Respiratory activity extracted from wrist-worn reflective photoplethysmography in a sleep-disordered population

PAPER

Respiratory activity extracted from wrist-worn reflective photoplethysmography in a sleep-disordered population

To cite this article: Gabriele B Papini *et al* 2020 *Physiol. Meas.* **41** 065010

IOP Publishing Physiol. Meas. 41 (2020) 065010 https://doi.org/10.1088/1361-6579/ab9481

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

Respiratory activity extracted from wrist-worn reflective photoplethysmography in a sleep-disordered population

Gabriele B Papini[1,2,3], **Pedro Fonseca**[1,2], **Merel M van Gilst**[1,3], **Jan WM Bergmans**[1,2], **Rik Vullings**[1] **and Sebastiaan Overeem**[1,3]

[1] Department of Electrical Engineering, Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, The Netherlands
[2] Philips Research, High Tech Campus, 5656 AE Eindhoven, The Netherlands
[3] Sleep Medicine Center Kempenhaeghe, P.O. Box 61, 5590 AB Heeze, The Netherlands

E-mail: g.p.papini@tue.nl

Keywords: photoplethysmography, respiratory activity, sleep, wearable

RECEIVED
23 March 2020

REVISED
4 May 2020

ACCEPTED FOR PUBLICATION
19 May 2020

PUBLISHED
2 July 2020

Abstract

Objective: Respiratory activity is an essential parameter to monitor healthy and disordered sleep, and unobtrusive measurement methods have important clinical applications in diagnostics of sleep-related breathing disorders. We propose a respiratory activity surrogate extracted from wrist-worn reflective photoplethysmography validated on a heterogeneous dataset of 389 sleep recordings. *Approach*: The surrogate was extracted by interpolating the amplitude of the PPG pulses after evaluation of pulse morphological quality. Subsequent multistep post-processing was applied to remove parts of the surrogate with low quality and high motion levels. In addition to standard respiration rate performance, we evaluated the similarity between surrogate respiratory activity and reference inductance plethysmography of the thorax, using Spearman's correlations and spectral coherence, and assessed the influence of PPG signal quality, motion levels, sleep stages and obstructive sleep apnea. *Main results*: Prior to post-processing, the surrogate already had a strong similarity with the reference (correlation = 0.54, coherence = 0.81), and reached respiration rate estimation performance in line with the literature (estimation error = 0.76± 2.11 breaths/min). Detrimental effects of low PPG quality, high motion levels and sleep-dependent physiological phenomena were significantly mitigated by the proposed post-processing steps (correlation = 0.62, coherence = 0.88, estimation error = 0.53± 1.85 breaths/min). *Significance*: Wrist-worn PPG can be used to extract respiratory activity, thus allowing respiration monitoring in real-world sleep medicine applications using (consumer) wearable devices.

1. Introduction

The assessment of respiratory activity is an important aspect in monitoring healthy and disordered sleep. For instance, it allows to distinguish central from obstructive apneas, i.e. respiratory events with and without respiratory effort (Berry *et al* 2015, Vandenbussche *et al* 2015). The gold standard to measure changes in respiratory effort during overnight polysomnography (PSG) is esophageal manometry. However, the obtrusiveness and sensitivity to movement-generated noise of this technique, make thoracoabdominal respiratory inductance plethysmography (RIP) the main respiratory effort indicator in daily practice. The standard RIP set-up employs two belts that measure volume variations generated by respiratory movement at the thorax and abdominal level.

Several studies showed that quantifying respiratory activity can improve unobtrusive sleep monitoring approaches based on surrogate measures. Long *et al* reported that thoracic respiratory activity information improved sleep/wake classification based on actigraphy, while Fonseca *et al* showed that including this information improved sleep stage classification based on heart rate variability (HRV) (Fonseca *et al* 2015, Long *et al* 2013). Bianchi *et al* reported that thoracic respiratory activity on its own can be used to estimate

obstructive sleep apnea (OSA) severity (Bianchi *et al* 2014). Even when respiratory activity was extracted as a surrogate, e.g. using ECG-derived respiration, it has been successfully used for automatic sleep staging and obstructive sleep apnea monitoring (Li *et al* 2018, Varon *et al* 2015, Sadr de Chazal 2019).

With the advent of consumer wearable devices, such as smart-watches and fitness trackers, research on their possible clinical use has increased as well, with a promise of reducing obtrusiveness as well as cost of sleep monitoring (Penzel *et al* 2018, Bianchi 2018). Most of these wearables employ wrist-worn reflective green light photoplethysmography (PPG). PPG assesses cardiovascular activity by optically measuring blood volume variations in the microvascular bed. The morphological variations of the PPG pulses contain information about the changes in the peripheral blood flow generated by various physiological processes, including respiration. PPG signals recorded at the wrist have usually a lower quality than at other body locations, such as earlobe or fingertip, however, they can be recorded via user-friendly devices that nowadays are widely accepted by the population (Tamura *et al* 2014, De Zambotti *et al* 2016) Therefore, wrist-worn reflective PPG is an attractive candidate for respiratory activity estimation during sleep in terms of unobtrusiveness, acceptability, and richness of physiological information. However, the literature regarding PPG respiration analysis mainly focuses on respiration rate estimation using transmissive PPG as part of clinical SpO2 sensors. Moreover, these methods are mostly evaluated on small datasets (less than 100 participants, with recording duration lower than one hour) (Charlton *et al* 2016, Dehkordi *et al* 2018, Karlen *et al* 2010). Charlton *et al* described most of these methods in a comprehensive review of respiration rate estimation from transmissive PPG. These methods mainly consist of the following steps: (1) extracting respiration signal surrogates based on distances between specific landmarks of the PPG pulses, (2) using these surrogates to estimate the average respiration rate with an epoch-by-epoch time or frequency analysis, and (3) post-processing the estimation (Charlton *et al* 2016). Even though respiration rate is a valuable health indicator (Braun 1990), it is only a single aspect of respiration and only partially represents respiratory activity during sleep and, especially, sleep-disordered breathing events.

A few researchers investigated the potential of using PPG to estimate respiratory activity as respiratory *effort*. They employed a limited amount of participants during a protocolized respiratory test or used pre-selected segments of the PPG signal (Addison 2016, Addison 2017, Khandoker *et al* 2013). During sleep, the PPG signal is influenced by various physiological conditions and events (such as different sleep stages, as well as sleep disorders and associated events) due to their effect on the respiratory activity, autonomic nervous system activity and arterial stiffness (Allen 2007, Hickey *et al* 2015). Datasets collected in real-world (clinical) settings, which contain a high variability of physiological conditions and events, are needed to validate the potential of respiratory activity extracted from PPG closer to its clinical application.

While clinical transmissive PPG and reflective PPG in wearables are based on the same physiological principle, they differ regarding the measured physical signal and the likelihood of artifacts (Tamura *et al* 2014). Therefore, findings and algorithms based on transmissive PPG need to be confirmed and validated also on reflective PPG. Hartmann *et al* investigated the infrared light reflective PPG sensor at several body locations, including the wrist (Hartmann *et al* 2019). However, their analysis focused on respiration rate estimation on short protocolized recordings and used a different PPG light than the more adequate green light PPG for wrist placement (Hartmann *et al* 2019, Maeda *et al* 2011). Unlike standard clinical transmissive PPG, wrist-worn PPG sensors are usually accompanied by an accelerometer. This additional sensor can help to discard part of the estimations influenced by movements, thus enhancing reliability and accuracy.

Here, we propose and characterize a respiratory activity surrogate obtained from overnight wrist-worn reflective PPG recordings (PPG-RAS). The PPG-RAS was obtained by interpolating the amplitude of the PPG pulses pre-selected using a previously published morphology-based pulse quality index (Papini *et al* 2018). The PPG-RAS was compared with the reference thorax RIP signal, i.e. the standard measurement of respiratory activity during sleep used in clinical practice. This signal was chosen as a reference because it is directly related to the intrathoracic pressure changes which generate the PPG respiratory component. In order to compare the morphology of PPG-RAS and reference respiratory activity signal, we quantified their similarity in the time and in the frequency domain, in addition to the standardly reported respiration rate performance. We analysed the effect of the motion level and the quality of the PPG signal on the similarity between the reference and the surrogate. Based on this analysis, we propose a post-processing methodology for the PPG-RAS to further enhance its reliability and similarity with the reference. We enriched our analyses by investigating the effect of different sleep stages and OSA severity on the measure, in a heterogeneous sleep-disordered population close to the envisioned unobtrusive sleep monitoring application. This population provides data with a wide spectrum of possible artifacts and confounding factors, in order to provide a well-rounded picture of the PPG-RAS for wearable sleep monitoring applications.

Table 1. Demographics of the participants expressed as average ± standard deviation [range]. The OSA severity classes are none with AHI< 5, mild with 5≤ AHI< 15, moderate with 15≤ AHI< 30, and severe with AHI> 30.

N [#] (male)	389 (245)
Age [years]	51 ± 15 [18 to −86]
BMI [kg/m^{-12}]	28 ± 5 [17 to −86]
Average heartrate [bpm]	62 ± 9 [44 to −92]
Average respiration rate [breaths/min]	15 ± 2 [10 to −24]
Total sleep time [min]	396 ± 80 [36 to −580]
Wake [min]	106 ± 73 [10 to −436]
REM [min]	66 ± 28 [0 to −157]
N1 [min]	54 ± 28 [4 to −201]
N2 [min]	210 ± 54 [19 to −408]
N3 [min]	65 ± 36 [0 to −236]
Apnea hypopnea index (AHI) [events/h]	15 ± 17 [0 to −108]
Obstructive apnea index [events/h]	1 ± 4 [0 to −56]
Hypopnea index [events/h]	12 ± 12 [0 to −71]
None/mild/moderate/severe OSA cases[#]	125/123/87/54
Participants with > 1 central apnea per hour [#]	69
Participants with > 1 mixed apnea per hour [#]	28

2. Materials and methods

2.1. Dataset

Recordings were selected from the SOMNIA database, a dataset from an ongoing data collection project at Sleep Medicine Centre Kempenhaeghe (Heeze, The Netherlands). The SOMNIA database includes unselected patients scheduled for a routine diagnostic PSG and is aimed to facilitate future research in sleep medicine (van Gilst *et al* 2019). The SOMNIA study was reviewed by the medical ethical committee of the Maxima Medical Center (Eindhoven, the Netherlands. File no: N16.074). All participants gave written informed consent.

