

Bayesian Fusion of Algorithms for the Robust Estimation of Respiratory Rate from the Photoplethysmogram

Tingting Zhu¹, Marco A. F. Pimentel¹, Gari D. Clifford², and David A. Clifton¹

Abstract—Respiratory rate (RR) is a key vital sign that is monitored to assess the health of patients. With the increase of the availability of wearable devices, it is important that RR is extracted in a robust and noninvasive manner from the photoplethysmogram (PPG) acquired from pulse oximeters and similar devices. However, existing methods of noninvasive RR estimation suffer from a lack of robustness, resulting in the fact that they are not used in clinical practice.

We propose a Bayesian approach to fusing the outputs of many RR estimation algorithms to improve the overall robustness of the resulting estimates. Our method estimates the accuracy of each algorithm and jointly infers the fused RR estimate in an unsupervised manner, with aim of producing a fused estimate that is more accurate than any of the algorithms taken individually. This approach is novel in the literature, where the latter has so far concentrated on attempting to produce single algorithms for RR estimation, without resulting in systems that have penetrated into clinical practice. A publicly-available dataset, Capnobase, was used to validate the performance of our proposed model. Our proposed methodology was compared to the best-performing individual algorithm from the literature, as well as to the results of using common fusing methodologies such as averaging, median, and maximum likelihood (ML).

Our proposed methodology resulted in a mean-absolute-error (MAE) of 1.98 breaths per minute (bpm), outperformed other fusing strategies (mean fusion: 2.95 bpm; median fusion: 2.33 bpm; ML: 2.30 bpm). It also outperformed the best single algorithm (2.39 bpm) and the benchmark algorithm proposed for use with Capnobase (2.22 bpm).

We conclude that the proposed fusion methodology can be used to combine RR estimates from multiple sources derived from the PPG, to infer a reliable and robust estimation of the respiratory rate in an unsupervised manner.

I. INTRODUCTION

Respiration rate (RR) is an important physiological variable measured in a wide range of clinical scenarios as it provides valuable diagnostic and prognostic information [1]. Measurement of RR is commonly performed by manually counting the chest movements of the individual over a period of approximately 30 - 60 seconds [2]. This practice is time-consuming and prone to large variation in RR estimates. Typically, the number of breaths occurring within 15 seconds is counted, and then multiplied by four to give an estimate in bpm, which is therefore quantised into units of 4 bpm. This approach is particularly problematic with younger children, who have fast breathing rates [3] and where the potential for error is substantial. A potential solution is to estimate RR from a physiological signal that is both routinely acquired

by mobile sensors (such as the electrocardiogram or the PPG from pulse oximeters), and which is modulated by respiration. Although poorly understood, several different physiological mechanisms cause modulation of the PPG [4], [5], which result in amplitude modulation (AM), baseline wander (BW), and modulation of beat-to-beat intervals (frequency modulation, FM). Recently proposed methods for extracting RR from segments of PPG signals rely on estimating RR from each modulation (the derived respiratory signals), and fusing the three estimates into a final RR value [4] [7]. However, the common “fusion” approaches of the RR estimates derived from the different modulations are typically very straightforward (resulting in poor overall fusion results), and which tend to reduce substantially the number of windows for which an estimate of RR is computed.

A. Related Works

The most commonly-used methods for fusing different respiratory estimates extracted from physiological signals are mean or median voting. However, they generally perform well only if there are a large number of respiratory signals available. Recently, the “Smart Fusion” approach was proposed by Karlen *et al.* [4], wherein they fused three respiratory modulation signals of RR estimates. These signals were estimated from the frequency content of respiratory-induced variation using an FFT-based approach, and which were then combined by taking a mean voting approach. Windows for which the three resulting RR estimates (one from each source: AM, BW, FM) had an associated standard deviation that exceeded 4 bpm were discarded, resulting a 41.5% data loss. An effective probabilistic approach to aggregate expert estimates was first proposed by Dawid and Skene [8] using the EM algorithm when the ground truth was not readily available. They applied their fusion method to classify the unknown state of health (i.e., “fit to undergo a general anaesthetic”) for 45 patients, given the decisions made by five independent anaesthetists. Raykar *et al.* [9] extended this approach to measure the diameter of a suspicious lesion on a medical image using a regression model, and optimised the parameters using an ML-based approach. Their assumption was that the discrepancies of the estimates of lesion diameter from different experts were noisy versions (with Gaussian distribution) of the true diameter. Here, we propose a probabilistic fusion method to improve RR estimates employing a Bayesian treatment of the EM algorithm. It measures the bias (i.e., inverse accuracy) and precision (i.e., inverse variance of noise) of the outputs of each RR estimation algorithm, which is used to infer an estimation of the fused RR values.

