Signal-Quality Indices for the Electrocardiogram and Photoplethysmogram: Derivation and Applications to Wireless Monitoring

Christina Orphanidou, Timothy Bonnici, Peter Charlton, David Clifton, David Vallance, and Lionel Tarassenko

***Abstract*—The identification of invalid data in recordings obtained using wearable sensors is of particular importance since data obtained from mobile patients is, in general, noisier than data obtained from nonmobile patients. In this paper, we present a signal quality index (SQI), which is intended to assess whether reliable heart rates (HRs) can be obtained from electrocardiogram (ECG) and photoplethysmogram (PPG) signals collected using wearable sensors. The algorithms were validated on manually labeled data. Sensitivities and specificities of 94% and 97% were achieved for the ECG and 91% and 95% for the PPG. Additionally, we propose two applications of the SQI. First, we demonstrate that, by using the SQI as a trigger for a power-saving strategy, it is possible to reduce the recording time by up to 94% for the ECG and 93% for the PPG with only minimal loss of valid vital-sign data. Second, we demonstrate how an SQI can be used to reduce the error in the estimation of respiratory rate (RR) from the PPG. The performance of the two applications was assessed on data collected from a clinical study on hospital patients who were able to walk unassisted.**

***Index Terms*—Battery life, electrocardiogram (ECG), photoplethysmogram (PPG), respiratory rate (RR), signal quality, telemonitoring, wearable sensors.**

I. Introduction

THE majority of telemonitoring methods proposed in the literature have still not found their way into widespread clinical use [1]. Challenges faced in the field of wearable sensor design include minimizing the weight and the size of the systems to increase comfort, mitigating against the corruption of recorded signal by motion artifact, and maximizing battery life while maintaining wearability [2].

Manuscript received December 18, 2013; revised April 27, 2014; accepted June 11, 2014. Date of publication July 23, 2014; date of current version May 7, 2015.

C. Orphanidou is with (OR) Oxygen Research Ltd, Limassol 3075, Cyprus, and also with the Institute of Biomedical Engineering, University of Oxford, Oxford, OX3 7BN, U.K. (e-mail: christina@oxygen-research.com).

T. Bonnici is with the Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, U.K. (e-mail: timothy.bonnici@ndm.ox.ac.uk).

P. Charlton is with King's College London, King's Health Partners, London, SE1 9RT, U.K. (e-mail: Peter.Charlton@gstt.nhs.uk).

D. Clifton and L. Tarassenko are with the Institute of Biomedical Engineering, University of Oxford, Oxford, OX3 7DQ, U.K. (e-mail: davidc@robots.ox.ac.uk; lionel@robots.ox.ac.uk).

D. Vallance is with the Nuffield Department of Surgery, John Radcliffe Hospital Headington, University of Oxford, Oxford, OX3 9DU, U.K. (e-mail: david.vallance@nds.ox.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2338351

Compared with conventional bedside monitoring, as used in intensive care units, the issue of identifying invalid data is of particular importance for wearable sensors, as data obtained from ambulatory patients, the patients most likely to benefit from the use of wearable sensors, are more likely to contain artifact than data from bed-bound patients [1]. Artifact-corrupted signal has been shown to lead to a large number of false alarms which can lead to the phenomenon of "alarm fatigue" whereby ward staff become desensitized to and ultimately ignore alerts from the system [3]. The Joint Commission, the body that accredits U.S. healthcare institutions, has recently issued a Sentinel Alert, highlighting the problem of alarm fatigue [4] and the ECRI Institute has listed alarm hazards in their annual *Top 10 Health Technology Hazards* report every year since 2010 [5]. A large number of artifact detection (AD) algorithms have been developed with the purpose of assessing the quality of physiological data. Recent publications propose algorithms for deriving signal quality indices (SQIs) either based on a single channel of physiological data or by fusing information from several channels. A comprehensive review of the development and utility of AD algorithms in critical care units (CCUs) was recently published in [6], which reviewed 80 AD approaches published between 1989 and 2012. The authors conclude that currently published AD techniques are highly specific to a particular clinical setting and require modification for validation and reuse in a different one. Most algorithms are hard coded to monitor specific data types and frequencies, which may limit their use. Furthermore, the fact that most commercially available monitors used in the reviewed studies have undisclosed built-in preprocessing algorithms imparts an unknown bias to the output of the AD algorithms.

In this paper, we propose an approach to automatically qualifying segments of electrocardiogram (ECG) and photoplethysmogram (PPG) collected from ambulatory patients via wearable sensors, using an SQI, extending work initially presented in [7]. The proposed algorithms are intended to provide real-time assessment of the suitability of ECG and PPG signals for deriving reliable heart rates (HRs). In order to make our algorithm usable in a range of clinical environments, it was developed and evaluated on data from different clinical settings and monitors. The output of the algorithm is given in a binary format, "good" (i.e., a reliable HR can be derived) and "bad" (i.e., a reliable HR cannot be derived) to simplify interpretation and facilitate applicability. In addition, we demonstrate the utility of the proposed SQI for improving the performance of wearable sensors by proposing strategies for: 1) power

saving in order to extend the battery lifetime of the sensors and 2) improving the estimation of respiratory rate (RR) from the PPG signal.

