets, number of events.

Events, classes	Training set	Test set	Total
IP artifacts, a_1	36	35	71
Motion artifacts, a_2	5	6	11
Apneic COBE, b_1	8	9	17
Non-apneic COBE, b_2	2	1	3
Low RR, c	2	3	5
Total, Ω_c	53	54	107
Positives, $P = b_1 \cup b_2$	10	10	20
Negatives, $N = a_1 \cup a_2 \cup c$	43	44	87
Total, Ω_c	53	54	107

1-second increments. In practice, only a subset of 10 negative events was used in each experiment on the training set to ensure an equal number of positive and negative events in this set.

8.4 Results

8.4.1 Creation of a dataset

The clinical reviewers annotated a total of $|\Omega_c| = 107$ potential COBEs. Table 8.2 and Figure 8.8 summarise the findings of this annotation process.

In Table 8.2, we present a contingency table which relates the labels attributed by neonatal experts before and after the video footage was reviewed.

In Figure 8.8, we show how the total number of annotated events was distributed between labels after the two stages of manual validation were completed. The analysis of potential COBEs by the clinical reviewers revealed that 82 out of 107 (76.6%) such occurrences were due to the presence of artefacts in the IP signal (11 due to motion and 71 due to poor signal acquisition). In the remaining 25 artefact-free events, a total of 20 COBEs were registered, comprising 17 apnoeic events and 3 shorter cessations of breathing which occurred with no presentation of an AOP episode (*i.e.* without comcomittant bradycardia or desaturation).

8.4.2 COBE detection

Figure 8.9 shows activity ($\alpha(t)$) and respiratory signals during a clinically validated episode of neonatal apnoea. We observe that motion events manifest as perturbations in the activity signal $\alpha(t)$. After motion events (denoted by $\alpha(t) > 0$) are excluded, the lower intensity changes in the camera-derived signal due to the respiratory cycle become apparent (as illustrated in the $\widetilde{\psi}$ plot of ψ rhythm subject to $\alpha(t) = 0$). During the

Label before video

Label after video	IP artifacts, $\mathbf{a_1}$	Motion artifacts, $\mathbf{a_2}$	Apnoeic COBEs, $\mathbf{b_1}$	Non-apnoeic COBEs, $\mathbf{b_2}$	$RR < 20$ bpm, $\mathbf{c}$	Lens covered	Probe off	No consensus	total
IP artifacts, $\mathbf{a_1}$	**66**	0	2	0	2	0	0	1	71
Motion artifacts, $\mathbf{a_2}$	5	**5**	1	0	0	0	0	0	11
Apnoeic COBEs, $\mathbf{b_1}$	0	0	**16**	0	0	0	0	1	17
Non-apnoeic COBEs, $\mathbf{b_2}$	3	0	0	**0**	0	0	0	0	3
$RR < 20$ bpm , $\mathbf{c}$	1	0	0	0	**4**	0	0	0	5
Lens covered	0	1	3	0	0	**0**	2	0	6
Probe off	0	0	0	0	0	0	**0**	0	0
No consensus	0	0	0	0	0	0	0	**0**	0
total	75	6	22	0	6	0	2	2	

Table 8.2
Confusion table of potential COBEs classified by clinical experts before and after the video footage was inspected.

apnoeic episode shown, $RR_{\widetilde{\psi}}$ falls from a resting rate of approximately 50 bpm to values below the 20 bpm criterion for a period of approximately 35 seconds.

By following the methodology described, it was possible to classify the true cessations of breathing. The performance of the proposed method for COBE detection on the training and test sets is reported in Table 8.3.

The majority of IP artefacts (a_1) in the test set were correctly identified as negative events (28 out of 35 events) and all the movement-related artefacts (a_2) in the same set were correctly identified (6 events). Regarding positive events (b_1 and b_2), 7 out of 10 COBEs in the test set were detected.

174
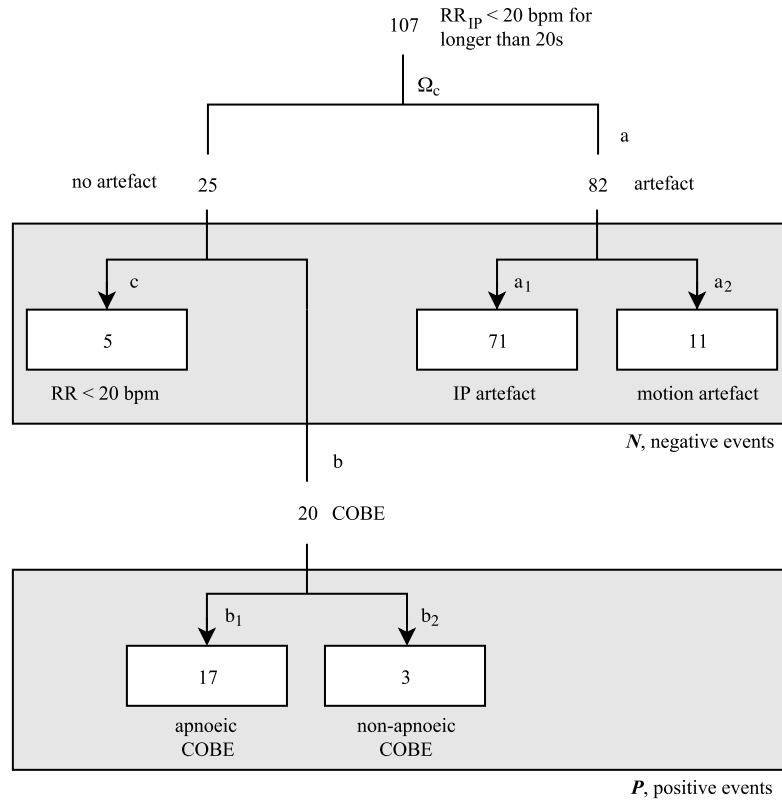
Number of events under each of the labels defined in Section 8.3.1 after manual annotation of the collected vital signs and video footage for the set of potential COBEs, Ω_c.

8.5 Discussion

A resting respiratory rate below 20 bpm is abnormal in premature infants (Figure 3.8b). Such a low rate could indicate a progression into an apnoeic event, and therefore, a reported respiratory rate below this value (20 bpm) should prompt immediate clinical attention. Our study in the Oxford NICU revealed that a high number of these occurrences (76.6%) were due to the presence of artefacts in the IP signal, the standard modality for monitoring respiration in neonatal units. This poses a clear problem. High false alarm rates may desensitize NICU staff and reduce reaction speed as a result of alarm fatigue

Table 8.3
Performance in COBE detection. Number of correctly identified events for each class. Figures are given for each subclass (a_1, a_2, b_1, b_2 and c) under the test set.

Training set		Test set						
P (/10)	N (/10)	P (/10)			N (/44)			
9	7	7			34			
		b_1 (/9)	b_2 (/1)	a_1 (/35)	a_2 (/6)	c (/3)		
		6	1	28	6	2		

175
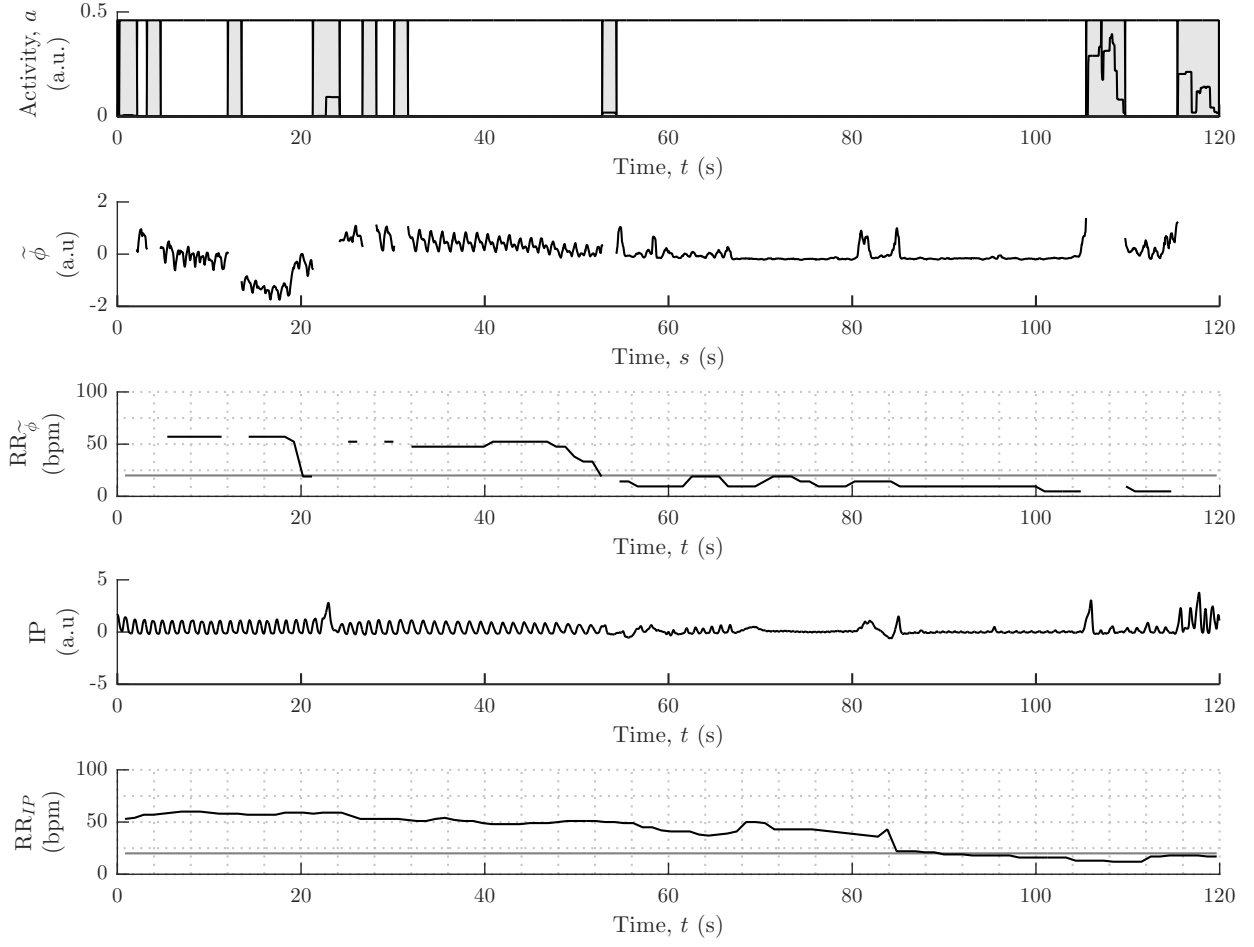
Figure 8.9

Segment of activity, camera-derived respiratory signal ($\tilde{\psi}$), IP and respiratory rates ($RR_{\tilde{\psi}}$ estimated from the camera-derived respiratory signal, and RR_{IP} provided by the patient monitor) during an apnoeic event. The segments in grey denote periods for which $SQI_\alpha = 0$. The second plot illustrates the application of the criterion defined for the deletion of segments corrupted by motion artefact as shown in the activity plot $\alpha(t)$. The intervals between motion events, where α is very low, correspond to periods of quiet sleep. A cessation of respiration (marked by both low α and $SQI_{\tilde{\psi}} < 0.9$) can be observed in the camera-derived respiratory signal from $t = 70$ s to $t = 105$ s. The tilde $\tilde{}$ denotes a variable linearly interpolated between windows for which $SQI_\alpha = 1$.

[Monasterio et al., 2012].

The poor quality of the IP signal is further evidenced by the presence of non-diagonal elements in the confusion matrix of Table 8.2. These elements account for events where the information provided by video monitoring led the team of reviewers to alter the label originally assigned.

Through the analysis of the video signals acquired from the inside of the cot, we classified the potential cessation of breathing events in the dataset. By identifying periods of high-frequency movement, it was possible to distinguish between subtle body movements associated with the respiratory effort from spurious motion inside the cot. Further analysis of the segments of low activity allowed us to classify breathing *versus* non-breathing behaviour.

A meaningful reduction in the number artefact-related events was achieved. All the movement-related artefacts (a_2) and the majority of IP artefacts (a_1) in an out-of-sample test set were correctly identified as artefactual detections. Given that the detection of true cessations is paramount in patient care, it is of critical importance that any increase in specificity is not achieved at the expense of the ability to detect positive events (b_1 and b_2). In the work presented here, 7 out of 10 COBEs in the test set were detected. On closer inspection, 2 of these cases were found to correspond to periods when the monitored infant was undergoing photo-therapy. This procedure involves prolonged exposure to LED light predominantly in the blue region of the spectrum (460 - 490 nm) up to an irradiation level of 30 - 40 $\mu W/cm^2/nm$. This causes bright regions in the acquired images to become saturated, thus decreasing SNR over these regions and depressing the $\alpha(t)$ signal.

8.6   Chapter conclusions

Our results show that the method presented here can successfully distinguish between artefactual and true decreases in RR in a range of critically low values for this variable (RR < 20 bpm). Although the sample size was small, the feasibility of our approach based on motion analysis was clearly demonstrated.

Conventional neonatal monitors could, therefore, benefit greatly from the collection of video signals. Specifically, the implementation of video-based COBE detection systems could potentially help reduce the high rate of false apnoea alarms in these units. Hence, we believe that our findings should motivate extensive validation of such techniques.

Chapter 9

Conclusions and further work

Contents

9.1 Main conclusions

9.1.1 Clinical relevance

There is a growing interest in non-invasive means of monitoring the physiological status. Camera-based monitoring of cardiorespiratory vital signals has recently been demonstrated in adults with the development of methods to estimate heart rate (HR), respiratory rate (RR), and changes in arterial oxygen saturation (spO_2). The goal of this thesis was to investigate the feasibility of camera-based monitoring of respiration in preterm infants during their stay in the neonatal intensive care unit (NICU). Specifically, our goals were (a) to validate respiratory measurements against a gold standard, and (b) to identify instances of clinically-validated cessation of breathing.

The observational study designed to achieve this goal included the recruitment of 30 infants of < 37 weeks postmenstrual age admitted to the high-dependency unit at the John Radcliffe Hospital in Oxford, from February 2014 to May 2015. Manual breath counts were obtained for analysis in 5 of these infants (17%). 8 infants (27%) had some form of cessation of breathing (apnoeic or non-apnoeic) as judged by a panel of neonatologists.

By applying the techniques developed here to the dataset of 165 manual breath counts, we estimated RR and inter-breath intervals (IBIs) from video data and the three signal sources typically available in neonatal practice (ECG, PPG, and IP). We excluded periods of poor signal quality, which occurred due to patient movement, nursing interventions,

or maternal touch, as identified by a motion-sensitive signal quality index (SQI). At all times when the infant was at rest $(73.1\%[1])$, we assessed the performance of RR and IBI derivation methods with respect to the reference respiratory signal ζ computed from manual breath counts.

The performance of the proposed methods for the extraction of IBI was assessed in terms of spectral coherence between the spectral structure of the derived IBIs and that of manual IBIs. We report our results in Table 7.1. Comparable values of statistically significant magnitude-squared coherence (SMSC) were found between camera-derived estimates of IBI and those derived from the IP signal and the amplitude analysis of the ECG- and PPG-derived signals.

We evaluated the performance of our RR estimation methods for two different frequency ranges: (a) RR $<$ 80 bpm (84.7% of windows) and RR $\geqslant$ 80 bpm (15.4% of windows). A summary of error analysis is shown in Table 7.2. Similar error rates were found between camera-based estimates (RR_ψ) and those derived from the analysis of the IP signal; with mean absolute error $MAE(RR_\psi) = 5.0$ bpm *versus* $MAE(RR_{IP}) = 4.0$ bpm, respectively, over the entire breathing range. The fused RR_ψ method showed marginally worse performance than RR_{IP} for RR $<$ 80 bpm, with $MAE(RR_\psi) = 4.7$ bpm *versus* $MAE(RR_{IP}) = 3.6$ bpm, yet slightly better performance for RR $\geqslant$ 80 bpm, with $MAE(RR_\psi) = 9.3$ bpm *versus* $MAE(RR_{IP}) = 10.7$ bpm. This behaviour was expected as chest motion artefacts which corrupt the IP waveform were naturally expected to also disrupt the visual cues upon which our algorithm relies to detect breathing movement.

We also found that the measurement of RR using indirect methods based on ECG or PPG faces a fundamental challenge at high breathing rates (above 80 bpm). We hypothesize that this is due to the fact that the sampling frequency of these techniques is limited by the cardiac beat. Therefore, to avoid aliasing, the respiratory frequency cannot exceed half of the instantaneous heart rate during the same time epoch. As shown in Figure 7.10, such an occurrence was not uncommon even in the quiet periods during manual counts. Based on this result, the good performance of non-invasive techniques based on ECG and PPG found in recent studies may not be transferable to the preterm population. To our knowledge, this is the first account of this problem.

The records of patients during critically low values of respiratory rate (RR $<$ 20 bpm) were examined for cessations of breathing (COBEs) within a 5-minute period of these alarms being triggered. Our results show that a high percentage of these occurrences (76.6%) were due to the presence of artefacts in the IP signal. This study allowed us to develop and test a camera-based algorithm for the detection of COBEs. Although the sample size was small, the feasibility of our approach was clearly demonstrated as (a) the majority of IP artefacts (28 out of 35 events) and all the movement-related artefacts (6 events) in the test set were correctly identified as negative events, and (b) 7 out of 10 true

---

[1]defined as the percentage of valid time windows divided by the number of windows available for analysis.

COBEs in the test set were detected. On closer inspection, 2 of these 3 cases were found to correspond to periods when the monitored infant was undergoing photo-therapy. It is therefore reasonable to hope that camera-based detection of COBEs could lead to early warning of episodes of apnoea of prematurity. We believe our findings should motivate extensive validation on a larger dataset.

During the creation of a dataset of potential COBEs, it was observed that in 20 out of the 107 (19%) episodes reviewed, the team of neonatologists changed the event label originally assigned from the analysis of the patient monitor traces when this evidence was supplemented with video footage collected during the episode (Table 8.2). In particular, we highlight an instance where an infant (female; gestational age = 30 weeks) presented an unreported episode of very long apnoea (VLA) which was only revealed upon the retrospective inspection of the video feed. These observations confirm the added value of video-based monitoring.

9.1.2 Advances in non-contact physiological monitoring

A persistent problem in camera-based monitoring is the low signal-to-noise ratio (SNR) of physiological signals. As a result, denoising methods to suppress signal components unrelated to physiological information have attracted research efforts. The method developed in Chapter 6 (τCA) is a novel approach that utilises the temporal characteristics of the signal of interest to improve the quality of the extracted respiratory signal ψ.

