Novel Continuous Respiratory Rate Monitoring Using an Armband Wearable Sensor

Nicholas Huang*,†, Menglian Zhou†, Dayi Bian, Pooja Mehta, Milan Shah, Kuldeep Singh Rajput and Nandakumar Selvaraj

Abstract—Photoplethysmography (PPG) and accelerometer (ACC) are commonly integrated into wearable devices for continuous unobtrusive pulse rate and activity monitoring of individuals during daily life. However, obtaining continuous and clinically accurate respiratory rate measurements using such wearable sensors remains a challenge. This article presents a novel algorithm for estimation of respiration rate (RR) using an upper-arm worn wearable device by deriving multiple respiratory surrogate signals from PPG and ACC sensing. This RR algorithm is retrospectively evaluated on a controlled respiratory clinical testing dataset from 38 subjects with simultaneously recorded wearable sensor data and a standard capnography monitor as an RR reference. The proposed RR method shows great performance and robustness in determining RR measurements over a wide range of 4–59 brpm with an overall bias of -1.3 brpm, mean absolute error (MAE) of 2.7±1.6 brpm, and a meager outage of 0.3±1.2%, while a standard PPG Smart Fusion method produces a bias of -3.6 brpm, an MAE of 5.5±3.1 brpm, and an outage of 0.7±2.5% for direct comparison. In addition, the proposed algorithm showed no significant differences (p=0.63) in accurately determining RR values in subjects with darker skin tones, while the RR performance of the PPG Smart Fusion method is significantly (P<0.001) affected by the darker skin pigmentation. This study demonstrates a highly accurate RR algorithm for unobtrusive continuous RR monitoring using an armband wearable device.

Index Terms—Wearable device, Photoplethysmography, Accelerometer, Respiration Rate Monitoring

I. INTRODUCTION

In recent years, development of wearable devices embedded with photoplethysmograph (PPG) and tri-axial accelerometer (ACC) has enabled unobtrusive remote monitoring of pulse rate and activity patterns as a cost-effective essential tool widely adapted into our day-to-day lives. While respiratory rate (RR) is a top predictor of serious illness such as cardiac arrest and sepsis [1], in contrast to the commonly measured pulse rate vital sign, RR measurement is still not available in most widely used wearable sensors. In clinical practice and out-of-hospital monitoring settings, RR has consistently been the least frequently measured vital sign [2]. Traditional stethoscope-based RR measurements are subjective and laborious, whereas nasal gas sampling lines or dual chest bands are obtrusive and uncomfortable. Additionally, while PPG or ACC sensing modes are unobtrusive and more commonly available, there also exists a technological gap in producing accurate RR using these signals. Even though there has been extensive literature on RR measurements using PPG or ACC, a practical wearable solution for continuous and accurate RR measurements still remains elusive.

PPG measures tissue blood volume change and is widely used for oxygen saturation and pulse rate monitoring. Meanwhile, the PPG waveform can be used for RR estimation since it is modulated by ventilation in three ways [3]: amplitude modulation (AM), frequency modulation (FM), and baseline wandering (BW). A number of signal processing methods including time-frequency analysis [4, 5], incremental-merge segmentation [3], and auto-regressive modeling [6] have been shown to extract respiratory modulations from PPG waveforms; methods such as deep learning have also been used to obtain RR from PPG directly [7]. However, a number of factors can affect the PPG-based RR estimate, including (i) in-band low frequency Mayer waves from bodily regulatory mechanisms, (ii) the Nyquist limit, where the respiratory surrogate signals are undersampled and aliased when the fundamental HR frequency is lower than two times that of the RR frequency, (iii) dark skin tones leading to very low PPG signal strength due to absorption of light by melanin pigmentation, and last but not the least, (iv) motion artifacts.