The issue of power consumption is a major obstacle to the wider clinical adoption of wearable sensors mainly due to the fact that signal processing algorithms are usually very computationally intensive. One approach for limiting the energy consumption of a sensor is to duty cycle, i.e., to turn off components of the device periodically [8]. Duty cycling strategies can be data-independent, where input data are not taken into consideration and a device is turned on or off after a predetermined time-period has elapsed, or data-dependent, where decisions to turn on or off components of the device are based on identifying a specific pattern in the input data [8]. To illustrate our approach, we have proposed an application where a single reliable HR value (estimated from a 10s segment of "good" data) is required in every 5-min window of data. While this rate is smaller than the desired continuous recording promised by wearable sensor developers, the majority of commercially available wearable sensors are still not capable of continuous recording and transmission of ECG and PPG over a 24 h period mainly due to short battery-life limitations [9]–[11]. Using today's technology one must accept that recording will be intermittent if nursing staff are not to be overburdened with recharging or changing batteries. In most clinical scenarios where a wearable sensor might be used (long-term monitoring of patients in their home environment or surveillance of patients on general hospital wards [2]), measuring HR every 5 minutes is far in excess of the current standard of care (expert consensus is that patients on general wards should be monitored every 6–12 hours for patients with normal vital signs [12] and hourly for patients with all but the most severely deranged vital signs [13]). Once a single segment of "good" data is identified, the device is turned off and is reactivated in time to take another reliable value in the next 5-min window. By implementing the proposed approach retrospectively on recordings obtained using wearable sensors, our aim is to study whether the saving in power consumption is associated with a loss of reliable vital-sign information.

The inclusion of quality metrics for improving the estimation of RR has received some attention recently with promising results [14], [15]. Methods proposed recently rely on data fusion of either multiple signals [14] or different sources of respiratory information from the PPG [15]. While the methods demonstrated a significant improvement in estimating RR, they either require the presence of additional signals or multiple channels of the same signal, which might not always be available in a telemonitoring setting. Furthermore, most approaches in the literature were not validated on data obtained from ambulatory patients but from static patients. Our proposed approach uses the SQI to improve the RR estimation using data from a single channel of PPG. When the SQI identifies a segment of PPG data as "bad," the RR derived from this segment is considered unreliable and is ignored. Our intention is, first, to demonstrate that by restricting the amount of artifact present in a segment of PPG signal we are able to reduce the error in the estimation of RR and, second, to associate any reduction in the error with any decrease in the frequency of obtaining an estimate.

II. Methods

A. Databases

For this study, ECG and PPG records taken from two different studies were used: the Physionet/Computing in Cardiology Challenge 2011 [16] and a clinical study on ambulatory hospital patients using wearable sensors, carried out in the John Radcliffe (JR) hospital in Oxford [9].

The Sana/Physionet database (denoted PCinC), made available freely as part of the Physionet/Computing in Cardiology Challenge 2011 [16], consisted of 1500 segments of standard 12-lead ECGs, sampled at 500 Hz with a 16-bit resolution, collected using conventional ECG monitors. For this study, only the first lead of ECG data was used.

The second database (denoted JR) was collected as part of a feasibility study investigating the suitability of commercially available wearable sensors for clinical use [9] and contains ECG, PPG, and respiratory inductance plethysmography (RIP) recordings. The patient population consisted of adult patients recruited from the acute medical, acute surgical, and care of the elderly wards at the John Radcliffe Hospital in Oxford who were able to move around unassisted. An ambulatory score ranging from 1 to 5 was recorded for each patient, with 1 being bed-bound and 5 being able to mobilize independently. The range of ambulatory score was 1–5 for the participating patients with a mean of 4.5 and a median of 5 [9]. Recordings were taken from three of the monitors used in this study: the Dyna-Vision DVM012S (RS-TechMedic, Langedijk, Netherlands), the Equivital EQ-02 LifeMonitor (Hidalgo, Swavesey, UK), and the Wrist Ox_2 3150 (Nonin Medical Inc., Plymouth, MN, USA). In all cases, soft silicone finger probes were used with the PPG sensors.

The Dyna-Vision DVM012S samples the ECG at 100 Hz with 16-bit resolution using conventional ECG leads attached to adhesive wet gel electrodes. The equipment also contains transducer circuitry, manufactured by Nonin Medical Inc., for recording PPG at 75 Hz with a 12-bit resolution.

The Equivital EQ-02 attaches to a belt worn around the patient's chest. ECG is sampled at 256 Hz and the RIP signal at 25.6 Hz. Additional transducers may be attached to the Equivital EQ-02. In this study, XPOD 3012LP modules (Nonin Medical Inc.) were used to sample PPG at 75 Hz, with 16-bit resolution.

