Respiratory Quality Index Design and Validation for ECG and PPG Derived Respiratory Data

Report for transfer of status
11th December 2015

Drew Birrenkott
Trinity College
University of Oxford,
Department of Engineering Science

Supervised by
Dr. David Clifton
Abstract

Patient vital signs have been widely recognized as an early indicator of catastrophic patient decline in the hospital. This has led to the widespread implementation of Rapid Response Teams (RRTs) and Early Warning Score (EWS) systems within hospitals. However, in order to both RRTs and EWS systems to be most effective, it is critical that patient vital signs are recorded accurately and often. However, despite findings that of the four major vital signs respiratory rate is the most crucial in predicting catastrophic decline, it is the most often not recorded. One possible solution to this is to develop a device that is capable of non-invasively monitoring respiratory rate. Two prime candidates for this are the photoplethysmograph and electrocardiogram. Both the PPG and ECG are widely used to measure heart rate in the hospital and because the heart rate is modulated in multiple ways by the respiratory rate, it is possible to extract a respiratory rate estimate from both the PPG and ECG. However, the overall modulations of the respiratory rate on the PPG and ECG are small meaning that often algorithms that are designed to detect them cannot tell when these modulations are actually present or not. This research presents a method for detecting when the respiratory modulations can be extracted from the PPG and ECG using respiratory quality indices (RQIs) which are implemented after extraction of the respiratory rate modulations but before the respiratory rate is estimated.

This research was performed using three different datasets of varying data quality. The first dataset was the CapnoBase dataset which contains PPG, ECG, and capnography for anesthetized patients. This dataset is considered the “gold standard” for comparison. The second dataset was the MIMIC II dataset which contains PPG, ECG, and transthoracic impedance (TTI) data for patients that are bedridden in the hospital. This was considered a “silver standard” for comparison. Finally, the last dataset was the PICRAM dataset which was collected for mobile patients and contained PPG and ECG waveforms in addition to clinician annotations for respiratory rate. This was considered the “bronze standard” for comparison. These datasets were used to test and validate four different RQIs: the FFT-RQI, the AR-RQI, the AutoCor-RQI, and the Hjorth Parameter RQI. The FFT-RQI is based on a signals Fast Fourier Transform and functions by calculating the area of the largest peak in the frequency domain compared to the rest of the signal. The AR-RQI works based on the autocorrelation function. The AR-RQI is calculated by finding the best performing AR model order and selecting the largest pole for that model order to represent the AR-RQI. The AutoCor-RQI is based on the autocorrelation function and seeks to find the area within the respiratory rate range where the signal best corresponds to itself where the higher the autocorrelation value is, the more likely it is to correspond to an actual respiratory wave. Finally, the Hjorth Parameter RQI is calculated using the third Hjorth parameter which is a measure of how sinusoidal a signal where the periodicity of a signal is expected to correspond to a more pronounced respiratory waveform.

The results of this research ultimately demonstrated that three of the four RQIs: the FFT-RQI, AR-RQI, and AutoCor-RQI, perform better than a signal quality index (SQI) applied prior to respiratory rate modulation extraction. Additionally, it was found that in general the RQIs were robust and performed well when different respiratory rate estimation algorithms were applied as the standard of comparison, when the data quality was reduced from CapnoBase to PICRAM, and for both the PPG and ECG signals. While this research demonstrates a promising future role
for RQIs in respiratory rate estimation from PPG and ECG data, further work is still needed in order to fuse the different RQIs to get a single respiratory rate estimate. Ultimately however, this work shows a lot of promise for creating a robust respiratory rate estimation algorithm that can be used in both hospital and clinical settings in the developing and developed world.
### Analysis of Respiratory Quality Indices

4.1 Overview ........................................................................................................ 41

4.2 Performance of Respiratory Quality Indices .................................................. 42
   4.2.1 Plots of RQI Performance ......................................................................... 42
   4.2.2 RQI Comparison Metric .......................................................................... 42

### Discussion

5.1 Overview .......................................................................................................... 45
   5.1.1 General RQI Performance ....................................................................... 45
   5.1.2 RQI Robustness ...................................................................................... 46
   5.1.3 Performance of RQI with Decreasing Data Quality ............................... 47
   5.1.4 RQI Value for PPG and ECG .................................................................. 48
   5.1.5 Limitations of the RQIs .......................................................................... 48

### Future Work

6.1 Conclusions ...................................................................................................... 50

6.2 Future Work ..................................................................................................... 51
   6.2.1 Fusion of Respiratory Rates Derived from Different Modulations ....... 51
   6.2.2 Integration with Vital-Sign Data Fusion Models ................................. 52
   6.2.3 Integration with Applications in m-Health for Developed and Developing Regions .......................................................... 53
   6.2.4 Project Timeline ..................................................................................... 54
Chapter 1

Introduction

1.1 Vital Signs Monitoring

For over 100 years, clinicians have been performing routine patient surveillance using four primary vital signs: blood pressure, heart rate, respiratory rate, and temperature \[1,2\]. The collection of vital signs has become so ingrained in the ethos of modern medicine that it is the second set of information collected during a patient examination, after an assessment of patient history \[3\]. The presence of abnormalities in vital signs during an examination leads to a rapid and focused response that takes precedence over the response to “normal” physiology \[3\].

1.1.1 Importance of Vital Signs

Despite the ubiquity and importance of collecting vital signs in the modern hospital setting, the likelihood ratios of abnormal vital-sign values being symptoms of specific pathophysiological problems is negligible \[4\]. Given this, the value of collecting vital signs is that they can be used in concert with other medical information to render a specific diagnosis more likely. For example, if it is likely that a patient has a certain acute illness and their vital signs indicate general malaise, it is far more likely that the patient has that disease than if they simply exhibit symptoms for the disease but do not have abnormal vital-sign values \[4\]. Additionally, the converse of this is supported in that if a patient has abnormal vital-sign values but no other systemic concerns or symptoms, it is unlikely to indicate the presence of acute illness \[4\]. Furthermore, for in-hospital patients with acute illnesses, abnormal vital signs are the most critical early predictor of catastrophic deterioration such as cardiopulmonary arrest \[5,7\].
Despite the growing presence of specific in-hospital cardiac arrest teams and technological advances in cardiopulmonary resuscitation methods, hospital patients who suffer from cardiopulmonary arrest still have a 50-80% chance of mortality \[8-10\]. However, it has been well documented that, for many patients, physiological derangements in vital signs are present up to eight hours before the cardiopulmonary event \[5-6-8-11-12\]. In a study by Schein et al. \[5\], for example, it was shown that 84% of patients showed some form of pathophysiological alteration in the eight hours preceding arrest. This suggests that, with proper recording and observation of vital signs, it is possible to provide early warning that a patient is at high risk of cardiopulmonary arrest, so that active measures to prevent it can be taken \[13\].

1.1.2 Clinical Response to Abnormal Vital Signs

Within the past 20 years, hospitals have begun organising rapid response teams (RRTs) to respond to critical care emergencies, such as severe vital sign derangements \[12\]. RRTs respond based on pre-set criteria of two forms: “calling criteria” and “early warning scores” (EWS) \[12\].

**Calling Criteria** Calling criteria systems are the more basic of the two systems and work by staff calling the RRT when one or more physiological indicators fall outside of a pre-set range \[12-14\]. In one calling criteria study in which the RRT could be called for either abnormal physiological variables or catastrophic deterioration, such as cardiac arrest, abnormal physiology was the primary reason for calling the RRT in 60% of cases; and in 49% of cases, the only reason for the RRT call was abnormal physiology \[14\]. This study showed an effective use of the RRT; however, it did not follow these patients to see if they had better hospital outcomes than patients with abnormal physiology that did not have the RRT called.

**Early Warning Scores (EWS)** EWS systems work on a point-based system where points are assigned to each vital sign based on the relative abnormality of its value. All points for the various vital signs are then added, and the RRT is called if the number of points exceeds a certain threshold \[12-15-17\]. One of the more commonly used EWS systems is the modified early-warning score system (MEWS), shown in Table \[11\] \[18\]. The MEWS system includes all four major vital signs as well as an AVPU score. AVPU measures a patient’s cognitive function by determining if the patient is “alert,” “responsive to vocal stimulus,” “responsive to
Table 1.1: Modified Early Warning Score System

<table>
<thead>
<tr>
<th>Score</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory Rate (min⁻¹)</td>
<td>&lt; 9</td>
<td>9 – 14</td>
<td>15 – 20</td>
<td>21 – 29</td>
<td>≥ 30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart Rate (min⁻¹)</td>
<td>&lt; 40</td>
<td>41 – 50</td>
<td>51 – 100</td>
<td>101 – 110</td>
<td>111 – 129</td>
<td>≥ 130</td>
<td></td>
</tr>
<tr>
<td>Systolic BP (mmHg)</td>
<td>&lt; 70</td>
<td>71 – 80</td>
<td>81 – 100</td>
<td>101 – 199</td>
<td>≥ 200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>&lt; 35</td>
<td>35 – 38.4</td>
<td>≥ 38.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurological</td>
<td>Alert</td>
<td>Reacting to Voice</td>
<td>Reacting to Pain</td>
<td>Unresponsive</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

pain,” or “unresponsive.” For MEWS, aggregate scores above a threshold of five are considered critical scores that require further medical intervention. In one retrospective study of patients previously admitted to the general ward of a district general hospital, it was found that patients that had a critical MEWS score at any time during their stay had an increased risk of death (odds ratio 5.4), a higher risk of ICU admission (odds ratio 10.9), and a higher risk of admission to a high-dependency unit (HDU) (odds ratio 3.3) [18]. Furthermore, while it was demonstrated that a critical MEWS score suggested greater risk of death, ICU, or HDU admission, the findings did not show increased risk if only a single vital sign value was abnormal. This was especially apparent for blood pressure and temperature [18]. Thus for EWS systems to be most effective, gathering reliable data for all of the vital signs included in the EWS scoring system is critical.

1.2 Importance of Respiratory Rate

In order for EWS systems to be most effective, the entire panel of vital signs (blood pressure, heart rate, respiratory rate, and temperature) must be collected. Despite this, respiratory rate is often not collected, not properly recorded, or is only recorded intermittently compared to other vital signs [11,19–21]. One study over a 48-hour period showed the frequency of respiratory rate measurement was 1.0 reading/day compared to 5.0 readings/day for blood pressure, 4.4 readings/day for pulse rate, and 4.2 readings/day for temperature [20]. This is particularly concerning because of the four vital signs, an elevated or reduced respiratory rate provides the earliest and most accurate warning for catastrophic deterioration [5,17,22–24]. In a study of 64 patients that suffered from catastrophic deterioration in the hospital, Schein et al. [5] found that 24 patients (38%) had an abnormal respiratory rate preceding cardiopulmonary arrest and an additional 14 patients (22%) suffered from a change in their respiration in addition to at least one other alteration (e.g. in pulse rate). This means that 60% of patients who suffered cardiopulmonary arrest presented with abnormal respiratory rate prior to arrest.
1.2.1 Critical Respiratory Rate Indicators

The typical adult respiratory rate is between 12 and 20 respirations per minute \([4,25]\). A reduced respiratory rate is known as bradypnea and an elevated respiratory rate is known as tachypnea \([26]\). While both can be signs of critical illness, tachypnea more commonly indicates critical illness than bradypnea \([27]\). Respiratory rates over 20 breaths per minute indicate that a patient is unwell while respiratory rates above 24 to 27 breaths per minute are an indicator of critical illness \([22,23]\). Furthermore, a respiratory rate that is three to five breaths per minute above a patient’s normal respiratory rate has been shown to be a sign of early distress \([2]\).