A representative example of RR values derived using a modified PPG smart fusion method is illustrated in Figure 1 alongside simultaneously measured HR values and half of the RR values for comparative purposes. Accurate tracking of RR values using this standard PPG Fusion RR algorithm is possible when the HR is at least two times greater than the RR ground truth. However, when the Nyquist sampling relationship between HR and RR is not met physiologically, accurate tracking of RR using a PPG (or ECG) signal is inherently impossible, leading to unacceptable error margins when the RR range is higher than half that of HR, as shown.

Tri-axial ACC measures acceleration in three orthogonal directions and is also widely integrated in wearable devices for fitness monitoring (such as step count) and motion analysis (such as gait, posture, balance). Meanwhile, ACC is also feasible for sensing respiration-induced body movement to estimate RR in real time. In recent years, adaptive line enhancer and spectral fusion [8], Kalman filter based fusion [9], principal component analysis (PCA) [10], and independent component analysis [11] methods have been developed for this purpose. Although ACC is not affected by Mayer waves, the Nyquist limit on low HR, or skin tone,
Fig. 1. The influence of the Nyquist relationship between heart rate (HR) and respiration rate (RR) frequencies is illustrated from a sample subject. The standard PPG Fusion algorithm (blue) is able to estimate RR (black) while the ground truth RR is below half of HR (green), but it produces clinically unacceptable levels of error margins when the ground truth RR exceeds half of HR, i.e., the Nyquist limit.

any periodic motions stronger than the respiration-induced motion could become interference and affect the ACC-based RR estimation. The most common ACC measurement site for estimating RR is the chest [9, 10] using a chest belt/strap or adhesive patch sensors [12], but they are not as comfortable or convenient as a wristband/armband for long-term use. The upper arm is a better measurement site for RR purpose, since not only is the chest wall movement better transmitted to the upper arm, but also interference motion is significantly less frequent in upper arm as compared to wrist site under daily living conditions.

This article first presents a novel method for continuous RR estimation using both PPG and ACC sensing from an upper armband wearable device. Secondly, the performance of the proposed RR algorithm is evaluated and validated by retrospective analysis on a diverse clinical dataset of 38 participants with gold standard capnograph reference RR measurements. Finally, the performance of a standard PPG fusion method implemented from the literature is also evaluated for comparative analysis.

II. MATERIALS AND METHODS

A. Data collection

This work retrospectively analyzes a collective data acquired from 38 healthy subjects, ages 18 to 70 years, by a clinical laboratory (Clinimark, Louisville, CO), involving the IRB-approved respiratory rate protocols. The respiratory study protocols, in one case, required the participants to be spontaneously breathing for about 15–25 minutes in sitting and supine body postures. In another case, the protocol involved a metronome breathing exercise session for a duration up to 40 min, where the subjects were guided to perform low (15, 10, 5 bpm) and high (15, 20, 25, 30, 35, 40, 45, 50 bpm) respiration rates. The participants followed cues as closely as possible with natural variations, such that some subjects were not able to execute the lower or upper corner values of the desired RR range.

In these respiratory study protocols, a reusable armband wearable sensor named Everion (Biovotion, Zurich) was worn on one or both of each subject’s upper arms to measure and record PPG and ACC signals at a 51.2 Hz sampling rate; simultaneously, either a mouthpiece attached with a sample line or a flow-by mask with large bore ventilation tubing was outfitted to each subject for measurement of EtCO2 by a FDA-cleared device (GE Healthcare Datex-Ohmeda S/5 monitor with compact airway module M-COVX). Prior to the test, the procedures were explained to the subjects, and subjects completed health assessment and consent forms. A pulse oximeter and an ECG (GE Healthcare S/5 Monitor with M-NESTPR module) were also applied to the subjects for safety monitoring.

B. RR Estimation Algorithm

The RR algorithm presented in this work involves simultaneous processing of PPG and ACC signals measured on an upper arm location using Everion and determining RR outputs in a parallel fashion. A PPG-based RR estimate is derived by extracting surrogate respiration waveforms from unique modulations in the PPG signal; concurrently, an accelerometer-based RR estimate is also derived by extracting the time-frequency spectrum from the ACC signal and performing peak-tracking. Finally, these two independent RR estimates are fused based on their quality. Due to the dynamic synthesis of multi-mode respiratory waveforms extracted from ACC and PPG signals, the proposed algorithm is denoted as A-P Synthesis.