For this study, we planned to have a cohort of patients at least 50% of whom were diagnosed with obstructive sleep apnea syndrome (with an approximately uniform severity distribution). Patients with a primary diagnosis of central sleep apnea and Cheyne–Stokes breathing were excluded, as well as recordings made during CPAP use. The final dataset consisted of 389 recordings, with the following primary diagnoses: obstructive sleep apnea syndrome (226 cases), insomnia (108 cases) and sleep movement disorders (45 cases). Most of the participants had additional sleep comorbidities. The demographic data of the dataset are reported in table 1.

The wrist-worn device employed in this research embedded a three-axial accelerometer and a reflective PPG sensor with a light source consisting of two green light LEDs. This device was developed by Philips for research purposes and it has been used in several other biomedical research efforts, e.g. on blood pressure, sleep and atrial fibrillation monitoring (Radha *et al* 2019, Fonseca *et al* 2017, Eerikäinen *et al* 2018) Measurement data included the following signals: wrist-worn green light reflective PPG (sampling frequency 32 Hz), wrist-worn accelerometry (128 Hz, dynamic range ± 8 g), and the thorax RIP signal (128 Hz). The RIP signal and the wrist-worn device signals were synchronized by aligning the IBIs extracted from PPG and ECG (included in the PSG recording with the reference respiratory activity). Although this type of synchronization is not ideal, it guarantees a synchronization error lower than one second (i.e. lower than the average IBI in our dataset). In addition to the physiological signals, we used the sleep and respiratory annotations made by expert sleep technicians from the Sleep Medicine Centre Kempenhaeghe (according to the AASM 2015 guidelines (Berry *et al* 2015)). The signals were trimmed based on the scored PSG. When present, the part of the signal before the lights were switched off in the test room was excluded. The thorax RIP signal was filtered with a 3rd order zero-phase band-pass Butterworth filter with cut-off frequencies of 0.05 Hz and 0.6 Hz (breathing range of 3 to 36 breaths/min) and downsampled to 10 Hz in order to match the respiratory activity surrogate described in the next section.

2.2. Extraction of the wrist-worn PPG respiratory activity surrogate

Our method is composed of four main parts: pulse segmentation, extraction of landmarks on the PPG pulses, pulse-by-pulse quality evaluation and calculation of surrogate respiratory activity from the reliable landmarks.

The PPG pulse segmentation was accomplished by adapting the technique described in our previous papers (Papini *et al* 2017, Papini *et al* 2018). In summary, the technique consisted of filtering the PPG signal

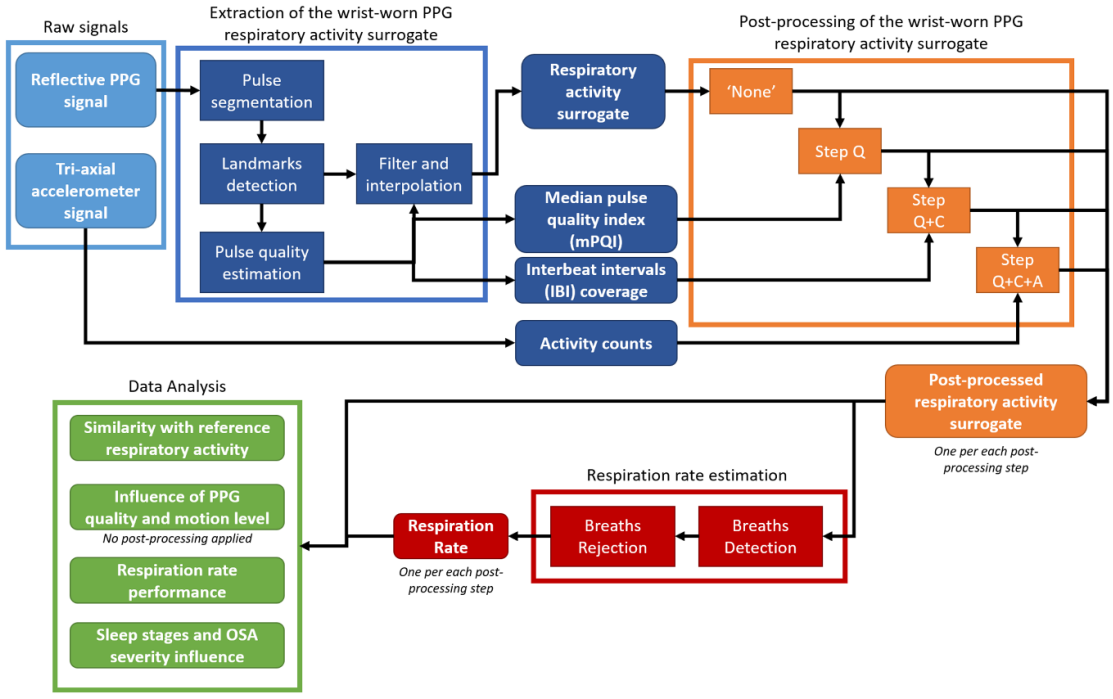

Figure 1. Flowchart of the proposed method. Each color identifies one of the steps described in the manuscript's sections.

in order to remove all the non-heart-rate components, finding the local minima of the filtered signal and using those to define the search area in which to find the start and the end of each PPG pulse. Finally, these points are used to segment a differently filtered version of the original PPG signal in order to include more cardiovascular information. Differently from our previous work, the second filtered PPG signal for this research was obtained using a 3rd order zero-phase band-pass Butterworth filter with cut-off frequencies of 0.05 Hz and 10 Hz in order to include also the respiratory frequency (0.05 Hz corresponds to 3 breaths/min).

Once the pulses were segmented, the time and amplitude location of four landmarks were obtained for each pulse (figure 2):

- foot: the starting point of the PPG pulse;
- end: the endpoint of the PPG pulse; normally corresponds to the foot of the next pulse;
- systolic peak: the maximum of the PPG pulse; and
- diastolic peak (also known as dicrotic wave peak): second inflection point after the systolic peak and before the end of the PPG pulse (first inflection point is the dicrotic notch).

Pulses, and their associated landmarks, were morphologically rated using the PPG quality index described in our previous research (Papini *et al* 2018): the quality index of each pulse was calculated by evaluating its similarity to a pulse template extracted from 15-minutes segment of the PPG signal itself. The template was obtained by using all the detected pulses in the segment, prior to estimating the pulse quality index. The template was considered valid only when the average of the non-null pulse quality indexes was above 0.9. In case the template was found to be invalid, the template was re-calculated by excluding the pulses with the 20% lowest pulse quality index, and the quality of all pulses in the PPG portions was re-assessed. If the template was still not valid, the algorithm used the template calculated in the previous segment of the PPG signal or, if there was no previous template available, a null pulse quality index was assigned to all pulses in the PPG signal portion. This quality index allows accounting for changes in the pulse morphology while at the same time allowing abnormal pulses to be identified, such as those generated by arrhythmic heartbeats. The PPG quality index calculation employed the same filters described in our previous work. For this research, a quality threshold of 0.6 (quality index from 0, low quality, to 1, high quality) was chosen based on the result of our previously published research in order to balance abnormal pulse rejection and sensitivity to sinus rhythm pulses (Papini *et al* 2018).

The landmarks can be used to calculate respiratory activity surrogates by interpolating either their amplitude, or the distance between consecutive landmarks (Charlton *et al* 2016). Each PPG pulse always had four associated landmarks and we relied on the pulse quality index to reject the landmarks wrongly assigned

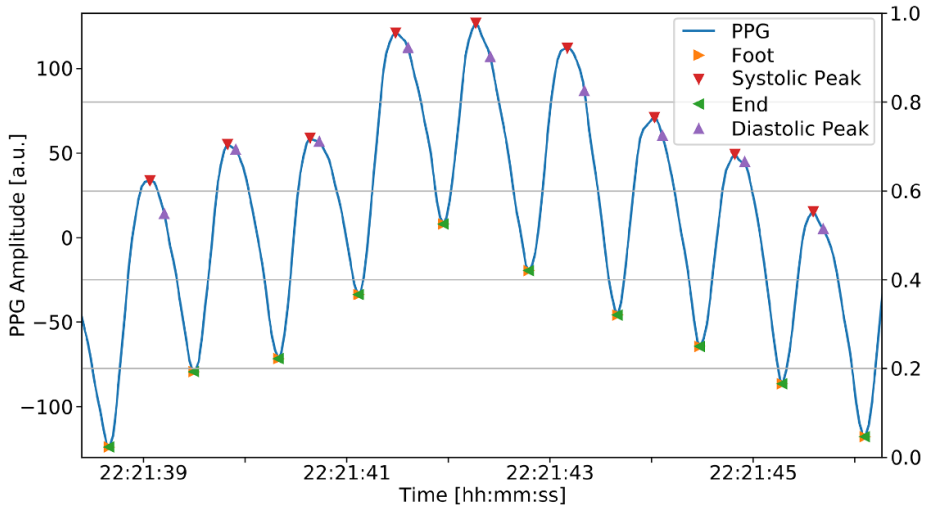

Figure 2. Wrist-Worn PPG showing the detected landmarks. The 'foot' and the 'end' are usually overlapping, however each of them can be missing because of the rejection of the pulse they belong to due, for instance, to artifacts.

to a corrupted PPG morphology. No procedure was applied to recover the rejected landmarks. We calculated surrogates by linearly interpolating at 10 Hz the amplitude of each landmark and all possible distances that can be obtained by pairing different landmarks. For instance, the surrogates related to the foot landmark were obtained by interpolating the foot amplitude values and distance in terms of amplitude and time between foot and end, foot and systolic peak, and foot and diastolic peak. In addition, we obtained surrogates by combining the surrogates of the same nature (i.e. interpolation of the amplitude values, of the amplitude distance and of the time distance) using principal component analysis and selecting the component with the largest eigenvalue. Finally, each surrogate was filtered with a 3rd order zero-phase bandpass Butterworth filter with cut-off frequencies of 0.05 Hz and 0.6 Hz (breathing rate of 3 to 36 breaths/min) (Charlton *et al* 2016). We opted for a 10 Hz sampling frequency in order to be consistent with the method of Long *et al* (2014) used in this research to process the respiratory activity signals (reported in section 2.4).