The Wrist Ox_2 3150 is a pulse oximeter worn on the wrist. It also samples the PPG at 75Hz, with 6-bit resolution. Given the differing capabilities of the monitors, it was not possible to capture all of the signals from all of the monitors.

ECG recordings were collected from 18 patients. Of these, 11 records (denoted $\mathrm{JR_{D-ECG}}$) were captured using RS TechMedic Dyna-Vision DVM012S monitors (123 h of recording) and 7 (denoted $\mathrm{JR_{E-ECG}}$) were captured using Equivital EQ-02 LifeMonitors (63 h of recording).

PPG recordings were collected from 7 patients using XPOD 3012LP transducers connected to the Equivital monitors

(denoted $\mathrm{JR_{E-PPG}}$) (63 h of recording) and from 12 patients using Nonin Wrist $\mathrm{Ox_2}$ 3150 pulse oximeters (denoted $\mathrm{JR_{N-PPG}}$) (131 h of recording).

RIP recordings were only gathered from the 7 patients who wore the Equivital monitors (denoted $\mathrm{JR_{E-RESP}}$) (63 h of recording).

B. Development of SQIs

The SQI we propose is intrinsically linked to the ECG and the PPG peak detectors used since an assessment of their performance is used in the development of the SQI. For R-peak detection, we used the widely accepted Hamilton and Tompkins algorithm [17], and for PPG pulse-peak detection, we used a three-point peak detector with a set of empirically determined thresholds.

To develop the ECG SQI, we used 1500 10-second segments of ECG, comprising 500 segments randomly drawn from each one of the PCinC, $\mathrm{JR_{D-ECG}}$, and $\mathrm{JR_{E-ECG}}$ databases. For the development of the PPG SQI, we used 1500 10-second segments of PPG, comprising 750 segments randomly drawn from each one of the $\mathrm{JR_{E-PPG}}$ and $\mathrm{JR_{N-PPG}}$ databases. All 3000 segments were manually labelled.

1) Labeling of ECG and PPG Signals: Labeling was carried out in three stages. In the first stage, two assessors categorized the 3000 samples based on the following rule: "An ECG or PPG segment is labeled as 'clinically usable' if a human expert can confidently derive a reliable HR from it, by counting the number of salient features (such as R-peaks or PPG pulse peaks) over fixed time intervals. Otherwise, it is labeled as 'clinically unusable." When there was disagreement between the two assessors, a third assessor reviewed the ambiguous samples and gave the decisive label. 991 ECG samples (66%) were ranked as "clinically usable," and 509 (34%) were ranked as "clinically unusable." In the case of the PPG, 828 samples (55.2%) were ranked as "clinically usable" and 672 (44.8%) as "clinically unusable." The third assessor reviewed 12.2% of the ECG data and 13.6% of the PPG data.

In the second labeling stage, the R-peaks and PPG pulse peaks of the "clinically usable" samples were manually identified by the two human experts using a custom annotation GUI written in MATLAB (MathWorks, Natick, MA, USA). As in the first stage of labeling, the third assessor reviewed annotated samples for which annotations of the first two assessors differed. The third annotation was then taken to be the correct one. 103 ECG samples (10.4%) and 127 PPG samples (15.3%) had to be reviewed by the third assessor.

The third labeling stage involved comparing the output of the R-peak and PPG pulse-peak detectors applied to the "clinically usable" ECG and PPG samples with the annotations obtained from the second stage of labeling. As a result of this process, the samples had a second label applied to them according to the following rule: "An ECG/PPG segment is labeled as being 'bad' if more than one R-peak/PPG pulse-peak is missed by the R-peak/PPG pulse-peak detector, or if more than one instance of artifact is mistakenly identified as an R-peak/PPG pulse-peak by the R-peak/PPG pulse-peak detector. Otherwise, it is labeled as 'good'."

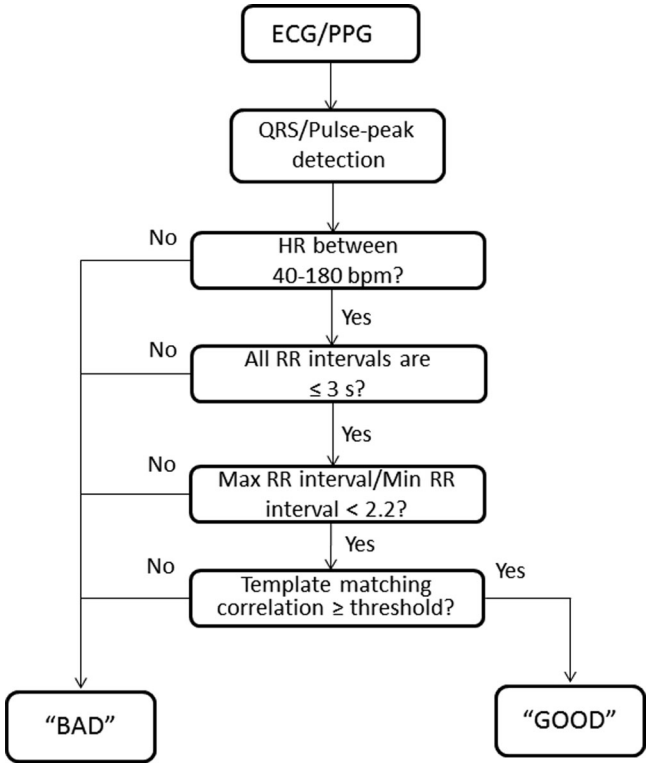

Fig. 1. Flowchart of SQI algorithm.

