Sinus or not

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

Gabriele B Papini[1,2,3], Pedro Fonseca[1,2], Linda M Eerikäinen[1,2], Sebastiaan Overeem[1,3], Jan WM Bergmans[1,2] and Rik Vullings[1]

1 Department of Electrical Engineering, TU/e, Den Dolech 2, 5612 AZ Eindhoven, The Netherlands

2 Philips Research, High Tech Campus, 5656 AE Eindhoven, The Netherlands

3 Kempenhaeghe Foundation, Sleep Medicine Centre, P.O. Box 61, 5590 AB Heeze, The Netherlands

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

September 2018

Abstract. *Objective:* Wrist-worn photoplethysmography (PPG) can enable free-living physiological monitoring of people during diverse activities, ranging from sleep to physical exercise. In many applications, it is important to remove the pulses not related to sinus rhythm beats from the PPG signal before using it as a cardiovascular descriptor. In this manuscript, we propose an algorithm to assess the morphology of the PPG signal in order to reject non-sinus rhythm pulses, such as artifacts or pulses related to arrhythmic beats.

Approach: The algorithm segments the PPG signal into individual pulses and dynamically evaluates their morphological likelihood of being normal sinus rhythm pulses via a template-matching approach that accounts for the physiological variability of the signal. The normal sinus rhythm likelihood of each pulse is expressed as a quality index that can be employed to reject artifacts and pulses related to arrhythmic beats.

Main results: Thresholding the pulse quality index enables a near perfect detection of normal sinus rhythm beats by rejecting most of the non-sinus rhythm pulses (positive predictive value 98-99%), both in healthy subjects and arrhythmic patients. The rejection of arrhythmic beats is almost complete (sensitivity to arrhythmic beats 7-3%), while the sensitivity to sinus rhythm beats is not compromised (96-91%).

Significance: The developed algorithm consistently detects normal sinus rhythm beats in PPG signal by rejecting artifacts and, first of its kind, arrhythmic beats. This increases the reliability in the extraction of features which are adversely influenced by the presence of non-sinus pulses, whether related to inter-beat intervals or to pulse morphology, from wrist-worn PPG signals recorded in free-living conditions.

Keywords: Photoplethysmography, PPG, Signal Quality, Morphology
Submitted to: *Physiol. Meas.*

1. Introduction

1.1. Physiological monitoring with photoplethysmography

Over the last decades wearable devices for general consumer applications, such as fitness trackers and smartwatches, are increasingly used in medical and diagnostic applications. The need for unobtrusive and inexpensive monitoring in certain areas of healthcare, such as cardiac and sleep monitoring, is important driving force for the development of medically graded wearable devices and algorithms (Russo et al. 2015, Eapen et al. 2016, Piwek et al. 2016). Photoplethysmography (PPG) is one of the most promising technologies for wearable medical monitoring (Allen 2007). The PPG sensor uses one or more LEDs to illuminate the skin and a photo-diode to measure the quantity of light that is either transmitted through (transmissive PPG) or reflected (reflective PPG) by the tissue. The acquired PPG signal describes the variation of blood volume in the micro-vessels that is caused by the pulsatile nature of the circulating blood. PPG technology is broadly researched because it can be used to derive a substantial number of physiologically relevant signals and features, such as respiration and heart rate variability (HRV) (Charlton et al. 2017, Lázaro et al. 2011, Allen 2007). In turn, these signals and features can be used to develop algorithms to monitor or classify various physiological phenomena. Reflective PPG is commonly applied in consumer devices, such as smartwatches, as it is easily embeddable in wrist-worn devices and enables continuous and relatively unobtrusive monitoring of cardio-circulatory activity. Several research studies have been conducted on wrist-worn PPG devices for medical applications with the intent to complement or replace more classical diagnostic techniques, such as polysomnography in sleep research and Holter ECG in cardiac-monitoring (Fonseca et al. 2017, Eerikäinen et al. 2017, Bonomi et al. 2016).

For several applications, it is necessary to evaluate the quality and the regularity of the PPG signal in order to correctly extract the desired information. For instance, the removal of artifacts and arrhythmic beats from the inter-beat intervals (IBI) is paramount for computation of accurate HRV features. The irregularities caused by artifacts and non-sinus rhythm beats can increase IBI variability or bias IBI analysis and, therefore, corrupt HRV estimation (Camm et al. 1996, Peltola 2012). In addition to their effect on HRV calculation, the presence of pulses from arrhythmic beats hampers the calculation of PPG morphology-related features, such as PPG-derived respiration, due to the different hemodynamical response as compared to normal sinus rhythm beats (Camm et al. 1996, Sološenko et al. 2017). Signal quality assessment is particularly important for PPG because this signal is more prone to be corrupted when compared, for example, with an ECG signal. The PPG signal can be distorted by motion artifacts, low blood perfusion or sub-optimal contact between the sensor and the skin (Allen 2007, Schäfer & Vagedes 2013, Tamura et al. 2014). At the same time, evaluating the quality of PPG is not a trivial task, as the morphology of the pulses is highly variable due to the broad spectrum of physiological parameters influencing the signal. The pulse morphology depends on the PPG measuring location, the pressure applied to the

sensor, the systemic and local blood pressure, cardiorespiratory dynamics, and factors influencing the mechanical properties of the vessels—such as arterial stiffness due to age or plaque buildup (Linder et al. 2006, Allen 2007, Lázaro et al. 2011, Tamura et al. 2014, Hickey et al. 2015). Moreover, the presence of pulses caused by arrhythmic beats such as premature ventricular contractions, can be hidden by adjacent sinus rhythm pulses (Sološenko et al. 2017, Gil et al. 2013).

1.2. Quality assessment in PPG signals

Several algorithms for beat detection and quality assessment of PPG signals are described in the literature. Some of them estimate the quality of each pulse by thresholding specific morphological features, such as the duration of the rising slope and the amplitude of the PPG pulse (Fischer et al. 2017, Sukor et al. 2011). Furthermore, the thresholds for quality assessment are often defined heuristically or experimentally and are not necessarily valid for every PPG recording set-up. For example, some signal amplitude-related thresholds require adjustment depending on the location or the type of PPG sensor (Allen 2007, Nijboer et al. 1981). Other algorithms assess the quality of each PPG pulse by comparing it with a pulse template derived from the PPG signal itself. Such a template is often obtained by averaging multiple PPG pulses and quality is then calculated by directly comparing each individual pulse with the template, typically using one or more matching metrics, such as the Pearson's correlation coefficient (Karlen et al. 2012, Orphanidou et al. 2015) or the dynamic time warping (DTW) distance (Li & Clifford 2012). Depending on the set of pulses considered, averaging the pulse signals can lead to a distorted template because of the variability in the morphology of the pulses (Boudaoud et al. 2005, Petitjean et al. 2014). Moreover, direct comparison between a pulse and a template could lead to misclassification of a sinus rhythm pulse as a corrupted pulse, for instance when pulse-morphology changes due to a postural change. In addition, metrics like Pearson's correlation coefficient evaluate only the strength of the linear relationship between two variables. Therefore, it might not be sensitive to morphology changes that do not affect the rise-decline behavior of the PPG pulses, such as arrhythmic phenomena, but which are relevant to identify. We recently proposed an algorithm that overcomes these drawbacks (Papini et al. 2017). This algorithm calculates the pulse quality by comparing each individual PPG pulse to a template obtained from a set of pulses via a DTW-based averaging. The pulse-template comparison is mediated by the DTW of the pulse onto the template in order to account for the physiological variability of the PPG signal. However, the warping of the PPG pulse is not constrained and could therefore over-adapt the pulses to the template. This limits its applicability when a high similarity between the non-sinus rhythm pulse and the template is present, e.g. when the pulses are generated by arrhythmic beats. In addition, similar to the algorithms in the literature that we are aware of, this algorithm was developed and tested only on transmissive PPG, conventionally used in medical pulse oximeters. Moreover, none of the algorithms presented in the literature are developed to reject arrhythmic pulses and most of them were not tested in the presence

of arrhythmias or in long-term recordings in free-living conditions such as in overnight recordings. Therefore, questions remain regarding the applicability and reliability in important real-world applications such as night-by-night sleep analysis and follow-up of patients after hospital discharge.

This paper describes an extended version of the algorithm presented previously (Papini et al. 2017). Here we describe an algorithm to assess the quality of individual PPG pulses, defined as the likelihood of being generated by normal sinus rhythm contractions. The quality index is assessed by comparing PPG pulse morphology to an adaptive template. The template is created via DTW barycenter averaging of several (DBA) pulses in order to be insensitive to physiological differences in pulses used to create the template and to guarantee a tailored template for every recording. The comparison between each pulse and the template is mediated by constrained DTW in order to account for the physiological variability of pulse morphology during sinus rhythm. The performance of the algorithm to classify normal sinus rhythm beats is tested on two datasets containing overnight free-living data from a wrist-worn reflective PPG device and analyzed in terms of beat detection and rejection rate of pulses caused by arrhythmic beats.