The best candidate for respiratory activity surrogate was selected based on the respiration rate estimation limits of agreement and bias (Charlton *et al* 2016), with the respiration rate calculated as described in section 2.4.

2.3. Post-processing of the wrist-worn PPG respiratory activity surrogate

The similarity between the PPG-RAS and the reference can be affected by the quality of the PPG signal and by motion levels (Papini *et al* 2018). These factors can be easily calculated in an epoch-by-epoch manner from our wrist-worn PPG device.

We derived an indicator of the PPG signal quality by calculating the median pulse quality index (mPQI) over 30-seconds epochs. The median was chosen instead of the average because the quality index distribution is not normal (bounded between 0 and 1 and tendentially skewed towards 1) (Papini *et al* 2018). Epochs with less than ten pulse quality estimations were considered invalid and excluded from further analysis. A second quality indicator was derived by calculating the ratio between the sum of the interbeat intervals (IBIs) divided by the duration of the epoch (IBI coverage), i.e. 30 seconds. The IBIs were calculated from the foot landmarks belonging to each normal sinus pulses (pulse quality index above 0.6). IBIs were excluded if longer than 2 seconds or shorter than 0.5 seconds because they are physiologically very unlikely during sleep (Mason *et al* 2007, Penzel *et al* 2003), or if the ratio of an IBI and its preceding value was larger than 1.5 because it was influenced by the presence of an ectopic beat (Papini *et al* 2019).

Movement levels were estimated using the signal measured with the accelerometer in the wrist-worn device. Motion was quantified in so-called activity counts, consisting of a sum of the three accelerometer signals, one for each axis, over a certain amount of time (30 seconds in our case) after the removal of the gravity acceleration component. They are commonly used in physical activity monitoring and, especially relevant for our case, for sleep and wake classification (Bonomi and Westerterp 2012, Long *et al* 2013).

The two PPG signal quality indicators and the activity counts can be thresholded to remove 30-seconds segments of respiratory activity surrogate considered less reliable or likely having been influenced by non-respiratory phenomena. Post-processed versions of the PPG-RAS were calculated by cumulatively excluding the 30-seconds epoch signal with an mPQI lower than 0.8 (step Q), an IBI coverage lower than

66% (step Q+C), and activity counts higher than 21 (step Q+C+A). The order of these three steps was decided based on the drop in coverage they caused, i.e. from the one reducing least to the one reducing most coverage (respectively, median coverage drops of 5%, 6%, and 17%). The mPQI threshold was chosen in order to guarantee that fewer than 50% of the pulses had a high probability to be non-normal sinus pulses (Papini *et al* 2018). The IBI coverage threshold allowed having most of the 30-seconds epochs covered to extract the respiratory activity surrogate. The activity counts threshold was chosen based on the 90th percentile of the activity counts during the N2, N3 and REM sleep stages, i.e. the stages with less movement in our dataset and according to literature (Allena *et al* 2009).

2.4. Respiration rate estimation

To extract breathing rate, respiratory activity signals were pre-processed and each breath landmark (i.e. beginning, peak, and end) located in order to calculate the required features. The pre-processing, described by Long *et al* (2014), consisted of a series of filters to remove the baseline of the signal via median filtering based on the distance between the beginning of the respiratory movement and its peak. The breath landmarks were detected using the changes in sign of the signal derivative. Landmarks were excluded if the sum of the time distance between the previous and successive landmarks was lower than the median time distance between landmarks across the entire recording. In addition, landmarks were excluded when their amplitude distance was lower than 15% of the median amplitude distance across the entire recording.

The respiration rate was calculated in breaths/min as the inverse of average time distance between the beginning and the end breath landmarks, i.e. breath lengths, on 30-seconds epochs. A respiration rate estimation was considered valid if a minimum of three breath cycles were present and rejected otherwise. The respiration rate was derived in the same manner for the reference respiratory activity signal and the PPG-RAS.

2.5. Data Analysis

2.5.1. Comparison of the respiratory activity signals

The reference respiratory activity signal and the PPG-RAS can have different morphology, depending for example on the presence of respiratory events, changes in limb position during long recordings and different baseline conditions differently affecting the amplitude of the two signals. For instance, a change in body position would change the range of motion of thoracic movement while influencing the PPG because of changes in local blood pressure, local variations in amount of venous blood and force with which the sensor is pressed on the skin (Reisner *et al* 2008). In addition, different participants might have different vascular mechanical properties that would influence the PPG signal relationship with breathing (Allen 2007, Tamura *et al* 2014, Addison 2017, Nilsson 2013). Therefore, we evaluated the morphological similarity between the reference respiratory activity signal and the PPG-RAS in the time and frequency domain using a windowed approach. We calculated the Spearman's correlation (ρ_s) and the average magnitude-squared coherence (C_{xy}) epoch-by-epoch – with an epoch size of 30 seconds – on windows of the signals centered on the current epoch. A window size of 30 seconds was chosen for ρ_s in order to reduce the influence of artifacts, while a window of 90 seconds was chosen for C_{xy} to be able to estimate the spectral properties of the signal using Welch's method. The short calculation windows minimize the effect of local changes on the similarity metric, making it more robust to, e.g. changes in body position or temporary loss of contact between the PPG sensor and the skin. Spearman's correlation was chosen instead of Pearson because the relationship between the two respiratory activity signals is monotonic and non-linear. Both windowed metrics were used in a similar evaluation approach as in literature by Widjaja *et al* (2012). The C_{xy} for a specific frequency f is calculated as

$$C_{xy}(f) = \frac{\mathrm{PSD}_{xy}(f)^2}{\mathrm{PSD}_{xx}(f) * \mathrm{PSD}_{yy}(f)} \quad (1)$$

where PSD^{xx} and PSD^{yy} are the power spectral densities of, respectively, the reference respiratory activity signals and the PPG-RAS, and PSD^{xy} is the cross power spectral density between the two. The parameter ρ^s ranges from -1 to 1 and gives an indication of the strength and direction of the monotonic relationship between the two variables. For each epoch, several ρ^s values were calculated by applying a maximum of 10 samples lag (i.e. 1 s) between the two respiratory activity signals; then the maximum value was taken as the final ρ_s value (Widjaja *et al* 2012). This was done to account for possible time-shift between the respiratory activity signals due to physiological or technical reasons, e.g. pulse arrival time and not perfect synchronization between sensors. On average, the pulse arrival time (when calculated as the time difference between the R-peak of an ECG heartbeat and the first derivative maxima of the PPG pulse rising slope at the wrist) is approximately 270 ms (Rajala *et al* 2018) but it may vary due to, for instance, body movements, arousals and the blood pressure dip occurring during sleep (Foo *et al* 2005, Katz *et al* 2003, Zheng *et al* 2016).

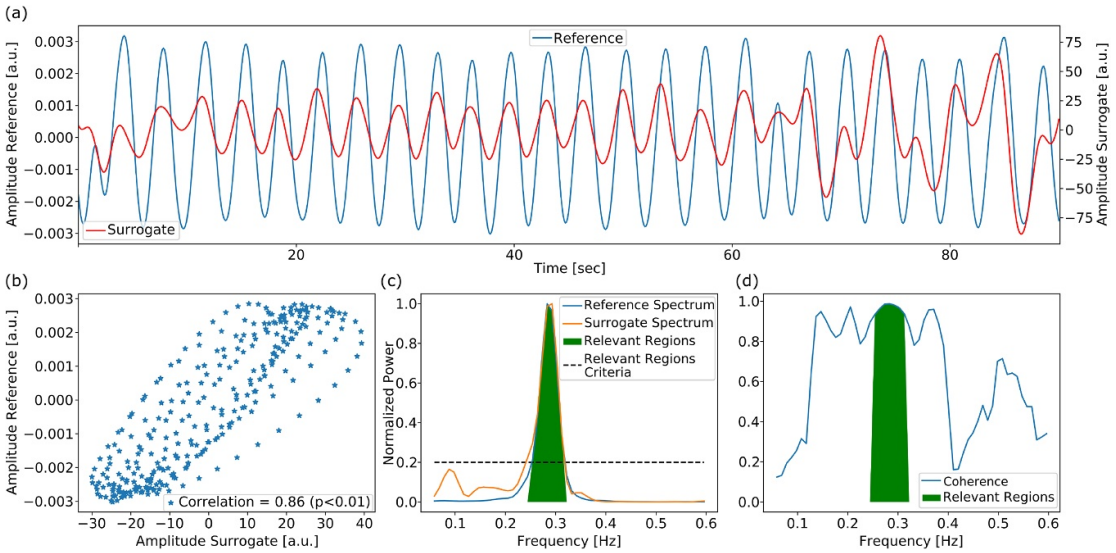

Figure 3. Example of a reference respiratory activity signal and a PPG-RAS with a strong similarity ($\rho_s = 0.86$, $C_{xy} = 0.96$). (a) The two respiratory activity signals in the C_{xy} calculation window (90 seconds), (b) scatter plot of the amplitude of the signals to show their correlation, (c) power spectrum of the signal highlighting the region used to calculate the average C_{xy}, (d) the spectral coherence between the two signals.

Although we could quantify and correct the lag between the respiratory activity signals, we cannot draw any physiological conclusion from it because, with our experimental setup, it not possible to disentangle its multiple physiological components and the misalignment due to signal synchronization technique used. The risk of overlapping the nth breath of a 30-sec window of the reference respiratory activity with the $(n \pm 1)$th breath of the corresponding 30-sec window of the PPG-RAS is minimal nevertheless, as the average breath length during sleep is approximately 3.75 seconds (respiration rate of 16 breaths/min) (Gutierrez *et al* 2016).