1.2.2 Physiological Antecedents to Abnormal Respiratory Rate

Baseline respiratory rate and breathing depth is controlled by the respiratory centre in the medulla oblongata, a portion of the brainstem that is responsible for autonomic bodily processes \([26]\). However, the baseline respiratory rate set by the respiratory centre can be altered by voluntary control from higher centres in the cerebral hemispheres and through nervous and chemical control throughout the body \([26]\). Nervous control of respiration is primarily through the phrenic and intercostal nerves which affect respiration by controlling the diaphragm and intercostal muscles, respectively \([26]\). Nervous control of respiration is primarily in response to pain, emotion, or anxiety \([26]\). Chemical control of respiration, which most commonly drives respiratory rates in critical illness, is through primary chemoreceptors in the medulla oblongata and peripheral chemoreceptors in the carotid and aortic bodies which act on signals from the bloodstream and send chemical signals to the respiratory centre to alter respiratory rate \([26]\). Generally, the chemoreceptors are activated in response to high levels of CO\(_2\) in the bloodstream (hypercapnia), low levels of O\(_2\) in the bloodstream (hypoxaemia), or acidosis \([26]\).

Based on this physiology, the reason that variation in respiratory rate is so sensitive and can serve as an effective proxy for future catastrophic deterioration is because respiratory rate is heavily influenced by many organ systems outside of the respiratory system. This means that any severe derangement in one of these systems is likely to have an effect on the respiratory system \([22]\). This is because one of the earliest signs of derangement in an organ system as result of hypoxia is lactic acid production \([28]\). Lactic acidosis, the production of lactic acid beyond the body’s normal ability to cope with it, in addition to ketoacidosis, renal failure, and
toxin ingestion, all cause a general class of acidosis called metabolic acidosis \[29\]. Metabolic acidosis is characterised by an increased presence of hydrogen ions in the bloodstream causing a drop in blood pH. The body’s physiological response to increased blood pH is to increase CO\(_2\) production \[22\]. However, this increased CO\(_2\) production ultimately leads to hypercapnia, to which the body responds by increasing respiratory rate \[22\]. Ultimately, this early biochemical pathway leading to tachypnoea is the reason that detection of increased respiratory rate can be used as an early predictor of future catastrophic deterioration.

### 1.3 Monitoring Respiratory Rate

Typically, respiratory rate is assessed by a clinician counting the rising and falling of a patient’s chest over a set period of time and using that to calculate a rate in breaths per minute \[3\]. This simple method suffers from two major faults that make it inaccurate. The first is the length of time that should be used to calculate the respiratory rate. The clinical standard is most often 30 seconds, but best practice indicates that the respiration should be counted over at least a full minute \[3\]. In one study that compared the length of time used for calculating respiratory rate, there was a significant difference in the calculated respiratory rates when nurses counted respiratory rate for 15 seconds and medical students counted respiratory rate for one minute \[30\]. The second major fault is that patient’s respiratory rates often increase when they realize that their respiratory rate is being observed. Thus, clinicians often make sure patients are unaware by pretending to take a radial pulse while counting respiratory rate \[26\].

There is a stark contrast between how respiratory rate is most often calculated compared to the other three major vital signs, all of which can be collected electronically. In a study of why nurses record respiratory rate more inaccurately and less frequently than other vital signs, three major reasons were given: lack of time, lack of knowledge, and lack of equipment \[19\].

**Lack of Time** Advances in medical technologies and the use of invasive procedures mean that patients are living longer than previously. As a result, the demand for beds is higher, especially for patients that would have previously been in ICUs or HDUs \[31\]. This overcrowding means that nurses cannot do more lengthy observations such as calculating respiratory rate \[19\].
Lack of Knowledge  As has been previously identified, respiratory rate is perhaps the most important vital sign in estimating a patient’s risk of catastrophic decline. Additionally, collection of respiratory rate is important when using an EWS system because all vital signs have to be recorded for an EWS to be effective. However, Hogan [19] found that the importance of these facts was under-emphasised to nursing staff. Furthermore, nursing students are often taught how to calculate respiratory rate, but are not taught why it is important to measure it [19]. This is further demonstrated in a study where a vital-sign chart that explicitly illustrated the importance of all vital signs was introduced for calculating EWS. Initially, at least one respiratory rate measurement was made per day for only 29.5 ($\pm 13.5\%$) of patients. After introduction of the vital signs chart, this number increased to 91.2 ($\pm 5.6\%$) [32].

Lack of Technology  Heart rate, blood pressure, and temperature are all often measured with electronic probes that can calculate these vital signs autonomously, as is oxygen saturation (SpO$_2$). However, in general hospital wards, equipment for calculating respiratory rate automatically is often not available [19].

1.3.1 Improving Respiratory Rate Monitoring

Current methods for measuring respiratory rate are inadequate because they are inaccurate and not recorded sufficiently often to be maximally useful. There is an urgent need to improve the methods for monitoring and recording respiratory rate, as this could substantially improve the ability to provide early warning of potential catastrophic deterioration. Thus, the ideal respiratory-rate monitoring system would continuously monitor all respiratory activity, including respiratory rate, respiratory depth, and the degree of gas exchange occurring, in an unobtrusive fashion [33]. Current continuous and autonomous methods for measuring respiratory rate include: transthoracic impedance, airflow measurement, and blood gas concentration measurements [33]. If all of these were used together, they would be capable of providing all of the desired information, but this could be too expensive to implement on a general hospital ward. However, one promising avenue is to modify the pulse oximetry method that is currently widely used, and expand it to be more reliable. Currently, pulse oximetry is able to provide SpO$_2$ percentage through the variable absorption of light by the blood [33]. Additionally, pulse oximetry is so widely used that it is becoming known as the “fifth vital sign” [4]. As a result,
the means to measure pulse oximetry is widely available throughout the hospital, including in general wards. Furthermore, pulse oximetry is capable of obtaining a pulse signal [34]. This is crucial because, due to the lungs physiological effects on the heartbeat, in many instances it is possible to obtain the respiratory rate from the pulse signal, as will be described later [34]. We will subsequently demonstrate how respiratory-rate estimation from such sources can be made sufficiently robust for use in clinical practice.

Given the importance of being able to calculate respiratory rate in an effective and non-invasive manner, the remainder of this report will focus on the methods that are being used to extract and validate the quality of the respiratory rate from information currently obtained through the photoplethysmogram (PPG) and electrocardiogram (ECG).
Chapter 2

Respiratory Rate Extraction Techniques

2.1 Current Technologies

While the clinical need for a technology that is capable of continuously monitoring respiratory rate is apparent, there are currently few methods that are able to do this reliably, comfortably, and cost effectively. While a number of technologies, including: spirometers, nasal thermocouples, transthoracic inductance, transthoracic impedance plethysmography, CO\textsubscript{2} capnography, and strain gauges, have all been used to monitor respiratory rate, they all require special equipment and may not be suitable for a general hospital setting \[35\,36\]. Of these, transthoracic impedance plethysmography (TTI) and CO\textsubscript{2} capnography are the most common clinically used methods; however, neither is ideal. TTI reliability is low due to electrode-skin impedance instabilities skin irritation caused by the electrode gel \[37\]. CO\textsubscript{2} capnography on the other had is invasive and difficult to setup and use quickly \[36\].

As a result of these challenges, an ideal scenario for continuous monitoring of respiratory rate would be to find a way to extract respiratory rate from physiological signals that are already widely collected for patients throughout the hospital. While many different electronic monitoring technologies are used throughout the hospital, two of the most ubiquitous technologies are the electrocardiogram (ECG) and the photoplethysmography (PPG) which are used for collecting heart rate data and oxygen saturation (Sp\textsubscript{O\textsubscript{2}}, PPG only). In fact these technologies are so ubiquitous that it has been recommended that all patients on the general ward should be monitored either continuously or intermittently with, at a minimum, either ECG or PPG \[38\]. The prevalence of both ECG and PPG in the hospital is of particular relevance as it
has been widely shown that the heart rate and circulatory system rhythms are physiologically modulated by the respiratory rate via responses from both the nervous system and through physical alterations in the thoracic cavity caused by respiration [39]. The following section of this report will describe the biological basis for these physiological modulations.

2.2 Pulmonary Modulation of the Cardiac System

In understanding the different organ systems in the body, they are often compartmentalized and simplified; however, the functioning of one organ system often has non-trivial effects on another. This is particularly the case with the cardiac and pulmonary systems. The cardiac and pulmonary systems are the two primary organ systems present in the thoracic cavity and function chiefly to transport oxygen and other nutrients into and carbon dioxide and other waste out of the body. As a result of this shared function, the cardiac and pulmonary systems are highly modulated by each other both through physical and nervous means.

2.2.1 Respiratory Modulation via Nervous Control

The nervous system, particularly the autonomic nervous system, is known to play a major role in the coordination of the pulmonary and cardiac systems [40]. One of the most prominent and widely studied interactions of the pulmonary and cardiac systems via the autonomic nervous system is respiratory sinus arrhythmia (RSA) [40]. RSA is a phenomenon by which the beat to beat heart rate interval is shortened during inspiration and lengthened during expiration. This manifests through activation of the pulmonary stretch receptors during inspiration. Once activated the pulmonary stretch receptors send inhibitor projections to the cardiac vagal neurones (CVNs) within the medulla oblongata of the brain. Normally, the CVNs send inhibitory signals to the heart to slow heart rate, but the signals from the pulmonary stretch receptors cause inhibition of the CVNs signals resulting in the elevated beat to beat rate that is observed in RSA [39]. It is hypothesized that the function of RSA is to improve gas exchange efficiency. This is because the instantaneous blood volume circulation is increased during the brief period where the heart rate is elevated which is ideal when inspiration is at its peak as the alveoli in the lungs will be maximally inflated and able to participate in maximum gas exchange [40]. This hypothesis was supported in a study which found that when artificially inducing RSA, O₂ up-
take was increased by 4% compared to when RSA was not present and O₂ update was reduced in a state of reversed RSA by 14% compared to when RSA was not present [40]. Importantly, the prevalence and magnitude of RSA has been shown to be easily affected and reduced by poor cardiopulmonary function and disease (such as coronary artery disease), old age (RSA is very prominent in infancy but declines after), and poor physical fitness (both athletes and people who routinely exercise have higher prevalence of RSA than people who do not exercise) [40].