2. Methods: Pulse quality index estimation algorithm

Figure 1 illustrates an overview of the algorithm; each of the steps will be described in the following sections. Regarding the abbreviations used, the subscript indicates the main characteristic of an element (e.g. Temp_{Ad} is the adapted template); PP indicates a pulse of a PPG signal (e.g. $\text{PP}_{10\text{Hz}}$ is a pulse of $\text{PPG}_{10\text{Hz}}$); the abbreviations ending in *s* have to be considered plural (e.g. $\text{PPs}_{10\text{Hz}}$ is the plural form of $\text{PP}_{10\text{Hz}}$).

2.1. PPG pre-processing, segmentation and beat localization

The raw PPG signal is filtered to obtain two versions with different frequency components:

- $\text{PPG}_{2.25\text{Hz}}$: a 3$^{\text{rd}}$ order zero-phase band-pass Butterworth filter with cut-off frequencies of 0.4 Hz and 2.25 Hz is used to remove high-frequency components related to noise and to the diastolic peak and the low-frequency components related to respiration and body position changes (Allen 2007) (figure 2(a)). This enhances the two local minima defining the start and the end of each pulse. The local minima are found according to:

$$\begin{aligned} l = \{t \mid &\text{sign } \dot{\text{PPG}}_{2.25\text{Hz}}(t) \neq \text{sign } \dot{\text{PPG}}_{2.25\text{Hz}}(t-1) \;\& \\ &\& \; \ddot{\text{PPG}}_{2.25\text{Hz}}(t) > 0 \;\& \; \text{PPG}_{2.25\text{Hz}}(t) < 0\}. \end{aligned} \tag{1}$$

  Where $\dot{\text{PPG}}_{2.25\text{Hz}}$ and $\ddot{\text{PPG}}_{2.25\text{Hz}}$ indicate, respectively, the first and second derivative of $\text{PPG}_{2.25\text{Hz}}$.

- $\text{PPG}_{10\text{Hz}}$: a 3$^{\text{rd}}$ order zero-phase band-pass Butterworth filter with cut-off frequencies of 0.4 Hz and 10 Hz is used to preserve the morphology of the PPG pulses which should still exhibit the component related to the diastolic peak (figure 2(a)).

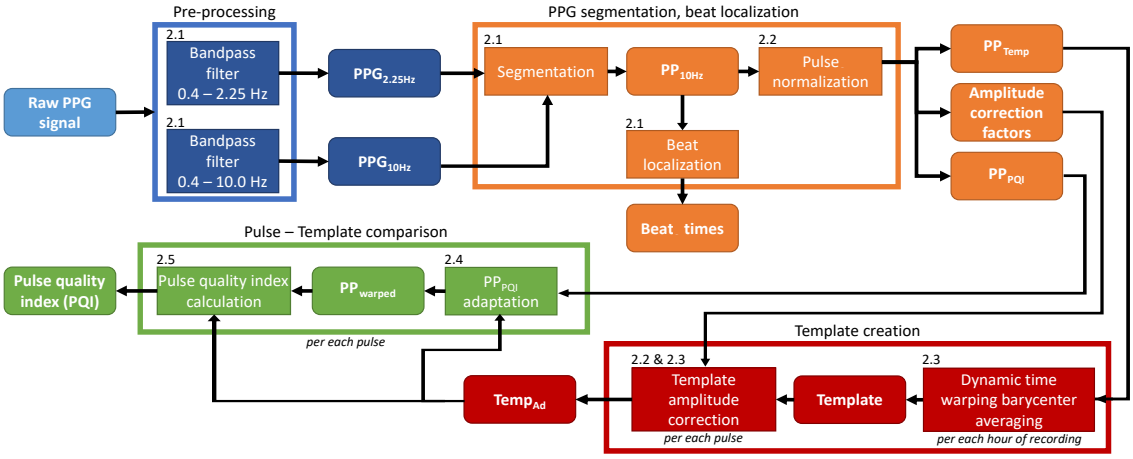

Figure 1. Overview of the algorithm. The flowchart represents the steps to obtain the beat locations and their associated pulse quality index (PQI) from the raw PPG signal. The rectangular blocks represent the operations and the blocks with rounded corners and bold text are the inputs/outputs for each operation. The sections of the paper corresponding to each operation block are reported above each of them. Regarding the acronyms in the figure: $\mathrm{PPG}_{2.25\mathrm{Hz}}$ and $\mathrm{PPG}_{10\mathrm{Hz}}$ are two filtered versions of the raw PPG signal (section 2.1), $\mathrm{PP}_{10\mathrm{Hz}}$ is a segment of $\mathrm{PPG}_{10\mathrm{Hz}}$ (section 2.1), $\mathrm{PP}_{\mathrm{Temp}}$ and $\mathrm{PP}_{\mathrm{PQI}}$ are two different normalization of $\mathrm{PP}_{10\mathrm{Hz}}$ used for template and quality calculation (section 2.2), $\mathrm{Temp}_{\mathrm{Ad}}$ is an amplitude adjusted version of the template (sections 2.2 and 2.3), $\mathrm{PP}_{\mathrm{warped}}$ is a dynamic time warped version of the $\mathrm{PP}_{\mathrm{PQI}}$ (section 2.4).

$\mathrm{PPG}_{10\mathrm{Hz}}$ is segmented in $\mathrm{PPs}_{10\mathrm{Hz}}$ using the local minima of $\mathrm{PPG}_{2.25\mathrm{Hz}}$ (figure 2(b)) according to the following equations:

$$
\begin{aligned}
l_i &= i^{th} \text{ local minimum time}, && (2)\\
\text{peak} &= \{t \mid \mathrm{PPG}_{2.25\mathrm{Hz}}(t) = \max \mathrm{PPG}_{2.25\mathrm{Hz}}(l_i, l_{i+1})\}, && (3)\\
\text{start} &= \{t \mid \mathrm{PPG}_{10\mathrm{Hz}}(t) = \min \mathrm{PPG}_{10\mathrm{Hz}}(l_i, \text{peak})\}, && (4)\\
\text{end} &= \{t \mid \mathrm{PPG}_{10\mathrm{Hz}}(t) = \min \mathrm{PPG}_{10\mathrm{Hz}}(\text{peak}, l_{i+1} + \tfrac{l_{i+1}-l_i}{4})\}, && (5)\\
\mathrm{PP}_{10\mathrm{Hz}} &= \mathrm{PPG}_{10\mathrm{Hz}}(\text{start}, \text{end}). && (6)
\end{aligned}
$$

The resulting $\mathrm{PPs}_{10\mathrm{Hz}}$ that have a duration above 1.5 sec or below 0.5 sec, corresponding to a heart rate outside normal sinus rhythm range of 40 to 120 beats per minute (bpm), are excluded from further analysis. The exclusion criteria are slightly broader than the normal resting HR range in adults (40 to 90 bpm) in order to retain HR changes that may be caused by conditions unrelated to cardiovascular conditions, such as obstructive sleep apnea (Penzel et al. 2003, Mason et al. 2007). The remaining $\mathrm{PPs}_{10\mathrm{Hz}}$ are upsampled to 1 kHz (in case recorded with a lower sampling frequency) using a cubic spline interpolation in order to localize more accurately their key points, such as their peaks. This procedure is commonly employed for ECG signals in order to better define the beat time location in the QRS complexes (Ellis et al. 2015). Since each PPG pulse is the hemodynamic consequence of a cardiac contraction, it is possible to associate a fiducial point of the pulse with a heartbeat. The algorithm detects the

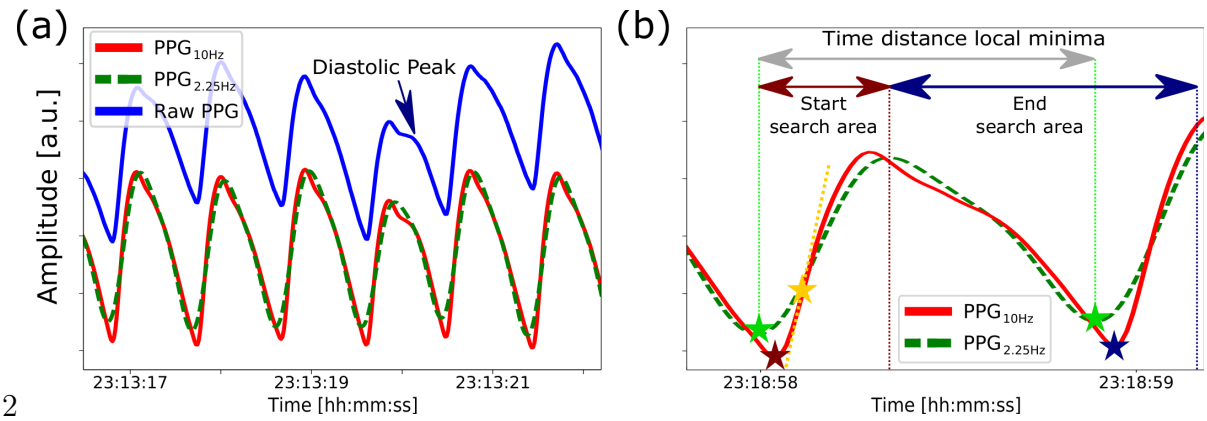

Figure 2. PPG signals and Pulses. (a): The raw PPG signal and the two filtered signals derived from it. (b): Segmentation of the $\mathrm{PPG}_{10\mathrm{Hz}}$ using the $\mathrm{PPG}_{2.25\mathrm{Hz}}$ minima (green stars); the red arrow indicates the search window for the start of the pulse (red star) and the blue arrow the search window for the end of the pulse (blue star). The yellow dotted line represents the maximum tangent to the rising slope of the $\mathrm{PP}_{10\mathrm{Hz}}$ and the yellow star indicates the point where is obtained. The yellow star is the fiducial point of the pulse associated to the heartbeat chosen in our research.

