

Robust Estimation of Respiratory Rate via ECG- and PPG-Derived Respiratory Quality Indices

Drew A. Birrenkott¹, Marco A.F. Pimentel¹, Peter J. Watkinson² and David A. Clifton¹

Abstract—Respiratory rate (RR) is one of the most informative indicators of a patient’s health status. However, automated, non-invasive measurements of RR are insufficiently robust for use in clinical practice. A number of methods have been described in the literature to estimate RR from both photoplethysmography (PPG) and electrocardiography (ECG) based on three physiological modulations of respiration: amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW). However, the quality of the respiratory information acquired is highly patient-dependent and often too noisy to be used. We address this by proposing respiratory quality indices (RQIs) that quantify the quality of the respiratory signal that can be extracted from each modulation from both PPG and ECG waveforms. Signal quality indices (SQIs) detect artefact in the ECG and PPG, which is relatively straight-forward. RQIs have a different role: they quantify if an individual patient’s physiology is modulating the sensor waveforms. We have designed four RQIs based on Fourier transform (RQI_{FFT}), autocorrelation (RQI_{AC}), autoregression (RQI_{AR}), and Hjorth complexity (RQI_{HC}). We validated the approach using PPG and ECG data in the CapnoBase and MIMIC II datasets. We conclude that the novel implementation of an RQI-based pre-processing step has the potential to improve substantially the performance of RR estimation algorithms.

I. INTRODUCTION

Deviations in normal respiratory rate (RR) are one of the most diagnostically useful indicators of a patient’s health status [1]. There is an urgent clinical need for algorithms that are capable of extracting RR from sensor waveforms that are already collected in the hospital setting. Two of these signals, the photoplethysmogram (PPG) and electrocardiogram (ECG), are ideal for such a purpose because they are collected for almost every patient and are modulated by respiration in multiple ways. RR can be estimated from both PPG [2] and ECG [3] using amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW). One challenge in using the AM, FM, and BW modulations is that the quality and prevalence of each modulation is highly patient-dependent. In particular, age, health status, level of physical activity, and hydration level have all been shown to have an effect on the quality of the respiratory signal that can be derived [4]. This variability is a major source of error for estimation algorithms which are not sufficiently robust for use in clinical practice. There is currently no way to determine the quality of the respiratory signal derived from each modulation.

¹Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, UK.

²Nuffield Department of Clinical Neurosciences, University of Oxford, UK.

Contact for Correspondence: drew.birrenkott@eng.ox.ac.uk

A solution to this problem is to derive a pre-processing algorithm that is capable of determining the presence of a respiratory signal and the strength of that signal among the noise components of the signal. This is a strategy that has been widely employed on an array of physiological signals, where the signal is assigned a signal quality index (SQI) that indicates the presence of artefact in the signal [5]. However, using SQIs to disregard segments of waveform that are known to be artefactual, and then using the non-artefactual data for RR estimation, still results in insufficiently robust RR estimates. In this work, we propose a novel pre-processing step, which we call respiratory quality indices (RQIs). These are based on the Fourier transform, autocorrelation, autoregression, and Hjorth complexity of signals. Each RQI has been designed specifically to estimate the quality of the respiratory modulation after it has been extracted from either the PPG or ECG. The notion of RQIs is distinct from SQIs because while the latter determine the presence of artefacts in the signal, the RQIs quantify the presence of the respiratory information in the AM, FM, or BW signals that have been extracted from the PPG or ECG.

II. METHODS

A. Data Sets

Data were obtained for this analysis from two publicly-available sources: CapnoBase [6] and MIMIC II [7]. Both data sets contained simultaneous PPG and ECG data, and either capnography or impedance plethysmography (IP) data, which are used to obtain a “gold standard” RR estimate.

1) *CapnoBase*: This data set was collected from 59 pediatric (median age: 9, range 1-17 years) and 35 adult (median age: 52, range: 26-76 years) patients undergoing elective surgery or routine anaesthesia [6]. We used the data set described in [2], which contains one high-quality, eight-minute segment with simultaneous PPG, ECG, and capnography waveform data for each of 42 patients (29 pediatric and 13 adult). Expertly annotated breaths from the capnography waveforms were used to define a reference RR estimate when using our method with the PPG and ECG.