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

²Departments of Biomedical Informatics and Biomedical Engineering, Emory University & Georgia Institute of Technology, USA.

Contact for Correspondence: tingting.zhu@eng.ox.ac.uk

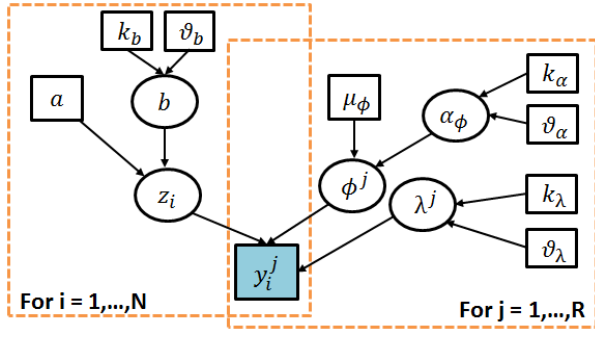


Fig. 1. Graphical representation of the BCLA model: y_i^j corresponds to a RR estimate provided by the j^{th} algorithm for the i^{th} window, with latent value z_i (the unknown true RR), the ϕ^j (bias of y), and the λ^j (precision of y). Furthermore, z_i is drawn from a Gaussian distribution with mean a and variance $1/b$, where a can be a function of feature vector \mathbf{x}_i . Let ϕ^j have a Gaussian distribution with mean μ_ϕ and variance $1/\alpha_\phi$. Let b , λ^j , and α_ϕ be drawn from Gamma distributions (denoted as Γ) with hyperparameters (k_b, ϑ_b) , $(k_\lambda, \vartheta_\lambda)$, and $(k_\alpha, \vartheta_\alpha)$, respectively.

II. THEORY

A. Bayesian Continuous-Valued Label Aggregator (BCLA)

Suppose that there are N windows of respiratory data estimated by R algorithms. Let $\mathbf{D} = [\mathbf{x}_i^T, y_i^1, \dots, y_i^R]$, where $i = 1 \dots N, j = 1 \dots R$. Here, \mathbf{x}_i is a column vector for the i^{th} window, and which contains d features as its elements. Let y_i^j correspond to the RR estimate provided by the j^{th} algorithm for the i^{th} window, and z_i represent the unknown true RR value for the i^{th} window. A graphical representation of the BCLA is shown in Fig. 1, where it may be seen that the \mathbf{x}_i are features that are used to influence the latent variable estimating the ground truth - typically, these are taken to be signal quality indices (SQIs) that indicate periods of noise in the time-series data presented to each RR-estimation algorithm [10].

B. EM with Maximum a Posteriori (MAP)

With the assumption that respiratory value are estimated independently for each window (as is the case), the likelihood of the parameters $\theta = \{\mathbf{w}, \lambda, \phi, \alpha_\phi, b, z_i\}$ for a given dataset \mathbf{D} can be formulated as:

$$P[\mathbf{D} | \theta] = \prod_{i=1}^N P[y_i^1, \dots, y_i^R | \mathbf{x}_i, \theta]. \quad (1)$$