largest first derivative in the rising slope of $\mathrm{PP}_{10\mathrm{Hz}}$ as the fiducial point (yellow star in figure 2(b)) since we found it to reproduce accurately the ECG-derived IBI (Nano et al. 2017).

2.2. Pulse normalization

The amplitude of the PPG pulses can vary significantly during a recording due, for instance, to a change in body position or to different pressure applied to the sensor (Linder et al. 2006, Tamura et al. 2014). Therefore, it is necessary to normalize the $\mathrm{PP}_{10\mathrm{Hz}}$ amplitude before creating a pulse template out of an ensemble of pulses. The normalized version of $\mathrm{PP}_{10\mathrm{Hz}}$ is calculated according to:

$$\text{amplitude} = |\max \mathrm{PP}_{10\mathrm{Hz}} - \min \mathrm{PP}_{10\mathrm{Hz}}|, \tag{7}$$

$$\text{shift} = (\max \mathrm{PP}_{10\mathrm{Hz}} + \min \mathrm{PP}_{10\mathrm{Hz}})/2, \tag{8}$$

$$\mathrm{PP}_{\mathrm{Temp}} = \frac{\mathrm{PP}_{10\mathrm{Hz}} - \text{shift}}{\text{amplitude}}. \tag{9}$$

Sudden substantial increases of the $\mathrm{PP}_{10\mathrm{Hz}}$ amplitudes are often representative of signal corruption. In these cases, the amplitude information is helpful in identifying segments that should have a low quality index. For this reason, a second normalized version of the $\mathrm{PPs}_{10\mathrm{Hz}}$ which is used for quality index calculation is computed as:

$$\mathrm{PP}_{\mathrm{PQI}} = \mathrm{PP}_{10\mathrm{Hz}} - \text{shift}. \tag{10}$$

The different normalization approaches complicate the estimation of the similarity between the template derived from the $\mathrm{PPs}_{\mathrm{Temp}}$ and each $\mathrm{PP}_{\mathrm{PQI}}$ because the first represents a normalized amplitude, while the second contains a varying amplitude. Therefore, in order to be able to compare template and $\mathrm{PPs}_{\mathrm{PQI}}$, it is necessary to adjust the template amplitude with a correction factor before each comparison between the template and $\mathrm{PP}_{\mathrm{PQI}}$. To derive the correction factors, a time series comprising all the

amplitudes obtained in (7) is stored. First, the algorithm removes from the amplitude time series all elements that have a value 50% higher or lower than the previous or the following value. Then the clipped amplitude time series is interpolated at 4 Hz using a cubic spline interpolation and filtered with a 3rd order zero-phase low-pass Butterworth filter with a cut-off frequency of 1.5 Hz. The cut-off frequency is chosen to confidently include physiological amplitude variations since these are not expected to occur faster than the respiration frequency (a frequency range from 4 to 60 breaths per minute, i.e. from 0.067 Hz to 1 Hz was considered) (Charlton et al. 2017). Finally, the filtered signal is resampled at the same time locations of the original amplitude time series. In this way, the algorithm suppresses abnormal amplitude variations from each PP_{PQI} while retaining the low-frequency physiological amplitude variations. The cleaned amplitude time series is used as a series of correction factors; the adjusted templates ($Temp_{Ad}$s) are obtained by multiplying the template by the corresponding correction factors for each comparison with the PP_{PQI}.

2.3. Template creation

The literature describes several examples of template extraction from cardiovascular signals such as ECG and ballistocardiogram for the purpose of beat classification and signal quality estimation (Orphanidou et al. 2015, Brüser et al. 2013, Redmond et al. 2012, Clifford 2002). In our algorithm, the quality assessment of each PPG pulse is also based on a template matching approach: each PP_{PQI} is compared with an amplitude adjusted template describing the morphology of an uncorrupted and regular PPG pulse, i.e. a pulse free from movement artifacts and caused by a normal sinus rhythm beat. The pulses in a PPG signal are a clear example of a time series that changes in shape: the pulse morphology changes according to normal physiological variation, such as heart rate (HR), vascular tone and body position (Linder et al. 2006, Allen 2007). For this reason, our algorithm calculates the pulse template by means of DBA (Petitjean et al. 2014). This allows the time series to be averaged by iteratively decreasing the DTW distance between an initial template and each individual pulse (figure 3). In each iteration,

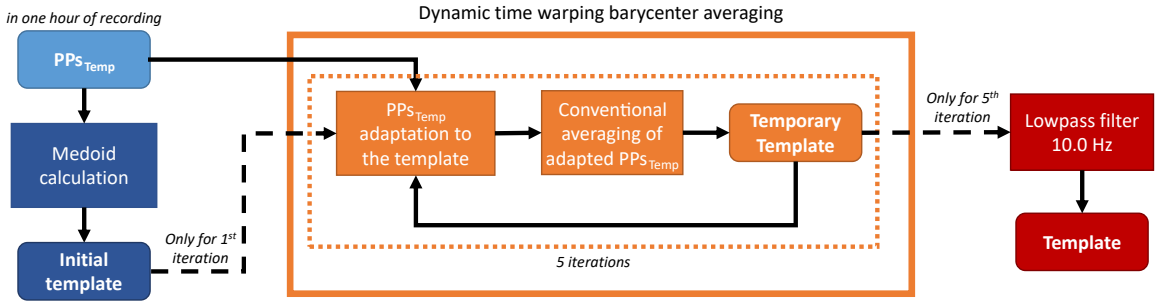

Figure 3. Template creation. The flowchart shows the steps to obtain, for each hour of PPG recording, a template from the PP_{Temp}s. The format of the blocks indicates their type as in figure 1. The PP_{Temp}s adaptation to the template is mediated by DTW, similar to the PP_{PQI}s adaptation described in section 2.4.

the resulting averaged time series is used as the initial template for the next iteration. The DBA is initialized with the medoid of the PPs_{Temp} as initial template and this

initial template is refined during five iterations (Petitjean et al. 2014). The number of iterations is chosen empirically to ensure the computation a pulse template with a correct morphological representation of the pulses, in an efficient computation time. The resulting pulse template is filtered using a 3rd order low-pass Butterworth filter with a cut-off frequency of 10 Hz to remove high-frequency components introduced by the DBA. This guarantees that the template has the same frequency components as the PPG segments used to obtain it (PPs_{Temp}) and of the PPG segments used to calculate the quality index (PPs_{PQI}). A separate template is calculated for each one-hour segment of the PPG recording. The choice of one hour segments represent a compromise between a sufficient amount of uncorrupted pulses while keeping computational time within reasonable limits. However, the algorithm is also able to handle shorter segments, for instance of eight minutes, as described in Papini et al. (2017).

Figure 4 illustrates an example where DBA outperforms a conventional averaging procedure. In this case, the subset of pulses used has a varying morphology caused by the presence of a venous pulsation at the beginning of some of the PPG pulses (e.g. in PP_{Temp} 2) (Shelley et al. 1993). The morphological variety is remarkably broad (the area highlighted in green) and conventional averaging does not yield a pulse rising slope characteristic of an actual PPG pulse. Instead, the DBA maintains the overall shape of the PPG pulses and, in addition, includes the influence of the venous pulsation. Aligning the pulses before conventional averaging would improve the results, but would also require removal of the part of the PPG pulses related to the venous pulsation and, therefore, part of the physiological information present in the signal. In our algorithm, it is not necessary to account separately for the pulse misalignment because the DBA accounts for it (pulse adaptation step in figure 3). Thanks to this, the DBA can operate on segments with different lengths, eliminating the need to zero-pad or trim segments to a fixed length before aligning and finally averaging them.