τCA has several advantages over conventional methods for solving linear source separation based on independent or principal component analysis (ICA or PCA).

From a physiological standpoint, τCA replaces the independence criterion of conventional ICA with a criterion based on the temporal structure of the signal (periodicity). This is more consistent with the clinical intuition about this biological rhythm; physiologists define respiratory rate as the respiratory frequency at rest and so it is not defined over time segments for which breathing movements are aperiodic.

From a technical standpoint, the changes in pixel intensity due to breathing can be subtle, particularly in small preterm subjects. As a result, the high-variance components extracted by PCA may exclude important components with respiratory-driven morphology due to their low amplitude with respect to signal components driven by large body motion.

Finally, even if the respiratory signal is successfully obtained after using either technique (ICA or PCA), it is not possible to determine which of the separated sources corresponds to this signal without custom criteria. Often periodicity measures are used for ranking the extracted components and eliminating less-informative components. In τCA, the periodicity term in the definition of the prediction error Q (in Equation 6.2) means that the most periodic component is explicitly sought. This feature is very helpful, especially for automating the selection of the respiratory component.

In this thesis, we used τCA to improve the quality of the extracted respiratory signal in multi-channel recordings from multiple regions-of-interest (ROIs). Nevertheless, this framework is general and can serve as a basis for suppressing non-physiological signal components in multi-channel recordings obtained using other monitoring modalities.

9.2 Further work

The remainder of this Chapter is devoted to open questions and potential fields of study concerning the extraction of camera-derived neonatal respiratory signals and their analysis.

Investigations into τCA

In Chapter 6, we introduced a novel approach to the problem of source separation based on the works of Saul et al. [2001] and Sameni [2008]. Although the method is an intuitive extension of conventional source separation techniques to periodic or pseudo-periodic sources, further quantitative studies are required for comparing its performance against that of classical methods.

In further studies, simulated data formed by artificial mixtures of periodic and Gaussian signals (such as that proposed in Chapter 6) may be used to evaluate the performance of this method. Some consideration should be given to the subject of which performance metrics are suitable for this task. These should express desirable properties of a source separation method, namely (a) the ability to isolate the desired signal in as few components as possible, (b) some measure of periodicity of the selected component (*e.g.* spectral purity or ratio between the power in the respiratory range over the entire frequency range) and (c) computational cost.

In the case study of Chapter 7, τCA was used for isolating the respiratory signal from a set of recordings of pixel intensity over multiple ROIs. Some aspects of this work need refinement. Perhaps the most urgent is the definition of the lag parameter range $\tau \in \left[\tau_{min}, \tau_{max} \right]$, which is associated with the range of plausible respiratory rate values. On the one hand, insufficient sampling of the respiratory range means that the set of cross-correlation matrices may not be diagonalisable. On the other hand, each τ requires the computation of $m \times m$ cross-correlation matrices $C_{xx}(\tau)$, where m is the number of channels. Even when subject to the joint diagonalisation procedure adopted (JADE), the diagonalisation of a large set of C_{xx} matrices may become costly. We hope to extend our approach in Chapter 7, wherein the range of τ was set to the range of respiratory frequencies in neonates, by one that optimises this trade-off. We intend to develop methods for tracking the respiratory frequency in order to restrict τ to possible values based on previous measurement windows.

Finally, beyond the problem of period tracking, issues concerning the choice of the number of effective ROIs, m, and their dimensions also need further consideration.

Towards continuous monitoring

In analysing the data collected from the John Radcliffe Hospital NICU, we focused our respiratory analysis on two problems for which clinical ground truth was available. In the future, we would like to extend this work to achieve continuous monitoring at the patient's bedside. Accurate subject detection from the background is crucial to successfully conduct continuous monitoring using cameras. In this work, we used a simple detector based on skin-like appearance to obtain the ROI from which the respiratory signal was extracted. Whilst this approach was effective in the quiet periods in which manual counts were scheduled (Chapter 7) or potential apnoeic episodes occurred (Chapter 8), there is a need for more robust subject detection to achieve continuous estimation over longer periods of time.

Many investigations into remote sensing of vital signs in adults have performed automatic selection of ROIs using well-established face detection routines or manually cropped ROIs. These operate in conjunction with a feature tracker to automatically update the ROIs in consecutive video frames (Poh et al. [2010]; van Gastel et al. [2016] among others). In the NICU environment, this approach is ill-suited for the detection of the subject of interest due to constant changes in light conditions, complex backgrounds, differences in skin tone and pose, body occlusions, and frequent interaction with the clinical team [Villarroel et al., 2014].

As an alternative to conventional processing pipelines, recent developments in deep learning research have greatly improved the classification performance in challenging segmentation scenarios. Architectures such as convolutional neural networks (CNNs) provide an efficient framework for encoding certain properties of natural images, such as stationarity of statistics and local pixel dependencies. We highlight a study by Chaichulee et al. [2017] in our group, in which we presented a multi-task CNN model for patient detection and skin segmentation. The robustness of the resulting classifier with respect to several skin tones, subject pose, subject locations, and illumination variations was demonstrated in the subset of frames of the NICU dataset presented in Chapter 3 of this thesis.

Towards a remote respiratory assessment of the neonate

In this work, we aimed to reproduce the RR measurements derived from manual breath counts performed as part of patient care in the NICU. Alongside respiratory rate, as part of breathing auscultation, clinical staff are required to observe a number of signals. These include inspiratory volume, effort, pallour, and altered breathing patterns, such as Cheyne-Stokes respiration [Niimi et al., 2013]. Regional asymmetries in breathing motion, such as abdominal *versus* thoracic or right *versus* left thorax contrasts, may also be noted.

Normally, the inspiratory expansion of the thorax and abdomen is almost synchronous. However, if the upper airway is partially obstructed, phase differences between movements of the thorax and abdomen may arise [Prisk et al., 2002]. This phenomenon, known as thoracoabdominal asynchrony, is a common finding in infants whose chest wall compliance is elevated or exacerbated by respiratory disease or respiratory muscle weakness [Strömberg and Nelson, 1998]. Conventional contact-based techniques have limited capacity to perform these measurements and cannot adequately measure regional asymmetries.

The use of video monitoring in conjunction with accurate body segmentation could permit the analysis of local respiratory signals. Cameras can be used to quantify the motion of the thorax and abdomen independently, including their speed, and volumetric deformation. This means that through camera-based monitoring, not only could continuous measurement of RR be achieved, but some domain knowledge of human observations, such as a richer description of breathing patterns and tidal volumes, would also be retained. In this context, recent work by our group [Tarassenko et al., 2014] in a dataset of dialysis patients has led us to construct maps of regional values of heart rate and respiratory rate based on the spatial analysis of the power distribution (as estimated by auto-regressive models) for each frame. These maps provide a visualisation of the location and strength of the heart rate and respiratory rate information in different body regions. This would be especially useful in the context of monitoring respiratory asymmetries in neonates.

Towards a remote neonatal apnoea monitor

When accompanied by bradycardia and/or oxygen desaturation, cessations of breathing are serious clinical events that need immediate medical attention and investigation for associated pathology. The possibility of extracting additional vital signs using the same camera-based device by combining different designs of illumination, camera optics, and signal processing algorithms would therefore offer significant advantages. The ability to obtain camera-derived estimates of HR and spO$_2$ (in addition to RR) during periods of breathing cessations would leverage the system presented here from a COBE detector to an apnoea detector. The continuous monitoring of HR has already been demonstrated in two NICU patients [Villarroel et al., 2014] while recent advances in remote pulse oximetry have shown that changes in the spO$_2$ of neonates can also be detected under strictly-controlled experimental, conditions [van Gastel et al., 2016]. In conjunction with the work presented in this thesis, these preliminary results pave the way for truly non-contact detection of neonatal apnoea. A camera-based apnoea monitor would constitute a significant improvement over existing surveillance systems and should serve as an incentive for future research into this topic.

Appendix A

Human ventilatory response

The purpose of the experimental protocol described in Section 3.2 was to acquire data from subjects undergoing large variations in their vital signs while both subject motion and light conditions were strictly controlled. This appendix presents some topics in respiratory physiology relevant to our understanding of how the protocol adopted for the hypoxia and hypercapnoea sub-studies are responsible for the observed variations.

Hypoxia and hypercapnoea are defined in terms of the partial pressure (measured in mmHg or kPa) of a given gas species in a mixture of inspired gases. Hypoxia corresponds to an abnormally low partial pressure of oxygen, pO_2, and hypercapnoea to an abnormally high partial pressure of carbon dioxide, pCO_2. At sea level, O_2 constitutes roughly 21% of the atmospheric gases, so its partial pressure is 21% of 760 mmHg (the atmospheric pressure at sea level), or 160 mmHg. CO_2 constitutes only 0.04%, so its partial pressure is 0.04% of 760 = 0.3 mmHg.

In the next sections, we summarise the effects of short-term exposure to hypoxia and hypercapnoea in young healthy adults with a normal respiratory physiology. This discussion has informed our protocol design, particularly our decisions regarding the time intervals and magnitude of the ascending hypercapnoeic stimuli necessary to evoke a meaningful change in respiratory effort (Figure 3.1). In most of the literature reviewed on this topic, RR and V_T are often not recorded in favour of ventilation ($\dot{V}_E$), a measure of respiratory flow, often expressed in l min^{-1}, which is more commonly adopted as a measure of the stress induced by hypoxia or hypercapnoea.

A.1   to Hypoxia

In the range of hypoxia limited by ethical constraints, the hypoxic ventilatory response (HVR) is mediated by the chemoreflex, sensitive to arterial CO_2 via changes in $[H^+]$. Gen-

erally, HVR is measured either under varying or fixed CO_2 tensions (poikilocapnoea, PH, or isocapnoea, IH, respectively). The former case was tested under the hypoxia protocol. Whether in IH or PH, hypoxia is known to trigger a biphasic response characterised by an rapid increase in $\dot{V}_E$ followed by a progressive decline over the following 20 minutes, attributed to (among other factors) a desensitisation of peripheral chemorreceptors to hypoxia [Duffin, 2007; Pedersen et al., 1999]. During PH in particular, the acute increase in $\dot{V}_E$ described above is blunted; V_T does not increase as much as in IH and the respiratory rate response is practically abolished. This is easily understood if one considers that in PH, inspired CO_2 is not controlled, thus subjects can effectively increase CO_2 elimination without resorting to an increase in ventilation [Steinback and Poulin, 2007].

A.2   to Hypercapnoea

The situation with hypercapnoea is less complicated. It is widely acknowledged that the hypercapnoeic ventilatory response is also time dependent and can be divided into two phases. A rapid first phase, mediated by the peripheral chemoreflexes generated in the carotid bodies, acts within 8 - 26 seconds after the onset of the hypercapnoeic stimulus. The second phase is slower (65 - 180 seconds) but accounts for a greater part of the increase in minute ventilation (70 - 88%) versus 12 - 30% during the first phase. This component is attributed to the effect of hypercapnoea on the central receptors at medullar level, which are sensitive to seric levels of p_{CO_2} [Tansley et al., 1998]. Some authors have noted that these two phases alone do not account for the increase in minute ventilation during sustained hypercapnoea as ventilatory response can last up to 2 hours after the start of hypercapnoeic exposure [Khamnei and Robbins, 1990].

Conclusions

In conclusion, whilst exposure to poikilocapnoec hypoxia is not associated with a change in ventillatory effort, the sustained increase in ventilation with hypercapnoea is well-documented. These observations make the euoxic hypercapnoeic protocol ideal for testing monitoring techniques under fast breathing regimes, as they allow for the assessment of the sensitivity our methods in detecting incremental changes in respiratory rate.

Appendix B

Data pre-processing in the NICU dataset

B.1   Synchronisation between different devices

The PPG, ECG and IP signals displayed by the patient monitor were acquired using different devices described in Section 3.3.3.1. The task of recording the time points at which measurements are taken is left to the clock inside each device, even before data is relayed to the patient monitor. This gives rise to a delay between the raw signals collected by different devices (PPG the from oxygen saturation foot probe and ECG and IP from the chest electrodes). Such a delay also applies to the physiological parameters derived at device level, *e.g.* heart rate from ECG (HR_{ECG}) and from PPG (HR_{PPG}), and thus it needed to be measured and compensated for before signals from different sources could be compared.

It is important to understand whether the observed time differences were due to a fixed delay or a clock drift. As the relationship between the salient points in the ECG and PPG waveforms is made complex by its dependence on the pulse transit time between the left heart and the PPG site [Smith et al., 1999] (not to mention other haemodynamic variables which vary even under normal physiology [Li et al., 2008]), we turned to the HR signals mentioned above to measure the time difference between time stamping by both sources.

First, we resampled the HR_{PPG} estimates at the times of HR_{ECG} estimates. The cross-correlation using MATLAB's *circshift* algorithm was then computed between 5-min segments of these two signals and recomputed after 2.5 min shifts. For each of these windows, we found the peak cross-correlation value and the time lag between the two

5-minute shifted segments that generated this peak (λ_{peak}).

Analysis of the $\lambda_{peak}(t)$ plots, where t is a discrete time axis with 2.5 min increments, revealed a constant profile around $\lambda_{peak} = 1$ s, with spurious ± 1 s deviations (equivalent to 1 sample) around this value (HR_{PPG} delayed in relation to HR_{ECG}). The next step was then to align both series along the time axis by applying a constant $1s$ shift to the entire HR_{PPG}, spO_2 and PPG records for that session. Our assumption is that a fixed 1 s delay exists due to the lag introduced by the devices while small variations around this value are attributed to the inherent variability associated with measuring heart rate with two different sources.

We note that while this observation accounts for the delays observed in most sessions, a few exceptions occur in sessions in which the Philips oxygen saturation probe was in use. In this configuration, the delays observed between HR_{PPG} and HR_{ECG} signals could be as high as $10s$. This is not a problem after cross-correlation but it is nevertheless interesting to note that different instruments can can cause a considerable time shift. This highlights the need for careful synchronisation even when data sources are fed to the same patient monitor.

B.2   Parsing the key logs

Manual measurements of respiration were available for the last 5 study subjects in the NICU dataset (PN 26-30). This data consisted of the timestamps of breaths detected manually during respiratory assessment by clinical staff and logged electronically using a keylogger.

This appendix describes the processes used to parse the keylogger files produced for each of these sessions in order to (a) identify counting cycles, (b) eliminate key presses associated to UI operation, and (c) remove aborted counting cycles.

To achieve (a), key logs were split into individual counting cycles by defining a maximum interval between keystrokes belonging to the same cycle (8 s). As a result of this parsing, 209 manual counting cycles were identified in the 14 recording sessions on 5 study subjects for which the manual respiratory measurements had been logged (Figure B.1a).

Manual measurements were taken during periods for which the infants were stable and at rest. This may explain the discrepancy in the total number of measurementes performed from session to session evidenced in Figure B.1a.

In our analysis in Chapter 7, we were careful to only include counting cycles carried to completion. Thus, cycles composed of fewer than 38 keystrokes[1] (see Figure B.1b)

---

[1]Although our protocol specified 40-breath counts, a considerable number of counting cycles was found with a total of $n_c = 38$ or $n_c = 39$ breaths (Figure B.1b). We volunteer that a mismatch of a mere couple of breaths is likely to be due to human error, rather than the count being aborted for one of the

(a)   (b)
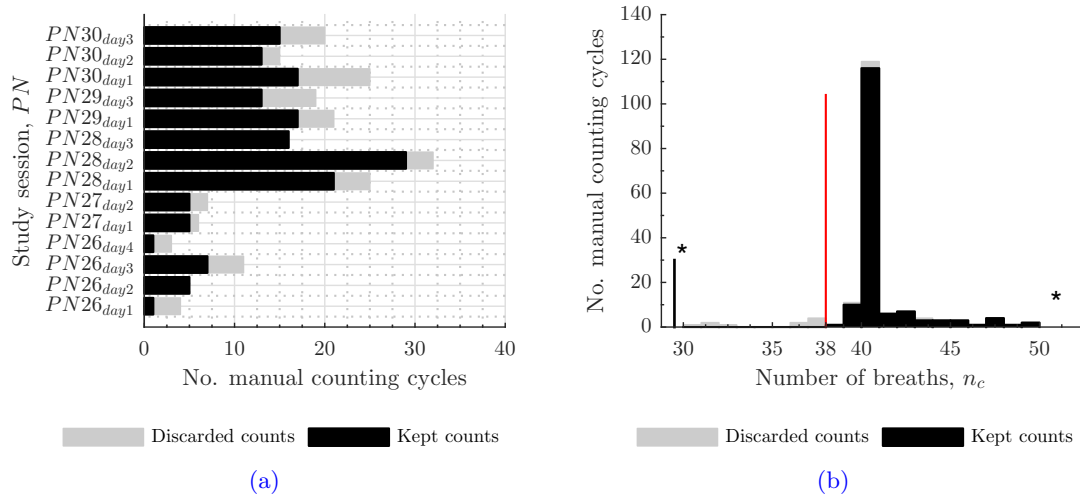
Figure B.1

Effects of the selection criteria defined on (B.1a) the distribution of the total number of manual counting cycles over the study sessions of NICU subjects PN 26 - 30, and (B.1b) the distribution of the total number of breaths over manual counting cycles . While most counts include upwards of 40 breaths, some shorter counts were concluded before this number was reached. Discarded counts with over 38 breaths were due to the absence of camera records in the course of the respiratory assessment.

Table B.1

Confusion matrix of complete manual counting cycles (with upwards of 38 breaths) *versus* camera status.

Manual counts	Camera off	Camera on	Total
< 38 breaths	0	39	39
≥ 38 breaths	5	**165**	170
Total	5	204	**209**

were considered to have been either a result of casual UI operation (case (b)) or aborted by the attending clinical staff due to excessive wriggling, infant motion or other factors which rendered breathing motion imperceptible (case (c)).

Manual counts performed when the video recording apparatus was not in use were also discarded. Table B.1 presents a breakdown of counts under each of these categories. After both selection criteria were applied, a total of 165 manual counts were considered for further analysis. Due to potential errors in subject detection via skin segmentation, further 5 counts were excluded after visual inspection of the video footage revealed that mother or nurse contact with the infants was seen at some point during these manual counts.

---

reasons specified in the study protocol.

Appendix C

Mathematical Methods

Contents

C.1 Spectral coherence analysis

Spectral coherence allows one to determine the relationship between two time series as a function of frequency.

Whilst spectral coherence analysis is a powerful tool, one must exercise caution when interpreting coherence values. In practice, the presence of coupling between two signals at a given frequency is often asserted by means of a comparison between the value of the coherence function at that frequency against a threshold value, which is customarily set to 0.5.