1) PPG-based RR algorithm: Respiration rate is estimated from the PPG signal by first extracting surrogate respiratory signals such as amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW). Each surrogate signal is sampled at pulse-related peaks in the PPG waveform, which are located using a peak-finding method. The surrogate signals are then interpolated to achieve uniform sampling rate and filtered to isolate the respiratory component. Breath-related peaks in the respiratory signal are identified again using a peak-finding method, with an adaptive peak-to-trough threshold to remove erroneous peaks. An estimate of RR for each surrogate is then calculated by taking the reciprocal of the average of the most recent 30 inter-breath intervals with low motion activity. A quality metric is calculated for each estimate based on the irregularities of the corresponding surrogate respiratory signal. That metric is then used to combine the estimates through a weighted average, resulting in one final PPG-based RR estimate.

2) ACC-based RR algorithm: Under daily living conditions, a subject’s upper arm may have different angles to the chest depending on posture at different times. Hence, we firstly project the three orthogonal 30-sec band-pass filtered ACC waveforms to the first principal axes acquired from PCA analysis, which is considered to be the ventilation-induced movement direction and is updated every 5 sec. Note that for each segment (30-sec waveform), we check the
second-by-second motion activity (MA) level and select only sections of the waveform whose MA is below a threshold to exclude interference from other motion. If less than 15 sec has MA below the threshold, then that segment is skipped. Otherwise, from the projected principal waveform, the FFT spectrum and its kurtosis score are derived. The FFT spectrum of each segment forms the time-frequency spectrum as shown in Figure 2, bottom panel, from which we can see a clear RR trace which changes over time. For real-time RR estimation, spectrum peak tracking (SPT) is applied to the time-frequency spectrum since RR normally changes gradually and because each segment’s waveform largely overlaps with the previous one.

The SPT starts with a global search on the FFT spectrum, whose highest peak location is chosen as the first RR estimation. From there, each new spectrum is searched for a dominant peak around the previous RR estimate. If no dominant peak is found, then the previous RR is used for up to 3 continuous segments; beyond 3 segments, a new global search is conducted. During SPT, if any segment’s Kurtosis score is below a threshold or if the variance of the principal waveform exceeds a threshold, we suspect the subject’s arm is at moderate/high movement, hence ACC is not suitable for RR estimation. During these segments there are no outputs for the ACC-based RR algorithm as shown in Figure 6. Simultaneously, the quality of this RR estimate is derived from the variance of the principal waveform of this segment.

3) PPG and ACC Combination: Once both the PPG and ACC estimates of RR are computed, the one with the highest quality is selected as the final RR output. If neither of the methods produce an estimate with high enough quality, then outage is recorded.

C. Performance and Statistical Analyses

Validation of the PPG Smart Fusion and A-P Synthesis models was performed through post-processing analysis of the PPG and ACC raw data. In this retrospective post-processing analysis of multiple clinical datasets, the online performance was carried out uniformly using MATLAB computing platform [13] by providing data in 1-second intervals. Performance of each method was assessed by comparing to reference RR on a second-by-second basis. Error values were calculated as model prediction of RR minus reference; thus, positive values are overestimates, while negative values are underestimates. Recording positions with no reference values were ignored, and any unphysiological abrupt jumps in reference values were removed (total only 3.8% of the reference points).

Average mean absolute error (MAE) refers to the average error across subjects, while aggregate MAE refers to averaging across all samples. Bias was calculated as average signed error, and 95% limits of agreement (LoA) were computed as the average error ±1.96 * standard deviation. Outage was calculated as the amount of reference data with no
valid predicted RR value divided by the total amount of reference data. Statistical comparisons between demographic categories of subjects, including Fitzpatrick scale, gender, age, and BMI, were performed using two-sample t-tests over average MAE values of the study cohort, grouped according to the demographic variable of interest.