This rule is deliberately conservative: if more than one R-peak/PPG pulse-peak in a 10-s sample is missed, or more than one noisy peak is mistakenly identified as an R-peak/PPG pulse peak, the HR value ultimately derived from this sample will have a large error.

2) SQI Algorithm: A flowchart of the SQI algorithm is shown in Fig. 1. The different steps are explained in detail in the following section.

a) Feasibility rules: The first step of the SQI algorithm is to perform R-peak/PPG pulse-peak detection on a sample and to compare the output of the detector with a set of physiologically relevant rules. The following three rules are applied sequentially, and if any is not satisfied, the sample is classified as "bad."

1) **Rule 1:** The HR extrapolated from the 10-s sample must be between 40 and 180 beats per minute (bpm). (Though it is theoretically possible to have HRs outside of these values, this is the physiologically probable range of HR for the adult population likely to use wearable sensors.)
2) **Rule 2:** The maximum acceptable gap between successive R-peaks/PPG pulse-peaks is 3 s. (This rule ensures no more than one beat is missed.)
3) **Rule 3:** The ratio of the maximum beat-to-beat interval to the minimum beat-to-beat interval within the sample should be less than 2.2. (This is a conservative limit since we would not expect the HR to change by more than 10% in a 10-s sample. We use a limit of 2.2 to allow for a single missed beat.)

If all the three rules are satisfied, an adaptive QRS/PPG pulse-waveform template matching approach is used, as explained next.

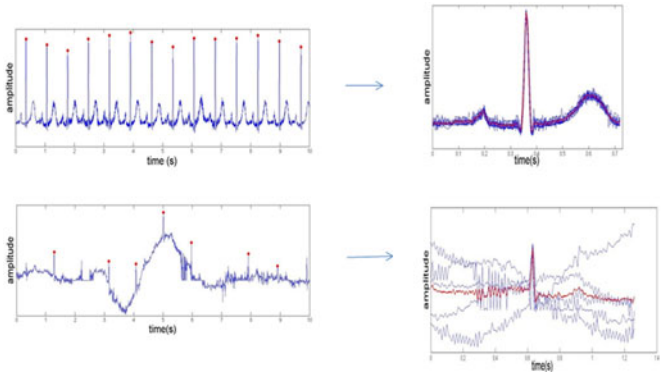

Fig. 2. QRS templates for "good" ECG (top) and "bad" ECG (bottom) taken from the test set. The "good" ECG has very regular QRS complexes (shown on the top right superimposed on each other) resulting in a high average correlation with the template. The "bad" ECG has irregular QRS complexes (shown on the bottom right) resulting in a low-average correlation with the template.

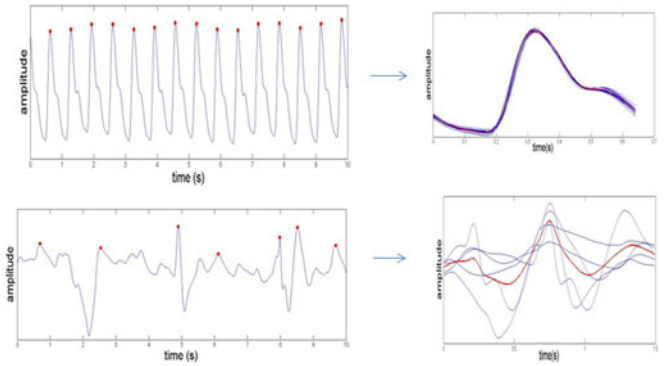

Fig. 3. PPG pulse-waveform templates for "good" PPG segment (top) and "bad" PPG segment (bottom) taken from the test set. The "good" PPG has very regular pulse-waveforms (shown on the top right superimposed on each other) resulting in a high average correlation with the PPG pulse-wave template. The "bad" PPG has irregular pulse waveforms (shown on the bottom right) resulting in a low-average correlation with the PPG pulse-wave template.

b) Adaptive template matching: Template-matching approaches have been used in the past for identifying ventricular ectopic beats [18] and heartbeats [19] in the ECG and for signal quality assessment of the PPG [20]. Regardless of the actual morphology of the QRS complexes or PPG pulse waveforms in a given ECG or PPG sample, template matching searches for regularity in a segment, which is an indicator of reliability (since a segment contaminated by artifact would be irregular in morphology). Our approach is as follows:

1) Using all the detected R-peaks/PPG-pulse peaks of each ECG/PPG sample, the median beat-to-beat interval is calculated.
2) Individual QRS complexes/PPG pulse waves are then extracted by taking a window, the width of which is the median beat-to-beat interval, centered on each detected R-peak/PPG-pulse peak.
3) The average QRS template is then obtained by taking the mean of all QRS complexes in the sample. Similarly, the mean PPG pulse-wave template is obtained by taking the mean of all PPG pulse waves in the sample. The correlation coefficient of each individual QRS complex with the average QRS template is then calculated. Similarly, the correlation coefficient of each individual PPG pulse wave with the average PPG pulse-wave template is calculated.
4) The average correlation coefficient is finally obtained by averaging all correlation coefficients over the whole ECG/PPG sample.