2) *MIMIC II*: This data set contains patient records for over 25,000 patients in four ICU units: medical, surgical, cardiac, and cardiac surgery [7]. For this analysis, only records containing waveform data were used. This subset consisted of 1017 adult patients (median age: 66, range: 18-91 years). For consistency with CapnoBase, a single eight-minute segment containing PPG, ECG, and IP (or the subset of these when not all three signals were recorded) was extracted between the 60th and 68th minute after recording

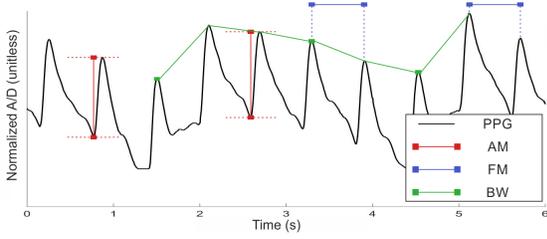


Fig. 1. Examples of the AM (red), FM (blue), and BW (green) respiratory modulations of the PPG.

began for each patient. The 60 minute extraction delay ensured that any artefact that arose at the start of the monitoring period was omitted. The IP was used to obtain a reference RR estimate; however, since the IP waveforms were not expertly annotated, in instances where the IP waveforms were noisy, the reference estimate was likely to be wrong. To overcome this, two benchmark RR estimation algorithms (ARSpec [8] and an FFT-based algorithm [2]) were used to assess the RR from the IP data. When the algorithms agreed within ≤ 2 breaths/min, the estimates were averaged to give the reference RR and when the algorithms differed by > 2 breaths/min, it was assumed that a reliable reference RR could not be obtained, and the segment was not analysed.

B. Respiratory Waveform Extraction

Three respiration waveforms representing the three respiration modulations, AM, FM, and BW, were extracted from the PPG (Figure 1) and ECG (Figure 2) using peak-trough detection in the time domain [9]. The peaks are represented as a series of pairs, $(t_{pk,i}, y_{pk,i})_{i \dots N_{pk}}$, as are the troughs, $(t_{tr,i}, y_{tr,i})_{i \dots N_{tr}}$. The respiratory amplitude modulation (AM) is the height of the signal from the peak to the corresponding trough, $y_{AM} = (t_{pk,i}, y_{pk,i} - y_{tr,i})_{i \dots N_{pk}}$, where t_i is taken to be the time of the peak pair. The frequency modulation (FM) is the change in time interval between peaks, $y_{FM} = (t_{pk,i}, t_{pk,i+1} - t_{pk,i})_{i \dots N_{pk}-1}$. The baseline wander (BW) is the envelope of the original signal, $y_{BW} = (t_{pk,i}, y_{pk,i})_{i \dots N_{pk}}$. These modulations were obtained for each 8-minute data sample by segmenting the sample into 15 non-overlapping windows of 32s length. Each modulation was extracted for every 32s window resulting in 3 respiratory waveforms (AM, FM, and BW) for each window of the original data. Each respiratory waveform was then filtered with a 5th-order Butterworth IIR bandpass filter between 0.83 and 1 Hz and subsequently downsampled to 4 Hz.

C. Respiratory Quality Indices

The RQIs were designed to determine if respiratory information was contained in the AM, FM, and BW modulations from the 32s segments of data. Because of the short time window, it was assumed that the respiratory signal would be stationary. From this assumption, four RQIs based on the Fourier transform, autocorrelation, autoregression, and Hjorth complexity were designed for each modulation signal.

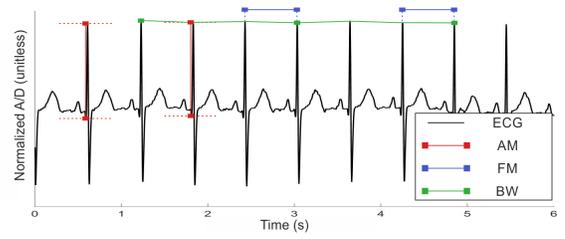


Fig. 2. Examples of the AM (red), FM (blue), and BW (green) respiratory modulations of the ECG.

1) *Fast Fourier Transform RQI*: The RQI_{FFT} was derived using the power spectrum of the FFT and was calculated as:

$$RQI_{FFT} = \frac{\sum_{i=1}^5 X_{PS}(m_i)}{\sum_{m=0.1}^{1.0} X_{PS}(m)}$$

where, $\sum_{m=0.1}^{1.0} X_{PS}(m)$ is the sum of the power spectrum within the physiologically relevant respiratory range from 0.1 to 1 Hz, and $\sum_{i=1}^5 X_{PS}(m_i)$ is the sum of the series of the five largest, continuous points within the respiration range where m_1 is the first point and m_5 the last point in that five point sequence. Thus, the RQI_{FFT} represents the maximum frequency component, which is assumed to be the respiratory frequency, divided by the sum of all the frequency components within the physiologically relevant respiratory range.