However, to ascertain if a given coherence value is *statistically* significant (*i.e.* that the null hypothesis of zero coherence between the input signals can be rejected), a higher threshold may be desired. In fact, this value relies not only on the desired significance level but also on the probability density that generated the coherence estimates, which in turn are heavily dependent on the specific method employed for the estimation of the signal power spectra.

This section introduces spectral coherence analysis and describes how we arrived at such a threshold from results in selected literature on the estimation of spectral coherence using modern spectral techniques based on Welch's method.

Spectral coherence

Analytically, the *complex* coherence function between two discrete time series $x(n)$ and $y(n)$ is defined by [Marple, 1986][1]:

$$\gamma(f) = \frac{\mathcal{P}_{xy}(f)}{\sqrt{\mathcal{P}_{xx}(f)}\sqrt{\mathcal{P}_{yy}(f)}} \tag{C.1}$$

where $\mathcal{P}_{xy}(f)$ is the cross-spectral density function between $x(n)$ and $y(n)$, and $\mathcal{P}_{xx}(f)$ and $\mathcal{P}_{yy}(f)$ are the power spectral densities of $x(n)$ and $y(n)$ respectively.

Hence, the magnitude-squared coherence (often referred to simply as spectral coherence) is a real-valued function given by

$$\begin{aligned}
\gamma^2(f) &= |\gamma(f)|^2 \\
&= \frac{|\mathcal{P}_{xy}(f)|^2}{\mathcal{P}_{xx}(f)\mathcal{P}_{yy}(f)}
\end{aligned} \tag{C.2}$$

As can be seen from Eq.C.2, γ^2 is the frequency domain analogue of the correlation coefficient routinely used in linear regression (see Penny [2000] for a summarised proof). Similarly, γ^2 is strictly positive and confined to $0 < \gamma^2 < 1$.

Under assumptions of signal ergodicity and stationarity, $\gamma^2(f)$ is equal to unity if and only if a linear relationship is observed between the amplitudes of two signals at the same frequency component f. To that effect, γ^2 can be regarded as an index of signal coupling in the frequency domain.

Statistical analysis

In practice, any power spectral density functions calculated numerically carry a variance inherent to the estimation process. In the context of the methods just described, this variance deviates the estimated coherence from its expected value.

Thus, nonzero coherence values may arise at some frequency bins even in the case of complete decoupling between two time series (*i.e.* expected coherence = 0). This problem calls for statistical tests to determine a threshold level in the magnitude of estimated coherence above which the relationship between the two series at a given frequency is deemed significant.

The process whereby one finds this value can be based on either theoretical arguments (grounded on known mathematical equalities) or empirical ones, driven by the time series data. The common objective is to determine the distribution of the coherence estimator

---

[1]And obviously not defined at frequencies k where $\mathcal{P}_{xx}(f) = 0$ or $\mathcal{P}_{yy}(f) = 0$.

under the null hypothesis that no frequency correlation exists between the two time series.

The empirical approach relies on the use of bootstrapping techniques [Faes et al., 2004]. Upon randomly reordering one of the time series for which coherence was computed, it makes intuitive sense that the resulting series should likely have null coherence with the other series. Algorithmically, after the random resampling of one of the series the coherence between these series is computed. This operation is repeated a large number of times to obtain a distribution of coherence estimates for all frequencies for uncorrelated time series. From this estimated distribution, a coherence threshold is derived given a specified significance level [Faes et al., 2004].

As an alternative to these approaches based on surrogate data, given knowledge of the true probability distribution of the estimated spectral coherence estimates for each f (henceforth $\widehat{\gamma}^2(f)$) one may derive a coherence threshold over which the hypothesis of zero coherence between the signals can be rejected in a more principled manner.

This leads to a probabilistic definition of the threshold value $(\gamma^2_{thresh}(f))$ in terms of the probability of Type I errors (α):

$$P(\widehat{\gamma}^2(f) > \gamma^2_{thresh}(f) \mid \gamma^2(f) = 0) = \alpha \tag{C.3}$$

A theoretical expression for the probability distribution $P(\widehat{\gamma}^2(f) \mid \gamma^2(f) = 0)$ has been derived by Koopmans [1995] for the particular case of Gaussian input signals and in the case of non-overlapping segments. Koopmans demonstrated that if the coherence spectra are computed using Welch's method [Welch, 1967], the estimated spectra follow a χ^2-distribution with $2K$ degrees of freedom for each frequency bin f in the case of K independent non-overlapping segments. This derivation yielded the following relation between γ^2_{thresh} and α:

$$\gamma^2_{thresh}(\alpha) = 1 - \alpha^{\frac{1}{K-1}} \tag{C.4}$$

Originally proposed by Welch, the use of overlapping segments efficiently reduces the variance of the spectral estimates by using a larger part of the original signal. The next section addresses how Equation C.4 can be modified to permit its use when coherence spectra are computed using overlapping segments.

Relation to Welch's Overlapped Segment Averaging Spectral Estimation

In practice, the computation of the true power spectral density of a signal through the computation of the discrete-time Fourier transform of the samples of the process can be numerically unstable. To mitigate this, algorithms based on the Welch's method are widely used to estimate PSD functions [Welch, 1967]. These algorithms work by splitting the signal into a number of segments N_{Welch} each of length T_{Welch} and overlapped by a D_{Welch} amount of time (which may be null). The spectra for each segment is computed with a Fourier transform, and then averaged to produce a global estimate of the power spectral density of the signal. For a wide-sense stationary process, the PSD estimates of the different segments of the time series provide uncorrelated estimates of the true PSD and thus averaging reduces the variability of this estimate. However, because the independence between segments is lost, the distribution of $\gamma^2_{thresh}(f)$ is no longer the one found above (in Equation C.4). Recent years have seen different approaches to address this matter. In a recent study by Gallet and Julien [2011], a modification of the relation by Koopmans was proposed for taking into account the effect of the overlapping procedure. It was proposed that for the case of overlapping segments, the threshold level $\gamma^2_{thresh}(\alpha)$ can be obtained by replacing K for $\widetilde{K}$ in Equation C.4.

$$\gamma^2_{thresh}(\alpha) = 1 - \alpha^{\frac{1}{\widetilde{K}-1}} \tag{C.5}$$

where $\widetilde{K} = \frac{K}{c_w(D)}$ with $c_w(D)$ and ρ_w defined as

$$c_w(D) = 1 + 2\sum_{j=1}^{K} \frac{K-j}{K} \rho_w^2(jD) \tag{C.6}$$

$$\rho_w(M) = \frac{\sum_{t=0}^{L-M-1} w_L(t)w_L(t+M)}{\sum_{t=0}^{L-1} w_L^2(t)} \tag{C.7}$$

where w_L is the windowing function (*e.g.* Hanning) used in the computation of the modified periodogram for each of the overlapping segments .

Summary

We obtain in Equations C.4 and C.5 the desired relation between γ^2_{thresh} for a set significance level α and the parameters used for estimation of the coherence spectrum via the Welch's overlapped segment averaging method. The use of a lower $\widetilde{K}$ accounts for the reduction of the degrees of freedom of the χ^2-distribution describing the coherence values for a given frequency that results from using overlapping segments in estimating cross and auto-spectra.

C.2 Auto-regressive spectral analysis

Spectral methods are used to describe how the power of a time series is distributed with frequency.

In this thesis, parametric methods have been applied to estimate the spectral content of physiological signals. These are especially suited for signals with low signal-to-noise ratio, as often happens with biophysical signals, since they offer a significant improvement in frequency resolution when compared to traditional (non-parametric) methods, such as the Fast Fourier Transform (FFT). Another advantage of parametric methods is that they avoid spectral leakage due to the implicit windowing that occurs when processing the periodogram [Kay and Marple, 1981].

By definition, parametric methods rely on estimating the parameters of a model chosen *a priori* to describe the observed signal. Autoregressive (AR) methods are the most widely used parametric methods for power spectral analysis and the ones we use here. The reasons for selecting AR models over traditional Fourier approaches, are well-documented [Pardey et al., 1996].

The general framework of auto-regressive analysis assumes that the generation of an equally-spaced real-valued signal $x(n)$ can be described by an all-pole linear filter driven by a white noise process $e(n)$, as shown in Equation C.8.

$$x(n) = \sum_{k=1}^{N} a_k x(n-k) + e(n) \tag{C.8}$$

where a_k, $k = 1, \cdots, N$, are the coefficients of the N$^{\text{th}}$ order AR model, n is the discrete time index and $e(n)$ is a zero-mean white noise with variance σ_e^2.

As can be seen from Equation C.8, such a description of $x(n)$ is equivalent to expressing $x(n)$ as a linear regression on itself with an associated prediction error $e(n)$ (*i.e.* the error between the actual value and the predicted value) assumed to be normally distributed with zero mean.

Several algorithms are available for estimating the a_k coefficients of the AR model. Here, these parameters were determined by solving the Yule-Walker equations in order to minimise the total sum of least squares errors as given by both the forward and the back prediction error formulae (Kay and Marple [1981],p.1388-1399), while using Burg's recursion [Burg, 1975] for optimisation.

Equation C.8 can be expressed in the z-domain as

$$X(z) = \frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}} X(z) + E(z) = H(z)E(z) \tag{C.9}$$

using a transfer function $H(z)$ defined as

$$H(z) = \frac{1}{1 - \sum_{k=1}^{N} a_k z^{-k}} = \frac{z^N}{\prod_{k=1}^{N}(z - z_k)} \tag{C.10}$$

As shown in Equation C.10, the denominator of $H(z)$ can be factorised into N terms. Each of these terms has a complex root z_i corresponding to a pole of the transfer function $H(z)$. Since $H(z)$ has no finite zeros, the AR model is an all-pole model.

The poles occur in complex-conjugate pairs and (in the discrete case) their coordinate position in the z-plane with respect to the unit circle determines the magnitude of the frequency response of the all-pole filter. The resonant frequency of each spectral peak is given by its phase angle, by substituting Equation C.11 in C.9.

$$z = e^{j2\pi \frac{f}{f_s}} \tag{C.11}$$

with the phase angle defined as $\theta = 2\pi \frac{f}{f_s}$ for a pole at frequency f. The magnitude of the pole indicates the signal strength at that frequency.

Finally, given a set of estimated AR parameters $\{\hat{a}_k\}_{k=1,...,N}$, an estimate of the power spectrum can be found by taking the z-transform and applying the convolution theorem as in Proakis [1975]:

$$\mathcal{P}_{xx}(f) = \frac{\hat{e_N}}{\mid 1 - \sum_{k=1}^{N} \hat{a_k} e^{-j2\pi fk} \mid^2} \tag{C.12}$$

where $\hat{e_N}$ is the total least-squares error.

C.3 Notes on the Gaussian filter

This section is concerned with the practical aspects of Gaussian filtering. When smoothing signals using Gaussian kernels, one needs to establish the relation between the standard deviation of the filter (σ) and its frequency cut-off k_c. This happens because while the filter specifications are given in terms of k_c, the implementation is generally in terms of σ. In this section, we derive expressions for the filter magnitude and power responses, obtain the desired relationship and interpret its meaning in the context of anti-aliasing filters.

The one-dimensional Gaussian filter has an impulse response function given by [1]

$$g(x) = \sqrt{\frac{a}{\pi}} e^{-ax^2}, a \in \mathbb{R}^+ \tag{C.13}$$

---

[1]Although the expressions presented here are written for the continuous case, we will later see how they can be applied to the discrete case.

Equivalently, as the Gaussian filter can be expressed in terms of its standard deviation $\sigma = \frac{1}{\sqrt{2a}}$, we can write

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}} \tag{C.14}$$

Considering the integral transformation defined by[1]:

$$f(x) \xrightarrow{\mathcal{F}} \mathcal{F}(k) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi i k x}\mathrm{d}x \tag{C.15}$$

and where k is the ordinary (in this case, spatial) frequency, this kernel has a transform given by:

$$\mathcal{G}(k) = e^{-\frac{\pi^2 k^2}{a}} \tag{C.16}$$

or

$$\mathcal{G}(k) = e^{-\frac{k^2}{2\sigma_k^2}} \tag{C.17}$$

with σ_k defined as $\sigma_k \equiv \frac{1}{2\pi}\sqrt{2a}$ and related to the kernel σ through:

$$\sigma\,\sigma_k = \frac{1}{2\pi} \tag{C.18}$$

Based on Equation C.17, we can compute the filter magnitude response as:

$$|\mathcal{G}(k)| = \sqrt{Real^2(\mathcal{G}(k)) + Im^2(\mathcal{G}(k))} = e^{-\frac{k^2}{2\sigma_k^2}} \tag{C.19}$$

Thus, we obtain the well-known result that the magnitude response of a Gaussian filter retains its Gaussian shape.

For the purpose of designing an appropriate anti-aliasing filter, we seek a relationship between a given cut-off frequency k_c and the corresponding magnitude and power response of the filter at this frequency. Given an arbitrary cut-off value $\frac{1}{c}$ in the magnitude spectrum, Equation C.19 yields

$$k_c = \sqrt{2\ln c}\,\sigma_k \tag{C.20}$$

---

[1]$\mathcal{F}(k)$ is usually called the Fourier transform of $f(x)$.

By substitution of Equations 5.6 and C.18 in Equation C.20, we obtain

$$SF k_s = 2\sqrt{2\ln c}\,\frac{1}{2\pi\sigma} \tag{C.21}$$

The Gaussian kernel is continuous and extends from $x \in (-\infty, +\infty)$. In practice, digital images are processed by discrete filters. The sampled Gaussian filter (Equation C.22) is a commonly used discrete equivalent of the Gaussian kernel that implements a sampled version of Equation C.13 with $x = nT_s, n=0,1,..,N$, where T_s is the sampling period of the filter an N is the filter size in samples.

$$g(n) = \sqrt{\frac{a}{\pi}}e^{-a(nT_s)^2}, a \in \mathbb{R}^+ n = 1, ..., N \tag{C.22}$$

Converting σ in Equation C.21 to sample units ($\frac{\sigma}{k_s}$), one obtains

$$SF = \sqrt{2\ln c}\,\frac{1}{\pi\sigma} \tag{C.23}$$

which is the relationship we seek. It relates (a) the scale factor used for decimation (SF), (b) the filter magnitude response at the cut-off frequency ($\frac{1}{c}$), and (c) the filter σ one should select in order to observe the Nyquist criterion. The inverse proportionality between SF and σ as evidenced by Equation C.23 means that for stronger decimation (lower scale factor) a broader Gaussian kernel is required.

We note that, in Equation C.23, σ is expressed in sample units. We have opted to preserve the term σ for notational simplicity though it should not be confused with the σ in previous equations in this section, where it is given in physical units.

One question remains related to the choice of c. While for common linear filters, such as Butterworth or Chebyshev designs, there are generally accepted definitions of cut-off frequencies, for Gaussian kernels, several definitions co-exist and are used in practice. From a statistics point of view, k_c of a Gaussian filter can be regarded as the standard deviation of the magnitude response of this kernel (σ_f). Alternatively, k_c can also be defined as the frequency corresponding to the Full Width Half Maximum, that is, the frequency at which the magnitude response is reduced to $\frac{1}{2}$ (*ie.* $c = 2$). In digital signal analysis, it is common to use the frequency at which the response in the power spectrum P_{xx} is halved, or reduced to 3 dB, as the filter cut-off. Given that $P_{xx}(k) = \mathcal{G}(k)\mathcal{G}^*(k) = \mathcal{G}^2(k)$, this corresponds to setting $c = \frac{1}{\sqrt{2}}$. This is the definition we adopt here. Figure C.1 illustrates the difference in the magnitude response of a Gaussian filter $|\mathcal{G}(k)|$ for these three definitions of k_c.

A Gaussian kernel is parametrised by its σ and size (in samples). Generally, these two quantities are related as the choice of filter size should depend on σ (upon which the

Figure C.1
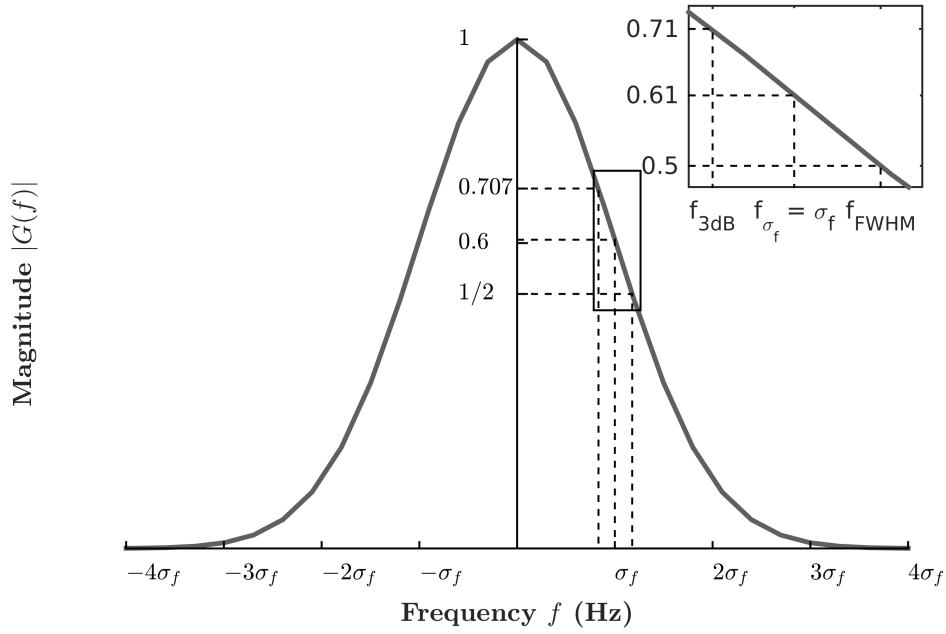Magnitude response of the Guassian filter $|\mathcal{G}(k)|$ and its value for different definitions of cut-off frequency.

filter frequency cut-off k_c also depends, as we have already seen). The knowledge that a Gaussian PDF approaches 0 at about three standard deviations of the mean (99% of the distribution density falls within 3σ from the mean), led us to adopt a filter size of 3σ.

Appendix D

Segmentation of photoplethysmographic pulses

Photoplethysmography measures blood volume changes over the microvascular supply of a transilluminated measurement site, generally a finger or an ear lobe. In general, these signals bear great morphological resemblance to the arterial blood pulse (ABP), the pressure waveform (depicted in Figure D.1) which forms with the sudden ejection of blood into the aorta and propagates throughout the arterial tree.

The ABP waveform is rich in physiological information. In addition to its spectral content, modulated by the pulsatile nature of the cardiac output, the respiratory cycle, and vasomotor, Traube Hering Mayer, and thermoregulatory activity [Allen, 2007], it carries morphological information in the shape of easily recognisable waveform features.