Figs. 2 and 3 show examples of the average QRS complex and PPG pulse wave, respectively, template creation from morphologically regular and irregular samples of signal.

c) Training, validation, and assessment: In the training stage of algorithm development, the value of the average correlation coefficient was optimized in order to determine the threshold, which gives the optimum cutoff between "good" and "bad" segments. We used 1000 of the 1500 ECG samples (chosen at random) to optimize the threshold of the average correlation coefficient. We tested values of the average correlation coefficient ranging from 0.5 to 1 and averaged the performance of the algorithm over five cycles of cross validation in order to determine the threshold which best matched the manual classification. The remaining 500 of the 1500 samples were used in the testing stage. If the average correlation coefficient of a given test sample was lower than the determined threshold, it was classified as "bad," and if it was higher than the determined threshold, it was classified as "good." The performance of the algorithm was then assessed by calculating the sensitivity (proportion of "bad" signal correctly identified as such) and specificity (proportion of "good" signal correctly identified as such).

C. Applications

1) Power-Saving Strategies: We investigated two power-saving strategies in order to evaluate the effect of duty cycling on the number of reliable HR estimates we obtain. The power consumption was approximated as being proportional to recording time, which is a reasonable assumption particularly in the case of the PPG where the central processing unit (CPU) consumes minimal power compared with the pulse oximeter LEDs and transmitter [21]. We calculated the saving in recording time achieved by following one or other of the two strategies described below and examined it against the number of reliable HR estimates obtained. To illustrate this application, we have postulated a scenario in which clinicians applying the monitor to patients wish to record an accurate HR value every 300 s (5 min) and that deriving the HR from 10 s of "good" data is acceptable.

The two proposed duty-cycling recording strategies are:

Strategy A: Stop recording as soon as a reliable segment of signal is found. If no reliable segment is obtained after 150 s of continuous recording, stop recording for a further 150 s. If a reliable segment is obtained, stop as soon as the segment occurs and do not record for the remainder of 300 s. (If, for instance, a reliable segment is first found within the 50–60 s window, stop recording for 240 s.) Start recording again (see Fig. 4).

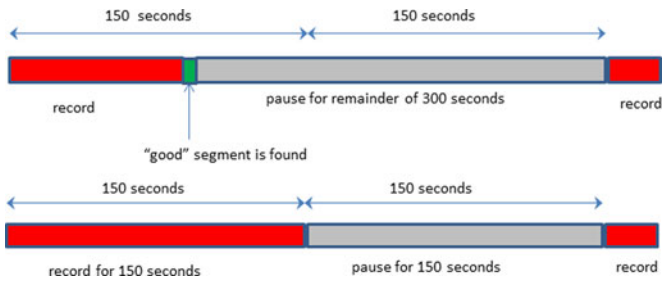

Fig. 4. Strategy A. This strategy works in 300 s cycles. If a reliable segment is recorded in the first 150 s of recording, we stop as soon as the segment is recorded and pause for the remainder of 300 s (top diagram). If no reliable segment is obtained in the first 150 s of recording we stop recording and pause for 150 s (bottom diagram). (If, for instance, a reliable segment is first found within the 50–60 s window, stop recording for 240 s.) At the end of the 300-s cycle, we start recording again.

Fig. 5. Strategy B. As soon as a "good" segment of data is recorded, we pause for 290 s. At the end of the 290 s, we start recording again.

Strategy B: Stop recording as soon as a reliable ECG or PPG segment occurs. Do not record for 290 s. Start recording again after a 290 s gap (see Fig. 5).

We tested the strategies retrospectively on the entire $\mathrm{JR_{D-ECG}}$, $\mathrm{JR_{E-ECG}}$, $\mathrm{JR_{E-PPG}}$, and $\mathrm{JR_{N-PPG}}$ databases and calculated the recording time and the number of 5 min windows in which an accurate HR could be derived. The latter statistic was calculated by partitioning the recordings in nonoverlapping 5-min windows and then determining the total number of 5-min windows that contained *at least one* reliable 10-s segment of ECG or PPG data that was selected by the strategy under consideration. The SQI was calculated for 10-s windows, advanced by 1 s.