It is assumed that y_i^1, \dots, y_i^R are conditionally independent given the feature \mathbf{x}_i (i.e., each algorithm is independent). The likelihood of parameters θ for a given input set \mathbf{D} can be written using Bayes' theorem as follows (defined in Fig. 1):

$$\begin{aligned} P[\theta | \mathbf{D}] &\propto P[\mathbf{D} | \theta] \cdot P[\theta] \\ &= \Gamma(\alpha_\phi | k_\alpha, \vartheta_\alpha) \Gamma(b | k_b, \vartheta_b) \times \\ &\quad \left[\prod_{j=1}^R \mathcal{N}(\phi^j | \mu_\phi, 1/\alpha_\phi) \Gamma(\lambda^j | k_\lambda, \vartheta_\lambda) \right] \times \\ &\quad \left[\prod_{i=1}^N \mathcal{N}(z_i | a, 1/b) \prod_{j=1}^R \mathcal{N}(y_i^j | z_i + \phi^j, 1/\lambda^j) \right]. \quad (2) \end{aligned}$$

where \mathcal{N} denotes a Normal distribution, $\mathcal{N}(c | \mu, 1/\alpha)$, and where Γ denotes a Gamma distribution $\Gamma(t | k, \vartheta)$. Parameter values θ may be estimated using a MAP approach, which maximises the log-likelihood of the parameters, i.e., $\arg\max_{\theta} \{\log P[\theta | \mathbf{D}]\}$. This can be performed by equating the gradient of the log-likelihood of each element in θ to zero to give:

$$\frac{1}{\lambda^j} = \frac{1}{N + 2(k_\lambda - 1)} \left[\sum_{i=1}^N (y_i^j - \phi^j - z_i)^2 + \frac{2}{\vartheta_\lambda} \right] \quad (3)$$

$$\mathbf{w} = \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^N \mathbf{x}_i z_i \quad (4)$$

$$\phi^j = \frac{1}{N + \frac{\alpha_\phi}{\lambda^j}} \left[\sum_{i=1}^N (y_i^j - z_i) + \mu_\phi \left(\frac{\alpha_\phi}{\lambda^j} \right) \right] \quad (5)$$

$$\frac{1}{\alpha_\phi} = \frac{1}{R + 2(k_\alpha - 1)} \left[\sum_{j=1}^R (\phi^j - \mu_\phi)^2 + \frac{2}{\vartheta_\alpha} \right] \quad (6)$$

$$z_i = \frac{\sum_{j=1}^R [(y_i^j - \phi^j) \lambda^j] + (\mathbf{x}_i^T \mathbf{w}) b}{\sum_{j=1}^R \lambda^j + b} \quad (7)$$

$$\frac{1}{b} = \frac{1}{N + 2(k_b - 1)} \left[\sum_{i=1}^N (z_i - \mathbf{x}_i^T \mathbf{w})^2 + \frac{2}{\vartheta_b} \right] \quad (8)$$

This can be solved using the EM algorithm in a two-step iterative process: (i) the E-step estimates the expected *true* RR estimates for all windows, $\hat{\mathbf{z}}$, as a weighted sum of the provided estimates, and can be calculated using equation (7); and then (ii) the M-step is based on the current estimation of $\hat{\mathbf{z}}$ given \mathbf{D} . The model parameters can be updated using the above equations in a sequential order until convergence.

III. METHOD

A. Data Description

The Capnobase dataset [11] was collected during elective surgery and routine anaesthesia. It consists of PPG recordings and capnometry data ($F_s = 300$ Hz), from 59 children (median age: 9, range: 1-17 years) and 35 adults (median age: 52, range: 26-76 years). We used the set as described in [4], which has 42 recordings of 8 minutes duration (336 minutes in total) containing reliable recordings of breathing. The capnometric waveform was used as the reference for RR estimates [4]. Only open-source data were used, and so ethical approval is irrelevant for our numerical study.