These features are introduced at specific events during the cardiac cycle and are transmitted, damped or distorted, as the pulse travels through the arterial system. Among the commonly reported waveform features in a pulse cycle are (a) the onset of the arterial blood pulse, (b) the diastolic peak, (c) the percussion peak, and (d) the dichrotic notch [Elgendi, 2012]. Not only do these features provide valuable insights into the haemodynamic properties of the arterial tree and circulating blood [O'Rourke, MF and Kelly RP, Aviolo, 1992] but can also inform strategies for the segmentation of cardiac pulses and thus beat-by-beat analysis of this waveform.

The detection of these fiducial marks from the ABP waveform is an active area of research. This appendix provides a comparative assessment of three algorithms available in the public literature for beat pulse segmentation in large-scale ABP waveforms and in doing so substantiates our choice of method to perform this task in Chapter 4. We have selected for this analysis three methods which have undergone systematic performance analysis. We sought to ascertain whether there were measurable differences in the PPG
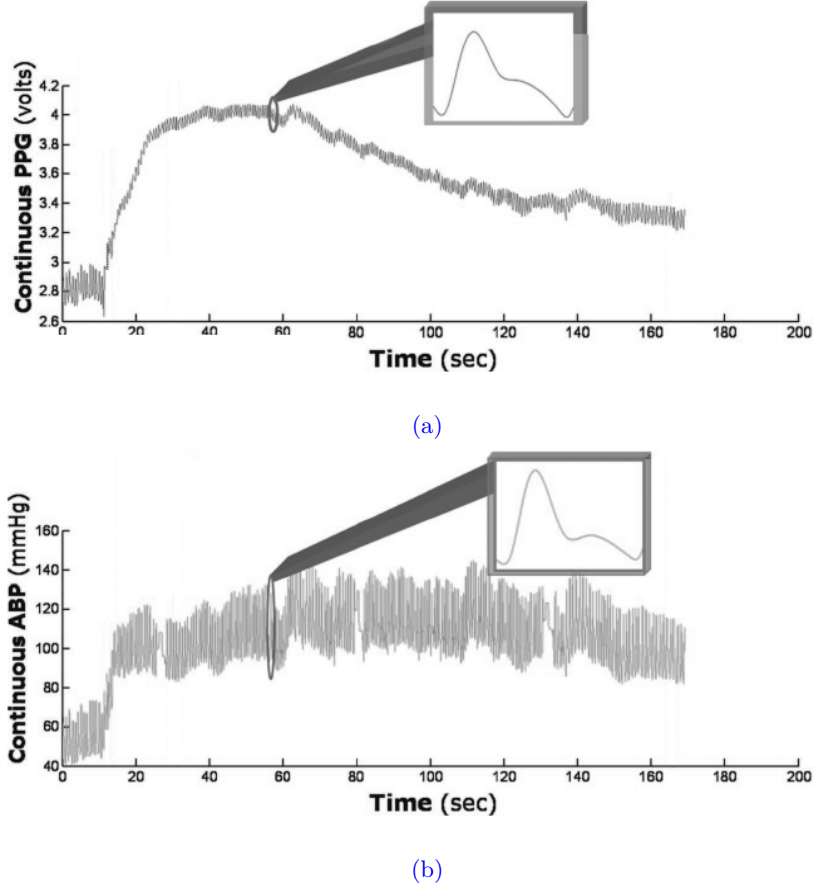
PPG (D.1a) and ABP (D.1b) signals, measured in an ipsilateral hand, using a laboratory prototype PPG sensor and a Finapres® (Finapres Medical Systems, Amsterdam, The Netherlands), respectively. The similar profile of the two waveforms is illustrated in the inlays. Data collected in a laboratory setting by Reisner et al. [2008] and reproduced here with permission.

wave segmentations provided by these algorithms. Specifically, we assessed the accuracy of (a) pulse peak detection and (b) pulse onset detection, *i.e.* the detection of the main fiducial points reported in pulse wave analysis.

Overview of ABP beat detection

Most algorithms for feature detection in physiological signals can be subdivided into two processing stages. First, a pre-processing stage emphasizes the desired components of the signal morphology. This stage traditionally relies on digital filters implementing signal derivatives, FIR or IIR filters although recent algorithms also employ wavelets and filter banks. A decision stage then follows to determine whether a given candidate is a true feature based on a set of user-specified rules. This architecture is a common one, which can be found, for example, in most QRS detection algorithms in electrocardiographic studies (see, *e.g.* Pan and Tompkins [1985]).

One the first effective attempts at locating arterial pulses in continuous ABP signals

is found in Zong et al. [2003]. In this study, an integration filter with a kernel size that approximates the width of ABP pulses was applied to enhance the up-slope in the anachrotic phase of the pulse while suppressing the rest of the pressure waveform. Subsequent adaptive thresholding on this signal (allied to local search strategies) allowed detected pulse onsets to be placed in close proximity to the onsets annotated in ECG traces from the MIT-BIH Polysomnographic database [Goldberger et al., 2000; Ichimaru and Moody, 1999] (99.31% sensitivity on 368,364 beats) and accurate (< 20 ms) onset detection in 96.41% of 39,848 beats from manual annotations.

Conversely, the pulse segmentation algorithm by Li et al. [2010] explored the properties of the ABP signal derivatives. The authors noted that onsets and systolic peaks in the original ABP waveform were related to zero-crossing points in the first-order derivative of the (low-pass) filtered ABP signal in the following way; zero crossings in the first derivative that corresponded to onsets in the original signal occurred before a zero-crossing in the second derivative (*i.e.* on the inflection point in the original signal) whilst those corresponding to systolic peaks occurred only after this inflection. Once a beat was detected in this way, the authors successfully found the dichrotic notch between that beat and the preceding one using similar derivative analysis.

The performance of this algorithm for the detection of the three fiducial points was carried out on both open-access [Aboy et al., 2005; Goldberger et al., 2000; Ichimaru and Moody, 1999; Iyengar et al., 1996] and in-house databases (to evaluate dichrotic notch detection). As for beat detection, Li et al's method achieved an average sensitivity of 99.43% and a positive predictive power of 99.45%.

Using algorithms originally developed for computer vision, Karlen et al. [2012] pursued a segmentation strategy based on Iterative-End-Point-Fit [Borges and Aldon, 2004]. In summary, the up-slopes and down-slopes in ABP were identified by drawing line segments delimiting segments over which the gradient of the ABP is consistent. Systolic peaks were then identified as the endpoints of each up-slope. The Complex Systems Laboratory (CSL), Portland USA [Aboy et al., 2005] dataset acquired at a paediatric intensive care unit was used for performance evaluation. The algorithm achieved a sensitivity of 98.93% and positive predictivity of 96.68% on 16,155 beats.

Regarding pre-processing of the ABP signal, all approaches described above applied a filtering step to suppress the high frequency noise in this waveform. Zong et al. employed a 2nd order low-pass filter. Li et al. opted for a 3rd order low-pass Bessel filter. In the work of Karlen et al., the signal is filtered by a 20th order high pass linear phase, FIR filter, which uses a Kaiser window with a β equal to four and a cut-off frequency of 0.05 Hz (3 bpm).

Methods

We implemented three methods above and analysed their results on the finger-PPG traces of a subset of 12 volunteers who underwent the Hypercapnoea protocol described in Section 3.2. A total of 8.62 hours of PPG records for their sessions were processed in non-overlapping 30-second windows.

(a)

(b)

(c)
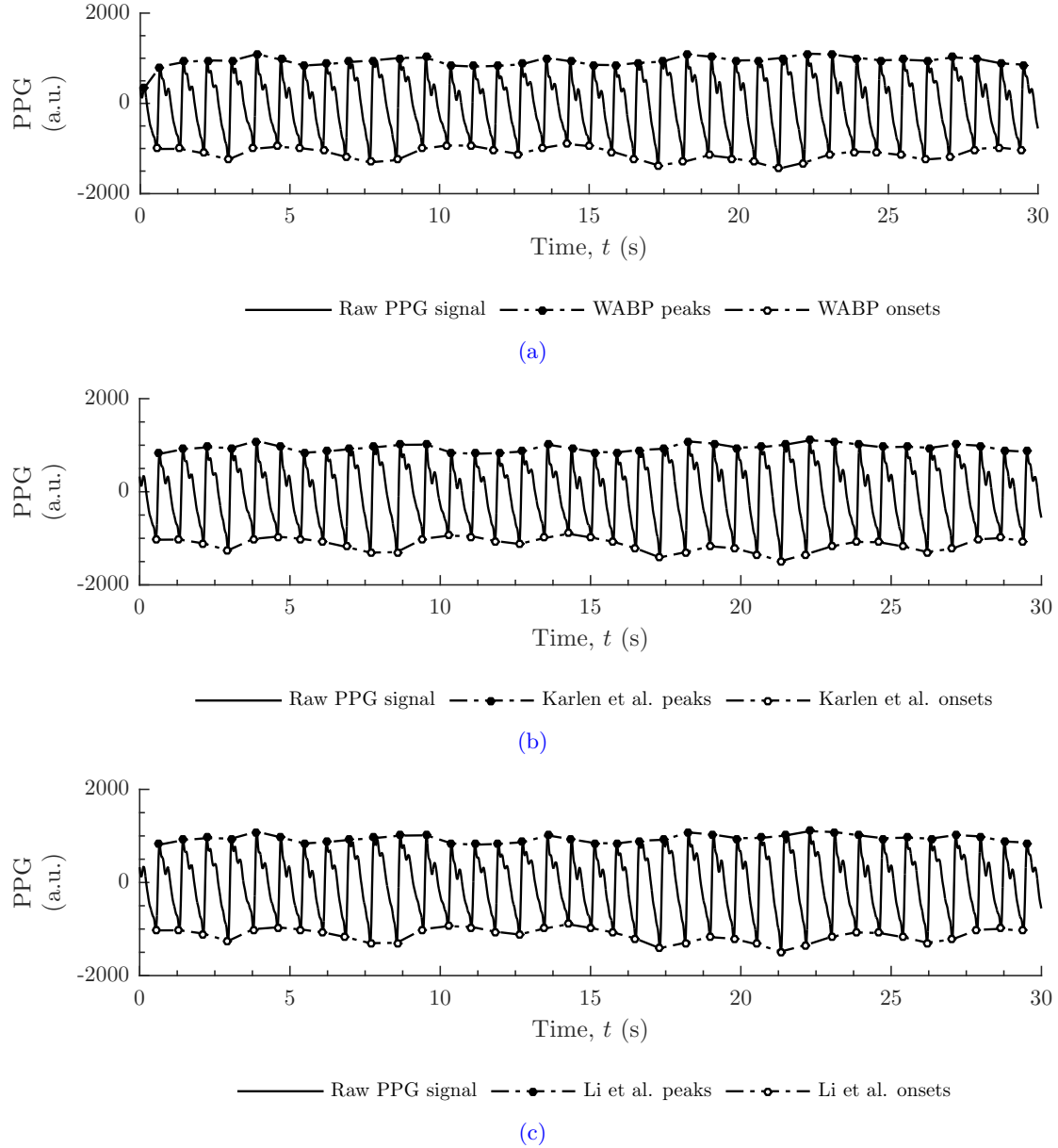
Figure D.2
Example of PPG waveform segmented by WABP (D.2a), Karlen et al. (D.2b) and Li et al. (D.2c) for a sample 30s time window. Detected peaks and onsets marked by dots and circles, respectively.

A note should be made regarding the WABP peak locations reported. In its original form, the detector by Zong et al. segments ABP pulses by estimating the location of their onsets. As peak locations are essential to the extraction of respiratory rhythms
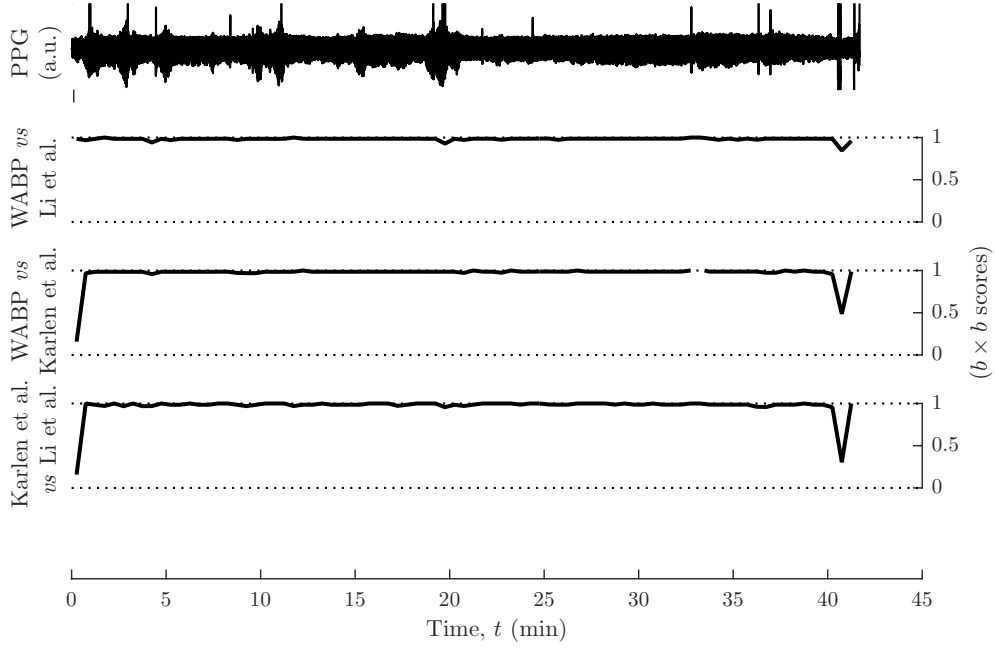
201

Figure D.3

Comparison of the three ABP onset detectors for a 42-minute-long study session. (Top) raw PPG signal (Middle to bottom) b×b scores computed for three pairings of three methods of ABP segmentation tested. The statistics shown were computed over 30 s time windows with no overlap.

from amplitude-modulated plethysmography, we have computed peak locations from WABP onset locations in the following way; each time window was partitioned in $n + 1$ disjoint intervals so that the endpoints of each segment were given by the time instants corresponding to the n WABP onset locations and the two window endpoints. From the set of local maxima present in each of these time intervals we have selected the WABP PPG peak as the one which occurs first in each interval. If an interval had no local peaks, no peak was considered.

Results

Figure D.2 illustrates the performance of the three algorithms on a sample time window. Both peak and onset matching was performed using a standard beat comparison algorithm (b×b developed by [Moody, 2012]). When testing the agreement between detected events, one needs to define a window for temporal matching. Standards for accurate pulse detection have been set by American National Standards Institute [2012](AAMI) so that meaningful comparisons between pulse locations returned by different algorithms may be performed. Following their guidelines, we set the length of this window to 150 ms. Table D.1 reports b×b scores for peak and onset locations for each method pairing (for a total of three possible combinations) for a random selection of 12 study sessions in the Hypercapnoea dataset. These are the same sessions analysed in Chapter 4. To examine the time-dependence of the scores presented in Table D.1, Figure D.3 compares the results of ABP onset detection on a sample record. In both cases, the statistics reported were

| | | b×b onset comparison | | b×b peak comparison | |
Method	No. beats	Li et al.	Karlen et al.	Li et al.	Karlen et al.
WABP (2003)	32552	0.95 (0.07)	0.93 (0.14)	0.95 (0.06)	0.93 (0.14)
Li et al. (2010)	30523	-	**0.96 (0.14)**	-	**0.96 (0.14)**
Karlen et al. (2012)	29974	-	-	-	-

Table D.1

Total number of beats (onsets) detections and b×b scores for ABP peak and onset locations returned for all pairings of three methods for ABP segmentation tested on a subset of 12 study sessions in the Hypercapnoea dataset. The statistics reported were computed over 30 s time windows with no overlap.

computed over 30 s time windows with no overlap.

Discussion

As evidenced in Table D.1, there is a clear agreement between both peak and onset locations computed by the three algorithms. This agreement is marginally higher when segmentations provided by Karlen et al, and Li et al delineator are compared, both in peak (0.96 ± 0.14) and onset (0.96 ± 0.14) b×b scores.

In the absence of ground-truth data, we cannot conclude which of the proposed techniques presents the highest accuracy for this dataset. We can however hypothesize that given the high agreement scores (> 0.95) the detected peaks and onsets are placed in close proximity regardless of our choice of algorithm. Since these fiducial points are the primitives in the computation of the respiratory modulators of PPG, we conculde that the choice of any one algorithm for ABP beat detection among the ones tested should not impact the results of the subsequent extraction of these rhythms.

A closer look at Figure D.3 reveals how agreement scores are distributed over the course of a typical study session. The low b×b scores at the start and end of the recording are due to to equipment setup and removal. These findings lend support to the hypothesis that the inclusion of artefactual PPG traces in our analysis may be responsible for some lack of agreement between the beat segmentation methods tested, when compared with b×b scores in previous studies. Hence, improvements on the scores reported in this appendix could potentially be achieved if windows with noisy PPG traces were excluded from the beat comparison, *e.g.* by establishing a minimum signal quality threshold value for inclusion in our analysis. The reason this has not been performed is this study is the following. Most signal quality measures for the PPG are based on the detection of abnormalities in the normal pulse morphology. Usually, morphological SQIs require a signal template of the repeating unit in the signal (beats in the case of ABP waves). Hence, the computation of an effective measure of PPG signal quality is intrinsically linked to the output of a given beat detector, whose performance is the very object of our study in this appendix.

Conclusions

The main findings in this appendix (summarised in Table D.1) show that any two of the three algorithms under comparison for ABP segmentation yield a similar ratio of synchronously beats according to the recommendations of the American National Standards Institute (ANSI/AAMIEC571998) when applied to the PPG traces collected in 12 Hypercapnea sessions. Further analysis highly suggests that discrepancies in the performance of these detectors may be due to noise in the raw PPG.

In conclusion, the peaks and onset locations (and resulting segmentation of ABP) provided by the three algorithms are in good agreement, and therefore, their use for the purpose of computing PPG SQIs or respiratory rhythms should yield similar results. Given the relatively similar principle of operation (based on signal derivatives) shared by these techniques, similar computational costs in their computation are expected. In this thesis, we have applied Karlen et al.'s method for PPG pulse segmentation, which showed marginally faster running times on the machine used for data analysis.

Appendix E

Screening of potential COBEs: a sensitivity analysis

Motivation

In Chapter 8, we sought to label the respiratory pauses present in the dataset of neonatal data described in Section 3.3.

Previous studies have shown that nursing documentation of apnoea is inaccurate when compared with event monitoring using electronic records [Razi et al., 1999]. In an observational study of 20 nonventilated infants of 28 to 33 weeks' gestational age, 99 % true apnoea episodes ($n = 211$) were recorded on bedside monitors but only 7.6 % of apnoea occurrences were recorded by nursing staff [Amin and Burnell, 2013]. The increased sensitivity of bedside monitors comes at the expense of specificity as only 50.4 % of apnoea events recorded by bedside monitors were confirmed as true apnoea.