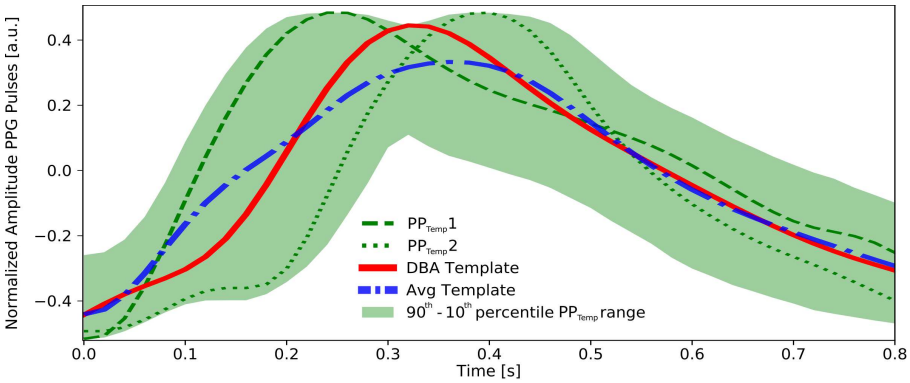

Figure 4. Template and PPs_{Temp}. The green area encloses the pulse morphology within the 90th and the 10th percentile of the PPs_{Temp}' silhouettes. The morphology varies from classical pulse ($PP_{Temp}1$) to pulse with a venous pulsation ($PP_{Temp}2$). The template obtained by DBA (DBA Template) is closer to a sinus rhythm pulse than that obtained by conventional averaging (Avg Template).

2.4. PP_{PQI} adaptation

The proposed algorithm uses DTW to derive $\mathrm{PP_{warped}}$ from the warping of $\mathrm{PP_{PQI}}$ to the template calculated for the one hour of PPG signal they belong to (figure 5). In contrast with our previous algorithm (Papini et al. 2017), we now apply two constraints to the DTW in order to avoid a distortedly warped $\mathrm{PPs_{PQI}}$ (Sakoe & Chiba 1978). One constraint limits the maximum time warping to 0.3 sec and the other limits the number of matching points to a maximum of 3. The first constraint is derived heuristically and corresponds approximately to two times the standard deviation of the inter-pulse intervals in our datasets. The second constraint is chosen in order to be able to transform a $\mathrm{PP_{PQI}}$ from 0.5 sec to 1.5 sec, and vice versa, and to guarantee a balanced warping to each point of the pulses. In this way, the entire spectrum of pulses, with a duration within the boundaries set in the PPG signal segmentation, can be covered. Finally, the $\mathrm{PP_{warped}}$ are filtered using the same approach as for the template, in order to remove high-frequency components introduced by the DTW (figure 5) and to enable the comparison with the $\mathrm{Temp_{Ad}}$ to calculate the pulse quality index.

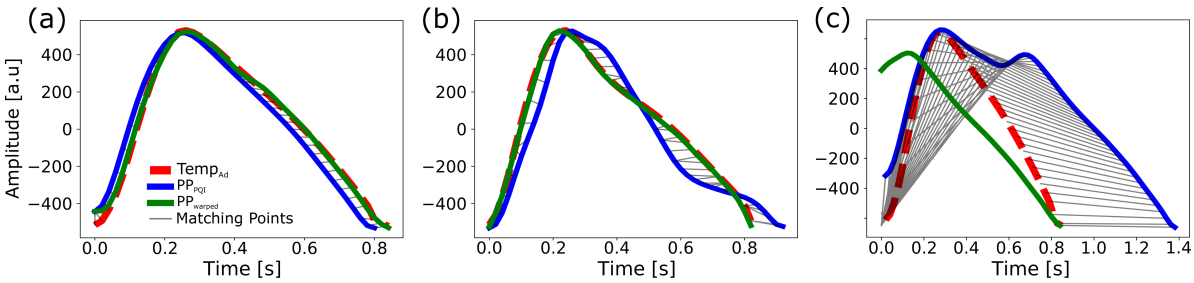

Figure 5. $\mathrm{PP_{PQI}}$ adaptation examples. (a) sinus rhythm pulse similar to the template, (b) sinus rhythm pulse different from the template, (c) pulse composed by a sinus rhythm pulse merged with a pulse caused by a premature ventricular contraction.

2.5. PQI calculation

The DTW produces a warped pulse that matches best, within the constraints, with the derived pulse template. However, some differences between them remain, for two reasons: first of all, because the DTW modifies the time locations of each point, but not their values, and second, because of the DTW constraints used. In this way, part of the morphological discrepancies between $\mathrm{PP_{PQI}}$ and $\mathrm{Temp_{Ad}}$ remain in the $\mathrm{PPs_{warped}}$. These residual differences are used by the algorithm to calculate the quality index (PQI) of each PPG pulse. An overview of the procedure is shown in figure 6 and the details are explained in the following equations.

First, the algorithm finds the indices of each sample of $\mathrm{PP_{warped}}$ with a difference higher than 10% with respect to the corresponding samples of the $\mathrm{Temp_{Ad}}$ (11).

$$\mathrm{UP} = \left\{ i \in \mathbb{Z}^{+} \middle| \left| \frac{\mathrm{PP_{warped}}(i) - \mathrm{Temp_{Ad}}(i)}{\mathrm{Temp_{Ad}}(i)} \right| > 0.1 \right\}. \tag{11}$$

Where $\mathbb{Z}^{+}$ denotes the set of positive integers ranging from one to the number of samples of $\mathrm{PP_{warped}}$. These indices allow the classification of each point of $\mathrm{PP_{warped}}$ as unmatched (UP) or matched (MP) to the relative $\mathrm{Temp_{Ad}}$ points. The 10% threshold

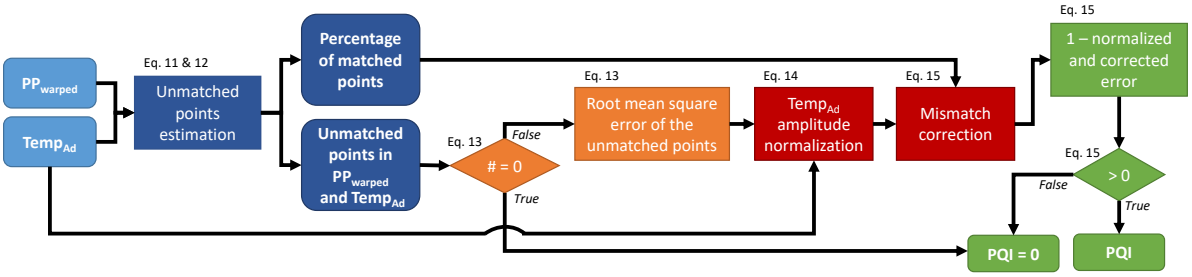

Figure 6. PQI calculation. The flowchart shows the steps to calculate the pulse quality index. The blue blocks represent the search for pulse-template dissimilarities, the orange blocks the quantification of the dissimilarity as an error, the red blocks the correction of the error and the green blocks the calculation of the PQI. The format of the blocks indicates their type as in figure 1. The numbers on top of the blocks indicate the corresponding equation.

is chosen heuristically and lower threshold values appear not to drastically influence the algorithm's performance. Instead, thresholds above 20% are not recommended in order to avoid decreasing the ability to correctly calculate the PQI. By using a 10% threshold, the algorithm can account for small deviations that the DTW cannot fully compensate for, and can treat them as residual physiological variations. The quantity of unmatched and matched points, respectively N_{UP} and N_{MP}, are used to calculate the percentage of matching points as:

$$N_{\mathrm{MP\%}} = N_{\mathrm{MP}}/(N_{\mathrm{MP}} + N_{\mathrm{UP}}). \tag{12}$$

The second step consists in calculating the root mean square error of the unmatched points in respect to $\mathrm{Temp_{Ad}}$ according to:

$$\mathrm{RMSE_{UP}} = \begin{cases} 0 & \text{if} \quad N_{up} = 0 \\ \sqrt{\frac{\sum_{j=1}^{N_{\mathrm{UP}}} [\mathrm{Temp_{Ad}}(\mathrm{UP}_j) - \mathrm{PP_{warped}}(\mathrm{UP}_j)]^2}{N_{\mathrm{UP}}}} & \text{if} \quad \text{otherwise.} \end{cases} \tag{13}$$

The $\mathrm{RMSE_{UP}}$ represents the pulse-template dissimilarity, but it depends on the amplitude of $\mathrm{Temp_{Ad}}$, which varies for each $\mathrm{PP_{warped}}$-$\mathrm{Temp_{Ad}}$ comparison. In fact when N_{UP} is not zero, the higher the amplitude of $\mathrm{Temp_{Ad}}$, the higher the $\mathrm{RMSE_{UP}}$ is, regardless of the actual dissimilarity between $\mathrm{Temp_{Ad}}$ and $\mathrm{PP_{warped}}$. Therefore, the algorithm normalizes the $\mathrm{RMSE_{UP}}$ for the amplitude of $\mathrm{Temp_{Ad}}$:

$$\mathrm{RMSE_{norm}} = \mathrm{RMSE_{UP}}/(\max(\mathrm{Temp_{Ad}}) - \min(\mathrm{Temp_{Ad}})). \tag{14}$$