In addition to RSA, it has been hypothesized by one study that the respiratory system modulates venous return to the thoracic cavity through vasoconstriction of the veins in the periphery. It is thought that this activation occurs through the autonomic nervous system causing an increase in pressure in the peripheral veins due to increased sympathetic nervous system activity during inspiration. Ultimately, this leads reduced blood volumes in the systemic veins during inspiration compared to during exhalation [41].

2.2.2 Respiratory Modulation via Physical Mechanisms

In addition to modulation via the nervous system, the major changes in pressure and size in the thoracic cavity caused by respiration have effects that alter the cardiac system. There are three major modulations caused by these physical changes: (1) decreased stroke volume (left ventricular output), (2) increased blood flow from the periphery to the thoracic cavity, and (3) changes in heart position within the thoracic cavity [39,42].

(1) Decreased Stroke Volume The decreased stroke volume noted during inspiration is caused by several factors. The primary factor is that the negative pressure in the thoracic cavity that is created causes distension of the pulmonary vessels which leads to blood pooling in these vessels [39,43]. Additionally, the pulmonary capillaries are compressed due to the expansion of the alveoli causing reduced return to the left ventricle and the expansion of the right ventricle during inspiration causes impingement in the left ventricle by the intra-ventricular septum [39].

(2) Increased Blood Flow into the Thoracic Cavity This is caused by two mechanisms. The first is due to the decrease in intrathoracic pressure. The pressure drop in the intrathoracic cavity causes an equivalent drop in pressure in the intrathroacic blood vessels causing increased flow to those vessels. Additionally, the lowering of the diaphragm causes ab-
dominal compression and increases in pressure in the intra-abdominal veins thus exacerbating
the pressure differential and leading to blood flow into the thoracic cavity [39].

(3) Changes in Heart Position within Thoracic Cavity The proximity of the lungs
to the heart means that during inspiration when the lungs are expanding, they cause a shift in
the positioning of the heart within the thoracic cavity. MRI studies of this effect have found
that movement is on average $12.4 \pm 5.9$ mm downward in the cranio-caudal axis, $4.3 \pm 3.7$ mm
in anterior direction on the anterior-posterior axis, and $2.0 \pm 2.1$ mm rightward on the left-right
axis. Additionally, some rotational effects were seen; however, this was highly variable [42].

2.3 Observable Modulations of Respiratory System in ECG
and PPG Waveforms

Previous research has shown that both the ECG and PPG are capable of acquiring the res-
piratory signal from the cardiac signal [44–48]. However, these two technologies use different
physiological measurements to acquire this information and thus it is as a result of different
modulations that the respiratory rate can be derived from the ECG and PPG.

2.3.1 Acquisition of Respiratory Rate: ECG

The ECG functions by converting the ionic current produced by the body during the heart
cycle into electron current that can be recorded by electrically conductive electrodes placed
on the surface of the body [49]. Ultimately, the voltage drop across the thorax caused by the
heart beat is detected and can be used to monitor all phases of the heart cycle and cardiac
abnormalities such as the presence of arrhythmia, ischemia, and infarction [49]. In order for
the ECG to be used to monitor respiration, the respiratory modulations on the heartbeat must
manifest as observable changes in the heart’s electrical signal. Currently, it is believed that two
of the physiological modulations previously noted: cardiac movement within the thoracic cavity
and RSA cause observable changes in the ECG that can be used for monitoring respiratory
rate. The way that these manifestations affect the ECG can be seen in Figure 2.1.
Figure 2.1: Respiratory modulations on the ECG. RWA (red) and RPA (green) represent cardiac movement modulations and RSA (blue) is the neurologically driven respiratory sinus arrhythmia.

Cardiac movement within the thoracic cavity As has previously been described, the heart moves within the thoracic cavity during respiration [42]. In addition to this effect, the electrodes used to measure the ECG change in their position relative to the heart which also has an effect on the overall ECG [46]. Ultimately it is believed that these two effects cause an observable modulation on the mean electrical axis (MEA) of the ECG and in addition to effects caused by the changing thoracic impedance due to the lungs filling with air cause a change in both R-wave peak amplitude (RPA) and R-wave area (RWA) [46,50,51]. Initial methods looking at using either RPA or RWA focused on using comparisons of multiple lead ECG data to acquire the respiratory signal [46,52]; however, recent work has focused on extracting the respiratory information using only a single lead ECG [44,50,53].

Respiratory sinus arrhythmia (RSA) As RSA is a nervous system response that physically alters when the AV node in the heart initiates a heartbeat, the RSA modulation on the ECG is obtained by detecting the peak of each QRS complex of the ECG and using these points to obtain the heart rate variability. Due to the respiratory rate synchronized instantaneous cycling of bradycardia and tachycardia observed in the instantaneous heart rate, it is possible to use the ECG to detect respiratory rate by plotting the length of each R-R interval at the instant in time that it was observed [44]. The validity of detecting RSA from ECG as a means for detecting respiratory rate has been widely reported [44,45,54].
2.3.2 Acquisition of Respiratory Rate: PPG

The PPG functions by measuring the light reflectance or transmission of red (660 nm) and infrared (940 nm) light from an LED to a photoelectric cell \[39,46,55\]. Reflectance PPG places the LED and photoelectric cell side by side anywhere on the body, but most often on appendages such as the arm, while transmission PPG places the LED on opposite sides of the body on small appendages such as the finger and earlobe \[49,55\]. While the PPG is most often used to measure SpO\(_2\) due to the varying extinction coefficients of haemoglobin and oxyhaemoglobin in the infrared spectrum, it can also be used to obtain a heartbeat signal through observation of the absorption of red light through time \[49\]. This is because the light absorbing coefficient of the haemoglobin in blood is higher than the surrounding tissues. Thus as the volume of blood increases in the circulatory system due to a heartbeat, it causes a decrease in the amount of light that passes through the body and is detected by the photoelectric cell giving a measurable heart rate signal \[55\]. One point of contention with this method is what part of the circulatory system the predominant observation comes from. Some authors claim that it is primarily the arterial blood volume that gives rise to the observed effect \[49\] while others note the importance of venous blood volume \[56\]. It is however most likely that both play a role and the effects of each are based on probe location \[40,57\].

Regardless, through the physiological modulations of the heartbeat discussed previously, the PPG heart rate signal can be used to acquire the respiratory signal through three techniques: respiratory-induced amplitude variation (RIAV), respiratory-induced intensity variation (RIIV), and respiratory-induced frequency variation (RIFV) (Figure 2.2) \[47\].

**Respiratory-induced amplitude variation (RIAV)**  RIAV is the noted decrease in peak-to-trough length during inspiration compared to exhalation \[47,48\]. The RIAV observed in the PPG signal is caused by the decreased stroke volume output by the heart during inspiration. As previously mentioned, during inspiration, the rate of cardiac filling is lessened resulting in a small cardiac output which results in the observable decrease in peak-to-trough length during inspiration as compared to during exhalation \[39,43,47,48\].

**Respiratory-induced intensity variation (RIIV)**  The RIIV can be observed in the PPG as the fluctuation of the peak amplitude heights in the PPG \[48,58\]. The RIIV signal is
not thought to be directly linked with the heartbeat itself, but rather is a unique physiological
effect on the PPG that is correlated to respiratory rate through the downstream physiological
effects of the changing pressure in the thoracic cavity [59]. Ultimately, it is thought that the
RIIV is the result of one or a combination of three effects [59]. The first is the increase in venous
return to the thoracic cavity due to both the increased pooling of blood in the pulmonary veins
and the increased flow out of the abdominal veins due to compression of the abdominal cavity
by the diaphragm [39,55]. The second effect is due to changes in arterial transmission [59,60].
The final effect is due to the nervous system directed vasoconstriction of the venous system
leading to reduced blood volume during inhalation [41,59].

**Respiratory-induced frequency modulation (RIFV)** Similar to the effect of RSA
on the ECG, RIFV is the difference in instantaneous heart rate observed in the PPG due to
respiratory sinus arrhythmia [39,47,48]. It can be monitored in the PPG in the same way by
noting the differences in the time length between peaks in the PPG [47,48].

2.4 Methods of Respiratory Rate Measurement

While respiration has numerous modulations which allow it to be observed on the ECG (RSA,
RPA, and RWA) and PPG (RIAV, RIIV, and RIFV), actually extracting these measures and
obtaining a medically useful respiratory rate is more difficult. Many different methods have been used to try to extract one or more of these modulations from the ECG and PPG including: digital filters, short-time fast Fourier transform, wavelet decomposition, autoregression, time-frequency spectral estimation, principle component analysis, and correntropy spectral density. The following section outlines each of these methods and discuss their previous applications to extracting respiratory information from either or both the ECG and PPG.

**Digital Filters** One of the simplest methods for obtaining the respiratory rate, particularly and most often, the RIIV from the PPG has been digital filtering. This is because the RIIV represents a unique respiratory signal in the DC region of the PPG while the cardiac signal lies in the AC region \[61\]. As a result, it has been possible to extract both the cardiac signal and the RIIV signal from the PPG by using different digital filters to remove the desired signal from the noise \[62\]. In one of the first instances of using digital filtering to detect respiratory rate from PPG, it was found that it was possible to extract the cardiac signal using a bandpass filter and then based on the estimated heart rate, one of three lowpass filters with varying cutoff frequencies could be used to detect the respiratory signal \[62\]. Most often methods using digital filters have extracted the respiratory signal from the PPG using either the fast Fourier transform or simple peak detection \[57,62\]. Through time-series analysis techniques, digital filters have been used to extract respiratory rate from the RIAV, RIIV, and RIFV modulations of the PPG \[47,62,65\] and the RSA from the ECG \[60\].

**Short-Time Fast Fourier Transform (STFFT)** In addition to simple digital filtering methods, the STFFT has been used to estimate respiratory rate of the RIIV signal in the PPG \[67\]. One of the limitations of the FFT is that it can only detect if a certain frequency is present in a sample of data, it cannot detect where that frequency is present in the signal. The STFFT, in contrast, uses much smaller sliding windows and performs sequential FFTs on those windows. This allows for a much finer time resolution as well as the observation of long term trends when all of the data are plotted and viewed together \[67\].

**Wavelet Decomposition** Both the continuous wavelet decomposition and the discrete wavelet decomposition have been used to extract the respiratory rate from the ECG \[68,69\] and PPG \[36,70,71\]. Wavelet decomposition has been widely used in signal processing as it allows
the time-frequency unfolding of signals in the time domain \[36\]. The wavelet decomposition method works by cross-correlating the input signal with a wavelet function of a given length and shifting that function the entire length of the input signal. The wavelet function is then stretched and the process is repeated. This is done repeatedly and ultimately allows for a finer understanding of the details of a signal to emerge which is particularly useful when dealing with signals where the long-term frequency is not necessarily uniform (i.e. the respiratory rate over a long period of time will not remain the same) \[72\].