As the NICU dataset comprises over 400 hours of monitoring, it was impractical to label all such occurrences through visual inspection of the cardiorespiratory and/or camera signals. There is an absence of well-established signal processing methods to classify neonatal apnoea/hypoapnoea episodes in the physiological signals recorded with patient monitors. We therefore adopted an hybrid approach, whereby the IP signal was first screened using an heuristic method to identify all potential cessations in the NICU dataset. These in turn were validated by a team of neonatal experts.

The question of whether all actual cessations are correctly identified using this approach (*i.e.* its sensitivity) is an important one as it impacts our ability to adequately train and test the camera-based detection methods developed subsequently. Assuming the clinical review team has ideal performance characteristics for this task, the sensitivity

of our approach is bound by the sensitivity of the algorithm used for screening the IP signal.

The algorithm used in Chapter 8 was based on a set of simple rules implemented in the early warning system built into patient monitors. A minimum threshold was set for the derived respiratory rate (RR_{IP}) for breathing to be considered to be present in the IP signal ($RR_{IP} < 20$ bpm for a period of at least 20 seconds). The application of this criterion may not capture short respiration pauses, whose duration is insufficient for RR to reach the threshold to set off the monitor alarms but which nevertheless warrant clinical attention.

In this appendix we investigate the sensitivity of our approach in a subset of randomly selected study sessions and provide further justification for the criteria adopted.

Methods

We compare cessations of breathing as detected by three different methods.

Monasterio et al.

Monasterio et al. [2012] identified apnoea-related desaturations in preterm infant admissions in the Physionet's MIMIC II dataset using a combination of spO_2 and ECG events. Desaturation events were defined as those intervals where $spO_2 < 90\%$. These events were then considered to be apnoea-related if the following conditions were fulfilled: within the interval of 300 s before the desaturation event (a) the HR decreased at least 10 beats·min^{-1} and (b) the minimum HR was < 130 beats·min^{-1}. In agreement with the authors, we refer to these occurrences as apnoea-related events (AREs).

Lee et al.

In previous work on a in-house dataset of very low birth weight patient admissions to the University of Virginia NICU, Lee et al. [2012] developed an apnoea detection algorithm that filtered the cardiogenic artefact from the IP signal. After further processing of this signal, a probability of apnoea was calculated from the 2 s standard deviation of this signal using a Fermi function. Cessations of breathing lasting n seconds were then defined as the time intervals for which the integral under the probability curve exceeded n seconds. This method has been successfully applied for the detection of apnoea events [Lee et al., 2012], the stochastic modelling of central apnoea events [Clark et al., 2016], and for the quantitative analysis of periodic breathing Mohr and Vergales [2015] and very long apnoea events [Mohr et al., 2015] (> 60 s).

We refer to cessations of breathing events detected in this way as WAD-20 [Lee et al., 2012] (an event with a 'weighted apnoea duration' of at least 20 seconds).

Patient monitor alarms

We perform this comparison against the highly-sensitive apnoea detection alarms built into patient monitors. These monitors are equipped with two alarms: (a) amber alerts and (b) red alerts. An amber alert indicates a low priority alarm (*e.g.* a respiration alarm limit violation), whilst a red alarm indicates a high priority patient alarm such as a potentially life-threatening situation (*e.g.* asystole). The apnoea alarm is triggered when the patient respiratory rate drops below 20 bpm: for values up to 10 bpm below this threshold an amber alert is registered, which is replaced by a red alert if RR drops more than 10 bpm below this threshold. In our analysis, a potential COBE was recorded when the RR decreases below 20 bpm for at least 20 s, in line with the definition provided by the American Academy of Paediatrics [Committee on Fetus and Newborn, 2003], which states that apnoeic episodes are defined as absent airflow in excess of 20 s.

We use the acronym COBE to designate cessations of breathing events detected using these criteria[1].

Results

To implement Lee et al. [2012]'s methodology, we had to recalibrate the parameters used in the Fermi function to convert values of 2-second standard deviations into apnoea probabilities for applicability in our dataset. For convenience, the parameters in the formula were adjusted through the study of two sessions, one with many labelled apnoeas (PN 023) and one with no apnoeas (PN 007, day one), as shown in Figure E.1.

The detected AREs, WAD-20s and COBEs are shown in Figures E.2 and E.3. A total of 113 COBEs (17 of which were classified as apnoeic events), 185 WAD-20s, and 1427 AREs were detected.

Discussion

We regard the high number of ARE detections in Figure E.2 with some scepticism. Whilst this approach may appear to be the most sensitive, a couple of points should be mentioned. Monasterio et al. [2012] consider events without an associated bradycardia or desaturation to be unrelated to apnoea. This is an unproven hypothesis. In fact, bradycardia occurs only in 10 % of apnoeic events with duration of 10 to 14 s, 34 % of apnoeas lasting 15 to 20 s, and 75 % of apnoeas that last > 20 s [Zhao et al., 2011]. In addition, when bradycardia is present, it may occur after the oxygen desaturation that is associated with apnoea[2], with a recent study demonstrating the onset of oxygen desaturation to be prior to that of bradycardia by a median interval of 4.2 s [Poets,

---

[1]for consistency with the nomenclature adopted in the main body of this thesis.

[2]In fact, as shown in Figure 2.3a, this is generally the case in episodes of central apnoea.
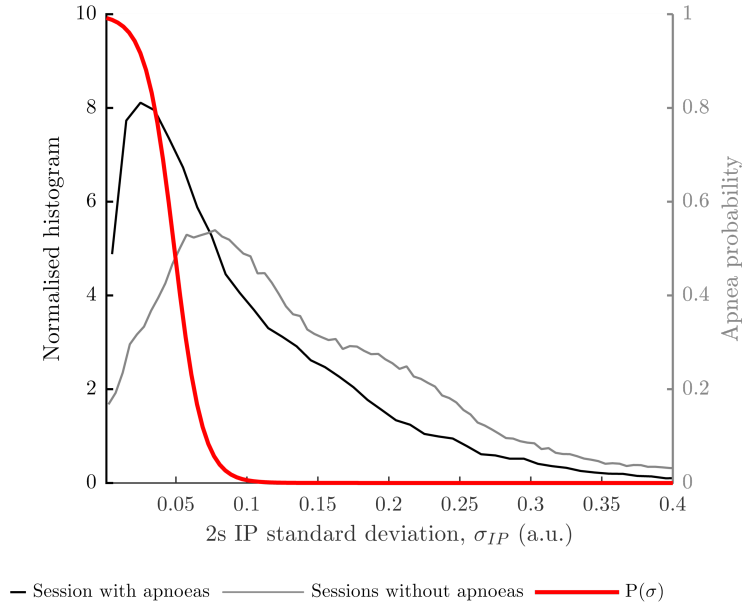
2010]. This is at odds with criteria (a) and (b) employed by the authors which search for bradycardia occurring in the 300 s before the onset of the desaturation.

A qualitative analysis of Figures E.3a and E.3b reveals a better agreement between the results provided by the two remaining approaches for the subset of sessions studied. There are however some discrepancies in the detections yielded by these two methods; most notably some events trigger a COBE but no WAD-20s, although the opposite is more frequent (185 WAD-20s *versus* 113 COBEs), and no WAD-20s were found for the true apnoeic events in PN002 (4th day), PN003 (1st day), both events in PN013 and PN030 (2nd day; 100 minutes into the session record) identified by the clinical team. While we do not exclude the possibility that, due to the fact that only COBEs were labelled, the converse would be observed (*i.e.* that the clinical review of the 185 WAD-20s events revealed true apnoeas which did not trigger a COBE) the fact that the criteria for COBE detection more closely resembles the respiratory alarms in patient monitors provides some justification for their use in our study of cessation of breathing in Chapter 8.

Conclusion

In this appendix we aimed to assess the sensitivity of the COBE detector used in Chapter 8. Ideally this question would be answered through performance analysis on a dataset of NICU physiological data with annotated apnoeas. Unfortunately, such datasets are not available in the public domain. In the absence of ground truth data, we proceeded in a manner similar to that adopted in Appendix D: we implement several methods and

Apnoea-related events or AREs detected in a subset of 32 sessions in the NICU dataset. Crosses indicate potential cessations detected. Confirmed apnoeic events (after clinical validation) are shown as red circles.
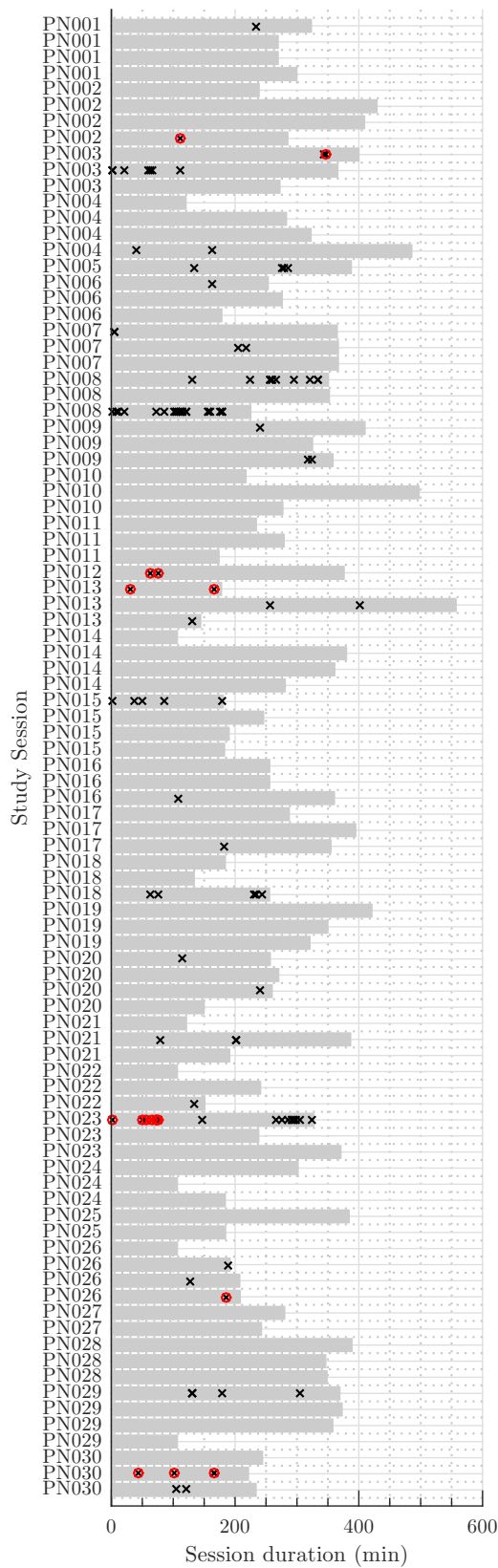
(a) COBEs.

(b) WAD-20s

Figure E.3

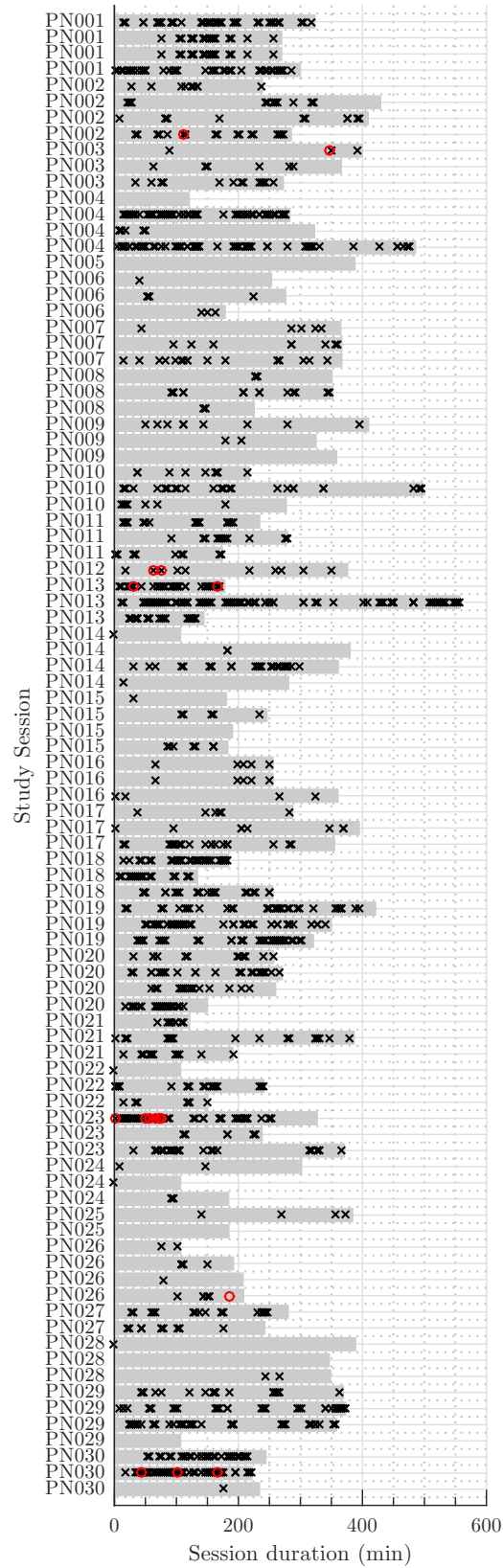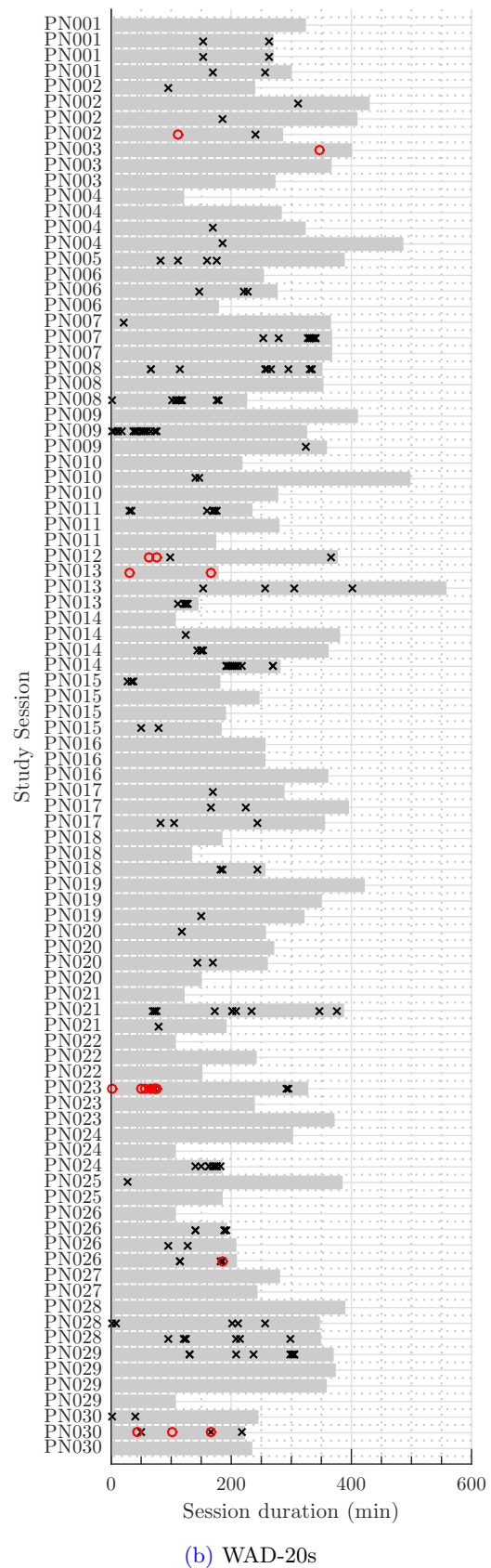Potential cessations of breathing or COBEs (E.3a) and WAD-20s (E.2) detected in a subset of 32 sessions in the NICU dataset. Crosses indicate potential cessations detected. Confirmed apnoeic events (after clinical validation) are shown as red circles.

compare the agreement between their detections in the expectation that their failure modes will not overlap.

We observed that the detections provided by Monasterio et al. [2012] generate many false positives for the detection of cessation of breathing. A better agreement is seen between COBE and WAD-20 detections. While there is no absolute agreement between these two approaches, we adopted the COBE method in our analysis of cessation of breathing as a good compromise between reproducibility and clinical relevance.