In case the error is larger than the amplitude of $\mathrm{Temp_{Ad}}$, the $\mathrm{PP_{warped}}$ has certainly a poor quality, therefore its PQI is set to zero. The $\mathrm{RMSE_{norm}}$ has the disadvantage that it takes into account only the magnitude of the mismatch and not the number of the points mismatching. This characteristic negatively influences the PQI because it makes PQI insensitive to situations where there is a high number of unmatched points (low $N_{\mathrm{MP\%}}$) in combination with a low $\mathrm{RMSE_{norm}}$. The algorithm accounts for this by dividing the $\mathrm{RMSE_{norm}}$ by $N_{\mathrm{MP\%}}$. This ratio can be higher than 1, resulting in a negative PQI. Since there is no reason to have a PQI lower than zero, PQI is set to 0 in case the ratio is larger than 1, i.e. $\mathrm{RMSE_{norm}}$ larger than $N_{\mathrm{MP\%}}$.

$$\mathrm{PQI} = \max\{0, 1 - \mathrm{RMSE_{norm}}/N_{\mathrm{MP\%}}\}. \tag{15}$$

The algorithm calculates a single PQI varying from 0—low-quality, highly corrupted or highly influenced by arrhythmia—to 1—high quality, uncorrupted and belonging to normal sinus rhythm—for each pulse in a one-hour segment of the PPG recording.

In case the average of the non-null PQIs is below 0.9 in an one-hour segments of the recording, the PQI is considered invalid, the algorithm excludes the pulses with the 20% lowest PQI, recalculates the template and re-assesses the quality of all pulses in that segment. If the PQI is still not valid, the template calculated in the previous hour of the recording is used. In case there is no template available, a PQI of zero is assigned to all the pulses in the segment.

3. Materials and validation methods

3.1. Datasets

Two datasets are used to assess the performance of the PQI as measure of normal sinus rhythm likelihood of a pulse. The first dataset (Heart Health Sleep dataset, HHS) consists of 26 overnight recordings from 16 healthy adults and is described by Fonseca et al. (2017) as *Validation set 1*. The subjects in the HHS datasets did not have any previous history of cardiovascular disorders and the subjects were assumed to have normal sinus rhythm. The second dataset (Atrial Fibrillation dataset, AF) consists of 16 24-hour recordings from 16 patients, of which four had continuous atrial fibrillation, one atrial flutter and eleven sinus rhythm with premature contractions. This dataset is described by Eerikäinen et al. (2017); for this research, only the overnight part of each recording is considered in order to have a stationary recording condition, as in most literature (Aboy et al. 2005, Karlen et al. 2012, Li & Clifford 2012, Fischer et al. 2017). The start and end time in bed were annotated by a researcher by analyzing the accelerometer signal. The HHS and AF datasets contain ECG (respectively from lead II and 12-lead) and wrist-worn PPG signal recorded simultaneously. In addition, the ECG data in the AF dataset was visually analyzed by a clinical expert using an automated software to label every beat either as sinus rhythm, AF, premature atrial contraction, premature ventricular contraction, artifact, or unknown. For the HHS dataset, the beats in the ECG signal were automatically detected using a QRS detector and post-processing localization algorithm described in Fonseca et al. (2014). The ECG signals comprise 769935 beats in the HHS dataset and 437337 beats in the AF dataset. The AF dataset comprises 276041 sinus rhythm beats and the arrhythmic beats comprise 9778 premature atrial contractions (PAC), 8060 premature ventricular contractions (PVC), 107485 beats during atrial fibrillation (AFib), 35873 beats during atrial flutter (AFlu) and 100 unknown/artifact. During the benchmarking of the algorithm, the unknown/artifact beats are treated in the same way as arrhythmic beats because of their limited amount and since both groups shoud anyway be rated with a low quality index.

3.2. Validation procedures

The current version of the algorithm differs from that described in Papini et al. (2017) in the two constraints employed in the pulse adaptation (section 2.4). The constraints do not refer to specific pulse types (e.g. uncorrupted, arrhythmic or artifact) or to a specific pathology. Given the large number of beats used to validate the algorithm, and the independence of the constraints from arrhythmic/artifact influences on the pulse morphology, we are convinced that there is no risk of overfitting to the validation datasets. The algorithm is tested regarding sinus rhythm beat detection capability and rejection of arrhythmic beats. Its performance is evaluated in comparison with a reference: either manually or automatically annotated beats from the ECG signal. Moreover, the performance is compared with the correlation derived pulse quality index described in Clifford et al. (2015) and summarized in the next section.

3.2.1. Correlation derived pulse quality index An algorithm to derive a pulse quality index based on the Pearson's correlation coefficient between the PPG pulses and a template has been publicly released during the PhysioNet/Computing in Cardiology Challenge of 2015. The algorithm takes as input a PPG signal in addition to the pulse segmentation points, and outputs a quality index for each pulse. The algorithm first computes a template from all the pulses and then uses this template to calculate the correlation-based pulse quality index of the same pulses ($\mathrm{PQI_{Corr}}$). In our research, the correlation-based algorithm takes as input the $\mathrm{PPs_{Temp}}$ and uses these pulses for template creation and quality assessment. The beat locations within the pulses are defined as in our algorithm, i.e. the point with the maximum slope in the rising part of the PPG pulse. The correlation-based algorithm uses a standard ensemble averaging to derive the template and it re-samples the pulses to match the number of samples of the template before calculating the quality index. The $\mathrm{PQI_{Corr}}$ is validated in the same way as the PQI, allowing us to benchmark the performance of our proposed algorithm with respect to a commonly used approach described in the literature.

3.2.2. Sinus rhythm beat detection The algorithm described in this manuscript gives as an output the beat times and the PQI associated with each pulse. The performances of the beat detection algorithms are calculated by comparing the beat times derived from PPG to the annotated beat times from ECG as in Karlen et al. (2012) and Aboy et al. (2005). Prior to the performance evaluation, the beat times derived from the PPG signal are synchronized with those derived from the ECG signal. This is done by maximizing the cross-correlation of the IBIs obtained from both signals (Fonseca et al. 2017). Sensitivity and positive predictive values (SEN, PPV) are employed to quantify the beat detection rate of normal sinus rhythm beats (ANSI-AAMI 1998). In order to be able to compute these metrics, true detections are determined by defining a maximum time difference allowed between the reference beat from ECG and a detected beat from PPG as in Aboy et al. (2005), Karlen et al. (2012), Papini et al. (2017). Every missed detection is considered a false negative while every detection not belonging to any

detection window is considered a false positive. SEN and PPV are then calculated as:

$$\text{SEN} = \frac{\text{True detections}}{\text{True detections} + \text{Missed detections}}, \tag{16}$$

$$\text{PPV} = \frac{\text{True detections}}{\text{True detections} + \text{Not belonging detections}}. \tag{17}$$

The detection window for ECG-PPG beat matching is set to 125 ms in order to account for the pulse transit time variations that can generate local time differences between PPG and ECG beats even after the global synchronization step, e.g. due to body position change (Foo et al. 2005). PQI is used to discard pulses with a PQI lower than a predefined threshold. SEN and PPV are calculated for different thresholds in order to determine the operating characteristics of the algorithm regarding detection of beats annotated as sinus rhythm with respect to the others, i.e. arrhythmic beats, artifacts or corrupted pulses. The assumption is that by increasing the quality index threshold the PPV should increase, because more erroneous detections, i.e. artifacts and non-sinus rhythm beats, will be rejected. SEN and PPV are not calculated for the subjects with continuous AFib and AFlu because they do not have any beats labeled as sinus rhythm, and the number of true detections cannot be calculated.