2) *Autocorrelation RQI*: The autocorrelation RQI (RQI_{AC}) was calculated by taking the autocorrelation of each window defined as:

$$r_k = \frac{\frac{1}{N-1} \sum_{n=1}^{N-k} (x(n) - \bar{x}) * (x(n+k) - \bar{x})}{c_0}$$

where r_k is the autocorrelation value, c_0 is the sample variance, N is the total length of the sample, \bar{x} represents the mean of the sample, and k represents the sample lag. The autocorrelation was taken for every sample lag within the possible respiratory range, $5 \leq k \leq 40$ (0.1 to 1 Hz), and the RQI_{AC} was defined as the maximum autocorrelation value within that range.

3) *Autoregression RQI*: The Autoregression RQI (RQI_{AR}) was found by taking the autoregression of each window using the Yule-Walker equation for model orders $m = 1 \dots 30$. The ideal model order was selected using Akaike's Information Criterion (AIC) and the RQI_{AR} was determined to be the autoregression pole with the largest magnitude within a frequency range from 0.083 to 1 Hz.

4) *Hjorth Complexity RQI*: The Hjorth Complexity RQI (RQI_{HC}) was found by calculating the third Hjorth Parameter, "complexity," which is a measure of the sinusoidality of a signal, with the Hjorth Complexity of the perfect sinusoid being 1 [10]. The RQI_{HC} was taken to be the value of the Hjorth Complexity, where better respiratory signals were expected to have values closest to 1.

D. Evaluation and Comparison

The RQIs were evaluated by using the existing ARSpec algorithm [8] to estimate RR for each of the three modu-

lations extracted from each window of PPG or ECG data. The absolute difference in the respiratory estimate between this and the reference RR estimate (from either capnography or IP, depending on the data set) was taken as a proxy for the quality of the respiratory information contained for a particular modulation in a given window. Each RQI was evaluated independently of this calculation and scaled such that the RQI values expected to indicate the best quality respiratory information (i.e., the RR estimation that most closely matched the reference RR value) were closest to 1 and the worst were closest to 0. The performance of each RQI was independently evaluated by sequentially calculating the mean absolute error (MAE) and standard error (SE) of the absolute difference of the estimate (from the existing ARSpec algorithm) and reference estimate (from capnography or IP) as the windows with the lowest RQI values were discarded. In practice, the MAE and SE were calculated for the entire data set (where every modulation and every window for each data set were analysed independently), and then the windows with the lowest RQI values were sequentially discarded, with the MAE and SE being recalculated after each discarding step. This process was repeated by discarding increments of 1% of the data, from lowest-RQI windows to highest.

In addition to evaluating the RQIs against each other, the performance of the RQIs was compared to an SQI defined in [11]. This represents the current state-of-the-art, in which RR estimation is performed on windows of data with high SQI values. We also considered the “ideal” performance scenario, where the MAE and SE are incrementally calculated when the absolute difference of the ARSpec estimate and reference estimate is used to identify and discard the poorest performing windows. This represents the best performance that could be achieved if the reference is available. In practice, the reference is unavailable-hence the need to estimate RR from the PPG or ECG-data.

III. RESULTS

Overall, a total of 1890 windows (100% of possible windows) were analysed for the CapnoBase PPG data, 1886 (99.8%) of CapnoBase ECG data, 32858 (71.8%) of MIMIC II PPG data, and 35784 (78.2%) of MIMIC II ECG data. Windows were discarded prior to analysis for three reasons: (1) if, for the given window, either the reference waveform (capnography or IP) or the PPG/ECG was not available or did not contain usable data (i.e., a flat line signal); (2) if a reliable reference RR could not be obtained (especially relevant in the noisy MIMIC II data set); or (3) if the peak and trough detectors detected ≤ 2 peaks/troughs, indicating an unusable set of respiratory modulations for a window.

The performance of the RQIs can be best understood by observing the change in the MAE as increasingly higher quality data is retained while lower quality data is discarded. Table I gives the difference of the MAE for 100% and 50% data retention ($MAE_{100} - MAE_{50}$) and the standard error of the difference of the MAE in breaths/min (br/min).