**Autoregression** Autoregressive modelling works on the principle of using a certain number of previous data points to explain the current data point. In essence, it is a linear prediction where the current value is modelled as a sum of a set number (p) of the preceding values \[73\,74\]. The result of an autoregressive model is a number of poles which represent the dominant frequencies in a signal. Using this information, and with proper pre-processing, the highest magnitude poles (the poles that are most dominant in the signal) can be used to express the respiratory rate \[73\]. A simple autoregressive model has been used to extract RIIV information from the PPG \[73\] and the RSA and RPA from the ECG \[75\]. Furthermore, more computationally advanced methods have used autoregression as the core of their respiratory rate estimation algorithms including ARxCor \[34\] and ARSpec \[76\].

**Time-Frequency Spectral Estimation** Time-frequency spectral estimation, specifically variable-frequency complex demodulation (VFCDM) has been used to extract the RIIV signal from the PPG \[77\]. VFCDM is a two-step process. The first step is to decompose the signal into sinusoidal modulations using complex demodulation. The second step is to use the calculated centre frequencies from the previous step as the backbone to obtain the entire frequency spectrum \[78\].

**Principal Component Analysis (PCA)** PCA is method that is most often used for identifying patterns in data and reducing the dimensionality of large, multidimensional data sets \[79\]. However, by separating individual heartbeats from either the PPG or ECG, it is possible to obtain a feature matrix which contains all the individual heartbeats in which PCA can be conducted. By doing this, the dimensionality of the heartbeats is reduced and the principle component (PC), the axis that contains most of the variation, can be extracted \[79\].
This technique has been used successfully to extract the RIIV and RIFV from the PPG [79] and the RPA and RWA from the ECG [80, 81].

**Correntropy Spectral Density (CSD)**  
CSD is one of the most recent techniques used for extracting the respiratory rate. CSD provides improved resolution in the frequency spectrum compared to standard power spectral density methods [82]. The method works using correntropy, a correlation function that can provide information on higher-order statistics [82]. The method has recently been used to predict the heart and respiratory rates from PPG data [82].

### 2.5 The Respiratory Quality Index

Currently, the process of detecting the respiratory rate from either the PPG or ECG is a four step process. The first step is the acquisition of the data from the PPG or ECG monitor. The second step is an optional step that looks to discard poor quality data due to a variety of factors including: detached leads, low signal to noise ratio, and poor lead placement among others. Many different algorithms for both PPG and ECG have been derived to achieve this aim [58, 83–85]. The third step is extracting the relevant time series features based on the previously described modulations of the respiratory rate on the ECG (Figure 2.1) and PPG (Figure 2.2). The fourth step is applying one of the suites of respiratory rate detection algorithms described in the previous section. Of note is the fact that some of these detection algorithms are designed to work for specific respiratory modulations (i.e. RIIV) and as a result feature extraction is not necessary for these methods; however, many of these methods can be adapted for all six of the noted respiratory modulations if a feature extraction step is first implemented.

This methodology for extracting the respiratory rate however has a number of shortcomings related to the quality of the respiratory signals that can be obtained for any one of the modulations described. Particularly, the modulations are often very subtle and even under ideal circumstances are hard to detect [77]. Furthermore, for clinical populations, ideal circumstances are nearly impossible to achieve and often the signals are corrupted with noise artifacts [77]. To compound these challenges, it has been widely found that the specific respiratory modulations on the ECG and PPG are patient specific and it is hard to predict which modulation will be most prevalent for a particular patient. For example one study found that the PPG modulations
that worked best for patients was dependent on a multitude of factors including gender, body position, and respiratory rate [48]. Further research on RSA as a physiological phenomenon has shown that its prevalence is highly dependent on pre-existing conditions health conditions, age, hydration levels, and a patient’s level of physical activity [40].

Ultimately, these shortcomings suggest that for any respiratory rate extraction algorithm used on its own, even the best-performing or most-sophisticated algorithm, extraction of the respiratory rate from the PPG or ECG for all patients may not be possible. This shortcoming has led researchers to try to account for this by pursuing multi-parameter and smart fusion methods that are capable of taking respiratory rate estimations from multiple different modulations and merging them into a single respiratory rate [47]. However, the techniques currently used have further pitfalls in that the fusion algorithm used evenly weights all respiratory rate estimations and only discards estimations if they are outliers compared to the other estimations. One solution to this problem is to try to quantify the quality of the respiratory data that can be acquired for each specific patient for each specific respiratory modulation on the ECG and PPG. Ultimately, this would implement a step in the process of detecting the respiratory rate between the third (feature extraction) and fourth step (respiratory rate estimation). After all of the various features have been extracted, the quality of the respiratory signal obtained for each one can be obtained through a series of respiratory quality indices (RQIs). Using the results of these RQIs, certain data windows or modulations for certain patients can simply be removed from further analysis due to poor quality before they are used to predict a respiratory rate or they can be used to predict respiratory rate but be given a low weight in fusion models compared to other estimations to account for the fact that the confidence in the quality of the estimation is low. The remainder of this work will focus on the design and verification of a respiratory quality index methodology to be implemented for both PPG and ECG data.
Chapter 3

Dataset and RQI Methods

3.1 Datasets

The research in this report is based on results obtained from three different datasets: CapnoBase, MIMIC II, and PICRAM. Each of these datasets contains PPG and ECG waveform data as well as a respiratory rate “standard” signal from which a reference respiratory rate measurement may be used for comparison. The different respiratory standards of comparison used in the datasets represent a decreasing level of accuracy and confidence. The respiratory rate available in CapnoBase is based on measuring capnography and has been validated and had each respiration annotated by an expert rater, it is considered the “gold standard” of comparison because each respiration has been explicitly annotated. The respiratory rate available in MIMIC II is impedance plethysmography which is considered a “silver standard” because the data has not been labeled and may be imperfect due to challenges such as high noise levels or instances of detached leads. Finally, the respiratory measure available in the PICRAM dataset represents a clinician’s tabulation of a patient’s breathing rate at a given time. This is considered a “bronze standard” as respiratory rates calculated by clinicians are known to be less accurate than other measures. Despite these challenges, these three databases have been used because they increasingly represent the less controlled setting that is present in most hospital wards. CapnoBase was collected on patients that were anesthetized, MIMIC II was collected on patients that were bedridden and immobile, and PICRAM was collected on patients that were mobile within the hospital unit. Thus, CapnoBase is considered an ideal scenario and is used as an initial validation of the RQI. MIMIC II is used as a large scale validation in an ICU.
or HDU setting, and PICRAM is used as a validation in a step-down or outpatient ward where
the level of patient monitoring using advanced equipment is minimal.

3.1.1 Capnobase

The data contained in the CapnoBase dataset was collected by researchers at the University
of British Columbia, Vancouver, Canada. The dataset used in this research is part of a larger
dataset which includes capnography data collected from clinical cases, simulation data, as well
as the respiratory rate from the PPG dataset used in this research. The respiratory rate from
the PPG dataset contains raw waveforms for capnography, PPG, and ECG data. Additionally,
it contains respiratory rate from capnography, instantaneous heart rate from ECG, R peaks
from ECG and waveform peaks from PPG and ECG all of which is validated by an expert rater.
This data was collected for 94 patients (59 paediatric and 35 adults) undergoing elective surgery
or routine anesthesia. Of the 94 patients, high-quality eight minute segments of waveform data
are available from 42 randomly selected patients comprising 29 paediatric cases and 13 adult
cases [47].

3.1.2 MIMIC II

The data contained in the MIMIC II dataset was collected from 2001 to 2007 at the Beth Israel
Deaconess Medical Center, Boston, MA, USA in four ICU units within the hospital: medical
ICU, surgical ICU, cardiac ICU, and cardiac surgery ICU. While the MIMIC II dataset contains
records for over 25,000 adult patients, the data used for this research were only those instances
where waveform data was available for a patient, this amounted to a total of approximately
2,800 records. Furthermore, only patient recordings that included TTI, ECG, and PPG could
be used in this research which ultimately left a subpopulation of 1017 patients for analysis. The
median age of this subset of the patient population where age was recorded (1009 instances)
was 66 with a minimum age of 18 and a maximum age of 91.4. Additionally, for this subset,
55.2% of the population was male and 44.8% of the population was female where gender was
recorded (1009 instances) [86,87].
3.1.3 PICRAM

The Post Intensive Care Risk Alerting and Monitoring (PICRAM) dataset was collected between 2012 and 2014 at the John Radcliffe Hospital, Oxford, UK. Patients enrolled in PICRAM were initially in the ICU and were followed after discharge from the ICU to a step-down ward. The data used in this study comprise only the data collected after patients were discharged to the step down ward. The data that are available for each patient in the step down ward includes both PPG and ECG waveforms and the patient’s vital signs as collected by the nursing staff, usually once every hour. In total, 91 patients were monitored with a mean age of 61.48±16.83. 64.84% of patients were male and 35.16% were female. For this study, eight minute waveform segments were collected for the first 25 patients enrolled in PICRAM by locating the point on the waveforms where nurse observations of respiratory rate were taken and using the four minutes before and four minutes after these measurements to create the eight minute segment and using the nurse observation as the standard of comparison. Because the respiratory rate was almost always monitored more than once for each patient during their hospital stay, in total 474 eight minute segments and 501 eight minute segments were extracted for the 25 patients from the PPG and ECG signals respectively. The discrepancy in the number of segments for PPG and ECG arises because in some instances either PPG or ECG waveform data was not available when a respiratory rate measure was taken by a clinician.

3.1.4 Acquisition of Data and Respiratory Rate Standards

While CapnoBase, MIMIC II, and PICRAM all contain PPG and ECG waveform data for large patient populations, each one is represented by a different “prototypical” patient. In CapnoBase, the patients are anesthetized, in MIMIC II they are bedridden, and in PICRAM, they are mobile. Additionally, each dataset uses a different respiratory rate standard of comparison. As a result, acquisition of the data and derivation of the standard respiratory rate estimation was different for each dataset.

**CapnoBase**  The CapnoBase dataset contained a single eight minute segment containing capnography, PPG, and ECG for each patient. The eight minute segment was sectioned into 15 non-overlapping 32 second windows for each patient. The “gold standard” respiratory rate from
CapnoBase was derived individually for each window by summing the number of respiratory annotations made by the rater within that window and using it to calculate respirations/minute.

**MIMIC II** For the 1017 patient records used from MIMIC II, one eight minute segment of data for each patient was extracted by taking the first eight minute segment of data after the first hour of recording (i.e. the segment for each patient is from the 60th to the 68th minute). This was done because often the initial setup of the equipment on the patient can lead to poor quality data over the course of the first hour of monitoring. In the same way as the CapnoBase dataset, these eight minute segments were split into 15 non-overlapping 32 second windows for each segment. The “silver standard” respiratory rate was derived for these windows using the TTI waveform. This was done by using ARxCor [34] and ARSpec [76] to estimate the respiratory rate from the TTI. These two estimates were compared and in instances where their estimates disagreed by $\geq$ 2 breaths/minute, the 32 second segment was removed for analysis under the assumption that a reliable standard respiratory rate estimate could not be acquired. If the disagreement in the two estimates was $\leq$ 2 breaths/minute, the two values were averaged and used as the “silver standard” respiratory estimate for the window.