3.2.3. Rejection of arrhythmic beats Each arrhythmic beat in the AF dataset has a reference label of PAC, PVC, AFlu or AFib. The capability of the algorithm to assign a low quality index to each arrhythmic beat is tested as the sensitivity to each of these types of beats (SEN_{Arr}). SEN_{Arr} is calculated in the same way as the sensitivity to sinus rhythm beats by now using the arrhythmic beats as positive class instead of sinus rhythms beats. The assumption is that by increasing the PQI threshold the number of arrhythmic beats rejected should increase and consequently the SEN_{Arr} should decrease due to the lower arrhythmic beat detection (true positives). The presence of an arrhythmic beat can also influence the morphology of a previous PPG pulse by, for instance, deforming its diastolic phase (Sološenko et al. 2017, Gil et al. 2013). Therefore, we determine a subject-specific time threshold as the median of the time distances between each peak and the corresponding end of each pulse in the complete recording (to account for the subject variability of the PPG signal). If two PPG beats are closer in time than the subject-specific threshold, the same label as that assigned to the subsequent PPG beat is also assigned to the first one. Subjects with continuous AFib or AFlu do not have any beats labeled as sinus rhythm, therefore their SEN_{Arr} is calculated separately from the other arrhythmic subjects. The four AFib subjects are treated differently from the rest because no sinus rhythm is present and, therefore, it is not possible to assess deviations from sinus rhythm. The AFlu subject is also treated separately because one form of the pathology reduces the ventricular contraction variability on the contrary to the other arrhythmias (Scholz et al. 2014). The pulses during atrial flutter are extremely regular and are therefore not distinguishable by simple morphological comparison. By isolating subjects with AFib and AFlu, the sensitivity to arrhythmic beats can be calculated over the entire AF dataset without being biased

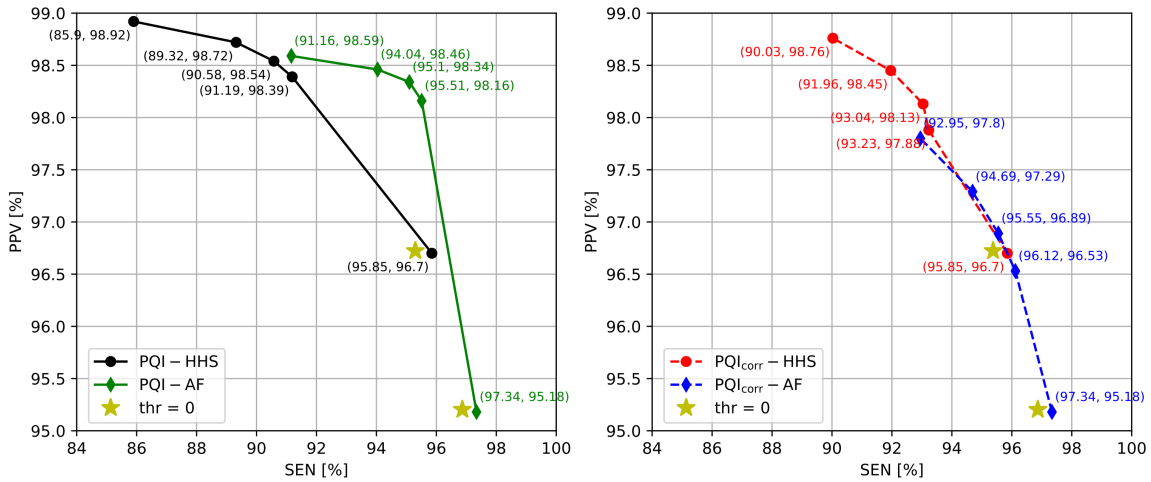

Figure 7. Sinus rhythm beat detection. The sensitivity and positive predictive value obtained using the PQI (left) and the PQI_{Corr} (right). Each curve shows the results for one of the datasets; the marker types are dataset-specific. The results start for a quality threshold equal to zero (stars in the figure) and are reported for threshold going from 0.5 to 0.8 with 0.1 step. The exact sensitivity and positive predictive value for each point are in parenthesis.

by the different type of arrhythmic populations, i.e. continuous as in AFib and AFlu, versus intermittent PAC and PVC. Nonetheless, for these five subjects, the sensitivity to arrhythmic beats is calculated separately for each pathology for the overall number of arrhythmic beats. In addition, the overall quality indices values are calculated to describe the algorithm's behavior for these two arrhythmia cases and to compare it with the average quality indices for premature contractions and sinus rhythm beats using the Mann & Whitney (1947) test. For the recordings with premature contractions, SEN_{Arr} is calculated for the overall number of arrhythmic beats, as the average of each subject's SEN_{Arr} and for the overall number of PAC and PVC beats.

4. Results

4.1. Sinus rhythm beat detection

The PPV and SEN for detecting normal sinus rhythm beats in the HHS and AF datasets calculated for PQI and PQI_{Corr} thresholds ranging from 0.0 to 0.8 are reported in figure 7. For both pulse quality indices, increasing the threshold of beat rejection increases the positive predictive value at the expense of sensitivity. While the positive predictive value is comparable in both datasets, the sensitivity is lower in the HHS dataset. The beat detection algorithm performs well, also without thresholding the PQI; the positive predictive value with a PQI threshold of 0.0 is around 97% and 95% in the HHS and the AF datasets, respectively. By increasing the PQI threshold, the positive predictive value increases consistently by around 2% and 3% in the two datasets with a moderate decrease in sensitivity. For a PQI threshold of 0.6, nearly all the beats are successfully matched with the reference beats from the ECG.

Table 1. Distribution of PQI and PQI_{Corr} for each beat type.

	AFib	AFlu	PAC+PVC	Sinus Rhythm
PQI avg±std	0.55±0.38†	0.98±0.11∗	0.20±0.28†	0.95±0.14
PQI_{Corr} avg±std	0.68±0.26†	0.98±0.08∗	0.57±0.29†	0.95±0.10

Statistically significantly lower (†)/higher (∗) than Sinus Rhythm beats ($p<0.01$).

Table 2. SEN_{Arr} [%] for the overall number of beats, per-subject average (avg±std) and overall number of beats for arrhythmia cases for different thresholds (Thr).

	Thr	0.0	0.5	0.6	0.7	0.8
PQI	Overall	46.92	6.90	4.90	3.75	2.65
	Per-subject	53.64 ± 25.54	13.73 ± 11.74	11.38 ± 10.66	9.53 ± 8.93	4.06 ± 4.56
	PAC	62.43	11.00	7.92	6.16	4.40
	PVC	28.24	1.97	1.25	0.84	0.55
	AFib	80.70	55.88	51.80	44.06	30.68
	AFlu	98.40	97.34	97.13	96.78	95.78
PQI_{Corr}	Overall	46.92	30.03	25.10	19.62	12.53
	Per-subject	53.64 ± 25.54	35.99 ± 22.10	31.89 ± 22.63	29.05 ± 22.63	25.77 ± 22.48
	PAC	62.43	39.69	33.50	25.95	15.74
	PVC	28.24	18.40	14.98	11.99	8.66
	AFib	80.70	63.47	56.91	48.34	36.45
	AFlu	98.40	97.90	97.70	97.50	97.16

4.2. Arrhythmic beat rejection

The overall quality index distributions are reported in table 1. The beats belonging to the subjects with continuous Afib and Aflu have, respectively, significantly lower and higher quality indices compared to sinus rhythm beats. In table 2 the SEN_{Arr} is reported for the same quality index thresholds as used in the normal sinus rhythm beat detection performance evaluation. For each quality index threshold, the SEN_{Arr} is lower for the PQI when compared with the PQI_{Corr}, independently of how the metric is evaluated.

5. Discussion

In this paper, we describe and evaluate an algorithm to detect normal sinus rhythm beats from PPG signals by means of a pulse quality-based rejection of arrhythmic beats and corrupted pulses. With respect to the detection of normal sinus rhythm beats, the PQI calculated by comparing the PPG pulses and the signal specific templates achieves a performance which is on par with the literature. Using a threshold to reject beats based on the PQI leads to a good agreement with beats obtained from ECG, which is a more noise resilient signal compared to the PPG signal for beat detection.

The sensitivity to sinus rhythm beats we obtained is lower compared to our previous work and to results reported in literature (Papini et al. 2017, Fischer et al. 2017, Karlen et al. 2012, Aboy et al. 2005). However, this can be attributed to the datasets used for validation. An artifact present in the PPG signal is often not visible in the ECG signal, because the latter is less influenced by movement artifacts and because the PPG sensor often temporarily loses contact with the sensed area. For instance, limb-

worn devices are more subject to body movements than the thorax or abdomen where ECG is conventionally mounted. In addition, the lack of sinus rhythm annotation and the completely automatic detection of the reference beats make the HHS dataset sub-optimal to test the sinus rhythm detection capabilities of our algorithm. The annotation methodology of the HHS dataset can cause the inclusion of some artifacts of the ECG signal or arrhythmic beats as sinus rhythm beats. However, it is legitimate to assume that the large majority of the beats detected in the ECG signals of the HHS dataset are sinus rhythm beats because the dataset consists of healthy subjects, with no primary history of cardiovascular disorders or using any substance that could have influenced the cardiovascular activity (Fonseca et al. 2017). In addition, the ECG beat detector has been published with good results and used in other research works (Radha et al. 2017, Fonseca et al. 2018). Nevertheless, the results obtained on the HHS dataset should not be considered on their own, but only as further evidence of the algorithm's performance besides what was obtained on the correctly annotated AF dataset. Despite this apparent decrease in performance, it should be emphasized that unlike earlier work, this research focused on free-living recordings of healthy subjects and of patients with cardiac arrhythmias. The recording conditions introduce a higher physiological variability in the PPG signals due to the long recording duration and to the unconstrained setup. Moreover, the number and different pathologies of the measured subjects enriches the morphological variability of the PPG signal. The reference used and the type of recordings contribute to a lower sensitivity to sinus rhythm beats, but the algorithm is able to provide—even for higher PQI thresholds—an average sensitivity higher than 90% together with positive predictive values close to 99%. In terms of sinus rhythm beat detection, the PQI performance is comparable to what is obtained using the PQI_{Corr}. However, the positive predictive value is consistently higher for the PQI, especially for the AF dataset. This performance difference can be attributed to a higher specificity in assessing the quality of the arrhythmic beats: the PPG pulses associated with arrhythmic beats are more distinguishable, and hence, easier to reject when rated with PQI than with PQI_{Corr}. In fact, nearly all the arrhythmic beats are rated with a PQI below 0.6 while PQI_{Corr} distribution has the same trend for arrhythmic and sinus rhythm beats (figure 8).