Bibliography

Aarts L., Jeanne V., Cleary J. P., Lieber C., Nelson J. S., Bambang Oetomo S., and Verkruysse W. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - A pilot study. *Early Human Development*, 89(12):943–8, 2013. 2, 33, 40, 52

Abbas A. K. and Leonhardt S. Intelligent neonatal monitoring based on a virtual thermal sensor. *BMC Medical Imaging*, 14(9):1–13, 2014. 52

Abbas A. K., Heimann K., Jergus K., Orlikowsky T., and Leonhardt S. Neonatal non-contact respiratory monitoring based on real-time infrared thermography. *Biomedical Engineering Online*, 10(1):1–17, 2011. 30, 52

Aboy M., McNames J., Thong T., Tsunami D., Ellenby M. S., and Goldstein B. An automatic beat detection algorithm for pressure signals. *IEEE Transactions on Biomedical Engineering*, 52(10):1662–1670, 2005. 64, 200

Abu-Shaweesh J. M. and Martin R. J. Neonatal Apnea: What's New? . *Pediatric Pulmonology*, 43(10):937–44, 2008. 14

Adams J. A., Zabaleta I. A., and Sackner M. A. Hypoxemic events in spontaneously breathing premature infants: etiologic basis. *Pediatric Research*, 42(4):463–71, 1997. 15

Addison P. S., Watson J. N., Mestek M. L., Ochs J. P., Uribe A. a., and Bergese S. D. Pulse oximetry-derived respiratory rate in general care floor patients. *Journal of Clinical Monitoring and Computing*, 29(1), 2014. 79

Afsar F. S. Skin care for preterm and term neonates, 2009. 29

Akselrod S., Gordon D., Ubel F. A., Shannon D. C., Berger A. C., and Cohen R. J. Power spectrum analysis of heart rate fluctuation: a quantitative probe of beat-to-beat cardiovascular control. *Science*, 213(4504):220–222, 1981. 26, 27, 76

Al-Naji A. and Chahl J. Remote respiratory monitoring system based on developing motion magnification technique. *Biomedical Signal Processing and Control*, 29:1–10, 2016. 39, 41

Albiol A., Torres L., and Delp E. Optimum color spaces for skin detection. *Proceedings of the International Conference on Image Processing 2001*, pages 3–5, 2001. 106

Alinovi D., Cattani L., Ferrari G., Pisani F., and Raheli R. Spatio-temporal video processing for respiratory rate estimation. *IEEE Instrumentation and Measurement Society*, 2015. 39, 41

Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1—-39, 2007. 27, 198

Allison R. D., Holmes E. L., and Nyboer J. Volumetric dynamics of respiration as measured by electrical impedance plethysmography. *Journal of Applied Physiology*, 19 (1):166–173, 1964. 21

American National Standards Institute . *Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms: ANSI/AAMI EC57*, 2012. 202

Amin S. B. and Burnell E. Monitoring apnea of prematurity: Validity of nursing documentation and bedside cardiorespiratory monitor. *American Journal of Perinatology*, 30(8):643–648, 2013. 205

Anand K. J. S. and Scalzo F. M. Can adverse neonatal experiences alter brain development and subsequent behavior? *Biology of the Neonate*, 77:69–82, 2000. 29

Arthur D. and Vassilvitskii S. k-means++: The advantages of careful seeding. *Proceedings of the 18th annual ACM-SIAM Symposium on Discrete algorithms*, pages 1027–1035, 2007. 112

Askanazi J., Silverberg P. A., Foster R. J., Hyman A. I., Milic-Emili J., and Kinney J. M. Effects of respiratory apparatus on breathing pattern. *Journal of Applied Physiology*, 48(4):577–580, 1980. 20

Association A. L. Respiratory distress syndrome of the newborn fact sheet, 2006. 13

Aubert E., Seps B., and Beckers F. Heart rate variability in athletes. *Sports Medicine*, 33(12):889–919, 2003. xv, 27

Baharestani M. M. An overview of neonatal and pediatric wound care knowledge and considerations, 2007. 29

Bai Y., Li W., and Yeh C.-h. Design and implementation of an embedded monitor system for body breath detection by using image processing methods. *2010 Digest of Technical Papers International Conference on Consumer Electronics*, pages 2–4, 2010. xv, 35, 36, 37

Baker S. and Matthews I. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004. 32

Baltag O. Microwaves Doppler transducer for noninvasive monitoring of the cardiorespiratory activity. In *IEEE Transactions on Magnetics*, volume 44, pages 4484–4487, 2008. 2, 29

Baranoski G. V. G. and Krishnaswamy A. *Light & Skin Interactions: Simulations for Computer Graphics Applications*. Elsevier, 2010. 100

Barnhart H. X., Haber M. J., and Lin L. I. An overview on assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics*, 17(4):529–569, 2007. 74

Barrett K. E. and Ganong W. F. *Ganong's review of medical physiology*, volume 17. 2010. 9, 11

Bartula M., Tigges T., and Muehlsteff J. Camera-based system for contactless monitoring of respiration. *IEEE Engineering in Medicine and Biology Society Annual Conference*, pages 2672–2675, 2013. 36

BBCNews . 'Microsoft Kinect fastest-selling device on record'. http://www.bbc.co.uk/news/business-12697975, 2011. 37

Behar J., Oster J., Li Q., and Clifford G. D. ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Transactions on Biomedical Engineering*, 60(6):1660–1666, 2013. 66

Berthiaume Y., Voisin G., and Dagenais A. The alveolar type I cells: the new knight of the alveolus? *The Journal of Physiology*, 572(3):609–10, 2006. 8, 9

Bilmes J. a. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, 4(510):126, 1998. 111

Bishop C. *Pattern Recognition and Machine Learning*. Springer, 2007. 109

Borges G. A. and Aldon M. J. Line extraction in 2D range images for mobile robotics. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 40(3):267–297, 2004. 200

Bousefsaf F., Maaoui C., and Pruski A. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control*, 8(6):568–574, 2013. 32, 33, 35

Brand J. and Mason J. A comparative assessment of three approaches to pixel-level human skin-detection. *Proceedings of the 15th International Conference on Pattern Recognition*, 1:1056–1059, 2000. 105, 106

Brooks L. J., DiFiore J. M., Martin R. J., Baird T. M., Colton T., Corwin M. J., Crowell D., Davidson Ward S. L., Hoppenbrouwers T. T., Hufford D. R., Hunt C. E., Keens T. G., Lister G., Neuman M. R., Ramanathan R., Schafer S., Silvestri J. M., Tinsley L., Weese-Mayer D. E., and Willinger M. Assessment of tidal volume over time in preterm infants using respiratory inductance plethysmography. *Pediatric Pulmonology*, 23(6): 429–433, 1997. 24

Brouillette R. T., Morrow A. S., Weese-Mayer D. E., and Hunt C. E. Comparison of respiratory inductive plethysmography and thoracic impedance for apnea monitoring. *The Journal of Pediatrics*, 111(3):377–383, 1987. 2, 29, 163

Burba N. and Bolas M. Unobtrusive measurement of subtle nonverbal behaviors with the Microsoft Kinect. *IEEE VR Workshop on Ambient Information Technologies*, 2012. 37

Burg J. P. Maximum entropy epectral analysis. *Proceedings of the 37th Annual International SEG Meeting*, 6:0, 1975. 83, 193

Burt P. and Adelson E. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. 38

Byrnes J. *Unexploded ordnance detection and mitigation*. Springer, 2009. 30

214

Cai J. and Goshtasby a. Detecting human faces in color images. *Image and Vision Computing*, 18(1):63–75, 1999. 106, 108

Cardoso J.-F. High-order contrasts for Independent Component Analysis. *Neural Computation*, 11(1):157–192, 1999. 129

Cardoso J. and Souloumiac A. Jacobi angles for simultaneous diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1):161–164, 1996. 128

Carroll J. L. and Agarwal A. Development of ventilatory control in infants. *Paediatric Respiratory Reviews*, 11(4):199–207, 2010. 15

Cattin P. Digital Image Fundamentals - Biomedical Image Analysis. Technical report, MIAC, University of Basel, Basel, 2010. xvii, 104

Cennini G., Arguel J., Ak K., and Leest A. V. Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optics Express*, 18(5):4867–4875, 2010. 31, 34

Chadha I. A. Neonatal resuscitation: current issues, 2010. 16

Chaichulee S., Villarroel M., Jorge J., Arteta C., Green G., Mccormick K., Zisserman A., and Tarassenko L. Multi-task Convolutional Neural Network for patient detection and skin segmentation in continuous non-contact vital sign monitoring. In *IEEE Face and Gesture Recognition*, 2017. 182

Chambrin M.-C. Alarms in the intensive care unit: how can the number of false alarms be reduced? *Critical Care*, 5(4):184–188, 2001. 164

Charlton P. H., Bonnici T., Tarassenko L., Clifton D. A., Beale R., and Watkinson P. J. An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram. *Physiological Measurement*, 37(4):610–626, 2016. xv, 25, 28, 96

Chen W., Bouwstra S., Oetomo S. B., and Feijs L. Intelligent design for neonatal monitoring with wearable sensors. *Design*, (January):1–24, 2010. 29

Chon K. H., Dash S., and Ju K. Estimation of respiratory rate from photoplethysmogram data using time-frequency spectral estimation. *IEEE Transactions on Biomedical Engineering*, 56(8):2054–2063, 2009. 79

Clark M. T., Rusin C. G., Hudson J. L., Lee H., Delos J. B., Guin L. E., Vergales B. D., Paget-Brown a., Kattwinkel J., Lake D. E., and Moorman J. R. Breath-by-breath analysis of cardiorespiratory interaction for quantifying developmental maturity in premature infants. *Journal of Applied Physiology*, 112(5):859–867, 2012. 161

Clark M. T., Delos J. B., Lake D. E., Lee H., Fairchild K. D., Kattwinkel J., and Randall Moorman J. Stochastic modeling of central apnea events in preterm infants. *Physiological measurement*, 37(4):463–484, 2016. 206

Clifton D., Douglas G. J., Addison P. S., and Watson J. N. Measurement of respiratory rate from the photoplethysmogram in chest clinic patients. *Journal of Clinical Monitoring and Computing*, 21(1):55–61, 2007. 79

color reference system C. C. colorcodex.com, 2016. xvii, 105

Committee on Fetus and Newborn . Apnea, sudden infant death syndrome, and home monitoring. *Pediatrics*, 111(4):914–917, 2003. 14, 166, 207

Cretikos M. A., Bellomo R., Hillman K., Chen J., Finfer S., and Flabouris A. Respiratory rate: the neglected vital sign. *Medical Journal of Australia*, 188(11):657–659, 2008. 35

Cross K. W. and Oppe T. E. The respiratory rate and volume in the premature infant. *The Journal of Physiology*, 116(2):168–174, 1952. 10

Crowe J. A. and Damianou D. The wavelength dependence of the photoplethysmogram and its implication to pulse oximetry. *14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6:2423–2424, 1992. 33

Cui W. J., Ostrander L. E., and Lee B. Y. In vivo reflectance of blood and tissue as a function of light wavelength. *IEEE Transactions on Biomedical Engineering*, 37(6): 632–9, 1990. 140

Cunningham S., Deere S., Elton R. A., and Mcintosh N. Neonatal physiological trend monitoring by computer. *International Journal of Clinical Monitoring and Computing*, 9:221–227, 1992. 14, 15

Cysarz D., Zerm R., Bettermann H., Frühwirth M., Moser M., and Kröz M. Comparison of respiratory rates derived from heart rate variability, ECG amplitude, and nasal/oral airflow. *Annals of Biomedical Engineering*, 36(12):2085–2094, 2008. 76, 97

Daly J., Monasterio V., and Clifford G. D. A neonatal apnoea monitor for resource-constrained environments. *Computing in Cardiology 2013*, pages 321–324, 2012. 165

Dash S., Shelley K. H., Silverman D. G., and Chon K. H. Estimation of respiratory rate from ECG, photoplethysmogram, and piezoelectric pulse transducer signals: A comparative study of time-frequency methods. *IEEE Transactions on Biomedical Engineering*, 57(5):1099–1107, 2010. 79

De Chazal P., Heneghan C., Sheridan E., Reilly R., Nolan P., and O'Malley M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Transactions on Biomedical Engineering*, 50(6):686–696, 2003. 76

De Groote a., Wantier M., Cheron G., Estenne M., and Paiva M. Chest wall motion during tidal breathing. *Journal of Applied Physiology*, 83(5):1531–1537, 1997. 2, 36

De Paoli A. G., Davis P. G., Faber B., and Morley C. J. Devices and pressure sources for administration of nasal continuous positive airway pressure (NCPAP) in preterm neonates. *Cochrane Database of Systematics Reviews*, (1), 2008. 17

DeBoer S. L. *Emergency Newborn Care*. Trafford Publishing, 2004. 9

Dempster A., Laird N., and Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 39(1):1–38, 1977. 111, 118

Deshmane A. V. False arrhythmia alarm suppression using ECG, ABP, and photo-plethysmogram. Master's thesis, Massachusetts Institute of Technology, 2009. 66, 70

Di Fiore J. M. Neonatal cardiorespiratory monitoring techniques. *Seminars in Neonatology*, 9(3):195–203, 2004. 22

Dice L. R. . Measures of the amount of ecologic association between species. *Ecology*, 26 (3):297–302, 1945. 115

Dlugokencky E. and Tans P. Earth System Research Laboratory Global Monitoring Division, 2015. 21

Dobrev D. and Daskalov I. Two-electrode telemetric instrument for infant heart rate and apnea monitoring. *Medical Engineering and Physics*, 20(10):729–734, 1999. 75, 76

Dransfield D. A., Spitzer A. R., and Fox W. W. Episodic airway obstruction in premature infants. *American Journal of Disease in Children*, 137(5):441–443, 1983. 14

Duffin J. Measuring the ventilatory response to hypoxia. *The Journal of Physiology*, 584 (1):285–293, 2007. 185

Dunn O. J. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961. 83

Eckberg D. L. Sympathovagal balance a critical appraisal. *Circulation*, 96(9):3224–3232, 1997. 26

Eichenwald E. C., Aina A., and Stark A. R. Apnea frequently persists beyond term gestation in infants delivered at 24 to 28 weeks. *Pediatrics*, 100(3):354–9, 1997. 163

Elgammal A., Muang C., and Hu D. *Encyclopedia of Biometrics*, chapter Skin Detection - a Short Tutorial, pages 1–10. Springer-Verlag, 2009. 103

Elgendi M. On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews*, 8(1):14–25, 2012. 198

Ersch J., Roth-Kleiner M., Baeckert P., and Bucher H. U. Increasing incidence of respiratory distress in neonates. *Acta Paediatrica, International Journal of Paediatrics*, 96 (11):1577–1581, 2007. 1

Evans N. J. and Rutter N. Development of the epidermis in the newborn. *Biology of the neonate*, 49(2):74–80, 1986. 29

Everingham M. R., Sivic J., and Zisserman A. Hello! My name is... Buffy" – Automatic naming of characters in TV video. *Proceedings of the British Machine Vision Conference*, pages 92.1–92.10, 2006. 32

Faes L., Pinna G. D., Porta A., Maestri R., and Nollo G. Surrogate data analysis for assessing the significance of the coherence function. *IEEE Transactions on Biomedical Engineering*, 51(7):1156–1166, 2004. 191

Fairchild M. D. *Color appearance models*. Wiley & Sons, 2005. 102

Farah A. and Saatchi R. An evaluation of thermal imaging based respiration rate monitoring in children. *American Journal of Engineering and Applied Sciences*, 4(4): 586–597, 2012. 30

Fei J. F. J., Zhu Z. Z. Z., and Pavlidis I. P. I. Imaging breathing rate in the CO_2 absorption band. *27th IEEE Annual Conference on Engineering in Medicine and Biology*, 1:700–5, 2005. 30

Fernando S. and Haan G. D. Feasibility of contactless pulse rate monitoring of neonates using Google Glass. In *Mobile Health*, volume 1, pages 3–7, 2015. 2, 41

Fitzpatrick . The validity and practicality of sun-reactive skin types, I through IV. *Archives of Dermatology*, 124(6):2–4, 1988. 46

Fleming S., Tarassenko L., Thompson M., and Mant D. Non-invasive measurement of respiratory rate in children using the photoplethysmogram. *Proceedings of the Annual International Conference of IEEE Engineering in Medicine and Biology Society*, 2008: 1886–1889, 2008. 79

Fleming S., Thompson M., Stevens R., Heneghan C., Plüddemann A., MacOnochie I., Tarassenko L., and Mant D. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. *The Lancet*, 377(9770):1011–1018, 2011. 14

Fleming S. G. and Tarassenko L. A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram. *International Journal of Biological and Life Sciences*, 2(4):233–237, 2006. 79

Folke M., Cernerud L., Ekström M., and Hök B. Critical review of non-invasive respiratory monitoring in medical care. *Medical & Biological Engineering & Computing*, 41(4): 377–83, 2003. 20

Foo J. Y. a. and Wilson S. J. Estimation of breathing interval from the photoplethysmographic signals in children. *Physiological Measurement*, 26(6):1049–1058, 2005. 79

Frey U., Silverman M., Barabási a. L., and Suki B. Irregularities and power law distributions in the breathing pattern in preterm and term infants Irregularities and power law distributions in the breathing pattern in preterm and term infants. *Journal of Applied Physiology*, 85:789–797, 1998. 144

Friesen G. M., Jannett T. C., Jadallah M. a., Yates S. L., Quint S. R., and Nagle H. T. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions in Biomedical Engineering*, 37(1):85–98, 1990. 67

Gallacher D. J., Hart K., and Kotecha S. Common respiratory conditions of the newborn. *Breathe*, 12(1):30–42, 2016. 12, 13, 17

Gallet C. and Julien C. The significance threshold for coherence when using the Welch's periodogram method: Effect of overlapping segments. *Biomedical Signal Processing and Control*, 6(4):405–409, 2011. 192

Garbey M., Sun N., Merla A., and Pavlidis I. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering*, 54(8):1418–1426, 2007. 2

Garde A., Karlen W., Ansermino J. M., and Dumont G. A. Estimating respiratory and heart rates from the correntropy spectral density of the photoplethysmogram. *PLoS ONE*, 9(1):1–11, 2014. 80

Gerrits J. and Muntinghe O. Treatment of asthmatic bronchitis with drugs, administered as an aerosol; the profit of pneumotachography in judging its merits, trypsin a powerful new drug. *Acta Allergologica*, 9(2):147–58, 1955. 19

Gibson E. and Nawab U. Respiratory Distress Syndrome in Neonates, 2015. 13, 14

Gil E. and Bailon R. PTT variability for discrimination of sleep apnea related decreases in the amplitude fluctuations of PPG signal in children. *IEEE Transactions in Biomedical Engineering*, 57(5):1079–1088, 2010. 69, 70

Gil E., María Vergara J., and Laguna P. Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children. *Biomedical Signal Processing and Control*, 3(3):267–277, 2008. 71

Goldberger A. L., Amaral L. A. N., Glass L., Hausdorff J. M., Ivanov P. C., Mark R. G., Mietus J. E., Moody G. B., Peng C.-k., and Stanley H. E. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):215–220, 2000. 70, 200

Goldensohn E. S. and Zablow L. An electrical impedance spirometer. *Journal of Applied Physiology*, 14(3):463–4, 1959. 21

Goncharova I. I. and Barlow J. S. Changes in EEG mean frequency and spectral purity during spontaneous alpha blocking. *Electroencephalography and Clinical Neurophysiology*, 76(3):197–204, 1990. 71

Gonzalez R. and Woods R. *Digital Image Processing*. Prentice Hall, New Jersey, 2nd editio edition, 2002. 104, 106, 112

Grenvik A., Mcginley E., Millen E., Gooley W. L., Ph D., Safar P., and Ii F. G. G. P. Impedance pneumography. *Chest*, 62(4):439–443, 1972. 22

Gupta A. K. Respiration rate measurement based on impedance pneumography, 2011. 22

Guyton A. C., Hall J. E., and Fisher J. P. *Respiration: ventilation, circulation, & transport*. Saunders, 11 edition, 2012. xv, 2, 7, 9, 10, 14, 21, 45

Hakumäki M. O. Seventy years of the Bainbridge reflex. *Acta physiologica Scandinavica*, 130(2):177–185, 1987. 12

Hansen J. T. and Koeppen B. M. *Netter Atlas of Human Physiology*. Saunders, 2002. 7

Harris C. and Stephens M. A combined edge and corner detector. *Proceedings of 4th Alvey Vision Conference*, 1988. 39

He T., Clifford G., and Tarassenko L. Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing and Applications*, 15(2):105–116, 2006. 68

Healthwise . Respiratory system in a preterm ifant, 2014. xv, 8

Heath P. T., Balfour G., Weisner A. M., Efstratiou A., Lamagni T. L., Tighe H., O'Connell L. A. F., Cafferkey M., Verlander N. Q., Nicoll A., and McCartney A. C. Group B streptococcal disease in UK and Irish infants younger than 90 days. *Lancet*, 363(9405): 292–294, 2004. 17