2) Improving the Estimation of RR From PPG signals: For the development of this application, the $\mathrm{JR_{E-RESP}}$ data were used in order to obtain a "gold standard" RR. Since both the $\mathrm{JR_{E-ECG}}$ and the $\mathrm{JR_{E-RESP}}$ data were collected using the belt worn around the chest, the locations of artifact in the two databases were highly correlated. We therefore excluded the $\mathrm{JR_{E-ECG}}$ database and used only the $\mathrm{JR_{E-PPG}}$ database, which, having been collected using a finger probe, would allow for a fairer comparison. The algorithm we used for estimating RR is based on the data fusion algorithm presented in [22]. While the algorithm presented in [22] was designed for ECG signals, we extended it to the PPG signal, the difference being the use of a PPG pulse-peak detector instead of a QRS detector. The RR estimation algorithm uses 60-s windows of data; therefore, we divided our test recordings into 60-s segments and classified the quality of each segment based on the duration of *continuous* "good" signal present in the window. We advanced windows by 10 s throughout the recording, which resulted in having a classification every 10 s. The 60-s segments were divided into the following classes, listed in terms of increasing signal quality: $\leq$60 s continuous artifact (i.e., all data), $\leq$40 s continuous artifact, $\leq$20 s continuous artifact, or no artifact. We calculated a RR from the PPG and the respiration signal (gold standard) and calculated the mean absolute error (MAE) and standard deviation over the whole database.

For each class of signal quality, we calculated the accuracy of the RR estimation and the frequency with which those estimates could be obtained.

III. Results

A. Classification Performance of SQI

The optimum threshold for the average correlation coefficient was found to be 0.66 for the ECG SQI and 0.86 for the PPG SQI. On the testing set, this corresponded to a sensitivity of 94% and specificity of 97% for the ECG SQI and a sensitivity of 91% and specificity of 95% for the PPG SQI. The ECG SQI matched, or was more conservative than, the manual annotation 99% of time, while the PPG SQI matched, or was more conservative than, the manual annotation 97% of the time.

B. Performance of Power-Saving Strategies

Table I shows the results of the retrospective application of the two proposed power-saving strategies versus continuous recording for the two ECG databases and two PPG databases it was tested on. It is worth mentioning that continuous recording did not result in a reliable HR being obtained in 100% of 5-min windows because recordings from ambulatory patients often had long periods of artifact. Specifically, in the $\mathrm{JR_{D-ECG}}$ database, a reliable HR was obtained in 77% of 5-min windows and in the $\mathrm{JR_{E-ECG}}$, a reliable HR was obtained in 94% of 5-min windows. For the PPG, in the $\mathrm{JR_{N-PPG}}$ database, continuous recording resulted in obtaining a reliable HR in 71% of 5-min windows and in the $\mathrm{JR_{E-PPG}}$ database, in 97% of 5-min windows.

As can be seen in Table I, the two proposed power-saving strategies significantly reduce the recording time with only minimal losses in terms of the number of windows for which reliable HR values are obtained. Strategy A, in particular, results in a reduction in recording time ranging from 79% to 94% for the four databases used with a loss in the percentage of the valid HR values obtained ranging from 4% to 5%, which is not likely to be clinically significant in the majority of cases when monitoring ambulatory patients.

Strategy B had marginally higher percentage of HR values obtained but less recording time saved. The percentage of recording time saved ranged from 61% to 89% for the four databases used. The differences in the performance of the two strategies compared to continuous recording were due to the difference in the quality of the signal recorded by each one of the different sensors. Important to note is that while the percentage of HRs missed is more or less the same for all databases, the reduction in recording time is quite varied, depending on the database (but significant for all).

TABLE I
PERCENTAGE RECORDING TIME AND PERCENTAGE OF HRS (COMPARED WITH CONTINUOUS RECORDING) OBTAINED FOR THE PROPOSED POWER-SAVING STRATEGIES

	% HRs Obtained			% Recording Time		
	Continuous	Strategy A	Strategy B	Continuous	Strategy A	Strategy B
JR_{D-ECG}	77	72	73	100	18	30
JR_{E-ECG}	94	90	91	100	6	11
JR_{N-PPG}	71	67	68	100	21	39
JR_{E-PPG}	97	93	92	100	7	11

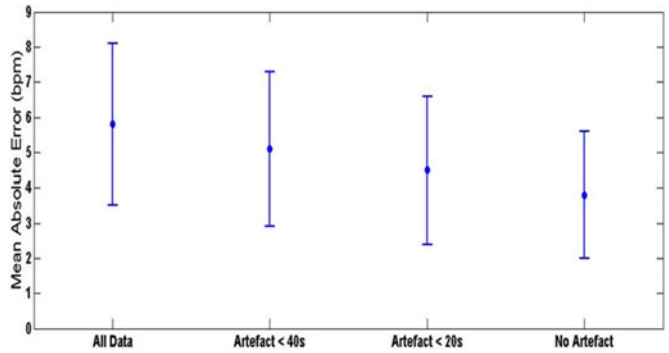

Fig. 6. Error-bar plot of MAEs (in bpm) in the estimation of RR for segments of data containing different durations of artifact (cumulative). As is evident, the MAE decreases with the duration of artifact present in the PPG.