TABLE I
DIFFERENCE BETWEEN MAE_{100} AND MAE_{50}

	CapnoBase		MIMIC II	
	PPG	ECG	PPG	ECG
RQI_{FFT}	4.17 ± 0.10	3.83 ± 0.12	3.49 ± 0.03	3.79 ± 0.03
RQI_{AC}	4.33 ± 0.10	4.95 ± 0.11	3.09 ± 0.03	3.12 ± 0.03
RQI_{AR}	3.49 ± 0.11	3.33 ± 0.12	2.36 ± 0.03	2.23 ± 0.03
RQI_{HC}	3.02 ± 0.11	3.00 ± 0.12	3.32 ± 0.03	3.44 ± 0.03
SQI	0.47 ± 0.12	0.76 ± 0.13	0.93 ± 0.03	0.49 ± 0.03
Control	5.61 ± 0.07	7.14 ± 0.08	8.41 ± 0.02	9.20 ± 0.02

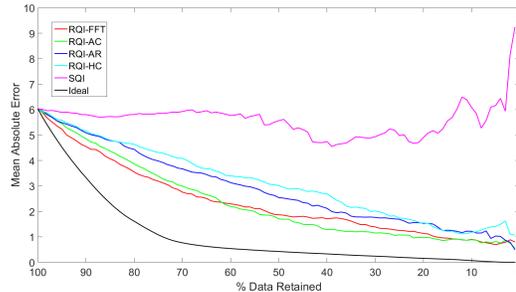


Fig. 3. RQI Performance on CapnoBase PPG data. Data were sequentially discarded in 1% increments based on the quality of the respiratory data as suggested by each individual RQI. The MAE was recalculated after each discarding step. Each RQI, RQI_{FFT} (red), RQI_{AC} (green), RQI_{AR} (blue), and RQI_{HC} (cyan), as well the SQI (pink) and the ideal (black), are represented.

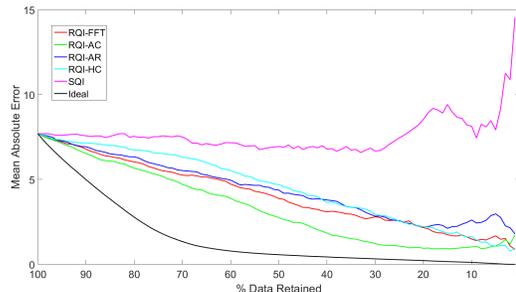


Fig. 4. RQI Performance on CapnoBase ECG data. Legend as Figure 3.

The largest difference between MAE_{100} and MAE_{50} for CapnoBase was seen when using RQI_{AC} for both PPG (4.33 ± 0.10 br/min) and ECG (4.95 ± 0.11 br/min). All RQIs substantially improved on the performance of the SQI for PPG (0.47 ± 0.12 br/min) and ECG (0.76 ± 0.13 br/min). Furthermore, the best performing RQI for MIMIC II was RQI_{FFT} for both PPG (3.49 ± 0.03 br/min) and ECG (3.79 ± 0.03 br/min), and all RQIs outperformed the SQI for PPG (0.93 ± 0.03 br/min) and ECG (0.49 ± 0.03 br/min).

The performance of each RQI at each 1% data discard increment is shown in Figure 3 (CapnoBase PPG), Figure 4 (CapnoBase ECG), Figure 5 (MIMIC II PPG), and Figure 6 (MIMIC II ECG). These plots show a steady trend in decreasing MAE for each RQI as the data with the poorest performing RQI values are discarded. Furthermore, these plots also indicate that each RQI outperforms the SQI regardless of how much data is discarded.

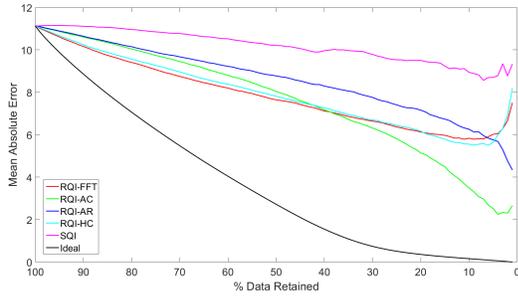


Fig. 5. RQI Performance on MIMIC II PPG data. Legend as Figure 3.

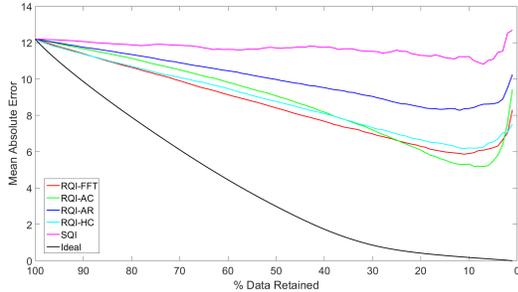


Fig. 6. RQI Performance on MIMIC II ECG data. Legend as Figure 3.

IV. CONCLUSION

One of the biggest challenges in estimating RR from PPG and ECG is ensuring the quality of the respiratory signal from which the estimation will be made. In this work, the goal was to address this challenge by proposing a novel pre-processing step to each respiratory modulation (AM, FM, and BW), termed a respiratory quality index. Four RQIs were designed and validated based on the assumption that for small timescales, the respiratory waveform is periodic.