The parameter C_{xy} ranges from 0 to 1 corresponding to the range between absent to complete spectral coherence between the signals for a specific frequency. Widjaja *et al* calculated the average C_{xy} among the frequencies surrounding the respiratory frequency with at least half of the maximum power. Instead, we selected the frequencies to be averaged from those that had, in the reference respiratory activity signal, a power higher than 20% of the maximum spectral power component. The calculation of the maximum and the selection of the other frequencies were within the physical limits of the breathing range (0.05 to 0.6 Hz corresponding to 3 to 36 breaths/min). This approach guaranteed that all the potentially relevant respiratory information in the frequency domain was included and not only the frequencies close to the respiratory frequency peak. The spectrum and C_{xy} were calculated with Welch's method with fast Fourier transform using a periodic Hann window on segments of 40 seconds, zero-padded to 1024 points and a 50% overlap. Windows with less than three segments were excluded from the analysis, i.e. the corresponding epoch was considered not valid. To give an idea of similarity metrics used, two examples taken from our data indicating strong and weak similarity between the reference respiratory activity signal and the PPG-RAS are shown in figures 3 and 4.

We evaluated the effect of the quality of the PPG signal and of the motion level on the similarity between reference and surrogate respiratory activity. For C_{xy}, these parameters were computed on the central 30 seconds of the 90 seconds window used for the parameter calculation. We grouped the epochs in intervals with a width of 0.1 (left-closed and right-open intervals) with respect to the mPQI and of the IBI coverage, e.g. mPQI$\in$[0.6–0.7) and mPQI$\in$[0.7–0.8). This way, we were able to visualize the trend of ρ_s and C_{xy} with boxplots of the created subgroups. We opted for a similar plot to describe the influence of the motion level; the epochs were separated using 0, 21, 70 and 387 activity counts as boundaries for the groups (also in this case left-closed and right-open intervals). The motion thresholds above 0 activity counts were based on the 90th percentile of the activity counts during the sleep stages N2+N3+REM, during N1, and during Wake, i.e. from the most to the least 'motionless' stages of sleep (Wilde-Frenz and Schulz 1983, Gori *et al* 2004).

Finally, we calculated the per-participant median values of ρ_s and C_{xy}, and 30-seconds epoch coverage when each step of our post-processing was applied and reported as median and interquartile range. The effect of sleep stages and OSA severity on the similarity was analysed by separating the pooled 30-seconds epochs according to the canonical sleep stages (i.e. Wake, REM, N1, N2, and N3) and OSA severity classes (i.e. normal with AHI<5, mild with $5\leq$ AHI<15, moderate with $15\leq$ AHI<30, and severe with AHI>30.

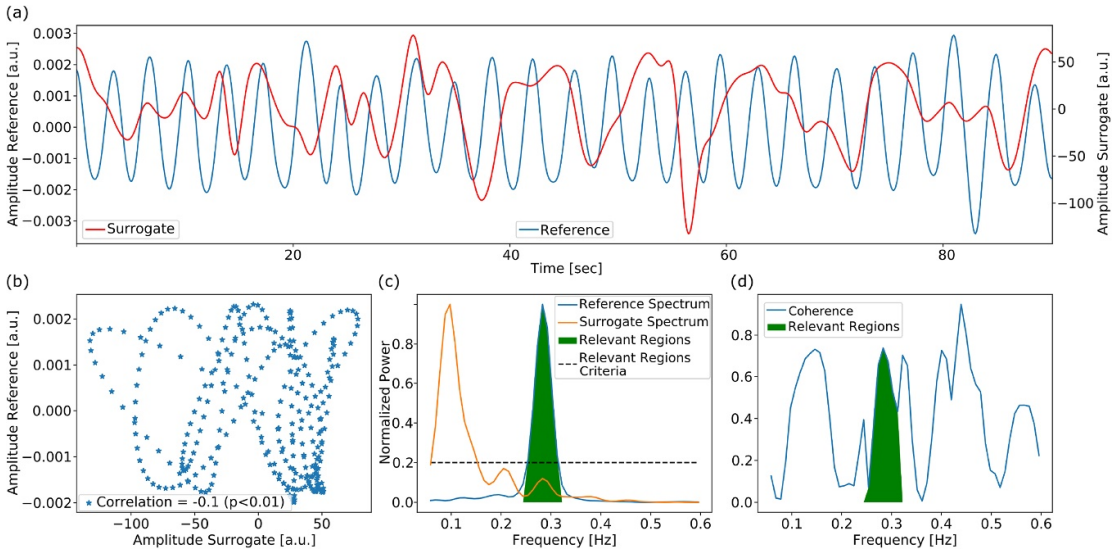

Figure 4. Example of a reference respiratory activity signal and a PPG-RAS with a weak similarity (ρ_s = -0.1, C_{xy} = 0.49). (a) The two respiratory activity signals in the C_{xy} calculation window (90 seconds), (b) scatter plot of the amplitude of the signals to show their correlation, (c) power spectrum of the signal highlighting the region used to calculate the average C_{xy}, (d) the spectral coherence between the two signals.

When analysing the effect of OSA, we excluded the Wake epochs since they do not contribute to the OSA severity estimation and they likely have low similarity. This analysis was done with and without our final post-processing step (step Q+C+A) in order to highlight its effect on the performance.

2.5.2. Respiration rate evaluation

Breathing rate estimation performance was evaluated similarly to the algorithms described by Charlton *et al* (2016). We calculated the per-participant average and standard deviation of the estimation error, and the coverage probability for 2 breaths/min (CP_2) with respect to the respiration rate estimated from the reference thorax belt (using the same breath detections methods). The CP_2 is defined as the percentage of estimations with an absolute error lower than 2 breaths/min (Barnhart *et al* 2007). The performance was calculated with and without each proposed post-processing step and median and inter-quartile range were reported.

2.5.3. Statistical analysis

The change in performance generated by each post-processing step was statistically tested against the performance of the original PPG-RAS signal using a Mann–Whitney rank test (Mann and Whitney 1947). Differences between the sleep stages and OSA severities were evaluated with a Kruskal–Wallis H test (Kruskal and Wallis 1952), followed by a post-hoc Conover's test with a Bonferroni correction to determine which group was statistically different (Conover and Iman 1979). The performance changes for each sleep stage and OSA severity generated by applying all three post-processing steps were evaluated with a Mann–Whitney rank test (Mann and Whitney 1947). Analyses were done with a 0.05 p-value threshold to establish statistically significant differences. We opted for non-parametric tests because the data were statistically different from a normal distribution using the Shapiro-Wilk test (Shapiro Wilk 1965).

3. Results

The respiratory activity surrogate with the best breathing rate estimation performance was obtained by interpolating and filtering the amplitude distance between the foot and the systolic peaks landmarks (table 2) We restrict further analysis of similarity with the reference respiratory activity measures only to this surrogate.

3.1. Comparison between reference and surrogate respiratory activity signals

The median and interquartile range per-participant of ρ_s and C_{xy} before post-processing are reported in the first row of table 3 (post-processing 'None'). For most participants, the value of ρ_s was on average above fair ($\rho_s > 0.4$ for more than 75% of the participants), while C_{xy} indicated a strong coherence between the spectrum of the reference and the PPG-RAS in the relevant frequencies of the reference signal.

Table 2. Overall respiration rate estimation bias, standard deviation and limits of agreement (bias± 1.96*standard deviation) for the three best performing respiratory activity surrogate candidates.

Pulse landmarks candidates	Bias [breaths/min]	Standard deviation [breaths/min]	Limits of agreement [breaths/min]
Amplitude distance foot-systolic peak	0.99	2.57	−4.05 — 6.03
Time distance foot-end	0.41	2.85	−5.18 — 6.00
Time distance foot-systolic peak	0.61	2.95	−5.17 — 6.39

Table 3. Per-participant coverage, median ρ_s, and median C_{xy} expressed as median [interquartile range], and the number of participants with at least one hour of valid epochs for different post-processing.

Post-processing steps	Per-participant coverage [%]	Per-participant ρ_s	Per-participant C_{xy}	Participants with > 1 h of valid epochs [#]
None	100.0 [100.0-100.0]	0.54 [0.44-0.64]	0.81 [0.72-0.88]	389
Step Q	95.4 [90.3-97.5]	0.57 [0.45-0.66]	0.83 [0.74-0.89]	387
Step Q+C	89.9 [81.1-93.8]	0.59 [0.47-0.67]	0.85 [0.76-0.90]	386
Step Q+C+A	75.3 [64.1-84.2]	0.62 [0.50-0.72]	0.88 [0.82-0.92]	384

Post-processing steps: None = no post-processing, step Q = mPQI > 0.7, step Q+C = step Q & IBI coverage > 66%, step Q+C+A = step Q+C & activity counts > 21.
Statistically different ($p < 0.05$) from the performance without post-processing (None).

Table 4. Per-participant respiration rate estimation performance in terms of average and the standard deviation and CP_2 expressed as median [interquartile range].

Post-processing steps	Per-participant coverage [%]	Per-participant average estimation error [breaths/min]	Per-participant standard deviation of estimation error [breaths/min]	Per-participant CP_2 [%]
None	87.4 [78.8-93.7]	0.76 [0.27-1.44]	2.11 [1.80-2.55]	70.4 [59.2-78.6]
Step Q	84.1 [74.3-91.7]	0.71 [0.22-1.37]	2.06 [1.73-2.45]	71.4 [60.7-79.4]
Step Q+C	80.1 [68.0-88.5]	0.62 [0.12-1.22]	1.96 [1.60-2.30]	73.7 [63.1-81.1]
Step Q+C+A	69.9 [54.8-79.4]	0.53 [0.05-1.12]	1.85 [1.44-2.19]	76.1 [66.1-84.2]

Post-processing steps: None = no post-processing, step Q = mPQI > 0.7, step Q+C = step Q & IBI coverage > 66%, step Q+C+A = step Q+C & activity counts > 21.
Statistically different ($p < 0.05$) from the performance without post-processing (None).