B. Respiratory Estimates and Fusion

RR was computed for 32-second windows, with successive windows having 29 seconds overlap. To extract the three respiratory-induced modulations (AM, BW, FM), PPG beat detection was performed using a segmentation algorithm [12]. The latter produces a series of maximum and minimum intensities for each pulse. As shown in Figure 2, the series of maximum intensities of the PPG pulses was used for extracting the BW timeseries. The (max-min) amplitude was

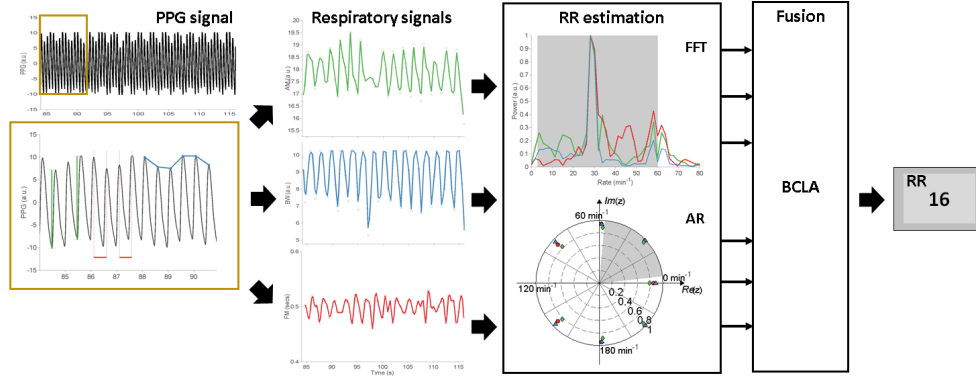


Fig. 2. Example of a 32-sec PPG window used for RR estimation. AM (green), BW (blue), and FM (red) respiratory modulations are extracted; for the FFT-based method, the power spectrum is calculated for each modulation using FFT, and the maximum power is selected within the physiologically-plausible RR range (grey area); for the AR-based method, the poles for each modulation are determined using an AR model, and the dominant pole within the plausible range of RR (grey area) is selected. Our proposed BCLA is then used to combine estimates, and provide a final “fused” value for that window.

used to derive the AM timeseries. The intervals between successive beats was used to extract the FM timeseries.

textbfRR was estimated using two different spectral approaches that have been used in the recent literature: Fourier analysis and autoregressive (AR) modelling. Spectral analysis requires evenly-sampled data, and so each timeseries (corresponding to BW, AM and FM) was first re-sampled at 4 Hz using linear interpolation. For the first method [4], the frequency spectra of the resulting respiratory signals were calculated. The frequency at which the maximum intensity of each spectrum is obtained within the frequency range of interest (corresponding to 3 to 60 bpm), was taken as the respiratory frequency (Figure 2). For the latter [6], [7], an AR model (of order 7) was fitted to each timeseries. The respiratory frequency was identified as that corresponding to the pole with the greatest magnitude within the plausible range of frequencies. We note that for each window, for each algorithm, three RR estimates were determined.

C. Evaluation and Comparison

The mean of the mean absolute error (MAE) and the standard error of the inferred RR estimates (from the PPG) across all subjects using the proposed fusion method, BCLA, were compared to the reference RR values (from capnography). BCLA was applied to the RR estimates extracted using the FFT-based, AR-based, and all (FFT+AR) algorithms. Additionally, the performance of the BCLA was compared to that of “Smart Fusion” [4], and with the best-performing (lowest-MAE) single algorithm (denoted “Best”). These were compared to other commonly used voting strategies, such as mean, median, and ML proposed in [9], using (i) RR estimates for all windows; (ii) only RR estimates for those windows following the criteria considered by Smart Fusion (denoted *); i.e., those windows where RR estimates with one standard deviation exceeding 4 bpm were discarded.

IV. RESULTS

The resulting values of the parameters and hyperparameters of the BCLA model for the Capnabase dataset are

TABLE I
BCLA PARAMETER VALUES FOR MODELLING CAPNOBASE

Symbol	Definition	Value
k_b	shape of Gamma distribution for b	3
ϑ_b	scale of Gamma distribution for b	0.006
μ_ϕ	mean of the bias distribution	variable †
k_α	shape of Gamma distribution for α_ϕ	5
ϑ_α	scale of Gamma distribution for α_ϕ	0.1
k_λ	shape of Gamma distribution for λ	3 ‡
ϑ_λ	scale of Gamma distribution for λ	0.02 ‡

N.B.: b is the precision for the estimate of the ground truth. α_ϕ is the precision for the estimate of the bias from ground truth. λ refers to signal-specific precisions. The values with ‡ are determined with the assumption that the RR estimates provided by the best modulation signal is ± 2 bpm away from the reference. The values with † are estimated from the median RR estimates provided by the algorithms.

described in Table I. The mean of the biases varied for individual subjects, and it was calculated using the median of the RR estimates across all windows.