**PICRAM** In the PICRAM dataset, the only respiratory rate estimate available for comparison to that derived from the PPG and ECG is based on clinician observations during a patient’s hospital stay. While this is the most unreliable means of acquiring a respiratory rate estimate [19], for datasets collected in general and step-down wards, it is most likely to be the only estimate available. Similar to CapnoBase and MIMIC II, eight minute segments of PPG and ECG data were extracted from the PICRAM dataset; however, these eight minute segments were extracted centered around a clinician estimate of respiratory rate. In practice, this meant that for every respiratory rate annotation made for a patient, the time of that annotation was used to extract the PPG and ECG waveforms for four minutes before the annotation and four minutes after the annotation. Unlike PICRAM and MIMIC II, this resulted in multiple eight minute segments being collected for each patient. The eight minute segments were again windowed into 15 non-overlapping 32 second windows; however, a unique respiratory rate estimate was not available for each of these windows. Based on the assumption that the variation in respirator rate would be low in an eight minute segment, the “bronze standard” clinician
respiratory rate estimate was assigned to each window within an eight minute segment.

3.2 RQI Model Development

One of the major challenges with respiratory monitoring from both the PPG and ECG is that the respiratory modulations of those signals may be of high quality for some patients but of low quality for other patients; however, \textit{a priori}, it is impossible to know which modulations will work best for which patients. Thus in order to use PPG and ECG to get the best estimation of respiratory rate, all three modulations for either PPG or ECG can be used and then combined to estimate a patient’s respiratory rate. However, a critical step in this process is quantifying the quality of the respiratory signal from each of these modulations using a respiratory quality index (RQI). In the following section, the development of four different RQI methods based on the fast Fourier transform (FFT), autoregression, autocorrelation, and the Hjorth parameters for both PPG and ECG are described.

3.2.1 Data Pre-Processing

Prior to implementing any of the RQI methods, three pre-processing steps were implemented. The first of the pre-processing steps was to create 32 second windows from the available data. Each dataset used (CapnoBase, MIMIC II, and PICRAM) has been processed so that for each patient an eight minute waveform segment is available. However, when predicting respiratory rate, the longer a segment is, the greater the reduction in real-time performance \cite{82}. Furthermore, the longer a window is the lower the likelihood of observing small timescale variations in respiratory rate \cite{36,67}. In order to address this, each eight minute waveform segment has been windowed into 15 unique, non-overlapping 32 second windows. Ultimately windowing the data on such a small timescale is necessary for two reasons. First, as mentioned, it increases the likelihood that the existence of small timescale variations in respiratory rate will be observed. Secondly, and most importantly, if it is assumed that the respiratory rate will not change dramatically during this timescale, the ideal respiratory rate can be modeled as a perfectly periodic function within the window, an assumption that is critical for the design of the various RQIs.

After creating the unique 32 second windows for each patient, the second pre-processing step was to extract the peaks and troughs of the PPG or the Q wave minimum and R wave
maximum of the ECG and use them to obtain the associated respiratory modulations. For the CapnoBase dataset, these points are provided and have been validated by an expert rater. For the MIMIC II and PICRAM databases, the PPG peaks and troughs are calculated using a delineator described by Li et al. [89] that has been adapted for PPG and is based on using a combinatorial analysis of the PPG waveforms their derivatives. For ECG, the R wave peaks are found using the Pan-Tompkins algorithm and the Q wave troughs are found by selecting the waveform minimum 200 ms prior to the R wave peak [90,91]. These peak and trough features are then used to extract three unique waveforms for PPG (RIAV, RIIV, and RIFV) and ECG (RSA, RPA, and RWA) based on the modulations illustrated in Figure 2.1 and Figure 2.2. The unique time series that is created for each waveform is then resampled onto a uniform grid of 4 Hz.

The final pre-processing step is to filter each unique PPG or ECG waveform. Each waveform is filtered using a 5th order passband Butterworth infinite impulse response (IIR) filter with a passband range of 0.83 Hz to 1 Hz (5 to 60 breaths/minute). An IIR filter was selected over a finite impulse response filter (FIR) because of the reduced passband ripple and improved stopband to passband cutoff in an IIR filter. Because the data has been windowed and is not being constantly updated, continuous filtering is not necessary making the IIR filter a plausible option. Furthermore, a Butterworth filter was selected over Chebyshev and Bessel filters as the flat magnitude in the passband was prioritized over a rapid rate of attenuation from the passband to the stopband.

The final result of the pre-processing steps is a 32 second window sampled at 4 Hz of one of the respiratory modulations. An example of the output of the pre-processing steps is shown in Figure 3.1 with both the raw respiration waveform and the filtered respiration waveform.

3.2.2 Data Quality Assessment

In addition to the data pre-processing steps, seven data quality steps were implemented to discard data windows that were not analysable. This process was not meant to be a standalone SQI but rather was meant to discard windows that were unusable. The first data quality step for PPG analysis was to discard patient samples that do not contain both the PPG waveform and a respiratory rate standard of reference or when either the PPG waveform or standard of
reference contains non-existent values. Equally, in ECG analysis, discard patient samples that do not contain both the ECG waveform and a respiratory rate standard of reference or when either the ECG waveform or standard of reference contains non-existent values.

The second quality step was to verify the quality of the respiratory rate estimation for the standard of comparison measurement for a given dataset. This was a step that was only conducted on the MIMIC II dataset as the respiratory annotations on the capnography for the CapnoBase dataset and the nurse observations of the respiratory rate for the PICRAM dataset were assumed to be correct. For the MIMIC II dataset however, because the respiratory rate had to be estimated from the IP waveform and collecting the IP waveform is subject to the same inconsistencies as collecting all other biological waveforms, the quality of the respiratory rate estimate had to be verified. This was done through using two respiratory rate detection algorithms based on autoregression methodologies, ARxCor [34] and ARSpec [76]. The respiratory rates given by these methodologies were compared and if the two estimations disagreed by more than two breaths/minute, the window was discarded under the assumption that a reliable standard of comparison could not be obtained. If the two estimations agreed within two breaths/minute, the two values were averaged and used as the standard of comparison.

The third quality step was discarding respiratory rates below what is physiologically possible. After the peak detection pre-processing step (described above) the number of peaks was calculated. If the number of peaks in a 32 second window was less than two, representing a likely
respiratory rate of less than four breaths/minute, the window was discarded as representing a respiratory rate below physiologically possible. The fourth and final quality step was conducted after the extraction of each PPG and ECG modulation. This step checked the waveforms of each modulation to ensure that the data vectors contained usable data as opposed to a data flat line or non-existent values.

3.2.3 Fast-Fourier Transform RQI

The first RQI method developed is based on the Fourier transform. The Fourier transform is a signal processing method used to describe the harmonic or frequency content within a time-series based waveform [92]. Where \( x(n) \) is a discrete time-signal, the discrete Fourier transform (DFT) can be defined as:

\[
X(m) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nm/N}
\]

where \( X(m) \) represents the output of the time-series waveform \( x(n) \) in the frequency domain. However, in most discrete signal processing based algorithms, the discrete Fourier transform is replaced by the equivalent fast-Fourier transform (FFT) due to its increased processing speed [92].

In this research, the FFT is used to create the FFT-RQI which is a method that assigns a numeric value to given window based on its FFT that defines how strongly that window exhibits a respiratory rate waveform. The FFT-RQI is calculated by first conducting FFT-RQI specific pre-processing on \( x(n) \), where \( x(n) \) specifically represents one 32 second window for one patient for one of the specific modulations (ECG: RSA, RPA, RWA, PPG: RIAV, RIIV, RIFV). Specifically, prior to taking the FFT of \( x(n) \), \( x(n) \) is zero-padded, where necessary, linearly detrended, and windowed using a Hamming window function. After pre-processing, the FFT of \( x(n) \) is taken yielding \( X(m) \). \( X(m) \) is used to calculate the maximum peak area (MPA) of the FFT which is calculated as the sum of the three largest continuous set of values of \( X(m) \) surrounding the largest value of \( X(m) \) within the respiratory frequency spectrum from 0.1 Hz to 0.75 Hz (6 breaths/min to 45 breaths/min). For example, the MPA could represented by one of the following sums where \( M \) represents the index of the largest value of \( X(m) \):
\[ MPA = X(m - 2) + \ldots + X(m) \]
\[ MPA = X(m - 1) + \ldots + X(m + 1) \]
\[ MPA = X(m) + \ldots + X(m + 2) \]

Following the calculation of the \( MPA \), the total area of the FFT within the physiological respiratory range, termed total respiratory area (\( TRA \)), is calculated as the sum of all values \( X(m) \) that fall between the frequency range of 0.1 Hz to 0.75 Hz. Using these two values, the FFT-RQI is calculated as:

\[ RQI_{FFT} = \frac{MPA}{TRA} \]

Ultimately, the FFT-RQI gives a value of zero to one where the closer the value is to one the stronger the dominant frequency within the respiratory range is. It is assumed that the larger dominant frequency is within the respiratory frequency range, the stronger the given respiratory modulation is for that window and the higher confidence there is in the respiratory rate estimation that is made by that window. An example of an FFT an the regions representing the \( MPA \) and \( TRA \) are shown in Figure 3.2.

**Figure 3.2:** Example FFT for one window of data, \( x(n) \). Red line is the FFT frequency spectrum. Solid black lines represent the region from 0.83 Hz to 1 Hz where the \( TRA \) is calculated. Dotted black lines represent the region \( MPA \) representing the largest three points clustered around the largest point in the FFT.
3.2.4 Autoregression RQI

In addition to the FFT-RQI, an RQI based on an autoregression function was designed. While both autoregressive modelling and the FFT can be used to derive the major frequency components of a waveform, autoregression is distinct from FFT because it provides a smoother, more exact interpretation of the frequency components and can be run on smaller time windows; however a major disadvantage of the autoregression function is determining the most advantageous model order to use [74].