The similarity was influenced by the quality of the PPG signal and by motion levels, as shown in figure 5. Most epochs had a high mPQI (0.7 or more for 89% of epochs) and an average IBI coverage (0.5 or more for 84% of epochs). The similarity increased along with the increase of the two factors describing the quality of the PPG signal. As expected from sleep recordings, most of the epochs consisted of low motion levels (91% of epochs had activity counts< 70). Motion level affected ρ_s and C_{xy} negatively: an increase in movement activity coincided with a decrease in morphological similarity. Already from a mild motion level (70> activity counts≥ 21) the median ρ_s and the median C_{xy} decreased more than 60% respective to the null motion level (activity counts< 21). For medium and high movement levels (387> activity counts≥ 70 and activity counts≥ 387) the similarity between the reference respiratory activity and the PPG-RAS was mostly weak (median $\rho_s < 0.3$). When post-processing was applied, the similarity between the respiratory activity signals improved, e.g. the first interquartile of the similarity metrics increased up to 14% (table 3). With each post-processing step, the similarity between the signals increased and coverage decreased.

3.2. Respiration rate performance

The results for respiration rate estimation with different steps of post-processing are summarized in table 4. The coverage without post-processing was loe Orphanidower than 100% because the respiration rate estimation was performed only for epochs with at least three breathing cycles. Similarly to ρ_s and C_{xy}, each post-processing step increased the performance (lower average and standard deviation of the estimation error, and higher CP_2) at the cost of epoch coverage. We measured a 30% drop in coverage in the lowest interquartile between no post-processing ('None') and when the final post-processing step ('step Q+C+A') was applied.

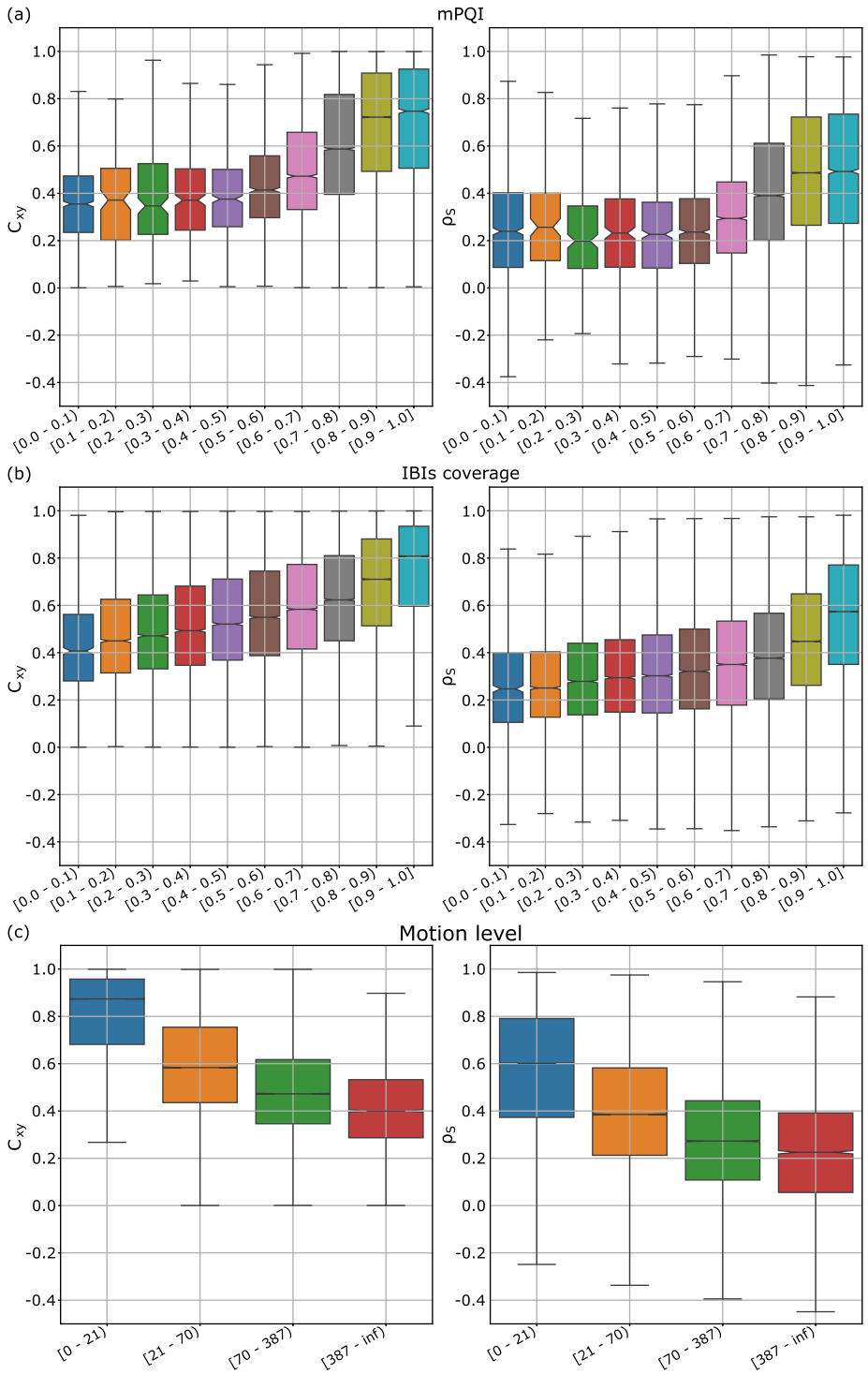

Figure 5. Epoch-by-epoch Spearman's correlation and spectral coherence for different values (*x*-axis) of (a) mPQI, (b) IBIs coverage and (c) motion level. The boxplots show median and interquartile range. The whiskers are 1.5 times the corresponding quartile, and points outside the whiskers are omitted since considered outliers.

3.3. Performance per sleep stage

Different degrees of similarity between the reference respiratory activity signal and the PPG-RAS were found when the 30-seconds epochs were grouped according to sleep stages (figure 6). The lowest performance was observed in Wake, followed by stage N1 sleep. All ρ_s and C_{xy} differences between sleep stages were statistically significant (Kruskal–Wallis H and Conover's tests $p < 0.05$). Applying the post-processing steps resulted in a statistically significant increase in performance for each sleep stage (Mann–Whitney rank test $p < 0.05$). Epochs scored as Wake and N1 had the largest increase in ρ_s and C_{xy} with, respectively, 28% and 12% increments of the median ρ_s and 23% and 11% increments of the median C_{xy}.

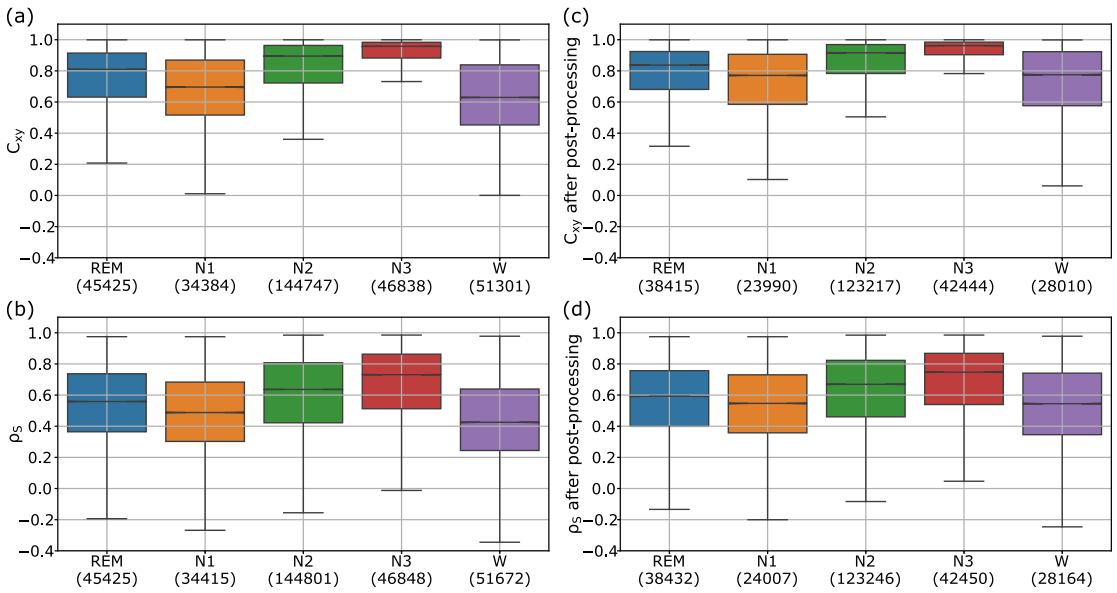

Figure 6. Epoch-by-epoch Spearman's correlation and spectral coherence for different sleep stages (in parenthesis the number of epochs). (a) and (b) before post-processing; (c) and (d) after the final post-processing step (step Q+C+A) were applied. The boxplots show the median and interquartile range. The whiskers are 1.5 times the corresponding quartile, and points outside the whiskers are omitted, since considered outliers.