TABLE II
NUMBER OF RRS OBTAINED FROM SEGMENTS WITH DIFFERENT DURATION OF ARTIFACT

Duration of Continuous Artifact	Number of RRs obtained
0s	6953 (31%)
≤20 s	12 064 (54%)
≤40 s	15 849 (71%)
≤60 s	22 172 (100%)

C. Improving the Estimation of RR From the PPG Signal

Fig. 6 shows the MAEs and standard deviations of errors in windows with different classes of signal quality. As expected, the accuracy in the estimation of RR is inversely proportional to the amount of artifact contained in the signal.

Table II shows the number of RRs obtained for each class of PPG signal quality. As expected, the number of valid RR estimates obtained decreases as the signal quality (expressed in terms of the duration of continuous artifact permitted) increases.

IV. DISCUSSION AND CONCLUSION

We have presented an SQI for the ECG and the PPG, which is intended to provide real-time assessment of the suitability of ECG and PPG signals for deriving reliable HR values. We evaluated the SQI on data from a range of wearable physiological sensors and achieved sensitivities and specificities of 94% and 97% for the ECG and 91% and 95% for the PPG. The proposed SQI is intrinsically linked to the peak detector used meaning that using other peak detectors would change the performance of the SQI. Comparing the performance of different peak detectors was beyond the scope of this study so we proposed our SQI algorithm using a well-known and widely accepted R-peak detector and a PPG-pulse peak detector which we found to have superior performance compared to other algorithms considered. Yet, during the labeling process we noticed that for a small percentage of segments the peak detectors would still underperform. This was mostly noticed in the case of the PPG signal which is more morphologically variable than the ECG signal. Better results would thus be expected if the PPG pulse-peak detection algorithm was improved (which would however require retraining of the algorithm).

We then proceeded to propose two applications of the SQI in order to improve the performance of wearable physiological monitoring systems. We first proposed a power-saving application utilizing the SQI in order to increase the battery lifetime of wearable monitors. Our proposed approach assumes that a single reliable HR value is required in every 5 min of ECG or PPG recording.

In particular, the device is turned off after a "good" segment of data is obtained and then on again after a predetermined period of time has elapsed. We presented two example strategies of this approach, which we applied retrospectively to recordings from ambulatory hospital patients and demonstrated that it is possible to significantly reduce the power consumption (by reducing the recording time) by up to 94% for the ECG and 93% for the PPG with a loss of reliable HRs obtained in the range of 4%–5%.

The proposed approach is based on the assumption that a single reliable HR measurement every 5 min is sufficient. While this frequency of HR measurement is far in excess of the current standard of care, there is still the danger of patient deterioration being missed because of intermittent recording. To alleviate this danger, the proposed strategies could be used as part of an intelligent scheme, in which, for the vast majority of the time during which HR (and perhaps all other vital signs) are normal, acquiring a single value every 5 min would be sufficient. However, if the HR value (or other vital signs) were indicative of patient abnormality, then the sampling rate could be increased, at the expense of battery life. This way, the risk of missed deterioration because of intermittent recording would be minimized. On the other hand, to preserve battery life, instead of using a binary ("good" or "bad") classification scheme, it would be possible to incorporate a fuzzy logic approach whereby measures of "goodness" are assigned to different segments of signal (based

on a predecided metric) and the cutoff between "good" and "bad" is lowered if the length of recording exceeds a predecided duration. The acuity of the patient population would always be the deciding factor on the scheme chosen so that the right balance is obtained between safe and accurate monitoring and battery preservation.

The second application involved using the SQI in order to improve the estimation of RR from PPG signals. We examined how the amount of artifact present in a window of data affected the performance of an RR estimation algorithm. Through this approach we demonstrated a flexible scheme, with a tradeoff between how often an estimate is required (which will depend on the acuity of the patient's health) and the reliability of the estimate. Similarly to the power-saving approach, this approach could be further extended by applying a fuzzy logic classification scheme and instead of varying only the duration of artifact present in a given segment of data, also vary the degree of "goodness" of a given segment in order to obtain the optimal tradeoff with how often an estimate is required and the reliability of the estimate. In any case, even if we apply strict quality restrictions and, hence, cause estimates to be made infrequently, the frequency will still be higher than the frequency of observations from a ward nurse (which may be only once or twice per day at low-acuity wards).

Our proposed SQI shows promising results in differentiating between "good" and "bad" segments of data obtained from ambulatory hospital patients using a range of wearable monitoring systems. Our approach concerns clinical settings where a single signal would be available. In settings where multiple signals are present, it is possible to further improve estimation of vital signs by fusing information from the different sources based on their respective quality indices as much recent research has demonstrated.

A limitation of our proposed method may be that in its current form, the algorithm is unlikely to perform well on patients exhibiting atrial fibrillation (AF), so further work is required in order to add flexibility to the algorithm.

Despite this limitation, the promising classification performance of the proposed SQI along with the significant reduction in power consumption demonstrated by the proposed power-saving strategies are steps toward integrating wearable sensing for patient monitoring into wider clinical practice.