The autoregression function is a means to predict the current value in a time series based on the past values from the series plus an error term. In essence, the autoregression function can be viewed as a set of autocorrelation function for every point \( x(n) \) in a series based on the previous \( x(n - i) \) terms and is defined as:

\[
x(n) = \sum_{i=1}^{M} a_i x(n - i) + e(n)
\]

where \( x(n) \) is the current value in the series, \( a_i \ldots a_M \) are the weighting coefficients, \( x(n - i) \ldots x(n - M) \) are the previous terms in the series, \( e(n) \) is the error term, and \( M \) is the model order which represents how many previous terms are used in predicting \( x(n) \) [74]. In practice, the weighting coefficients, \( a_i \ldots a_M \), are most often obtained using the Yule-Walker equation, defined as:

\[
a_{opt} = R^{-1}r
\]

where \( a_{opt} \) are the ideal weighting coefficients as set by \( R^{-1} \), the autocorrelation matrix and \( r \), the autocorrelation vector [74]. The ideal weighting coefficients, which define a function that recreates the observed signal \( x(n) \), can be used to define the transfer function \( H(e^{j\omega}) \) which can further be used to define \( X(\omega) \), the input sequence polynomial, the roots of which are the poles of the AR model which represent the dominant frequency components of the original signal \( x(n) \) [74].

Based on these features of the autocorrelation function, the autocorrelation RQI (AR-RQI) in a similar logic as the FFT-RQI seeks to assign a value to each window of data \( x(n) \) based on the strength of the dominant frequency component in that signal. However, one of the
shortcomings of autoregression is selecting the model order, \( M \). This is particularly pertinent in this instance when the AR-RQI will be applied uniformly to each \( x(n) \) where these \( x(n) \) represent data from different devices (ECG or PPG), different datasets, different patients, and different respiratory modulations. In order to address the possibility that no single model order would be suitable for this wide range of data, the autoregression for each window of data, \( x(n) \), was calculated for model orders \( M = 1 \) to 30. Then the ideal model order for each specific window, \( x(n) \), was selected by choosing the model order that returned the minimal Akaike’s Information Criterion (AIC). The AIC is defined as:

\[
AIC = \log(e \times (1 + \frac{2M}{N}))
\]

where \( e \) is the error term for the model, \( M \) is the model order, and \( N \) is the total number of data points in \( x(n) \) \(^{[74]}\). The AIC works to find the ideal model order for a given set of data by returning higher AIC values for large error terms and large model orders thus the ideal model order is one that has minimized the combination of the error and model order (Figure 3.3). Using the ideal model that is specifically determined for each \( x(n) \), the AR-RQI is determined by selecting the pole with the largest magnitude in the positive frequency range similar to the methodology used by Cazares et al. \(^{[93]}\) (Figure 3.4). The magnitude of this pole was used as the AR-RQI if the frequency of the pole fell within the 0.083 Hz to 1 Hz (5 breaths/min to 60 breaths/min) range and if it fell outside of this range, the AR-RQI was set to zero as it was determined that the noise of the signal outside of the respiratory range was likely too large to allow for an accurate respiratory rate estimation. In a similar fashion as the FFT-RQI, the values of the AR-RQI closest to one represent the signals where there is a dominating frequency within the respiratory range and likely represent the signals where the best respiratory rate estimations can be made.

### 3.3 Autocorrelation RQI

While the autoregression function can be thought of as a series of autocorrelation functions used to explain a single datapoint, it is also possible to define a simplified RQI using the autocorrelation function with lag times defined over the entire range of the signal. The autocorrelation
The ideal model order (represented by the red dot) is selected as the point where the AIC is minimized.

The function is defined as:

\[ r_k = \frac{c_k}{c_0} \]

where \( r_k \) is the autocorrelation value, \( c_0 \) is the sample variance, and:

\[ c_k = \frac{1}{N-1} \sum_{n=1}^{N-k} (x(n) - \bar{x}) \ast (x(n+k) - \bar{x}) \]

where \( N \) is the total length of the sample, and \( \bar{x} \) represents the mean of the sample. Ultimately, this function is scaled for unity such that when the lag of \( k \) is zero, the autocorrelation value is one indicating a perfect alignment of the two signals. For every other alignment the autocorrelation value is between zero and one. In the instance where a perfect sinusoid is present, when the lag, \( k \), is as long as the period of the sinusoid, the autocorrelation value would again reach one. It is this principle that the autocorrelation RQI (AutoCor-RQI) is based on. The AutoCor-RQI calculates the autocorrelation of a window \( x(n) \) for every lag time in the signal, from \( k = 0 \ldots (N - 1) \). Under the assumption that the respiratory signal is expected to be sinusoidal, the AutoCor-RQI selects the maximum autocorrelation within the lag range of \( k = 1.333 \text{ seconds} \) to \( k = 10 \text{ seconds} \) (6 breaths/min to 60 breaths/min). The AutoCor-RQI is then assigned to be the autocorrelation value of the specified \( k \) value because this represents the point where the signal most closely mirrors itself and thus the closer the value is to one at this point, the more sinusoidal the signal and the more likely it is that an accurate respiratory
Figure 3.4: Plot of the autoregression poles (plot includes conjugate pairs) for an example window, \( x(n) \). The AR-RQI is selected as the pole with the largest magnitude from the first two quadrants that falls within a range from 5 breaths/min to 60 breaths/min (blue star). In the event that the largest pole falls outside of that range, the AR-RQI is set to zero.

Rate can be extracted from this window (Figure 3.5).

### 3.3.1 Hjorth Parameter RQI

Hjorth parameters were first defined in 1970 as a set of descriptive statistics looking to define the amplitude and time pattern qualities of a signal [95]. Three Hjorth parameters have been defined: activity, mobility, and complexity. Activity, often referred to as the variance or mean power, is a measure of the squared standard deviation of a signal’s amplitude. Mobility, known as the mean frequency, is the standard deviation of the slope with reference to the activity. Finally, complexity is a measure of the “softness” of a curve where the “most soft” curve is the sinusoid which has a complexity value of unity [95]. The three Hjorth parameters can be calculated as follows:

\[
\text{Activity} = m_0 = f(t) \\
\text{Mobility} = \sqrt{\frac{m_2}{m_0}} = \frac{d(f)}{dt}
\]
Figure 3.5: Plot of the autocorrelation of one window $a(n)$. Autocorrelation value represents the alignment of the original signal and the lagged signal where unity represents perfect autocorrelation. The black bars represent the largest autocorrelation between a lag of 5.33 samples and 40 samples (which corresponds to 1.33 seconds and 10 seconds at a sampling rate of 4 Hz).

$$\text{Complexity} = \frac{\sqrt{m_4}}{\sqrt{m_2}} = \frac{\sqrt{m_2}}{\sqrt{m_0}} = d^2(f) dt^2$$

where $m_0$, $m_2$, and $m_4$ represent the zeroth, second, and fourth spectral moments, respectively of the signal and $f(t)$ represents the curve as a function of time [95]. Again making the assumption that the ideal respiratory waveform would be a perfect sinusoid, the Hjorth parameter of interest for the Hjorth RQI is complexity. Similar to the other RQIs, when the Hjorth complexity parameter is closest to one it represents the most sinusoidal signal. Thus for each window, $x(n)$, the windows with the highest Hjorth complexity are expected to be the most sinusoidal and thus are expected to be the most likely to give an accurate estimation of respiratory rate.

3.4 Validating the RQIs

Validating the RQIs involved two primary: (1) defining a standard of comparison based on previously implemented signal quality methods, specifically a pre-processing signal quality index (SQI), and (2) defining the methodology to compare the performance of each RQI and the SQI.
3.4.1 SQI Development

In order to verify that the RQIs developed improve the ability to determine what waveforms contain good respiratory information compared to those that provide poor respiratory information, a SQI was developed that is capable of removing low quality ECG and PPG data before the respiratory modulations are calculated. This allowed for the comparison of the data the SQI deemed to be of poor quality compared to the data that the RQI deemed to be of poor quality in order to determine if the RQIs were specifically able to remove poor respiratory data where the SQI was not.

The SQI implemented is based on a recent peak comparison algorithm using two peak detectors for the PPG and two R wave peak detectors for the ECG developed by Johnson et al. [88]. For both the PPG and ECG, one of the peak detectors was assigned as the standard and the other was assigned as the comparison detector which allowed for the calculation of an F1 score. The F1 score was defined as:

\[
F_1 = \frac{2 \times TP}{(2 \times TP) + FP + FN}
\]

where \(TP\) is a true positive and is defined as the event when both peak detectors found a peak within \(\pm 150\) ms of each other, \(FP\) is a false positive defined as an event where the comparison detector found peak that was not within \(\pm 150\) ms of a peak found by the standard, and \(FN\) was a false negative defined as an event were the comparison detector did not find a beat within \(\pm 150\) ms of a beat detected by the standard detector. The result of this calculation is that if all beats detected by the two peak detectors are true positives, the value of the F1 score will be unity; however, every false positive and false negative detection will reduce the F1 score providing an indication of a poor quality or very noisy signal.

3.4.2 RQI Comparison Methodology

In order to validate the RQIs among each other and against the SQI, a method of comparison that considered the performance of each RQI for a given window, \(x(n)\), to the known respiratory rate for that window given by the gold, silver, or bronze respiratory rate measure was necessary. This was done by using a panel of respiratory rate algorithms to estimate the respiratory rate
of each individual window, \( x(n) \), after the pre-processing step. The respiratory rate algorithms that were chosen were two of the most recent algorithms: ARxCor \[34\] and ARSpec \[76\] as well as an algorithm recently used based on an FFT \[47\]. Given the state of the art nature of these algorithms, and their varied methodologies, the assumption was made that these algorithms would make an accurate estimation of respiratory rate compared to the ground truth when the extracted respiratory waveform data was high quality and the estimations would be less accurate when the respiratory waveform was of lower quality. Based on this assumption, an accuracy measure could be applied to the entirety of the dataset being considered where the entirety of the dataset is every 32 second window for every patient for every modulation for one of the two initial waveform types (PPG or ECG) that made it through the pre-processing steps without being discarded. Three different accuracy measures that compare and sum the difference between a given respiratory rate algorithm (ARxCor, ARSpec, or FFT) compared to the gold, silver, or bronze standard were considered: root mean squared error (RMSE), the mean absolute error (MAE), and the median and interquartile range for the dataset. Ultimately, the accuracy measure that was implemented was the MAE. MAE was chosen as the optimal choice in this instance because it is less sensitive to outliers than the RMSE but still considers a measure of the average of the data where the median does not \[96\]. The MAE is defined as:

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |(\hat{x}_i - x_i)|
\]

where \( N \) is equal to the number of samples, and \( x_i \) is the predicted respiratory rate and \( \hat{x}_i \) is the known gold, silver, or bronze standard respiratory rate \[96\]. Following the calculation of the MAE for the entirety of a dataset (as previously described), data was sequentially discarded in increments of 5% of the data using the RQIs as the standard for removal. Because each RQI is designed to yield a value associated with data quality from zero to unity (with unity being high quality data), on the first removal step, the 5% of the data with the lowest value for each RQI was discarded. This process was done independently of the other RQIs as the goal was to compare the performance of each RQI to the others. What this means is that after the first 5% removal step there were four unique sets of data, one for each RQI, that contained the same number samples, but did not contain the exact same set of samples. The MAE for each of these new sets of data was then calculated under the assumption that if the RQI was capable
of finding poor quality data, the overall MAE for the new dataset should have decreased. This 5% removal and MAE recalculation step was then repeated in 5% increments until only 5% of the data remained. In addition to conducting this analysis for each RQI, it was also conducted for the SQI and for an ideal situation. The ideal situation was created simply by throwing out the 5% of the data with the largest absolute difference between the respiratory rate estimation algorithm and the gold, silver, or bronze standard. The closer an RQI is to the ideal situation, the closer it is to throwing out only the poorest quality data and keeping the best quality data on every increment.