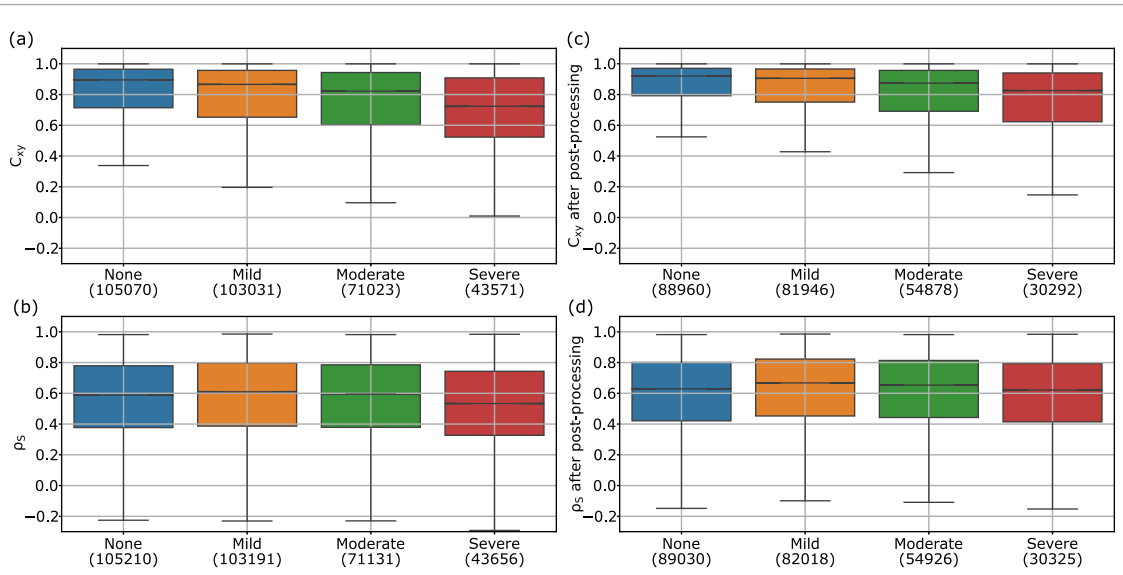

Figure 7. Epoch-by-epoch Spearman's correlation and spectral coherence for different OSA severities (in parenthesis the number of epochs). (a) and (b) before post-processing; (c) and (d) after the final post-processing step (step Q+C+A) were applied. The boxplots show the median and interquartile range. The whiskers are 1.5 times the corresponding quartile, and points outside the whiskers are omitted, since considered outliers.

3.4. Performance per OSA severity

We found a statistically significant difference in similarity between the reference respiratory activity signal and the PPG-RAS for each set of epochs grouped according the OSA severity class (Kruskal–Wallis H and Conover's tests $p < 0.05$) (figure 7). The participants without OSA had the higher C_{xy}; instead mild and moderate OSA cases had a higher ρ_s, especially after the post-processing was applied. For both similarity metrics, the epochs belonging to participants with severe OSA had the lowest values. N2 was the sleep stage most present in epochs grouped per OSA severity (over 50% of sleep time). The epochs belonging to severe OSA cases had the least amount of N3 (13% of sleep time) with respect to the other OSA severity classes (normal 18%, mild and moderate 16%). The performance for each OSA severity group significantly increased when the post-processing was applied (Mann–Whitney rank test $p < 0.05$). Epochs belonging to severe OSA had the largest increase in ρ_s and C_{xy} with 11% and 9% increments of median ρ_s and median C_{xy}.

4. Discussion

Extracting respiratory activity information from wrist-worn PPG could represent an important step towards new unobtrusive monitoring solutions for sleep and, especially, for sleep-disordered breathing conditions. Here, we propose a PPG-RAS: respiratory activity surrogate extracted from the wrist-worn reflective PPG signal. We assessed the morphological similarity between the reference thoracic respiratory activity signal and the PPG-RAS in the time and frequency domain using, respectively, the Spearman's correlation ρ_s and the spectral coherence C_{xy}. Results showed that PPG-RAS compares well with the reference thoracic respiratory activity signal. The proposed post-processing further strengthened performance using parameters that can be easily obtained in wrist-worn PPG devices, i.e. the quality of the PPG signal and the acceleration that is usually measured with this type of devices.

We found that PPG signal quality and motion levels influenced the performance of PPG-RAS. A low mPQI entails that most of the PPG pulses are rejected and, therefore, there are fewer data to calculate the respiratory surrogate. However, the mPQI is not only informative on the number of usable pulses, but also on the quality of the pulses included in the calculation. In fact, even when mPQI was just above the threshold used for pulse rejection (i.e. 0.6-0.7 interval), the similarity between the reference and the surrogate was weak ($\rho_s < 0.3$ and $C_{xy} < 0.5$, figure 5(a)). The IBI coverage complements the mPQI as a quality indicator because it is a direct measure of the number of pulses available for the surrogate extraction. The relationship between IBI coverage and performance (figure 5(b)) did not present the same sharp increase seen for the mPQI (figure 5(a)). This is because even when the IBI coverage was mediocre, there might have been enough high-quality pulses to accurately estimate the surrogate. The respiratory frequency range (upper limit 0.6 Hz) is generally lower than the heartbeat frequency range (lower limit 0.5 Hz), therefore even the exclusion of some pulses usually leaves enough points to describe the respiratory activity signal. The relationship between motion levels and respiratory activity signal similarity confirmed that the more quiet sleep is, the easier it is to accurately capture respiratory activity with a wrist-worn sensor (figure 5(c)).

Our three-step post-processing allows a trade-off between ensuring morphological similarity of the respiratory activity estimate with the reference, and coverage. Inevitably, some of the recordings ended up having a low number of epochs after applying all post-processing steps (table 3). This was expected given the dataset: the real-world nature of the recordings implies the presence of a higher number of artifacts in the collected signals when compared with a dataset acquired in a protocolized, well-controlled laboratory setting. It is also important to consider that a real-world clinical dataset of this size comports having a wide variety of characteristics, including participants with comorbidities or medication affecting the cardiovascular system, such as cardiac arrhythmias or hypertensive medication (Papini *et al* 2018, Allen 2007, Castaneda *et al* 2018). These factors may play a detrimental role when cardiovascular signals are used to extract other measurements, such as respiration in the current case. Nevertheless, the method and the post-processing proposed were able to derive a surrogate with a consistent match with the reference, as suggested by the interquartile range of the performance (table 3 and table 4).

To our knowledge, there is no previous work on reflective, green-light, wrist-worn PPG for the extraction of a respiratory activity surrogate and only one regarding respiration rate estimation with similar measurements. The device and method proposed by Renevey *et al* (2018) extracted respiration rate from 31 full-night sleep laboratory recordings belonging to seven healthy participants wearing two wrist-worn PPG devices. Their respiration rate estimation method had a mean absolute error of 0.8 [0.5–1.2] breaths/min (median and interquartile range). Our mean absolute errors were 1.50 [1.13–2.07] without post-processing, and 1.24 [0.87–1.72] breaths/min with full post-processing. Having multiple recordings from the same healthy participant, and even from the same measurement, lowered the variability of the dataset used by Renevey *et al* (2018) and most likely contributed to a lower estimation error with respect to our method. Besides, the result difference could have been enhanced by their manual selection of the recordings based on the quality of PPG and reference signals (56 recordings were initially collected) that was not performed on our dataset in order to provide performance close to the real application scenario. Apart from this research, the comparison with the literature is limited to methods tested with other types of setups and sensors, and results have to be interpreted with caution. Also Chang *et al* (2018) and Park and Lee (2019) proposed respiration rate estimation methods for wrist-worn green light PPG. Their methods, although interesting from a signal processing point of view, were evaluated on datasets considerably different from ours in terms of size, variability and similarity to the real-world application. The method proposed by Chang *et al* (2018) achieved higher respiration rate estimation performance than our method; however, they evaluated their method on 75 protocolized five-minutes recordings. Instead, we achieved similar performance to the method proposed by Park and Lee (2019) that was tested on 30 recordings lasting five minutes from five young healthy participants. Our method could estimate breathing rate to a level in line with the extensive literature using other types of PPG sensors even when post-processing was not applied. We were able to estimate the

respiration rate with high coverage especially considering the real-world dataset used (Charlton *et al* 2016). Using the post-processing steps allowed a significant increase in the respiration rate estimation performance, at the expense of the coverage. When our final post-processing step (step Q+C+A) was applied, the coverage drop came with a moderate increase in respiration rate performance (e.g. 14% for the CP_2 lowest interquartile). Therefore, especially for breathing rate estimation, choosing the post-processing steps should be done according to the application requirements, e.g. choosing to include the final post-processing step only when a high estimation accuracy is required.

We are not the first to propose a post-processing method in this field. However, most methods described in literature aim to improve breathing rate estimation and not the respiratory activity surrogate extraction per se (Charlton *et al* 2016). For instance, most of the methods proposed by Charlton *et al* either merge several estimations and exclude those with high disagreement (Karlen *et al* 2013) or smoothen the estimation based on previous values (Lazaro *et al* 2013). We tested these two approaches, finding that they did not improve respiration rate estimation performance of our method. It should be noted that these described methods were developed and tested for a different type of PPG sensor and a different application scenario. Here, we propose a different post-processing philosophy that does not rely on the output measurement itself. Instead, it is exclusively based on the characteristics of the PPG signal and the phenomena influencing it, and therefore independent of possible errors in the calculation of the output, whether it is respiratory activity or breathing rate.

Previously (Papini *et al* 2018), we showed that a quality index based on correlation (see Orphanidou *et al* 2014, used by Charlton *et al* 2016) tends to be less precise than our pulse quality index because it is less resilient to the presence of artifacts and, especially, of arrhythmic beats. In addition, we removed individual pulses with low quality after the PPG segmentation and only excluded epochs having mostly low-quality pulses during post-processing (by using the median quality instead of the average quality). This allowed epochs with only a few heavily corrupted pulses to still be used in the estimation of respiration without being influenced by low-quality pulses, thus increasing the performance with a low impact on coverage. Relying on a resilient pulse quality index and post-processing tuned specifically for wrist-worn PPG is particularly important for methods to measure patients overnight, since factors detrimentally influencing the estimation are likely to occur and an opportune coverage has to be guaranteed.

The PPG amplitude variation used to derive PPG-RAS is related to changes in the per-heartbeat cardiac output generated with each breath (Karlen *et al* 2013, Buda *et al* 1979): during inspiration, cardiac output decreases, while during expiration it increases. However, cardiac output is also driven by sympathetic and parasympathetic activity, with the former increasing it (Calvert and Lefer 2012). Therefore, the contribution of respiration activity to cardiac output is proportionally weaker during sleep stages where sympathetic activity is higher (or parasympathetic activity is lower). In addition, blood pressure and the vascular tone decrease with progression through the three non-REM sleep stages and from wake to sleep (Somers *et al* 1993, Tobaldini *et al* 2013), increasing the impact of respiratory activity on the PPG signal (Allen and Murray 1999, Shelley 2007). Therefore, we think that the significant influence of sleep stage on the performance of our method can be explained by differences in cardiovascular activity depending on each sleep stage. In fact, the increase in similarity from N1 to N3 non-REM sleep fits with the known autonomic changes over these sleep stages.