The results from this analysis highlight a few key findings. The first is the general usefulness of using RQIs. While the RQI_{AC} and RQI_{FFT} outperformed the other two RQIs, every RQI improved substantially on the ability of the SQI to distinguish between high-quality and low-quality respiratory signals. This suggests that the implementation of the RQI pre-processing step promises to improve the accuracy of RR estimation algorithms. Furthermore, improvements in the MAE for even modest amounts of data discarded based on RQI performance were seen. This indicates the robustness gained by using RQIs, as they are not only capable of distinguishing high-quality data from low-quality data as a binary estimate, but provide a continuous scale as to the quality of the respiratory data in each particular window.

Furthermore, the robustness of the RQIs extended to a number of varied settings. First and foremost, each RQI was applied to respiratory modulations obtained from both PPG and ECG data and each RQI performed similarly well on both. This suggests that the RQIs might be generally applicable to any small-timescale respiratory waveform (e.g., when estimating RR from video). The RQIs also demonstrated good performance as the data quality decreased.

This can be seen in the distinction of the CapnoBase and MIMIC II data sets. The CapnoBase data set was obtained under idealised circumstances for otherwise healthy patients undergoing routine anesthesia or elective surgery; however, the MIMIC II data set was collected on a much larger patient cohort in four different ICUs. As a result, quality of the MIMIC II data is much lower than that of the CapnoBase as evidenced by the much larger amount of data discarded prior to analysis and the higher MAE at 100% data retention. While RQI performance decreased slightly when applied to the MIMIC II data set compared to the CapnoBase data set, it exhibited the same general trends and was still able to effectively discard the poorest quality data.

The overall performance of each RQI vastly improved on the performance of the SQI; however, further improvement in the RQIs could be made by fusing each RQI into a single metric. The fused RQI is expected to outperform each individual RQI and obtain results more closely resembling the performance of the “ideal” metric.

ACKNOWLEDGMENT

DAB acknowledges support from the Rhodes Trust. MAFP acknowledges support from the Wellcome Trust under grant number WT 088877/Z/09/Z. DAC acknowledges support from the Royal Academy of Engineering, the EPSRC “Challenge Award,” and Balliol College, Oxford.

REFERENCES

- [1] M.A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris. Respiratory rate: the neglected vital sign. *Medical Journal of Australia*, 188(11):657–659, 2008.
- [2] W. Karlen, S. Raman, J.M. Ansermino, and G.A. Dumont. Multi-parameter respiratory rate estimation from the photoplethysmogram. *Biomedical Engineering, IEEE Transactions on*, 60(7):1946–1953, 2013.
- [3] L.S. Correa, E. Laciari, A. Torres, and R. Jane. Performance evaluation of three methods for respiratory signal estimation from the electrocardiogram. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 4760–4763.
- [4] F. Yasuma and J. Hayano. Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest*, 125(2):683–690, 2004.
- [5] S. Nizami, J.R. Green, and C. McGregor. Implementation of artifact detection in critical care: a methodological review. *Biomedical Engineering, IEEE Reviews in*, 6:127–142, 2013.
- [6] W. Karlen, M. Turner, E. Cooke, G.A. Dumont, and J.M. Ansermino. Capnobase: Signal database and tools to collect, share and annotate respiratory signals. In *Annual Meeting of the Society for Technology in Anesthesia (STA), West Palm Beach*, page 25.
- [7] M. Saeed, M. Villarroel, A.T. Reisner, G. Clifford, L.W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R.G. Mark. Multi-parameter intelligent monitoring in intensive care ii: a public-access intensive care unit database. *Crit Care Med*, 39(5):952–960, 2011.
- [8] S.A. Shah, S. Fleming, M. Thompson, and L. Tarassenko. Respiratory rate estimation during triage of children in hospitals. *Journal of Medical Engineering & Technology*, 39(8):514–524, 2015.
- [9] D.J. Meredith, D. Clifton, P. Charlton, J. Brooks, C.W. Pugh, and L. Tarassenko. Photoplethysmographic derivation of respiratory rate: a review of relevant physiology. *Journal of medical engineering & technology*, 36(1):1–7, 2012.
- [10] B. Hjorth. Eeg analysis based on time domain properties. *Electroencephalography Clin Neurophysiol*, 29(3):306–310, 1970.
- [11] M.A.F. Pimentel, M.D. Santos, D.B. Springer, and G.D. Clifford. Heart beat detection in multimodal physiological data using a hidden semi-markov model and signal quality indices. *Physiol Meas*, 36(8):1717–1727, 2015.