Helfenbein E., Firoozabadi R., Chien S., Carlson E., and Babaeizadeh S. Development of three methods for extracting respiration from the surface ECG: A review. *Journal of Electrocardiology*, 47(6):819–25, 2014. 26, 97

Henderson-Smart D. The effect of gestational age on the incidence and duration of recurrent apnoea in newborn babies. *Journal of Paediatrics and Child Health*, 17(4): 273–276, 1981. 15

Hering E. Uber Athembewegungen des Gefassystems. *Sitzungb d k Akad d W math naturw*, 60:829–856, 1869. 11

Hermansen C. L. and Lorah K. N. Respiratory distress in the newborn, 2007. 13

Hirsch J. a. and Bishop B. Human breathing patterns on mouthpiece or face mask during air, CO_2, or low O_2. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology*, 53(5):1281–90, 1982. 20, 45

Hjorth B. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970. 70, 71

Holme N. and Chetcuti P. The pathophysiology of respiratory distress syndrome in neonates. *Paediatrics and Child Health*, 22(12):507–512, 2012. 12

Hsu R. L., Abdel-Mottaleb M., and Jain K. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):696–706, 2002. 108

Huelsbusch M. and Blazek V. Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI. *Proceedings of SPIE Medical Imaging 2002*, 4683:110–117, 2002. 32, 34

HumanBenchmark . Reaction time statistics. website, 2016. 58

Hunt R. W. G. and Pointer M. R. *Measuring Colour: Fourth Edition*. Wiley & Sons, 2011. 102

Hunter R. and Harold R. *The Measurement of Appearance*. John Wiley & Sons, New York, NY, USA, 2nd edition, 1987. 101

Ichimaru Y. and Moody G. B. Development of the polysomnographic database on CD-ROM. *Psychiatry and Clinical Neurosciences*, 53(2):175–177, 1999. 200

Indic P., Paydarfar D., and Barbieri R. Point process modeling of interbreath interval: A new approach for the assessment of instability of breathing in neonates. *IEEE Transactions on Biomedical Engineering*, 60(10):2858–2866, 2013. 145, 161

Iyengar N., Peng C. K., Morin R., Goldberger a. L., and Lipsitz L. a. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *The American Journal of Physiology*, 271(4):R1078–R1084, 1996. 200

Janssen R., Wang W., Moço A., and de Haan G. Video-based respiration monitoring with automatic region of interest detection. *Physiological Measurement*, 37(1):100–114, 2016. 2, 40, 41

Jedynak B., Xheng H., Daoudi M., and Barret D. Maximum entropy models for skin detection. Technical report, Universite des Sciences et Technologies de Lille, France, Lille, 2002. 100

Johansson a. Neural network for photoplethysmographic respiratory rate monitoring. *Medical & Biological Engineering & Computing*, 41(3):242–248, 2003. 80

Johansson A. and Hök B. Sensors for Respiratory Monitoring. In *Sensors Applications*, pages 161–186. Wiley-VCH Verlag GmbH & Co. KGaA, 2008. 21

Johansson A., Oberg P. A., and Sedin G. Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique. *Journal of Clinical Monitoring and Computing*, 15(7-8):461–467, 1999. 79, 161

Johnston W. S. and Mendelson Y. Extracting breathing rate information from a wearable reflectance pulse oximeter sensor. *Proceedings of the Annual International Conference of IEEE Engineering in Medicine and Biology Society*, 7:5388–91, 2004. 79

Jones M. J. and Rehg J. M. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002. 101, 105, 106, 108

Jordão L., Perrone M., Costeira J. a. P., and Santos-Victor J. Active face and feature tracking. In *Proceedings of the International Conference on Image Analysis and Processing*, pages 572–576, 1999. 106

Jorge J., Villarroel M., Chaichulee S., Guazzi A., Davis S., Green G., McCormick K., and Tarassenko L. Non-contact monitoring of respiration in the neonatal intensive care unit. *12th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 286–293, 2017. 5

Kakumanu P., Makrogiannis S., and Bourbakis N. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122, 2007. 100

Kamal A. A., Harness J. B., Irving G., and Mearns A. J. Skin photoplethysmography–a review. *Computer Methods and Programs in Biomedicine*, 28(4):257–269, 1989. 27

Karlen W., Turner M., Cooke E., Dumont G. A., and Ansermino J. M. CapnoBase: Signal database and tools to collect, share and annotate respiratory signals. *Annual Meeting of the Society for Technology in Anesthesia*, page 25, 2010. 70

Karlen W., Ansermino J. M., and Fellow G. D. Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications. In *Proceeding of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 3131–3134. IEEE EMBS, 2012. 64, 70, 79, 200

Karlen W., Raman S., Ansermino J. M., and Dumont G. a. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7):1946–53, 2013. 80, 97, 148, 156

Kavanagh B. P., Sandler A. N., Turner K. E., Wick V., and Lawson S. Use of end-tidal pCO_2 and transcutaneous pCO_2 as noninvasive measurement of arterial pCO_2 in extubated patients recovering from general anesthesia. *Journal of Clinical Monitoring*, 8(3):226–230, 1992. 21

Kawato S. and Ohya J. Automatic skin-color distribution extraction for face detection and tracking. *5th International Conference on Signal Processing Proceedings*, 2, 2000. 106

Kay S. M. and Marple S. L. J. Spectrum analysis—A modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419, 1981. 193

Keogh E. and Ratanamahatana C. A. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386, 2005. 71

Khaled Z. B., Farges G., and Compiegne U. D. T. D. First approach for respiratory monitoring by amplitude demodulation of the electrocardiogram. *Proceedings of the 14th Annual International Conference of IEEE Engineering in Medicine and Biology Society*, pages 2535–2536, 1992. 75, 76, 96

Khamnei S. and Robbins P. a. Hypoxic depression of ventilation in humans: alternative models for the chemoreflexes. *Respiration Physiology*, 81(1):117–34, 1990. 185

Khan R., Hanbury A., Stöttinger J., and Bais A. Color based skin classification. *Pattern Recognition Letters*, 33(2):157–163, 2012. 103

Kinect M. Programming Guide. *Microsoft*, pages 1–33, 2011. 37

Klaessens J. H., van den Born M., van der Veen A., Sikkens-van de Kraats J., van den Dungen F. a., and Verdaasdonk R. M. Development of a baby friendly non-contact method for measuring vital signs: First results of clinical measurements in an open incubator at a neonatal intensive care unit. *Proceedings of SPIE Advanced Biomedical and Clinical Diagnostic Systems*, 8935, 2014. 2, 39, 41, 52

Koolen N., Decroupet O., Dereymaeker A., Jansen K., Vervisch J., Matic V., Vanrumste B., Naulaers G., Huffel S. V., and Vos M. D. Automated respiration detection from neonatal video data. *Proceedings of the 4th International Conference on Pattern Recognition Applications and Methods*, pages 164–169, 2015. 39, 41

Koopmans L. H. *The Spectral Analysis of Time Series*. Academic Press, 1995. 191

Kramer B. W., Kallapur S., Newnham J., and Jobe A. H. Prenatal inflammation and lung development. *Seminars in Fetal and Neonatal Medicine*, 14(1):2–7, 2009. 12

Kumar A. and Bhat B. V. Epidemiology of respiratory distress of newborns. *Indian Journal of Pediatrics*, 63(1):93–98, 1996. 1

Lai W.-H. and Li C.-T. Detecting faces in color images. Technical report, Department of Computing Science, University of Warwick, 1998. 106

Larsen P. D., Tzeng Y. C., Sin P. Y. W., and Galletly D. C. Respiratory sinus arrhythmia in conscious humans during spontaneous respiration. *Respiratory Physiology and Neurobiology*, 174(1-2):111–118, 2010. 18, 95

Lázaro J. and Gil E. Deriving respiration from the pulse photoplethysmographic signal. *Computing in Cardiology*, 38:713–716, 2011. 80

LeCun Y. and Bengio Y. Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361:255–258, 1995. 99

Lee H., Rusin C. G., Lake D. E., Clark M. T., Guin L., Smoot T. J., Paget-Brown A. O., Vergales B. D., Kattwinkel J., Moorman J. R., and Delos J. B. A new algorithm for detecting central apnea in neonates. *Physiological Measurement*, 33(1):1–17, 2012. 164, 206, 207

Lee J. and Chon K. H. Respiratory rate extraction via an autoregressive model using the optimal parameter search criterion. *Annals of Biomedical Engineering*, 38(10): 3218–3225, 2010. 79

Lee J. Y. and Yoo S. I. An elliptical boundary model for skin color detection. *Proceedings of the International Conference on Imaging Science, Systems, and Technology.*, 2002. 107

Lenz G., Epple E., and Heipertz W. Capnometry for continuous postoperative monitoring of nonintubated, spontaneously breathing patients. *Journal of Clinical Monitoring*, pages 245–248, 1991. 21

Leonard P., Grubb N. R., Addison P. S., Clifton D., and Watson J. N. An algorithm for the detection of individual breaths from the pulse oximeter waveform. *Journal of Clinical Monitoring and Computing*, 18(5-6):309–312, 2004. 79

Leonard P. A., Douglas J. G., Grubb N. R., Clifton D., Addison P. S., and Watson J. N. A fully automated algorithm for the determination of respiratory rate from the photoplethysmogram. *Journal of Clinical Monitoring and Computing*, 20(1):33–36, 2006. 79

Lewandowska M. Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity. *Proceedings of the Federated Conference on Computer Science and Information Systems*, pages 405–410, 2011. 33, 34, 139

Li B. N., Dong M. C., and Vai M. I. On an automatic delineator for arterial blood pressure waveforms. *Biomedical Signal Processing and Control*, 5(1):76–81, 2010. 64, 200

Li C., Cummings J., Lam J., Graves E., and Wu W. Radar remote monitoring of vital signs. *IEEE Microwave Magazine*, 10(1):47–56, 2009. 2, 29

Li M. H., Yadollahi A., and Taati B. A non-contact vision-based system for respiratory rate estimation. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014:2119–2122, 2014. 39

Li Q. and Clifford G. D. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological Measurement*, 33(9):1491–1501, 2012. 70, 71, 150

Li Q., Mark R. G., and Clifford G. D. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiological Measurement*, 29(1):15–32, 2008. 66, 67, 68, 69, 186

Lindberg L. G., Ugnell H., and Oberg P. A. Monitoring of respiratory and heart rates using a fibre-optic sensor. *Medical and Biological Engineering and Computing*, 30(5): 533–537, 1992. 79

Lovett P. B., Buchwald J. M., Sturmann K., and Bijur P. The vexatious vital: Neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. *Annals of Emergency Medicine*, 45(1):68–76, 2005. 17

Lukáč T., Púčik J., and Chrenko L. Contactless recognition of respiration phases using web camera. *Radioelektronica*, pages 1–4, 2014. 40, 41

Madhav K. V., Raghuram M., Krishna E. H., Reddy K. A., Venu Madhav K., Hari Krishna E., and Ashoka Reddy K. Monitoring respiratory activity using PPG signals by order reduced-modified covariance AR technique. In *Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering*, volume 1, pages 1–4, 2010. 69

Mandal N. Respirometers including spirometer, pneumotachograph and peak flow meter. *Anaesthesia & Intensive Care Medicine*, 7(1):1–5, 2006. 19, 20

Marple S. L. *Digital Spectral Analysis: With Applications.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1986. 190

Martin R. J. and Fanaroff A. A. Neonatal apnea, bradycardia, or desaturation: does it matter? *The Journal of Pediatrics*, 132(5):758–9, 1998. 15, 16

Martin R. J., Abu-Shaweesh J. M., and Baird T. M. Pathophysiologic mechanisms underlying apnea of prematurity. *Neoreviews*, 3(4), 2002. 15

Martinez M. and Stiefelhagen R. Breath rate monitoring during sleep using near-ir imagery and PCA. *International Conference on Pattern Recognition*, 2012. 37

Mason C. L. and Tarassenko L. Quantitative assessment of respiratory derivation algorithms. *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1998–2001, 2001. 76

Mathew O. P. Apnea of prematurity: pathogenesis and management strategies. *Journal of perinatology : Official Journal of the California Perinatal Association*, 31(5):302–310, 2011. 14, 17

MathWorks . MATLAB R2015a documentation for findpeaks.m, 2015. 111, 147

McClure C., Jang S. Y., and Fairchild K. Alarms, oxygen saturations, and SpO_2 averaging time in the NICU. *Journal of Neonatal-Perinatal Medicine*, 9(4):357–362, 2016. 66, 164

McKenna S., Gong S., and Raja Y. Modelling facial colour and identity with gaussian mixtures. *Pattern Recognition*, 31(12):1883–1892, 1998. 108

Meredith D. J., Clifton D., Charlton P., Brooks J., Pugh C. W., and Tarassenko L. Photoplethysmographic derivation of respiratory rate: a review of relevant physiology. *Journal of Medical Engineering & Technology*, 36(1):1–7, 2012. 27, 80

Mestha L. K., Kyal S., Xu B., and Edward L. Towards continuous monitoring of pulse rate in neonatal Intensive care unit with a webcam. *Institute of Electrical and Electronics Engineers*, pages 3817–3820, 2014. 2, 41

Milledge J. and Stott F. D. Inductive plethysmography–a new respiratory transducer. *Journal of Physiology*, 267(1):4–5, 1977. 23

Mohr M. and Vergales B. Very long apnea events in preterm infants. *Journal of Applied Physiology*, pages 558–568, 2015. 15, 206

Mohr M. A., Fairchild K. D., Patel M., Sinkin R. A., Clark M. T., Moorman J. R., Lake D. E., Kattwinkel J., and Delos J. B. Quantification of periodic breathing in premature infants. *Physiological measurement*, 36(7):1415–1427, 2015. 206

Monasterio V., Burgess F., and Clifford G. D. Robust classification of neonatal apnoea-related desaturations. *Physiological Measurement*, 33(9):1503–1516, 2012. 22, 70, 71, 161, 165, 176, 206, 207, 211

Moody G. B. and Mark R. G. Derivation of respiratory signals from multi-lead ECGs. *Computers in Cardiology*, 12:113–116, 1985. 25, 75, 76, 77

Moody G. The MIT_BIH Arrhythmia database on CD-ROM and software for use with IT, 1990. 64

Moody G. WFDB Applications Guide. Technical Report May, Harvard-MIT Division of Health Sciences and Technology, 2012. 202

Murthy V., Grove T., Harvey G., and Haywood L. Clinical usefulness of ECG frequency spectrum analysis. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 610–612, 1978. 68

Nakajim K., Matsumoto Y., and Tamura T. Development of real-time image sequence analysis for evaluating posture change and respiratory rate of a subject in bed. *Physiological Measurement*, 22(3):N21–8, 2001. 38

Nakajima K., Osa A., and Miike H. A method for measuring respiration and physical activity in bed by optical flow analysis. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2054, pages 2054–2057, 1997a. 38

Nakajima K., Maekawa T., and Miike H. Detection of apparent skin motion using optical flow analysis: Blood pulsation signal obtained from optical flow sequence. *Review of Scientific Instruments*, 68(2):1331, 1997b. 38

Nemati S., Malhotra A., and Clifford G. D. Data fusion for improved respiration rate estimation. *European Association for Signal Processing Journal on Advances in Signal Processing*, 2010:926305, 2010. 71, 81, 97

Niimi T., Itoh Y., Natori M., and Aoki Y. Apnea detection method for Cheyne-Stokes respiration analysis on newborn. *International Journal of Optomechatronics*, 7(2): 67–82, 2013. 182

Nilsson L., Johansson a., and Kalman S. Respiration can be monitored by photoplethysmography with high sensitivity and specificity regardless of anaesthesia and ventilatory mode. *Acta Anaesthesiologica Scandinavica*, 49(8):1157–62, 2005. 79

Nkadi P. O., Allen Merritt T., and Pillers D.-A. M. An overview of pulmonary surfactant in the neonate: genetics, metabolism, and the role of surfactant in health and disease. *Molecular Genetics and Metabolism*, 97(2):95–101, 2009. 9, 13

Null D. M. and Suresh G. K. *Assisted Ventilation of the Neonate*. Elsevier, sixth edition edition, 2017. 10

O'Brien C. and Heneghan C. A comparison of algorithms for estimation of a respiratory signal from the surface electrocardiogram. *Computers in Biology and Medicine*, 37(3): 305–314, 2007. 76

Ohta Y., Kanade T., and Sakai T. Color Information for region Segmentation. *Computer Graphics and Image Processing*, 13:222–241, 1980. 101

Oken E., Kleinman K. P., Rich-Edwards J., and Gillman M. W. A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC Pediatrics*, 3(1):6, 2003. 20

Oliver N., Pentland A., and Berard F. LAFTER: lips and face real time tracker. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997. 108

Olsson E., Ugnell H., Oberg P. a., and Sedin G. Photoplethysmography for simultaneous recording of heart and respiratory rates in newborn infants. *Acta Paediatrica*, 89(7): 853–861, 2000. 79

OpenClinica . OpenClinica. 53

Openshaw P., Edwards S., and Helms P. Changes in rib cage geometry during childhood. *Thorax*, 39(8):624–7, 1984. 8

O'Rourke, MF and Kelly RP, Aviolo A. *The arterial pulse*. Lea & Febiger, Malvern, PA, 1992. 198

Orphanidou C., Fleming S., Shah S., and Tarassenko L. Data fusion for estimating respiratory rate from a single-lead ECG. *Biomedical Signal Processing and Control*, 8 (1):98–105, 2013. 97, 148, 157

Pan J. and Tompkins W. A real-time QRS detection algorithm. *Biomedical Engineering, IEEE*, 32(3):230–236, 1985. 64, 67, 75, 77, 199

Pantalitschka T., Sievers J., Urschitz M. S., Herberts T., Reher C., and Poets C. F. Randomised crossover trial of four nasal respiratory support systems for apnoea of prematurity in very low birthweight infants. *Archives of Disease in Childhood - Fetal and Neonatal edition*, 94(4):F245–8, 2009. 17

Pardey J., Roberts S., and Tarassenko L. A review of parametric modelling techniques for EEG analysis. *Medical Engineering & Physics*, 18(1):2–11, 1996. 193

Park S. B., Noh Y. S., Park S. J., and Yoon H. R. An improved algorithm for respiration signal extraction from electrocardiogram measured by conductive textile electrodes using instantaneous frequency estimation. *Medical and Biological Engineering and Computing*, 46(2):147–158, 2008. 76

Parkash A., Haider N., Khoso Z. A., and Shaikh A. S. Frequency, causes and outcome of neonates with respiratory distress admitted to neonatal intensive care unit, national institute of child health, Karachi. *Journal of the Pakistan Medical Association*, 65(7): 771–775, 2015. 1