Wake epochs often contain movements and therefore changes in the PPG and thorax RIP signal which are independent of respiratory activity. In fact, when full post-processing was applied, the performance during awake periods significantly increased. However, the performance during Wake did not match the other stages even after post-processing, probably because awake periods have higher sympathetic activity compared to sleep (Tobaldini *et al* 2013). It is important to highlight that our recordings took place during 'natural' sleep, and as such, awake periods were different from previous studies on the topic where participants were simply asked to actively maintain a quiet behavior and sometimes even a fixed breathing rate and breathing depth (Lazaro *et al* 2013, Charlton *et al* 2016, Addison 2017, Hartmann *et al* 2019).

The lower morphological similarity of N1 in comparison with REM and the other non-REM sleep stages might be explained by high-intensity movements and periodic limb movements being more likely to occur during N1 (Allena *et al* 2009). This hypothesis is corroborated by the accelerometer-based estimations of motion, and by the percentage of annotated periodic limb movements during N1. In our dataset, the epochs containing such limb movement were 39% for N1 and 21% for REM, i.e. the sleep stage with the second most movements. Just like for wake, using our final post-processing step (step Q+C+A) significantly increased the similarity between reference respiratory activity and PPG-RAS during N1 but it remained significantly different from REM, i.e. the sleep stage with the closest performance. This might be due to the extended influence of periodic limb movements on the cardiovascular system occurring during non-REM sleep when compared with those occurring during REM (Allena *et al* 2009). In addition, N1 often follows

awakenings and, therefore, it might present the cardiovascular continuation of these events (Carskadon Rechtschaffen 2011).

Respiratory events have several hemodynamic consequences that might have affected the PPG-RAS. During these events, the cardiac output decreases due to intrathoracic pressure swings affecting the right and left ventricular functioning (Weiss *et al* 1996). When these events terminate, the blood pressure rises due to the increased cardiovascular resistance generated by the increased peripheral sympathetic activity (Tamisier *et al* 2018). In addition, the ends of the respiratory events are often associated with arousals and body movements (Thomas 2003) The decrease cardiac output might contribute to enhance the respiratory contribution to the cardiac output variation, thus increasing the PPG-RAS morphological similarity with the reference. However, the PPG-RAS might also be affected by a reduction of PPG amplitude occurring after the termination of respiratory events that it is not related to the respiratory activity (Gil *et al* 2008) (besides the detrimental effect of sympathetic activations and the movements). In mild and moderate OSA the positive effect of respiratory events on the similarity between the respiratory activity signals might have prevailed over the other negative effects in calculating ρ_s. This was not noticeable in severe OSA probably due to the lower amount of N3 sleep: N3 had the best similarity performance, therefore a lower presence of this stage offsets the ρ_s towards lower values. Differently from ρ_s, C_{xy} decreased along with the increase of OSA severity. This might be due to the increased spectral complexity of the respiratory activity introduced by respiratory events. These events are characterized by non-stationary breath length variations (e.g. due to normal breathing resumption) that might be not correctly quantified by this metric. However, further investigations, including hemodynamic and respiratory effort measurements, are required to confirm the effect of respiratory events on morphology alteration of PPG and, consequently, of the PPG-RAS.

The lowest percentage of N3 and the respiratory events prevalence are most likely the reasons behind severe OSA benefiting the most from the post-processing. Post-processing step decreases principally N1 sleep and preserves most of N3 (respectively 30% and 10% decreases), thus increasing the weight of sleep stage with higher similarity in the performance evaluation. Instead, the activity counts-based post-processing targets respiratory arousal movements, preserving the epochs with subtle movements that are less likely to affect the PPG-RAS extraction.

We think that the presence of OSA might be less relevant in determining the PPG-RAS performance than the different sleep stages since the similarity variation between OSA severities was lower than between the sleep stages. Therefore, PPG-RAS might be useful for the development of OSA monitoring algorithms and especially when they are provided with sleep stage information to contextualize the PPG-RAS reliability.

5. Conclusion

We proposed a method to derive a respiratory activity surrogate from a wrist-worn reflective PPG sensor, commonly used in consumer wearable devices. The surrogate was validated on a sleep-disordered population close to a 'real-world' sleep monitoring scenario with wearable devices. It showed good similarity with the reference respiratory activity and the proposed post-processing approach allowed further trade-offs between accuracy and a lower coverage of the recordings. We were able to characterize the method with respect to the quality of the PPG, movement levels, sleep stages and OSA severity. Understanding the strengths and limitations of the method in regard to these parameters is fundamental before wrist-worn PPG devices can be used to monitor respiration during sleep in clinical practice. Being able to derive a surrogate morphologically similar to the reference respiratory activity allows extracting respiratory activity characteristics other than just respiratory rate. For instance, it would be possible to assess the variability of the respiratory activity by deriving the standard deviation of the breath length or the relative changes in respiratory amplitude. These features could help to identify sleep phenomena, such as OSA, using wrist-worn PPG devices. Future work will focus on using the PPG-based respiratory activity signal and its associated features to develop monitoring algorithms for sleep apnea and related disorders.

6. Data availability

Data from the SOMNIA database will be made available under the conditions described in the published SOMNIA protocol (van Gilst *et al* 2019). Specific restrictions apply to the availability of the data collected with sensors not comprised in the standard PSG set-up, such as wrist-worn PPG, since these sensors are used under license and are not publicly available. These data will become available however from the authors upon reasonable request and with permission of the licensors.

7. Ethical statement

The SOMNIA study met the ethical principles of the Declaration of Helsinki, the guidelines of Good Clinical Practice and the current legal requirements. The study was reviewed by the medical ethical committee of the Maxima Medical Center (Eindhoven, the Netherlands. File number N16.074). The protocol for data analysis was approved by the Medical Ethical Committee of the Kempenhaeghe Hospital and by the Philips Institutional Review Board. All participants provided informed consent.

Acknowledgment

P F declares to be employed by Philips Research. The employer had no influence on the study and on the decision to publish. G B P is a PhD student, fully employed by the Eindhoven University of Technology, with a guest status with Sleep Medicine Centre Kempenhaeghe and Philips Research, in order to have access to data and tools within the collaboration. J W M B is an academic advisor at Philips Research. The other authors declare that they have no competing interests.

This work has been done in the IMPULS framework of the Eindhoven MedTech Innovation Center (e/MTIC, incorporating Eindhoven University of Technology, Philips Research, and Sleep Medicine Centre Kempenhaeghe), including a PPS-supplement from the Dutch Ministry of Economic Affairs and Climate Policy. The work was financially supported by STW/IWT in the context of the OSA+ project (Grant No. 14 619). The authors would like to thank J P van Dijk, R Krijn, and B Hoondert for their contribution to the SOMNIA database.

ORCID iDs

Gabriele B Papini https://orcid.org/0000-0002-5752-9226
Pedro Fonseca https://orcid.org/0000-0003-2932-6402
Sebastiaan Overeem https://orcid.org/0000-0002-6445-9836

References

Addison P S 2016 Respiratory modulations in the photoplethysmogram (dpop) as a measure of respiratory effort *J. Clin. Monitor. Comput.* **30** 595–602
Addison P S 2017 Respiratory effort from the photoplethysmogram *Med. Eng. Phys.* **41** 9–18
Allen J 2007 Photoplethysmography and its application in clinical physiological measurement *Physiol. Meas.* **28** R1
Allen J and Murray A 1999 Modelling the relationship between peripheral blood pressure and blood volume pulses using linear and neural network system identification techniques *Physiol. Meas.* **20** 287
Allena M, Campus C, Morrone E, De Carli F, Garbarino S, Manfredi C, Sebastiano D R Ferrillo F 2009 Periodic limb movements both in non-REM and REM sleep: relationships between cerebral and autonomic activities *Clin. Neurophysiol.* **120** 1282–90
Barnhart H X, Haber M J Lin L I 2007 An overview on assessing agreement with continuous measurements *J. Biopharma. Stat.* **17** 529–69
Berry R B, Brooks R, Gamaldo C E, Harding S M, Marcus C and Vaughn B V *et al* 2015 *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications Version 2* 2nd edn (Darien, IL: American Academy of Sleep Medicine)
Bianchi M, Lipoma T, Darling C, Alameddine Y and Westover P M 2014 Automated sleep apnea quantification based on respiratory movement *Int. J. Med. Sci.* **11** 796
Bianchi M T 2018 Sleep devices: wearables and nearables, informational and interventional, consumer and clinical *Metabolism* **84** 99–108
Bonomi A and Westerterp K 2012 Advances in physical activity monitoring and lifestyle interventions in obesity: a review *Int. J. Obesity* **36** 167–77
Braun S R 1990 Respiratory rate and pattern *Clinical Methods: The History, Physical, and Laboratory Examinations* 3rd edn Butterworths
Buda A J, Pinsky M R, Ingels Jr N B, Daughters G T, Stinson E B and Alderman E L 1979 Effect of intrathoracic pressure on left ventricular performance *New Engl. J. Med.* **301** 453–9
Calvert J W Lefer D J 2012 Overview of cardiac muscle physiology *Muscle* (Amsterdam: Elsevier) chap 6 57–66
Carskadon M A and Rechtschaffen A 2011 *Normal Human Sleep 'Principles and Practice of Sleep Medicine'* 3rd edn (Amsterdam: Elsevier Saunders St. Louis) pp 16–26
Castaneda D, Esparza A, Ghamari M, Soltanpur C and Nazeran H 2018 A review on wearable photoplethysmography sensors and their potential future applications in health care *Int. J. Biosensors Bioelectron.* **4** 195
Chang H-H, Hsu C-C, Chen C-Y, Lee W-K, Hsu H-T, Shyu K-K, Yeh J-R, Lin P-J Lee P-L 2018 A method for respiration rate detection in wrist ppg signal using holo-hilbert spectrum *IEEE Sens. J.* **18** 7560–9
Charlton P H, Bonnici T, Tarassenko L, Clifton D A, Beale R and Watkinson P J 2016 An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram *Physiol. Meas.* **37** 610
Conover W J and Iman R L 1979 On multiple-comparisons procedures *Los Alamos Sci. Lab. Tech. Rep. LA-7677-MS* pp 1–14
De Zambotti M, Godino J G, Baker F C, Cheung J, Patrick K Colrain I M 2016 The boom in wearable technology: cause for alarm or just what is needed to better understand sleep? *Sleep* **39** 1761–2
Dehkordi P, Garde A, Molavi B, Ansermino J M Dumont G A 2018 Extracting instantaneous respiratory rate from multiple photoplethysmogram respiratory-induced variations *Front. Physiol.* **9** 948