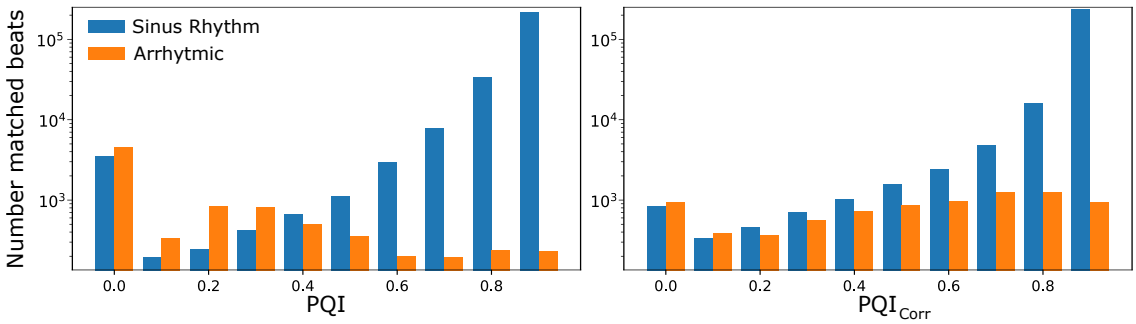

Figure 8. Quality index histograms. Distributions of the PQI (left) and PQI_{Corr} (right) for the beats matched with the reference in the AF dataset. The two classes are more easily separable using PQI than with PQI_{Corr}

The higher class separability between sinus rhythm and arrhythmic pulses using the PQI is reflected in the arrhythmic beat rejection results. The PQI is able, as desired, to decrease the overall sensitivity to arrhythmic beats—specifically to PACs and PVCs—to below 7% while with the PQI_{Corr} the SEN_{Arr} reaches 12% for a threshold of 0.8, i.e. close to the quality index suitable for sinus rhythm beats. Furthermore, this means that in order to have a proportion of arrhythmic beats lower than 12% using PQI_{Corr} one would have to sacrifice the sensitivity to sinus rhythm beats, making the SEN-SEN_{Arr} trade-off more disadvantageous than for PQI. In fact, PQI is able to guarantee, for every threshold, a lower presence of arrhythmic beats while maintaining a higher sensitivity to sinus rhythm beats. An example of the higher discriminative capacity of PQI can be seen in figure 5. The comparison between PP_{PQI} and the template in figure 5(c) by means of the Pearson's correlation coefficient overestimates the pulse quality by assigning a PQI_{Corr} of 0.89—indicating high similarity with the template—even though it is derived from the merging of a sinus rhythm PPG pulse with a premature ventricular contraction. Instead, the PQI correctly assigns a low quality index value to most similar pulse cases (in this example PQI is zero). Changing the similarity metric to a more sensitive one, such as the pulse-template distance derived via classical DTW (Li & Clifford 2012), can resolve the pulse quality overestimation, but at the same time it might increase the chance of underestimating the quality of uncorrupted pulses. For instance in figure 5(b), the difference between the pulse and the template is not due to signal corruption or abnormal HR behavior, but the DTW distance increases by 20% with respect to figure 5(a), where the difference between template and pulse is almost absent.

The SEN_{Arr} of the PQI is better than that for PQI_{Corr} when calculated across all beats as well as when calculated per-subject. In fact, the PQI offers a per-subject SEN_{Arr} three times lower with a twice as low standard deviation compared to the PQI_{Corr}, indicating a lower between-subject variability of the results. This does not only guarantee a higher rejection of arrhythmic beats in each subject, but also a higher consistency in the results, suggesting a higher generalizability of the PQI approach. Moreover, the SEN_{Arr} of PQI is similar among the two arrhythmic conditions under evaluation, i.e. PAC and PVC, while the PQI_{Corr} tends to have a larger difference in performance.

The rejection of arrhythmic beats is higher when the recordings are mainly characterized by PVCs rather than PACs. This can be explained by the morphological differences in the PPG signal for pulses generated by the PVC and PAC beats. A pulse related to a PVC often tends to be merged with the sinus rhythm pulse that precedes it, resulting in a single PP_{PQI} (e.g. figure 5(c)). This occurs less frequently in the case of PACs. In fact, in the PVC recordings, 99% of the beats preceding the actual PVC beats were assigned a PVC label as compared to only 55% of PAC beats. The PPs_{PQI} resulting from the merging of a sinus rhythm and PVC related pulses have a morphology that is significantly different than the generated templates and, therefore, are easier to correctly classify as low-quality pulses.

One of the limitations of the PQI is the inability to properly assess the normal sinus rhythm likelihood of every single beat/pulse in case an arrhythmic condition is continuously present in the PPG recording. However, in case of AFib the PQI is significantly lower than the PQI obtained for the sinus rhythm beats. This means that also when no normal sinus rhythm pulses are present, the algorithm can quantify the irregularity of the recording and assign a lower PQI to most of the pulses. This measure can, therefore, be used to exclude a subject completely from further analysis in case a PPG recording presents an overall low average PQI. The algorithm needs to be developed further to achieve a similar rejection rate during AFib as with PAC and PVC. One future development will be to separate the template creation from the PPs_{Temp} in case of overall low average PQI in order to increase the morphological difference between the arrhythmic pulses and the template. These expedients cannot be used in case of atrial flutter since this condition produces a significant reduction in contraction variability, which leads to pulses with an extremely regular morphology, as confirmed by an AFlu SEN_{Arr} higher than the SEN for every threshold (table 2). In fact, the PQI distribution of beats from the AFlu subject is significantly higher than the PQI for any other cardiac condition and sinus rhythm pulsations. Even though the PQI cannot be used to flat the potential presence of a condition in the case of AFlu, it behaves as expected: regular pulses are classified with a higher quality index. To correct for this drawback, a standard 24-hour Holter ECG monitoring could be used to confirm the suspicion of an atrial flutter.

Naturally, if arrhythmic beats or pulses are excluded when aiming to exclude pulses corrupted by movement artifacts, the resulting time series is not usable for arrhythmia detection. However, the information in the PQI can be used as an additional indicator for arrhythmia detection or to pre-select pulses deviating from the sinus rhythm.

Another limitation of our algorithm is that it is not suitable for online, real-time assessment of PPG pulse quality. However, by changing, for instance, the number of pulses required to derive the template, it could be adapted for such an application, with a trade-off between assessment delay (due to template build-up) and the power of the PQI in discriminating sinus rhythm and arrhythmic or corrupted beats.

6. Conclusion

We present a new beat detection algorithm for wrist-worn PPG signals. The algorithm allows the precise identification of normal sinus rhythm beats in a PPG signal recorded by a wrist-worn device. This is achieved by segmenting the signal into individual pulses and dynamically evaluating their quality via a template comparison approach. The algorithm can be employed to obtain inter-beat intervals or PPG pulse series comprising only of normal sinus rhythm beats by excluding artifacts and pulses corrupted by external or arrhythmic phenomena, and hence enabling, for instance, the calculation of heart rate variability features or PPG derived respiration. Moreover, the pulse quality index can be used according to the application needs: for instance, a higher quality threshold can be used when analyzing the morphology of the PPG puls; instead, a lower

threshold can be set in case of beat detection where the sporadic inclusion of a wrong beat can be corrected during post-processing of inter-beat intervals.

Future work on the algorithm will focus on a more application-specific pulse quality index calculation, for instance by weighing the mismatch error according to its location on the PPG pulse. This could help increasing the sensitivity to sinus rhythm beats in the cases where, e.g., only the rising slope of the PPG pulse is necessary for feature extraction.

Acknowledgments

This research was supported by STW/IWT in the context of the OSA+ project (No. 14619). G.B.P. would like to thank X.L. Aubert for contributing to the algorithm development. The authors would like to express their gratitude to L. Dekker and his team at the Catharina Hospital for help with collecting the patient data.

References

Aboy, M., McNames, J., Thong, T., Tsunami, D., Ellenby, M. S. & Goldstein, B. (2005), 'An automatic beat detection algorithm for pressure signals', *IEEE Transactions on Biomedical Engineering* **52**(10), 1662–1670.