3.5 Conclusion

This chapter described the RQI methods that have been derived to try to find poor quality respiratory waveforms after the respiratory rate has been extracted but ultimately before the waveform is used to predict respiratory rate. Additionally, this chapter described the three datasets these RQIs have been tested on: CapnoBase, MIMIC II, and PICRAM as well as the pre-processing, data quality, and validation measures that have been put in place to test these RQIs. In the following chapter, all of these elements will be put together to test the performance of the RQIs against their stated aim of being able to recognize poor quality respiratory data better than previously implemented measures.
Chapter 4

Analysis of Respiratory Quality Indices

4.1 Overview

The methods described in Chapter 3 were used to test and validate the four RQIs described. The analysis was conducted on each permutation of data presented in Table 4.1 where each column represents one of the selection input choices for any given analysis. This resulted in a total of 72 unique analyses being conducted along with 36 control analyses (using the SQI or the ideal situation as the “RQI”). As a result of the large number of analyses conducted, only a representative selection of data is presented.

However, in order to allow for the comparison between all of the analyses, the RQI Comparison Metric (RCM), a method that allows for the controlled comparison across all of the analyses is defined. This is followed by a presentation of the results of the RCM for every data analysis permutation.

Table 4.1: RQI Analysis Selection Input Choices

<table>
<thead>
<tr>
<th>Database</th>
<th>Input Waveform</th>
<th>RR Prediction Algorithm</th>
<th>RQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CapnoBase</td>
<td>Photoplethysmography (PPG)</td>
<td>ARSpec</td>
<td>Fast Fourier Transform (FFT)</td>
</tr>
<tr>
<td></td>
<td>(RIAV, RIV, RIFV)</td>
<td>ARxCor</td>
<td>AutoRegression (AR)</td>
</tr>
<tr>
<td>MIMIC II</td>
<td>Electrocardiogram (ECG)</td>
<td>Fourier Analysis</td>
<td>Autocorrelation (AutoCor)</td>
</tr>
<tr>
<td>PICRAM</td>
<td>(RPA, RWA, RSA)</td>
<td></td>
<td>Hjorth Parameter Complexity</td>
</tr>
<tr>
<td></td>
<td>SQI Control</td>
<td></td>
<td>Ideal Control</td>
</tr>
</tbody>
</table>
4.2 Performance of Respiratory Quality Indices

4.2.1 Plots of RQI Performance

Due to the large number of analyses run, only a subset of representative plots are shown. Each plot shows the MAE at each stage of the 5% increment of data discards for each RQI with all four RQIs and two control analyses represented on each plot. In each plot, the FFT-RQI is shown in red, the AR-RQI in blue, the AutoCor-RQI in green, the Hjorth Complexity RQI in cyan, the SQI control in magenta, and the ideal control in black. In each figure, the RQI is shown for PPG signals (left panel, all PPG modulations merged) and ECG signals (right panel, all ECG modulations merged) using ARxCor to predict the respiratory rate. Figure 4.1, Figure 4.2, and Figure 4.3 are conducted on the CapnoBase, MIMIC II, and PICRAM datasets, respectively.

![RQI Performance Plot](image)

**Figure 4.1:** RQI performance plot displaying the MAE values for each RQI as the poorest quality data is sequentially removed. RQIs are applied to the CapnoBase dataset and use the ARxCor respiratory rate estimation algorithm. The left panel shows PPG data for all PPG modulations combined into one dataset and the right panel shows the ECG data for all ECG modulations combined into one dataset.

4.2.2 RQI Comparison Metric

Because a total of 72 analyses were conducted using every possible permutation of the selection of the database, the input waveforms, the respiratory rate estimation algorithm, and the RQI,
designing a metric that allows for the controlled comparison of the results across all of these permutations was necessary. In order to achieve this, the RQI Comparison Metric (RCM) was created. RCM can be calculated as follows:

$$RCM = \frac{MAE_{RQI, 50\%}}{MAE_{Ideal, 50\%}}$$

where $MAE_{RQI, 50\%}$ is the mean absolute error of any given RQI after 50% of the data has been discarded and $MAE_{Ideal, 50\%}$ is the specific mean absolute error at 50% data removal calculated for the same dataset, waveform, and estimation algorithm as the RQI being considered. While the RCM could theoretically be chosen for any removal percentage, 50% removal was chosen as it represents the likely upper limit of the amount of data that the RQI might remove in an actual clinical setting. The RCM represents how much better the ideal control performs compared to the RQI at the 50% data removal level. For example, if the RCM is 2, the ideal control has a MAE 2x lower than the RQI in question. Furthermore, a perfect RCM would be defined as 1 and would indicate that for the RQI in question, the MAE between the ideal control and the RQI are the same. Because the RCM evaluates the performance of an RQI compared to the ideal performance for that particular model, it makes it suitable for comparing between
Figure 4.3: RQI performance plot displaying the MAE values for each RQI as the poorest quality data is sequentially removed. RQIs are applied to the PICRAM dataset and use the ARSpec respiratory rate estimation algorithm. The left panel shows PPG data for all PPG modulations combined into one dataset and the right panel shows the ECG data for all ECG modulations combined into one dataset.

Table 4.2: RQI Comparison Metric

<table>
<thead>
<tr>
<th>RQI</th>
<th>PPG</th>
<th>ECG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FFT</td>
<td>AR</td>
</tr>
<tr>
<td>CapnoBase, ARSpec</td>
<td>3.90</td>
<td>3.44</td>
</tr>
<tr>
<td>CapnoBase, ARxCor</td>
<td>7.36</td>
<td>7.37</td>
</tr>
<tr>
<td>MIMIC II, ARSpec</td>
<td>4.08</td>
<td>4.66</td>
</tr>
<tr>
<td>MIMIC II, ARxCor</td>
<td>2.47</td>
<td>2.75</td>
</tr>
<tr>
<td>MIMIC II, FFT</td>
<td>4.01</td>
<td>4.72</td>
</tr>
<tr>
<td>PICRAM, ARSpec</td>
<td>1.69</td>
<td>1.79</td>
</tr>
<tr>
<td>PICRAM, ARxCor</td>
<td>1.51</td>
<td>1.67</td>
</tr>
<tr>
<td>PICRAM, FFT</td>
<td>1.74</td>
<td>1.91</td>
</tr>
</tbody>
</table>

models. The RCM for all 72 analyses as well as the RCM calculated for the SQI control are shown in Table 4.2.
Chapter 5

Discussion

5.1 Overview

The panoply of different models that the RQIs were applied to, has allowed for a robust analysis of how the RQIs perform under varied conditions. In the following section, the goal is to answer four specific questions about the performance of the RQIs using the data presented in Chapter 4. The first question is how the RQIs perform overall. This question is the primary motivation for this study and is answered through looking at how the RQIs have performed compared to the SQI control as well as among each other. The second, more nuanced question, seeks to answer if certain RQIs perform preferentially better when one of the respiratory rate estimation algorithms is applied compared to a different algorithm. The third question is how the RQI performance is affected as the data quality becomes worse. This will specifically involve looking at the performance trend as the models progress from CapnoBase to MIMIC to PICRAM. The final question is whether the RQIs work better for analysing PPG or ECG data. Finally, after answering these questions, some of the limitations of the RQIs will be addressed.

5.1.1 General RQI Performance

The results in Figure 4.1, Figure 4.3, and Figure 4.3 indicate that the FFT-RQI, AR-RQI, and AutoCor-RQI outperform the SQI control in the data discard region of interest (around approximately the 50% data discard region). The Hjorth Complexity RQI, however, does not appear to outperform the SQI and is in general the worst performing RQI. These general trends seen in the graph are supported by the RQI Comparison Metric as the FFT-RQI, AR-RQI,
and AutoCor-RQI outperform the SQI control in every instance where the RQIs were run on the CapnoBase and MIMIC II datasets for the same waveform using the same respiratory rate estimation algorithm as the SQI control, except for two instances. These two instances occur when the AR-RQI is applied to the MIMIC II PPG data using the ARSpec and FFT respiratory rate estimation algorithms, when the AR-RQI has RCM equal to 4.66 and 4.72, and when the SQI control in both instances is 4.64 yielding an improvement of the SQI of only 0.03 and 0.09 respectively. Additionally, while the performance of the FFT-RQI, AR-RQI, and AutoCor-RQI outperform the SQI in the PICRAM dataset when using the ARxCor algorithm, the RQI algorithms are slightly outperformed by the SQI when the ARSpec and FFT algorithms are applied. Overall, however, these results shown in Chapter 4 provide substantial evidence for one of the critical arguments in support of RQIs, ultimately showing that even without proper data cleaning prior to respiratory-rate extraction, the addition of an RQI processing step has the ability to improve respiratory rate estimation substantially.

Interestingly, the average performance across all of the data analyses for the FFT-RQI, AR-RQI, and AutoCor-RQI was almost the same. The mean RCM performance across every analysis for the FFT-RQI was $3.37(\pm 2.01)$, for the AR-RQI, $3.53(\pm 1.89)$, and for the AutoCor-RQI, $2.84(\pm 1.19)$. Ultimately this suggests that, while performance in specific instances for some RQIs may be far better than for other RQIs, in general, of these three RQIs, each may add something to a more complex model.

5.1.2 RQI Robustness

The second question of interest is if some RQIs perform preferentially better when particular rate estimation algorithms are applied. It was observed that for the data in general this outcome is independent of the method used to estimate respiratory rate. For example, the mean RCM for the FFT-RQI on all instances where ARSpec was used was $3.07(\pm 1.16)$, for ARxCor, $3.98(\pm 3.24)$, and for FFT, $3.05(\pm 1.13)$. Furthermore, the very close agreement of the RCM values of the ARSpec and FFT respiratory extraction methods is very interesting, as these two extraction techniques use very different signal processing principles. However, the one caveat to this general finding is that the general trends that were observed for each RQI regardless of which respiratory rate algorithm was used in both the CapnoBase and MIMIC II datasets
were not seen in the PICRAM data set for any of the RQIs when ARSpec and FFT algorithms were used. The trend seen across the other two datasets did however hold when the ARxCor algorithm was applied. This suggests that in general the critical finding that regardless of which respiratory rate extraction method is used, the implementation of an RQI can add value to the respiratory rate estimation is true; however, there may be an upper limit to RQI performance based on data quality and that level of performance might be slightly higher when using ARxCor as opposed to ARSpec or FFT.