Eerikäinen L M, Bonomi A G, Schipper F, Dekker L R, Vullings R, de Morree H M Aarts R M 2018 Comparison between electrocardiogram-and photoplethysmogram-derived features for atrial fibrillation detection in free-living conditions *Physiol. Meas.* **39** 084001
Fonseca P, Long X, Radha M, Haakma R, Aarts R M Rolink J 2015 Sleep stage classification with ECG and respiratory effort *Physiol. Meas.* **36** 2027
Fonseca P, Weysen T, Goelema M S, Mst E I, Radha M, Lunsingh Scheurleer C, van den Heuvel L Aarts R M 2017 Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults *Sleep* **40** zsx097
Foo J Y A, Wilson S J, Williams G R, Harris M-A Cooper D M 2005 Pulse transit time changes observed with different limb positions *Physiol. Meas.* **26** 1093
Gil E, Mendez M, Vergara J M, Cerutti S, Bianchi A M Laguna P 2008 Discrimination of sleep-apnea-related decreases in the amplitude fluctuations of PPG signal in children by HRV analysis *IEEE Trans. Biomed. Eng.* **56** 1005–14
Gori S, Ficca G, Giganti F, Di Nasso I, Murri L and Salzarulo P 2004 Body movements during night sleep in healthy elderly subjects and their relationships with sleep stages *Brain Res. Bull.* **63** 393–7
Gutierrez G *et al* 2016 Respiratory rate variability in sleeping adults without obstructive sleep apnea *Physiol. Rep.* **4** e12949
Hartmann V, Liu H, Chen F, Hong W, Hughes S Zheng D 2019 Towards accurate extraction of respiratory frequency from the photoplethysmogram: Effect of measurement site *Front. Physiol.* **10** 732
Hickey M, Phillips J P Kyriacou P A 2015 The effect of vascular changes on the photoplethysmographic signal at different hand elevations *Physiol. Meas.* **36** 425
Karlen W, Raman S, Ansermino J M Dumont G A 2013 Multiparameter respiratory rate estimation from the photoplethysmogram *IEEE Trans. Biomed. Eng.* **60** 1946–53
Karlen W, Turner M, Cooke E, Dumont G and Ansermino J M 2010 Capnobase: Signal database and tools to collect, share and annotate respiratory signals *Annual Meeting of the Society for Technology in Anesthesia (STA)* pp 25
Katz E S, Lutz J, Black C and Marcus C L 2003 Pulse transit time as a measure of arousal and respiratory effort in children with sleep-disordered breathing *Pediatric Res.* **53** 580–8
Khandoker A, Karmakar C K, Penzel T, Glos M and Palaniswami M 2013 Investigating relative respiratory effort signals during mixed sleep apnea using photoplethysmogram *Ann. Biomed. Eng.* **41** 2229–36
Kruskal W H and Wallis W A 1952 Use of ranks in one-criterion variance analysis *J. Am. Stat. Assoc.* **47** 583–621
Lazaro J, Gil E, Bailon R, Minchole A and Laguna P 2013 Deriving respiration from photoplethysmographic pulse width *Med. Biol. Eng. Comput.* **51** 233–42
Li Q, Li Q, Liu C, Shashikumar S P, Nemati S and Clifford G D 2018 Deep learning in the cross-time frequency domain for sleep staging from a single-lead electrocardiogram *Physiol. Meas.* **39** 124005
Long X, Fonseca P, Foussier J, Haakma R and Aarts R M 2013 Sleep and wake classification with actigraphy and respiratory effort using dynamic warping *IEEE J. Biomed. Health Inf.* **18** 1272–84
Long X, Foussier J, Fonseca P, Haakma R and Aarts R M 2014 Analyzing respiratory effort amplitude for automated sleep stage classification *Biomed. Signal Process. Control* **14** 197–205
Maeda Y, Sekine M and Tamura T 2011 The advantages of wearable green reflected photoplethysmography *J. Med. Syst.* **35** 829–34
Mann H B and Whitney D R 1947 On a test of whether one of two random variables is stochastically larger than the other *Ann. Math. Statist.* **18** 50–60
Mason J W, Ramseth D J, Chanter D O, Moon T E, Goodman D B Mendzelevski B 2007 Electrocardiographic reference ranges derived from 79,743 ambulatory subjects *J. Electrocardiol.* **40** 228–34
Nilsson L M 2013 Respiration signals from photoplethysmography *Anesthesia Analgesia* **117** 859–65
Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D and Tarassenko L 2014 Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring *IEEE J. Biomed. Health Inform.* **19** 832–8
Papini G B, Fonseca P, Aubert X L, Overeem S, Bergmans J W M Vullings R 2017 Photoplethysmography beat detection and pulse morphology quality assessment for signal reliability estimation *39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 117–20
Papini G B, Fonseca P, Eerikäinen L M, Overeem S, Bergmans J W Vullings R 2018 Sinus or not: a new beat detection algorithm based on a pulse morphology quality index to extract normal sinus rhythm beats from wrist-worn photoplethysmography recordings *Physiol. Meas.* **39** 115007
Papini G B, Fonseca P, van Gilst M M, van Dijk J P, Pevernagie D A, Bergmans J W, Vullings R and Overeem S 2019 Estimation of the apnea-hypopnea index in a heterogeneous sleep-disordered population using optimised cardiovascular features *Sci. Rep.* **9** 1–16
Park C and Lee B 2019 Energy-efficient photoplethysmogram compression to estimate heart and respiratory rates simultaneously *IEEE Access* **7** 71072–8
Penzel T, Kantelhardt J W, Grote L, Peter J-H Bunde A 2003 Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea *IEEE Trans. Biomed. Eng.* **50** 1143–51
Penzel T, Schöbel C and Fietze I 2018 New technology to assess sleep apnea: wearables, smartphones, and accessories *F1000Research* **7**
Radha M *et al* 2019 Estimating blood pressure trends and the nocturnal dip from photoplethysmography *Physiol. Meas.* **40** 025006
Rajala S, Lindholm H and Taipalus T 2018 Comparison of photoplethysmogram measured from wrist and finger and the effect of measurement location on pulse arrival time *Physiol. Meas.* **39** 075010
Reisner A, Shaltis P A, McCombie D and Asada H H 2008 Utility of the photoplethysmogram in circulatory monitoring *Anesthesiol: The J. Am. Soc. Anesthesiologists* **108** 950–8
Renevey P, Delgado-Gonzalo R, Lemkaddem A, Verjus C, Combertaldi S, Leeners B, Dammeier F *et al* 2018 Respiratory and cardiac monitoring at night using a wrist wearable optical system *40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* pp 2861–4
Sadr N and de Chazal P 2019 A comparison of three ECG-derived respiration methods for sleep apnoea detection *Biomed. Phys. Eng. Express* **5** 025027
Shapiro S S and Wilk M B 1965 An analysis of variance test for normality (complete samples) *Biometrika* **52** 591–611
Shelley K H 2007 Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate *Anesthesia Analgesia* **105** S31–S36
Somers V K, Dyken M E, Mark A L Abboud F M 1993 Sympathetic-nerve activity during sleep in normal subjects *New Engl. J. Med.* **328** 303–7
Tamisier R, Weiss J W Pépin J L 2018 Sleep biology updates: hemodynamic and autonomic control in sleep disorders *Metabolism* **84** 3–10
Tamura T, Maeda Y, Sekine M and Yoshida M 2014 Wearable photoplethysmographic sensors-past and present *Electronics* **3** 282–302

Thomas R J 2003 Arousals in sleep-disordered breathing: patterns and implications *Sleep* **26** 1042–7
Tobaldini E, Nobili L, Strada S, Casali K R, Braghiroli A and Montano N 2013 Heart rate variability in normal and pathological sleep *Front. Physiol.* **4** 294
van Gilst M M *et al* 2019 Protocol of the somnia project: an observational study to create a neurophysiological database for advanced clinical sleep monitoring *BMJ Open* **9** e030996
Vandenbussche N L, Overeem S, van Dijk J P, Simons P J Pevernagie D A 2015 Assessment of respiratory effort during sleep: esophageal pressure versus noninvasive monitoring techniques *Sleep Med. Rev.* **24** 28–36
Varon C, Caicedo A, Testelmans D, Buyse B and Van Huffel S 2015 A novel algorithm for the automatic detection of sleep apnea from single-lead ECG *IEEE Trans. Biomed. Eng.* **62** 2269–78
Weiss J W, Remsburg S, Garpestad E, Ringler J, Sparrow D and Parker J A 1996 Hemodynamic consequences of obstructive sleep apnea *Sleep* **19** 388–97
Widjaja D, Varon C, Dorado A, Suykens J A Van Huffel S 2012 Application of kernel principal component analysis for single-lead-ECG derived respiration *IEEE Trans. Biomed. Eng.* **59** 1169–76
Wilde-Frenz J and Schulz H 1983 Rate and distribution of body movements during sleep in humans *Perceptual Motor Skills* **56** 275–83
Zheng Y, Poon C C, Yan B P Lau J Y 2016 Pulse arrival time based cuff-less and 24-h wearable blood pressure monitoring and its diagnostic value in hypertension *J. Med. Syst.* **40** 195