III. RESULTS

RR prediction from a representative example shown in Figure 1 is first examined to exhibit the strength of the A-P Synthesis method, whereas the traditional PPG Smart Fusion method produces clinically unacceptable error margins for higher RR ranges. A comparison of the time-frequency spectra of PPG and ACC signals of this same example is shown in Figure 2. The PPG spectrum is dominated by low frequency Mayer waves near 6 brpm (top panel). By high pass filtering above the Mayer wave frequency, this interference could be eliminated; however, it would also make low RR undetectable. In contrast, the accelerometer does not have the Nyquist frequency limitation, nor is it affected by Mayer waves as shown in Figure 2. As a result, the energy at RR is very prominent in the bottom time-frequency spectrum, which is derived from the ACC-fused signal.

The performance of our A-P Synthesis algorithm on the example recording is shown in Figure 3. In contrast with the standard literature algorithm, A-P Synthesis is able to accurately predict RR up to 50 brpm. Further, the proposed algorithm is less sensitive to low-frequency Mayer waves, as shown by little to no erroneous predictions near 6 brpm in this example. Further, the RR prediction during each paced breathing plateau is very stable and shows little noise after reaching the steady state true RR.

A summary of the performance of A-P Synthesis over all tested recordings from 38 participants (62 independent Everion recordings, and 84,002 samples) is shown in Table I. Our A-P Synthesis method produces a highly accurate RR predictions with an average MAE of 2.7 ± 1.6 brpm over the full range of 4-59 brpm, while the PPG Smart Fusion method produces an average MAE of 5.5 ± 3.1 brpm. From the spontaneous and metronome breathing tests conducted, the reference RR ranges from 4-59 brpm with a distribution of 19.1 ± 9.6 brpm. For a more prevalent range of 35 brpm or below, the average MAE decreases even further to 2.3 ± 1.4 brpm. The outage of the RR algorithm is only 0.3%, and the correlation between the prediction and the reference RR is much stronger than the PPG Smart Fusion Method, regardless of the RR evaluation range. Finally, the bias and the 95% LoA are much more accurate and narrow compared to the PPG Smart Fusion method as shown in Table I.

The MAE Performances of RR prediction using A-P Synthesis and PPG Smart Fusion methods are compared in Figure 4 between the recordings of darker (14.6%) and lighter skin tones. The PPG Smart Fusion method has significantly (p < 0.001) higher error for darker skin tones, while A-P Synthesis exhibits no significant differences (p = 0.63) in accurately predicting RR for darker or lighter skin tones.

A-P Synthesis algorithm’s MAE performances were further compared across other subject population demographics, and no significant differences (p > 0.05) among genders of female vs male, age groups of under 40 years vs over 40 years, and BMI of under 25 vs over 25 were found.

Bland-Altman plots comparing the performance of PPG Smart Fusion (top) with A-P Synthesis (bottom) are shown in Figure 5. Scatter plot markers are scaled by the square root of the number of samples within a bin for better visualization. Histograms of the mean of reference RR and model predicted RR are shown on the x-axes; histograms of error values are shown on the y-axes. Both histograms are normalized by their maximum values.

The distribution of error values in the PPG Smart Fusion model clearly shows that there is a negative bias leading to under prediction (-3.6 brpm average bias) of RR measurements with wide error margins. In contrast, the proposed A-P Synthesis
Synthesis algorithm is able to accurately estimate RR values for the entire wide range, as evidenced by a low bias of about one breath, absence of negative bias slope, very tight error margins, and relatively narrow limits of agreement.