Parra L. and Sajda P. Blind Source Separation via Generalized Eigenvalue Decomposition. *Journal of Machine Learning Research*, 4(7-8):1261–1269, 2003. 123, 129

Pavlidis I., Dowdall J., Sun N., Puri C., Fei J., and Garbey M. Interacting with human physiology. *Computer Vision and Image Understanding*, 108(1âĂŞ2):150 – 170, 2007. Special Issue on Vision for Human-Computer Interaction. 29, 30

Pearson K. On lines and planes of closest fit to systems of points in Space. *Philosophical Magazine*, 2(11):559–572, 1901. 126

Pedersen M. E., Fatemian M., and Robbins P. a. Identification of fast and slow ventilatory responses to carbon dioxide under hypoxic and hyperoxic conditions in humans. *The Journal of Physiology*, 521(1):273–87, 1999. 185

Penny W. D. Signal Processing course. Technical report, University College London, 2000. 190

Phung S. L., Bouzerdoum a. S., and Chai D. S. Skin segmentation using color pixel classification: analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):148–154, 2005. 101

Poets C. F. Interventions for apnoea of prematurity: A personal view. *Acta Paediatrica, International Journal of Paediatrics*, 99(2):172–177, 2010. 207

Poh M.-z., McDuff D. J., and Picard R. W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18 (10):10762–10774, 2010. 34, 139, 182

Poh M.-z., McDuff D. J., and Picard R. W. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions in Biomedical Engineering*, 58(1):7–11, 2011. 32, 33, 34, 35, 139

Poudel R., Nait-Charif H., Zhang J. J., and Liu D. Region-based skin color detection. *International Conference on Computer Vision Theory and Applications*, 2012. 100

Poynton C. *The CIE System of Colorimetry*. Morgan Kauffman Publishers, 2nd edition, 2012. 105

Prahl S. Tabulated molar extinction coefficient for hemoglobin in wate, 1998. 33

Pramanik A. K., Rangaswamy N., and Gates T. Neonatal respiratory distress: a practical approach to its diagnosis and management. *Pediatric clinics of North America*, 62(2): 453–69, 2015. 1

Primiano Jr F. Measurements of the respiratory system. *Medical Instrumentation: Application and Design*, pages 377–448, 2009. 19, 21, 23

Prisk G. K., Hammer J., and Newth C. J. L. Techniques for measurement of thoracoabdominal asynchrony. *Pediatric pulmonology*, 34(6):462–72, 2002. 183

227

Proakis J. Digital signal processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(4):392–394, 1975. 194

Prutchi D. and Norris M. *Design and development of medical electronic instrumentation: a practical perspective of the design, construction, and test of medical devices*. John Wiley & Sons, 2004. xv, 22, 23

Qian L. L., Liu C. Q., Guo Y. X., Jiang Y. J., Ni L. M., Xia S. W., Liu X. H., Zhuang W. Z., Xiao Z. H., Wang S. N., Zhou X. Y., and Sun B. Current status of neonatal acute respiratory disorders: A one-year prospective survey from a Chinese neonatal network. *Chinese Medical Journal*, 123(20):2769–2775, 2010. 1

Ravenswaaij-Arts C. M. A. V., Kollee L. A. A., Hopman J. C. W., Stoelinga G. B. A., and Geijn H. P. V. Heart rate variability. *Annals of Internal Medicine.*, 118(6):436–447, 1993. 27

Razi N. M., Humphreys J., Pandit P. B., and Stahl G. E. Predischarge monitoring of preterm infants. *Pediatric Pulmonology*, 27(2):113–116, 1999. 205

Redner A. R. and Walker H. F. Mixture densities, maximum. *Society for Industrial and Applied Mathematics*, 26(2):195–239, 1984. 111, 118

Reisner A., Shaltis P. a., McCombie D., and Asada H. H. Utility of the photoplethysmogram in circulatory monitoring. *Anesthesiology*, 108(5):950–958, 2008. xviii, 199

Roald N. G. *Estimation of vital signs from photoplethysmography*. PhD thesis, Norwegian University of Science and Technology, 2013. 34

Rockafellar R. T. and Wets R. J.-B. *Variational Analysis*, volume 317. Washigton University, 2010. 115

Rocker J. A. and Bachur R. G. Pediatric apnea: Epidemiology, 2010. 17

Rodriguez R. J. Management of respiratory distress syndrome: an update. *Respiratory Care*, 48(3):277–279, 2003. 13

Rubaltelli F. F., Dani C., Reali M. F., Bertini G., Wiechmann L., Tangucci M., and Spagnolo A. Acute neonatal respiratory distress in Italy: a one-year prospective study. Italian Group of Neonatal Pneumology. *Acta Paediatrica*, 87(12):1261–1268, 1998. 1

Sahindrakar P., de Haan G., and Kirenko I. Improving motion robustness of contact-less monitoring of heart rate using video analysis. Master's thesis, Eindhoven University of Technology, 2011. 33

Sameni R. *Extraction of fetal cardiac signals from an array of maternal abdominal recordings*. PhD thesis, University of Grenoble, 2008. 126, 181

Saul L., Allen J., Labs T., Ave P., and Park F. Periodic Component Analysis : An eigenvalue method for representing periodic structure in speech. *Annual Neural Information Processing Systems Conference 2000*, 13:807–813, 2001. 123, 126, 181

Scalise L. and Bernacchia N. Heart rate measurement in neonatal patients using a webcamera. *IEEE International Symposium on Medical Measurement and Applications*, pages 1–4, 2012. 2, 40

Schäfer A. and Kratky K. W. Estimation of breathing rate from respiratory sinus arrhythmia: Comparison of various methods. *Annals of Biomedical Engineering*, 36 (3):476–485, 2008. 76, 97

Schmitz G. *Video camera based photoplethysmography using video camera based photoplethysmography using ambient light.* PhD thesis, Eindhoven University of Technology, 2011. 32, 33, 35

Selman A. C., Mcdonald A. H., Kitney R. I., and Linkens D. A. The interaction between heart rate and respiration : Part 1 - Experimental studies in man. *Automedica*, 4: 131–139, 1982. 26

Sexson W. R., Gotshall R. W., and Miles D. S. Cardiothoracic variables measured by bioelectrical impedance in preterm and term neonates. *Critical Care Medicine*, 19(8): 1054–1059, 1991. 22

Shelley K. H., Jablonka D. H., Awad A. a., Stout R. G., Rezkanna H., and Silverman D. G. What is the best site for measuring the effect of ventilation on the pulse oximeter waveform? *Anesthesia & Analgesia*, 103(2):372—-7, 2006. 28, 32, 79

Sherwood A., Allen M. T., Fahrenberg J., Kelsey R. M., Lovallo W. R., and Van Doornen L. J. P. Methodological guidelines for impedance cardiography. *Psychophysiology*, 27 (1):1–23, 1990. 21

Shin M., Chang K., and Tsap L. Does colorspace transformation make any difference on skin detection? *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 1–5, 2002. 103, 107

Shwayder T. and Akland T. Neonatal skin barrier: structure, function, and disorders. *Dermatologic Therapy*, 18(2):87–103, 2005. 29

Sigal L., Sclaroff S., and Athitsos V. Skin color-based video segmentation under time-varying illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):862–877, 2004. 107

Simoncelli E. P., Freeman W. T., Adelson E. H., and Heeger D. J. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, 1992. 39

Simonsen K. A., Anderson-Berry A. L., Delair S. F., and Dele Davies H. Early-onset neonatal sepsis. *Clinical Microbiology Reviews*, 27(1):21–47, 2014. 17

Sinha S. K. and Donn S. M. Fetal-to-neonatal maladaptation. *Seminars in Fetal and Neonatal Medicine*, 11(3):166–173, 2006. 1

Smith R., Argod J., Pepin J., and Levy P. Pulse transit time: an appraisal of potential clinical applications. *Thorax*, 54:452–457, 1999. 186

Smith T. G., Balanos G. M., Croft Q. P. P., Talbot N. P., Dorrington K. L., Ratcliffe P. J., and Robbins P. A. The increase in pulmonary arterial pressure caused by hypoxia depends on iron status. *The Journal of Physiology*, 586(24):5999–6005, 2008. 45

Soriano M., Martinkauppi B., and Huovinen S. Skin detection in video under changing illumination conditions. *Proceedings of the Annual Conference of the IEEE Computer Vision and Pattern Recognition Society*, 1:839–842, 2000. 106

Sornmo L. and Laguna P. *Bioelectrical signal processing in cardiac and neurological applications*, volume 16. Academic Press, 2014. 72

Starr C. *Biology: Concepts and Applications.* Thomson Brooks Cole, 2005. 30

Steinback C. D. and Poulin M. J. Ventilatory responses to isocapnic and poikilocapnic hypoxia in humans. *Respiratory Physiology and Neurobiology*, 155(2):104–113, 2007. 185

Stokowski L. A. A primer on apnea of prematurity. *Advances in Neonatal Care*, 5(3): 155–174, 2005. 16

Storring , Moritz , and Andersen H. J. Skin color detection under changing lighting conditions. *7th Symposium on Intelligent Robotics Systems*, 1999. 102, 108

Strang G. *Linear Algebra and Its Applications*, volume 3rd. Brookes Cole, 1988. 127

Strömberg N. O. and Nelson N. Thoracoabdominal asynchrony in small children with lung disease-methodological aspects and the relationship to lung mechanics. *Clinical Physiology*, 18(5):447–456, 1998. 183

Sukor J. A., Redmond S. J., and Lovell N. H. Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiological Measurement*, 32(3):369–384, 2011. 70

Sun Y., Hu S., Azorin-Peris V., Greenwald S., Chambers J., and Zhu Y. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *Journal of Biomedical Optics*, 16(7):077010–077010, 2011. 32, 33, 34

Sun Y., Papin C., Azorin-Peris V., Kalawsky R., Greenwald S., and Hu S. Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam. *Journal of Biomedical Optics*, 17(3): 37005–37010, 2012. 34

Tagawa T., Tamura T., and Oberg P. A. *Biomedical Sensors and Instruments.* CRC Press, 2nd edition, 2011. 19

Takano C. and Ohta Y. Heart rate measurement based on a time-lapse image. *Medical Engineering Physics*, 29(8):853–857, 2007. 32, 33, 35

Tambe P., Sammons H. M., and Choonara I. Why do young children die in the UK? A comparison with Sweden. *Archives of Disease in Childhood*, 100(10):928–31, 2015. 17

Tan K. S., Saatchi R., Elphick H., and Burke D. Real-time vision based respiration monitoring system. In *IEEE Transactions on Biomedical Engineering*, volume 53, pages 770–774, 2010. 35, 37

Tansley J. G., Pedersen M. E., Clar C., and Robbins P. a. Human ventilatory response to 8 hours of euoxic hypercapnia. *Journal of Applied Physiology*, 84(2):431–4, 1998. 185

Tarassenko L., Villarroel M., Guazzi A., Jorge J., Clifton D. a., and Pugh C. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35(5):807–31, 2014. 32, 33, 35, 40, 139, 183

Terrillon J.-C. and Akamatsu S. Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images. *Vision Interface*, 1999. 103, 108

Thayer J., Sollers , J.J. I., Ruiz-Padial E., and Vila J. Estimating respiratory frequency from autoregressive spectral analysis of heart period. *IEEE Engineering in Medicine and Biology Magazine*, 21(4):41–45, 2002. 76

Tin W., Milligan D. W., Pennefather P., and Hey E. Pulse oximetry, severe retinopathy, and outcome at one year in babies of less than 28 weeks gestation. *Archives of Disease in Childhood. Fetal and Neonatal edition*, 84(2):F106–F110, 2001. 28

Tobin R. M., Pologe J. a., and Batchelder P. B. A characterization of motion affecting pulse oximetry in 350 patients. *Anesthesia and analgesia*, 94(1 Suppl):S54–S61, 2002. 164

Tortora G. J. and Derrickson B. *Principles of Anatomy and Physiology*. Number 14. John Wiles & Sons, 2014. 7

Traube L. Uber periodische Tatigkeitsanderungen des Vasomotorischen und Hemmungs-Nervenzentrums. *Nervencentrum. Centralblatt fur die Medicinischen Wissenschaften*, 3(56):881–885, 1865. 11

Travaglini A., Lamberti C., DeBei J., and Ferri M. Respiratory signal derived from eight-lead ECG. *Computers in Cardiology*, page 4, 1998. 76

Tsouri G. R., Kyal S., Dianat S., and Mestha L. K. Constrained independent component analysis approach to nonobtrusive pulse rate measurements. *Journal of Biomedical Optics*, 17(7):077011–1–077011–4, 2012. 34

Tutdibi E., Gries K., Bücheler M., Misselwitz B., Schlosser R. L., and Gortner L. Impact of labor on outcomes in transient tachypnea of the newborn: population-based study. *Pediatrics*, 125(3):e577–83, 2010. 13

van Gastel M., Stuijk S., and de Haan G. New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring. *Nature Scientific reports*, 6: 38609, 2016. 2, 182, 183

van Ravenswaaij-Arts C. M. A. Heart rate variability. *Annals of Noninvasive Electro-cardiology*, 118:354–381, 1993. 12, 97

Verkruysse W., Svaasand L. O., and Nelson J. S. Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, 2008. xv, 31, 32, 33, 35

Vezhnevets V., Sazonov V., and Andreeva A. A survey on pixel-based skin color detection techniques. In *GRAPHICON*, pages 85–92, 2003. 100, 101, 103, 104, 106

Villarroel M., Guazzi A., Jorge J., Davis S., Watkinson P., Green G., Mccormick K., and Tarassenko L. Continuous non-contact vital sign monitoring in the neonatal intensive care unit. *Healthcare Technology Letters*, 1(3):87–91, 2014. 2, 32, 33, 40, 182, 183

Viola P. and Jones M. Rapid object detection using a boosted cascade of simple features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:511–518, 2001. 32

Wadhwa N. and Rubinstein M. Phase-based video motion processing. *ACM Transactions on Computer Systems*, 2013. xv, 39

Warburton D, Stark AR T. H. Apnea monitor failure in infants with upper airway obstruction. *Pediatrics*, 60(5):742–4, 1977. 2, 29, 164

Warfield S. K., Zou K. H., and Wells W. M. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004. 114

Welch P. The use of fast Fourier transform for the estimation of power spectra. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967. 81, 191, 192

Wertheim D., Olden C., Savage E., and Seddon P. Extracting respiratory data from pulse oximeter plethysmogram traces in newborn infants. *Archives of Disease in Childhood - Fetal and Neonatal edition*, 94(4):F301–F303, 2009. 79, 161

West J. B. *Respiratory Physiology - The Essentials*, volume Ninth Edit. Lippincott, Williams & Wilkins, 2012. 10

Wieringa F. P., Mastik F., and van der Steen A. F. W. Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO2 camera" technology. *Annals of Biomedical Engineering*, 33(8):1034–41, 2005. 31, 32, 33, 35

Williamson J. R., Bliss D. W., Browne D. W., Indic P., Bloch-Salisbury E., and Paydarfar D. Using physiological signals to predict apnea in preterm infants. *Conference Record - Asilomar Conference on Signals, Systems and Computers*, pages 1098–1102, 2011. 165

Wilson S. L., Thach B. T., Brouillette R. T., and Abu-Osba Y. K. Upper airway patency in the human infant: influence of airway pressure and posture. *Journal of Respiratory, Environmental and Exercise Physiology*, 48(3):500–4, 1980. 14

Wootton C. *A Practical Guide to Video and Audio Compression - From Sprockets and Rasters to Macroblocks*. Elsevier Ltd, 2005. 56

World Health Organization . Acute respiratory infections in children: case management in small hospitals in developing countries., 1990. 17

Wu C. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11(1): 95–103, 1983. 111, 118

Wu H.-y., Rubinstein M., Shih E., Guttag J., Durand F., and Freeman W. T. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions in Graphics*, 31(4), 2012. 38, 41

Xia J. and Siochi R. A. A real-time respiratory motion monitoring system using KINECT: proof of concept. *Medical Physics*, 39(5):2682–5, 2012. 37

Xu L. and Jordan M. I. On convergence properties of the EM Algorithm for Gaussian mixtures. *Neural Computation*, 8(1):129–151, 1996. 118

Yang J., Lu W., and Waibel A. Skin-color modeling and adaptation. *Lecture Notes in Computer Science*, 1352:687–694, 1997. 107, 108

Yang M.-h. and Ahuja N. Gaussian mixture model for human skin color and its applications in image and video databases. *Proceedings of the International Society for Optics and Photonics*, 3656:458–466, 1998. 100, 106, 107, 108

Yi W. and Park K. Derivation of respiration from ECG measured without subject's awareness using wavelet transform. *Proceedings of the 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society and Engineering in Medicine and Biology Society*, 1:183–184, 2002. 76

Yoshiya I., Shimada Y., and Tanaka K. Evaluation for clinical of a hot-wire respiratory flowmeter. *Journal of Applied Physiology*, 47(5):1131–1135, 1979. 20

Yu M., Liou J., and Kuo S. Noncontact respiratory measurement of volume change using depth camera. In *34th Annual International Conference of the IEEE EMBS*, pages 2371–2374, 2012. 36, 37

Zarit B., Super B., and Quek F. Comparison of five color models in skin pixel classification. *Proceedings International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 1999. 106

Zhao F., Li M., Qian Y., and Tsien J. Z. Remote measurements of heart and respiration rates for telemedicine. *PLoS ONE*, 8(10), 2013. 32, 33, 34

Zhao J., Gonzalez F., and Mu D. Apnea of prematurity: From cause to treatment. *European Journal of Pediatrics*, 170(9):1097–1105, 2011. 207

Zhao L., Reisman S., and Findley T. Derivation of respiration from electrocardiogram during heart rate variability studies. In *Computers in Cardiology*, pages 53–56, 1994. 76

Zhou Y., Zheng Y., Wang C., and Yuan J. Extraction of respiratory activity from photoplethysmographic signals based on an independent component analysis technique: Preliminary report. *Instrumentation Science & Technology*, 34(5):537–545, 2006. 80

Zimova-Herknerova M. and Plavka R. Expired tidal volumes measured by hot-wire anemometer during high-frequency oscillation in preterm infants. *Pediatric Pulmonology*, 41(5):428–433, 2006. 21

Zong W. A robust open-source algorithm to detect onset and duration of QRS Complexes. *Computers in Cardiology*, 2003. 64, 67, 150

Zong W., Heldt T., Moody G., and Mark R. An open-source algorithm to detect onset of arterial blood pressure pulses. *Computers in Cardiology*, pages 259–262, 2003. 64, 65, 200