Allen, J. (2007), 'Photoplethysmography and its application in clinical physiological measurement', *Physiological Measurement* **28**(3), R1.

ANSI-AAMI (1998), *ANSI/AAMI EC57: Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithm.*

Bonomi, A. G., Schipper, F., Eerikäinen, L. M., Margarito, J., Aarts, R. M., Babaeizadeh, S., de Morree, H. M. & Dekker, L. (2016), Atrial fibrillation detection using photo-plethysmography and acceleration data at the wrist, *in* '2016 Computing in Cardiology Conference (CinC)', pp. 277–280.

Boudaoud, S., Rix, H. & Meste, O. (2005), 'Integral shape averaging and structural average estimation: a comparative study', *IEEE Transactions on Signal Processing* **53**(10), 3644–3650.

Brüser, C., Winter, S. & Leonhardt, S. (2013), 'Robust inter-beat interval estimation in cardiac vibration signals', *Physiological Measurement* **34**(2), 123.

Camm, A. J., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., Coumel, P., Fallen, E., Kennedy, H., Kleiger, R. et al. (1996), 'Heart rate variability: standards of measurement, physiological interpretation and clinical use', *Circulation* **93**(5), 1043–1065.

Charlton, P. H., Bonnici, T., Tarassenko, L., Alastruey, J., Clifton, D. A., Beale, R. & Watkinson, P. J. (2017), 'Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants', *Physiological Measurement* **38**(5), 669.

Clifford, G. D. (2002), Signal Processing Methods for Heart Rate Variability, PhD thesis, Department of Engineering Science, University of Oxford.

Clifford, G. D., Silva, I., Moody, B., Li, Q., Kella, D., Shahin, A., Kooistra, T., Perry, D. & Mark, R. G. (2015), The PhysioNet/Computing in Cardiology Challenge 2015: Reducing false arrhythmia alarms in the ICU, *in* '2015 Computing in Cardiology Conference (CinC)', pp. 273–276.
URL: *https://github.com/MIT-LCP/PhysioNetChallengePublic*

Eapen, Z. J., Turakhia, M. P., McConnell, M. V., Graham, G., Dunn, P., Tiner, C., Rich, C., Harrington, R. A., Peterson, E. D. & Wayte, P. (2016), 'Defining a mobile health roadmap for cardiovascular health and disease', *Journal of the American Heart Association* **5**(7).

Eerikäinen, L. M., Dekker, L., Bonomi, A., Vullings, R., Schipper, F., Margarito, J., de Morree, H. & Aarts, R. (2017), Validating features for atrial fibrillation detection from photoplethysmogram

under hospital and free-living conditions, *in* '2017 Computing in Cardiology Conference (CinC)', p. IN PRESS.

Ellis, R. J., Zhu, B., Koenig, J., Thayer, J. F. & Wang, Y. (2015), 'A careful look at ECG sampling frequency and R-peak interpolation on short-term measures of heart rate variability', *Physiological Measurement* **36**(9), 1827.

Fischer, C., Dömer, B., Wibmer, T. & Penzel, T. (2017), 'An algorithm for real-time pulse waveform segmentation and artifact detection in photoplethysmograms', *IEEE Journal of Biomedical and Health Informatics* **21**(2), 372–381.

Fonseca, P., Aarts, R. M., Foussier, J. & Long, X. (2014), 'A novel low-complexity post-processing algorithm for precise QRS localization', *SpringerPlus* **3**(1), 376.

Fonseca, P., den Teuling, N., Long, X. & Aarts, R. M. (2018), 'A comparison of probabilistic classifiers for sleep stage classification', *Physiological Measurement* **39**(5), 055001.

Fonseca, P., Weysen, T., Goelema, M. S., Møst, 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**(7), 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', *Physiological Measurement* **26**(6), 1093.

Gil, E., Laguna, P., Martínez, J. P., Barquero-Pérez, Ó., García-Alberola, A. & Sörnmo, L. (2013), 'Heart rate turbulence analysis based on photoplethysmography', *IEEE Transactions on Biomedical Engineering* **60**(11), 3149–3155.

Hickey, M., Phillips, J. P. & Kyriacou, P. A. (2015), 'The effect of vascular changes on the photoplethysmographic signal at different hand elevations', *Physiological Measurement* **36**(3), 425.

Karlen, W., Kobayashi, K., Ansermino, J. M. & Dumont, G. A. (2012), 'Photoplethysmogram signal quality estimation using repeated gaussian filters and cross-correlation', *Physiological Measurement* **33**(10), 1617.

Lázaro, J., Gil, E., Bailón, R. & Laguna, P. (2011), Deriving respiration from the pulse photoplethysmographic signal, *in* '2011 Computing in Cardiology', pp. 713–716.

Li, Q. & Clifford, G. D. (2012), 'Dynamic time warping and machine learning for signal quality assessment of pulsatile signals', *Physiological Measurement* **33**(9), 1491.

Linder, S., Wendelken, S., Wei, E. & McGrath, S. (2006), 'Using the morphology of photoplethysmogram peaks to detect changes in posture', *Journal of Clinical Monitoring and Computing* **20**(3), 151–158.

Mann, H. B. & Whitney, D. R. (1947), 'On a test of whether one of two random variables is stochastically larger than the other', *Ann. Math. Statist.* **18**(1), 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', *Journal of Electrocardiology* **40**(3), 228 – 234.e8.

Nano, M., Papini, G., Fonseca, P., Vullings, R., Overeem, S., Bergmans, J. & Aarts, R. (2017), Comparing inter beat and inter pulse intervals from ECG and PPG signals during sleep, *in* 'Poster session presented at 11th Biomedica Summit, May 9-10, 2017, Eindhoven, Netherlands, Eindhoven, Netherlands'.

Nijboer, J. A., Dorlas, J. C. & Mahieu, H. F. (1981), 'Photoelectric plethysmography-some fundamental aspects of the reflection and transmission methods', *Clinical Physics and Physiological Measurement* **2**(3), 205.

Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D. & Tarassenko, L. (2015), 'Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring', *IEEE Journal of Biomedical and Health Informatics* **19**(3), 832–838.

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, *in* '2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)', pp. 117–120.

Peltola, M. (2012), 'Role of editing of R-R intervals in the analysis of heart rate variability', *Frontiers in Physiology* **3**, 148.

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 Transactions on Biomedical Engineering* **50**(10), 1143–1151.

Petitjean, F., Forestier, G., Webb, G. I., Nicholson, A. E., Chen, Y. & Keogh, E. (2014), Dynamic time warping averaging of time series allows faster and more accurate classification, *in* '2014 IEEE International Conference on Data Mining', pp. 470–479.

Piwek, L., Ellis, D. A., Andrews, S. & Joinson, A. (2016), 'The rise of consumer health wearables: Promises and barriers', *PLOS Medicine* **13**(2), 1–9.

Radha, M., Zhang, G., Gelissen, J., de Groot, K., Haakma, R. & Aarts, R. M. (2017), 'Arterial path selection to measure pulse wave velocity as a surrogate marker of blood pressure', *Biomedical Physics & Engineering Express* **3**(1), 015022.

Redmond, S. J., Xie, Y., Chang, D., Basilakis, J. & Lovell, N. H. (2012), 'Electrocardiogram signal quality measures for unsupervised telehealth environments', *Physiological Measurement* **33**(9), 1517.

Russo, K., Goparaju, B. & Bianchi, M. T. (2015), 'Consumer sleep monitors: is there a baby in the bathwater?', *Nature and science of sleep* **7**, 147.

Sakoe, H. & Chiba, S. (1978), 'Dynamic programming algorithm optimization for spoken word recognition', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**(1), 43–49.

Schäfer, A. & Vagedes, J. (2013), 'How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram', *International Journal of Cardiology* **166**(1), 15 – 29.

Scholz, E. P., Kehrle, F., Vossel, S., Hess, A., Zitron, E., Katus, H. A. & Sager, S. (2014), 'Discriminating atrial flutter from atrial fibrillation using a multilevel model of atrioventricular conduction', *Heart Rhythm* **11**(5), 877 – 884.

Shelley, K. H., Dickstein, M. & Shulman, S. M. (1993), 'The detection of peripheral venous pulsation using the pulse oximeter as a plethysmograph', *Journal of Clinical Monitoring* **9**(4), 283–287.

Sološenko, A., Petrénas, A., Marozas, V. & Sörnmo, L. (2017), 'Modeling of the photoplethysmogram during atrial fibrillation', *Computers in Biology and Medicine* **81**(Supplement C), 130 – 138.

Sukor, J. A., Redmond, S. J. & Lovell, N. H. (2011), 'Signal quality measures for pulse oximetry through waveform morphology analysis', *Physiological Measurement* **32**(3), 369.

Tamura, T., Maeda, Y., Sekine, M. & Yoshida, M. (2014), 'Wearable photoplethysmographic sensorspast and present', *Electronics* **3**(2), 282–302.