The mean and standard error of the MAE for each fusion strategy are shown in Table II. The associated MAE results with * indicate that some windows were discarded when fusing estimates if they had a standard deviation greater than 4 bpm. An example of the percentage of windows used for each subject is shown in the lower plot of Fig. 3.

RR estimates fused via BCLA had lower MAE than those of the Smart Fusion method (i.e., Mean* in Table II) when using either the FFT-based, or the AR-based, or both combined (FFT+AR) extraction methods. In the case of using the FFT-based algorithm in [4], the BCLA achieved a MAE of 1.97 ± 0.40 bpm and used all available windows. This was smaller than the MAE of the Smart Fusion (2.02 ± 0.39 bpm), even when it disregarded data (and therefore did not produce estimates) from 41.5% of windows.

The difference in MAE for each subject (i.e., the MAE of the Smart Fusion subtracted the MAE of the BCLA) is shown in the upper plot of Fig. 3. A positive difference in MAE indicates an improvement of the BCLA in bpm whereas a negative difference in MAE implies the opposite.

TABLE II

COMPARISON OF MAE (AND STANDARD ERROR) ACROSS 42 SUBJECTS

Fusion	AR	FFT	FFT+AR
Mean	3.58±0.41	2.92±0.40	2.95±0.39
Best	3.31±0.41	2.39±0.43	2.39±0.43
Median	3.25±0.41	2.22±0.37	2.33±0.38
ML	3.12±0.42	2.03±0.38	2.30±0.42
BCLA	3.03±0.41	1.97±0.40	1.98±0.37
Mean*	3.26±0.41	2.02±0.39	2.22±0.41
Best*	3.12±0.41	1.93±0.39	1.56±0.35
Median*	3.17±0.41	1.97±0.39	1.98±0.43
ML*	3.06±0.42	1.84±0.39	2.11±0.43
BCLA*	2.99±0.41	1.86±0.39	1.81±0.39

N.B.: * indicates use of windows where the standard deviation of estimates was not above 4 bpm.

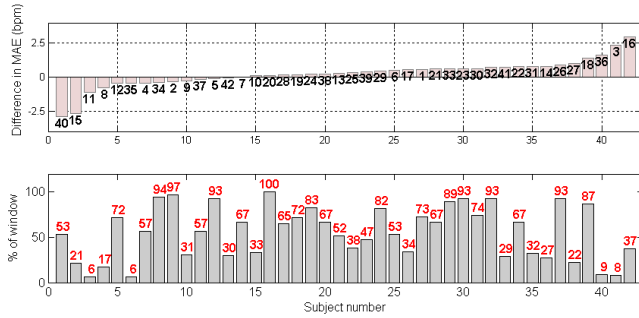


Fig. 3. Differences in MAE between Smart Fusion and the BCLA for all subjects are shown in the upper plot. A positive difference indicates an improvement of MAE using BCLA when fusing outputs from FFT+AR algorithms. The numeric labels corresponds to the subject IDs. The lower plot shows the percentage of windows considered when using the Smart Fusion for individual subjects. Note that BCLA uses all available windows.

MAE improved with BCLA (i.e., positive difference) for 66.7% of subjects; when windows were excluded if the standard deviation of RR estimates was greater than four, MAE improved for 81.0% of subjects.

Although the results of ML* and Best* outperformed those of BCLA*, the comparison was only based on 58.5% of the windows. When considering all windows, and thereby reporting RR estimates for all data, the BCLA was superior over other fusion strategies, demonstrating its ability to perform reliable fusion in the face of substantial noise. MAE values for AR-based algorithms were much worse than those of FFT-based algorithms in our study, and yet the BCLA was able to distinguish the quality of each algorithm's output based on their noise variance as well as their accuracy. This was true even in cases where there was a mixture of "good" and "bad" algorithms (i.e., FFT+AR). Modulation signals with high accuracy and low noise variance were favour by the BCLA, which demonstrates the ability of the BCLA to produce high-accuracy fused estimates (in an unsupervised manner) from a set of candidates, some of which may be individually poor estimates.