IV. DISCUSSION

This paper presents a novel method, A-P Synthesis, for continuous RR estimation using an armband wearable device, embedded with optical PPG sensors and a tri-axial accelerometer. For the first time, PPG-based RR estimation and ACC-based RR estimation are smartly combined to achieve far more accurate and robust continuous RR estimations over a wide range up to 4-60 bpm. The proposed algorithm also achieves the highest coverage and reliability from having two complementing methodologies, and shows no statistically significant difference in performances across different age ranges, skin tones, or BMI groups. This method is also computationally efficient and has been successfully integrated and validated on the Everion armband device for long-term RR estimation in real time.

The A-P Synthesis algorithm showcased an accurate and consistent RR estimation for a wide range of breathing rates, particularly including the higher range (35-60 bpm) above half of the heart rate. Such capability and performance of the proposed algorithm in convenient and unobtrusive wearable sensors allow for better capture of abnormal respiration rates in chronic and acute respiratory conditions including anxiety, congestive heart failure, asthma, lung disease, pneumonia, use of narcotics, and drug overdose. For comparison, the literature reports that current devices are capable of measuring RR only over a very narrow range of values, typically limited to the clinical RR range of a healthy person, which is around 10-30 bpm [14]. As examples, the Kick LL smartwatch under development directly limits RR between 12 and 30 bpm [15], while research with the Empatica E4 device classifies fast breathing as above a mere 15 bpm [16].

The present clinical testing data covers a wide range of skin tones (from Fitzpatrick scale of I to VI), ages (from 18 to 69 yrs old), and BMI (from 16.7 to 30.2). The study demonstrates no statistically significant difference in the RR estimation errors of A-P Synthesis across different age/skin tone/BMI groups. The green LED wavelength commonly used to obtain the PPG signal in wearable devices undergoes greater absorption for subjects with darker skin tones, which can in turn substantially influence the signal quality and affect the RR estimation error tremendously. Hence, the standard PPG Fusion method shows greater difficulty in predicting RR for subjects with darker skin tone, as shown in Figure 4. In contrast, however, A-P Synthesis is able to more accurately estimate RR regardless of the skin tone and BMI.

Part of the novelty of the A-P Synthesis algorithm is the adaptive use of both PPG and ACC signals in producing stable, reliable and accurate RR estimation using a uncommon, but convenient site of sensor attachment. The two signals capture unique dynamics of the respiratory system and have different failure modes complementing each other in certain conditions such as low perfusion and movements. This synergy is exemplified in Figure 6. High motion activity periods tend to produce an unreliable ACC derived respiratory signal;
however, a strong estimate of RR is still produced using the PPG signal and the proposed PPG synthesis module during these periods since the green PPG signal at the upper arm has been shown to be relatively robust to motion [17, 18]. On the other hand, in cases where the PPG signal is unreliable or even wholly absent due to poor skin contact or other factors, a reliable RR estimate can still be derived from the ACC signal alone using the proposed ACC synthesis module.

One of the distinct advantages of the Everion armband that enables this combination of signals is the location of the device. Given that the device is worn on the upper arm, chest wall movement during inhalation and exhalation can be very well captured by the tri-axial accelerometer under stationary conditions. Meanwhile, the green PPG signal strength and motion artifact resistance at the upper arm are better than at other peripheral sites such as the wrist and finger [18]. In addition, the location offers a more comfortable fit than devices worn on or adhered to the chest for RR measurements using ECG, capacitive sensors [19], and impedance sensors [20].

V. CONCLUSION

A novel A-P Synthesis algorithm is presented for continuous RR estimation using an armband wearable device, and the retrospective analysis of a clinical recorded dataset demonstrates a low level of RR prediction error for a wide range of respiration rates up to 60 bpm. The combination of ACC and PPG sensing modalities helps to take advantages of the strengths of each signal. The accelerometer provides robust direct RR estimation over a much wider range of breathing rates during relatively stationary conditions, while the PPG signal maintains a good RR measurement under moderate activity covering the vast majority of continuous patient monitoring scenarios. This accurate and consistent method of RR estimation in a comfortable wearable device is invaluable to healthcare applications ranging from personal fitness to long-term monitoring of patient health.

REFERENCES