REFERENCES

[1] G. Clifford and D. Clifton, "Annual review: Wireless technology in disease management and medicine," *Ann. Review Med.*, vol. 63, pp. 479–492, 2012.

[2] V. Nangalia, D. Prytherch, and G. Smith, "Health technology assessment review: Remote monitoring of vital signs-current status and future challenges," *Crit. Care*, vol. 14, no. 5, pp. 1–8, 2010.

[3] C. L. Tsien and J. C. Fackler, "Poor prognosis for existing monitors in the intensive care unit," *Crit. Care Med.*, vol. 25, pp. 614–619, 1997.

[4] The Joint Commission. (2013, Apr.), *Sentinel event alert issue 50: Medical device alarm safety in hospitals* [Online]. Available: http://www.jointcommission.org/sea_issue_50/

[5] ECRI. (2013). *Alarm safety resource site: Guidance and tools to help healthcare facilities improve alarm safety* [Online]. Available: https://www.ecri.org/Forms/Pages/Alarm_Safety_Resource.aspx

[6] S. Nizami, J. R. Green, and C. McGregor, "Implementation of artifact detection in critical care: A methodological review," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 127–142, Jan. 2013.

[7] C. Orphanidou, T. Bonnici, D. Vallance, A. Darrell, P. Charlton, and L. Tarassenko, "A method for assessing the reliability of heart rates obtained from ambulatory ECG," in Proc. BIBE, 2012, pp. 193–196.

[8] E. I. Shih, "Reducing the computational demands of medical monitoring classifiers by examining less data," Ph.D. dissertation, Dept. Elect. Eng. Comp. Sci., Mass. Inst. Technol., Cambridge, USA, 2010.

[9] T. Bonnici, C. Orphanidou, D. Vallance, A. Darrell, and L. Tarassenko, "Testing of wearable monitors in a real-world hospital environment: What lessons can be learnt?" in Proc. BSN, 2012, pp. 79–84.

[10] Philips. *M4840A/M4841A Technical Data Sheet* [Online]. Available: http://media.supplychain.nhs.uk/media/documents/N0889206/ Specification/31473_N0889206%20Intellivue%20Telemetry%20Datasheet.pdf

[11] Visi-Mobile. *FAQs: How is the Visi Mobile powered?* [Online]. Available: http://www.visimobile.com/visi-product-info/faqs/

[12] M. A. DeVita, G. B. Smith, S. K. Adam, I. Adams-Pizarro, M. Buist, R. Bellomo, R. Bonello, E. Cerchiari, B. Farlow, D. Goldsmith, H. Haskell, K. Hillman, M. Howell, M. Hravnak, E. A. Hunt, A. Hvarfner, J. Kellett, G. K. Lighthall, A. Lippert, F. K. Lippert, R. Mahroof, J. S. Myers, M. Rosen, S. Reynolds, A. Rotondi, F. Rubulotta and B. Winters "Identifying the hospitalised patient in crisis—A consensus conference on the afferent limb of rapid response systems," *Resuscitation*, pp. 375–382, 2010.

[13] Royal College of Physicians. *National Early Warning Score (NEWS)*. London, U.K.: Royal College of Physicians, 2012.

[14] S. Nemati, A. Malhotra, and G. D. Clifford, "Data fusion for improved respiration rate estimation," *EURASIP J. Adv. Signal Process.*, vol. 2010, p. 10, 2010.

[15] W. Karlen, S. Raman, J. M. Ansermino, and G. Dumont, "Multi-parameter respiratory rate estimation from the photoplethysmogram," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1946–1953, Jul. 2013.

[16] I. Silva, G. B. Moody, and L. Celi, "Improving the quality of ECGs collected using mobile phones: The physionet/computing in cardiology challenge 2011," *Comput. Cardiol.*, vol. 38, pp. 273–276, 2011.

[17] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database," *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 12, pp. 1157–1165, Dec. 1986.

[18] V. Krasteva and I. Jekova, "QRS template matching for recognition of ventricular ectopic beats," *Ann. Biomed. Eng.*, vol. 35, no. 12, pp. 2065–2076, 2007.

[19] H. L. Chan, G. U. Chen, M. A. Lin, and S. C. Fang, "Heartbeat detection using energy thresholding and template match," in Proc. EMBC, 2006, pp. 6668–6670.

[20] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Phys. Meas.*, vol. 33, no. 9, pp. 1491–1501, 2012.

[21] S. Rhee, B.-H Yang, and H. H. Asada, "Artifact-resistant power-efficient design of finger-ring plethysmographic sensors," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 7, pp. 795–805, Jul. 2001.

[22] C. Orphanidou, S. Fleming, S. A. Shah, and L. Tarassenko, "Data fusion for estimating respiratory rate from a single lead ECG," *Biomed. Signal Process. Control*, vol. 8, no. 1, pp. 98–105, 2013.

Authors' photographs and biographies not available at the time of publication.