V. CONCLUSIONS

The current model assumes the same distribution of noise variance for each an algorithm's estimates for each patient.

Future work will include use of SQIs for each sensor time-series; this will allow the model to switch behaviour between subjects, offering the possibility of further reducing the error in RR estimation. Furthermore, other features, such as age, can be incorporated in the model to adjust the mean of the estimated latent ground truth to fit a subject population better, noting that different age groups exhibit different respiratory rates.

We have demonstrated that the proposed fusion methodology can be used to combine RR estimates from multiple sources derived from the PPG, to infer a reliable and robust estimation of the respiratory rate in an unsupervised manner. No training data need to be held out for optimising the parameters of the BCLA, and no prior knowledge of the performance of each algorithm was given. The average time for fusing 900 estimates from six algorithms using BCLA was about 0.64 seconds, which allows our proposed Bayesian method to be used for real-time application to producing RR estimates that are more robust than any of the algorithms considered independently.

ACKNOWLEDGMENT

TZ and MAFP acknowledge the support of the RCUK Digital Economy Programme grant number EP/G036861/1 and an ARM Scholarship through Kellogg College. DAC is supported by the Royal Academy of Engineering and Balliol College, Oxford.

REFERENCES

- [1] L. Tarassenko and D.A. Clifton, Semiconductor Wireless Technology for Chronic Disease Management, Electronics Letters, s30, 2011.
- [2] World Health Organization: Fourth Programme Report, 1988-1989: ARI Programme for Control of Acute Respiratory Infections. Technical report, WHO, Geneva, 1990.
- [3] World Health Organization: IMCI technical seminar on Acute Respiratory Infection, 2001.
- [4] W. Karlen, S. Raman, J.M. Ansermino, and G.A. Dumont, Multi-parameter respiratory rate estimation from the photoplethysmogram, IEEE Transactions on Biomedical Engineering, vol. 60, no. 7, pp. 1946–53, 2013.
- [5] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D.A. Clifton, and C. Pugh, Non-contact video-based vital sign monitoring using ambient light and auto-regressive models, Physiological Measurement, vol. 35, no. 5, pp. 807–831, 2014.
- [6] C. Orphanidou, D.A. Clifton, S. Khan, M. Smith, J. Feldmar, and L. Tarassenko, Telemetry-Based Vital Sign Monitoring for Ambulatory Hospital Patients, IEEE EMBC, pp. 4650–4653, 2009.
- [7] C. Orphanidou, S. Fleming, S.A. Shah, L. Tarassenko, Data fusion for estimating respiratory rate from a single-lead ECG, Biomedical Signal Processing and Control, vol. 8, no. 1, pp. 98–105, 2013.
- [8] A.P. Dawid, A.M. Skene, Maximum likelihood estimation of observer error-rates using the EM algorithm, Journal of the Royal Statistical Society Series C Applied Statistics, vol. 28, no. 1, pp. 20–28, 1979.
- [9] V.C. Raykar, S. Yu, L.H. Zhao, G.H. Valadez, C. Florin, L. Bogoni, L. Moy, Learning from crowds, Journal of Machine Learning Research, pp. 1297–1322, 2010.
- [10] T. Zhu, N. Dunkley, J. Behar, D.A. Clifton, G.D. Clifford, Fusing Continuous-Valued Medical Labels using a Bayesian Model, 2015. arXiv:1503.06619
- [11] 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, Annual Meeting of the Society for Technology in Anesthesia, 2010.
- [12] B.N. Li, M.C. Dong, M.I. Vai, On an automatic delineator for arterial blood pressure waveforms, Biomedical Signal Processing and Control, vol. 5, no. 1, pp. 76–81, 2010.