5.1.3 Performance of RQI with Decreasing Data Quality

The third question addressed in this study was how the RQI would perform in cases of decreasing data quality. This was the reason that the three databases: CapnoBase, MIMIC, and PICRAM were chosen. CapnoBase represents a database with high quality signals and “gold standard” respiratory rate comparison, while MIMIC II represents a lower quality signal and “silver standard” of comparison, with PICRAM having the lowest quality signal and a “bronze standard” of comparison. Overall, the RQIs indicated that when considering only the CapnoBase and MIMIC II datasets, the RQIs did not perform any worse as the quality of the data decreased. In fact, it was found that in a number of instances, the RCM value for an RQI holding the waveform and respiratory rate estimation method constant actually decreased from CapnoBase to MIMIC.

However, when PICRAM is considered, the results are less clear. The one clear observation is that the performance of all of the RQIs is consistent among all three datasets when ARxCor is used as the respiratory rate estimation algorithm. This result is demonstrated in Figure 5.1 which plots the RCM trend for each ARxCor model in order from highest quality data (CapnoBase) to lowest quality data (PICRAM). An interesting finding is the reduction in RCM as the data quality decreases, particularly for the Hjorth Parameter RQI. However, this is likely due to the fact that as the data quality decreases, there is an increase in the MAE value for the ideal control. However, the reduction in the ability to estimate respiratory rate as the data quality decreases is to be expected, thus the MAE is expected to increase for noisier datasets. What is more important however is the trends that can be seen in Figures 4.1, 4.2, and 4.3 where the trends seen for each RQI appear to hold. Interestingly however this consistent
trend as seen for all RQIs for the ARxCor algorithm does not hold for the RQIs when either the ARSpec or FFT algorithms are applied suggesting again as was previously mentioned, the potential of an upper limit of performance given poor data quality for the RQIs when these algorithms are applied.

![Figure 5.1: Plot of the RCM trends across the three datasets (CapnoBase, MIMIC II, and PICRAM) in order of decreasing data quality for all RQI methods for both PPG and ECG using the ARSpec respiratory rate estimation algorithm.](image)

### 5.1.4 RQI Value for PPG and ECG

The last critical finding of this study is the comparison of the results of using the PPG and the ECG. While it is often anticipated that PPG data is of poorer quality than ECG data due to a higher risk of motion artifact, our results do not seem to suggest a poorer performance for the RQIs when applied to the PPG versus the ECG. This is a critical finding as it means that regardless of the input waveform, the importance of the RQIs is not diminished.

### 5.1.5 Limitations of the RQIs

While the RQIs appeared to function as hypothesised, there do appear to be a number of limitations. The first of these limitations is that even in the best scenario, the RCM was still 1.47 times worse at discriminating the highest-quality data from the lowest-quality data than the “ideal” control. While this is a clear improvement compared to just using an SQI alone, there is still substantial room to improve the performance of the method.
Furthermore, another limitation of the RQIs concerns the Hjorth Complexity RQI. This RQI in many instances appeared to perform the opposite of as expected, with the MAE increasing as more data were discarded. This may suggest that, despite initial assumptions, high-quality respiratory information is not purely sinusoidal but perhaps is better represented by a different periodic function such as a triangular wave or asymmetric wave. If this is the case, it raises interesting implications in basing the RQIs primarily on methods that perform best when the underlying information of interest is expected to be sinusoidal.
Chapter 6

Future Work

6.1 Conclusions

The results of this research have shown that it is possible to determine, without \textit{a priori} knowledge, the quality of respiratory signals derived from PPG and ECG using a novel set of respiratory quality indices. Importantly, this research tested four RQIs based on principles from Fourier analysis, autoregression, autocorrelation, and the Hjorth complexity parameter and determined that the FFT, AR, and AutoCor RQIs were able to more accurately classify respiratory signals as being of high or low quality than a signal quality index applied prior to extraction of the respiratory signal. This is significant because it supports the addition of an RQI step in the normal procedure of estimating respiratory rate. Furthermore, this research has not only demonstrated the value of implementing an RQI step but has shown that the FFT, AR, and AutoCor RQIs work effectively on multiple datasets of varying signal quality, that they work equally effectively using respiratory rates derived from both PPG and ECG, and that they work equally as well for three different respiratory-rate estimation algorithms: ARSpec, ARxCor, and FFT analysis. The research presented in this report represents the initial validation of these RQIs, and based on our positive results, suggest the value in further improving these methods and ultimately implementing them in a real-world setting.
6.2 Future Work

Future work on the RQIs described in this report will focus on bringing the benefits of high-quality respiratory rate estimations from non-invasive means such as the PPG and ECG to as wide a patient population as possible. The first stage of this work, described in Section 6.2.1 will focus on two primary tasks to improve the RQI methodology. The first is deriving decision criteria using machine learning methods to determine (based on results of the RQIs) when it is acceptable to derive the respiratory rate from an extracted respiratory waveform. The second is fusing the acceptable respiratory rate estimations for windows taken over an identical period (i.e., the respiratory rate derived from more than one modulation over the same window) into a single high-quality respiratory rate estimate for that given window. The second stage of this work, described in Section 6.2.2 will focus on integrating the high-quality respiratory rate estimation technology (described in Section 6.2.3) into current models using non-invasive (contact and non-contact) continuous vitalsign monitoring to predict clinical decline. Finally, the last stage of this research, described in Section 6.2.4 will be to implement the various models for predicting respiratory rates, based on their computational demand, into an “m-health” setting for both developed and developing regions of the world. The overall timeline for this project is outlined in Section 6.2.4.

6.2.1 Fusion of Respiratory Rates Derived from Different Modulations

In this research, a cut-off discarding 50% of the data was applied to calculate the RCM. However, this cut-off was arbitrarily determined and it is unreasonable to expect that discarding 50% of the respiratory data obtained will be effective in all cases, and in some cases the ideal cut-off may be higher or lower. However, a dataset-specific cut-off can be determined using machine learning methods for binary classification [97]. In order to ensure the best possible cut-off algorithm is implemented, a number of techniques including logistic regression, Bayesian time-series analysis, and support vector machines will be tested and compared. After implementation of a classification algorithm, a fusion method will be implemented that is capable of deriving a single respiratory rate estimate based on all those respiratory modulations that were classified as being useful in the previous step. Fusion models using the mean of the respiratory rates
from various modulations has already been implemented [47]; unfortunately, this model is overly simplistic and lacks robustness. However, more robust models have been applied to similar problems with different signals. For example, support vector machines have been used to fuse signal quality indices of acceptable electrocardiograms [84] and Bayesian continuous-valued labels aggregators (BCLA) have been used to fuse continuous-valued medical labels [98]. Furthermore, advances in methods such as probabilistic graphical models suggest new avenues for conducting data fusion.

6.2.2 Integration with Vital-Sign Data Fusion Models

Respiratory rate has been shown to be the most important vital sign in the early prediction of patient deterioration in the hospital; however, it is also the least often recorded vital sign as described in [1]. One of the reasons for the lack of recordings for respiratory rate is that there is no reliable non-invasive way to measure respiratory rate [19]. The lack of a reliable monitor for respiratory rate has been felt even more acutely as technologies conducting data fusion on continuous vital signs for early warning detection have advanced [99]. This suggests completion of the high-quality respiratory-rate estimation algorithm proposed here, if integrated with these fusion systems, could have substantial implications for patient monitoring. However, prior to any integration with such systems, the results of the respiratory fusion methodology and its effect on the vital sign fusion system previously defined [99] will be tested on the recently-released MIMIC III dataset which includes an additional 15,000 patients in addition to the 25,000 in the original MIMIC dataset [100].

Additionally, the planning for a clinical trial at the Oxford Churchill Hospitals ICU, called the Digital Health in a Connected Hospital study (DHiCH), is currently underway. This study will seek to use a camera to conduct non-contact vital sign measurements and fuse them into an early warning alarm algorithm. Again, the potential impact of implementing the respiratory rate estimation algorithm being proposed is large. One of the vital signs that can be collected via camera is the video PPG (vPPG) [86,101]. Because of the similarities of the PPG to the vPPG, it should be possible, with minimal alteration, to modify our respiratory rate algorithm to analyse the vPPG and serve as one of the inputs in the non-contact monitoring system.
6.2.3 Integration with Applications in m-Health for Developed and Developing Regions

Oxford University is currently designing a study to collect PPG data in clinics in Kenya using a mobile phone system interfaced with a pulse oximeter. This study offers a suitable avenue to expand the reach of the respiratory rate algorithm that has been proposed. Furthermore, the results from this study can be used as the starting point for designing a large m-health platform for monitoring respiratory rate, as shown in Figure 6.1. This shows the rough design of a three-tiered platform for predicting respiratory rate with each larger tier incorporating more heavyweight respiratory rate estimation methods that increase positive predictive value.

![Figure 6.1](image)

**Figure 6.1:** Diagram outlining one possible design for a large scale m-health platform for integrating all elements of the proposed respiratory rate estimation algorithm in a way that is accessible to both populations in developed and developing locations.

To illustrate how a platform like this might work in an m-health setting, the first tier, the sensor, is the “first line of defence.” The sensor would likely be a pulse oximeter with some basic computational functionality. The sensor would use a very lightweight respiratory rate estimation algorithm incorporating perhaps only one of the three RQIs defined in this study. The sensor would also have the functionality to estimate the likelihood that its estimation is accurate. If this likelihood is low, the sensor would pass the algorithm to a mobile platform (such as a tablet or smartphone) that incorporates a more heavyweight estimation algorithm, such as the calculation of all three high performing RQIs and a basic fusion method such as that presented by Karlen *et al.* [47]. If the mobile platform determines that the likelihood of its estimate being accurate is low, it will pass the signal to a central server via a cloud-based system.
where the most intensive algorithms, such as those defined in Section 6.2.1, can be implemented and the results then passed back down to either the sensor or the mobile platform. A system such as this defines an ideal configuration for use in both developed countries, where many signals can be passed to the cloud due to high levels of connectivity, and developing countries, where a respiratory estimation can be made even without reliable connectivity.

6.2.4 Project Timeline

<table>
<thead>
<tr>
<th>2015-2016</th>
<th>2016-2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>TT</td>
</tr>
</tbody>
</table>

- **Respiratory Fusion Methodology**
  - Develop Optimal Binary Classifier
  - Develop Optimal Fusion Algorithm

- **Vital Sign Monitoring Implementation**
  - MIMIC III Database Processing
  - MIMIC III Fusion Implementation
  - DIHICH Fusion Implementation

- **m-Health Monitoring Implementation**
  - Develop Pulse Oximeter Software
  - Pulse Oximeter Field Testing
  - Develop Mobile Phone Software
  - Mobile Phone Field Testing
  - Develop Cloud-Based Software
  - Cloud-Based Software Field Testing

- **Other**
  - Confirmation of Status
  - Writing of Thesis

**Figure 6.2:** Proposed Gantt Chart of DPhil Project Timeline.
Bibliography


