Benchmarking Photoplethysmography Peak Detection Algorithms Using the Electrocardiogram Signal as a Reference

Kevin Kotzen[1], Peter H Charlton[2], Amir Landesberg[1] and Joachim A. Behar[1]

[1] Technion Israel Institute of Technology, Haifa, Israel
[2] University of Cambridge, Cambridge, United Kingdom

Abstract

Introduction: *Photoplethysmography (PPG) is fast becoming the signal of choice for the widespread monitoring of sleep metrics obtained by wearable devices. Robust peak detection is critical for the extraction of meaningful features from the PPG waveform. There is however no consensus on what PPG peak detection algorithms perform best on nocturnal continuous PPG recordings. We introduce two methods to benchmark the performance of PPG peak detectors.* ***Methods:*** *We make use of data where nocturnal PPG and electrocardiogram (ECG) are measured synchronously. Within this setting, the ECG, a signal for which there are established R-peak detectors, is used as reference. The first method for benchmarking, denoted "Peak Matching", consists of forecasting the expected position of the PPG peaks using the ECG R-peaks as reference. The second technique, denoted "IHR-IPR Accuracy", compares the instantaneous pulse rate (IPR) extracted from the PPG with the instantaneous heart rate (IHR) extracted from the ECG. For benchmarking, we used the MESA dataset consisting of 2,055 overnight polysomnography recordings with a combined length of over 16,300 hours. Four open PPG peak detectors were benchmarked.* ***Results:*** *The "Pulses" detector performed best with a Peak Matching F1-score of 0.94 and an IHR-IPR Accuracy of 89.6%.* ***Discussion and conclusion:*** *We introduced two new methods for benchmarking PPG peak detectors. Among the four detectors evaluated, "Pulses" performed best. Benchmarking of further PPG detectors and on other data source (e.g. daytime recordings, recordings from patients with arrhythmia) is needed.*

1. Introduction

Robust photoplethysmography (PPG) peak detection is an important step in PPG based evaluation of sleep measures such as sleep staging from PPG. The time series generated from consecutive peak-to-peak timings of the PPG is defined as the instantaneous pulse rate (IPR). The IPR is known to be similar to the instantaneous heart rate (IHR) signal evaluated from the time intervals between consecutive R-peaks of the ECG [1]. The IPR signal is however only as reliable as the PPG peak detection algorithm used and therefore using the right peak detector is critical to good performance. There is however no consensus on which PPG peak detection algorithm performs best. In this work we developed two methods to benchmark PPG peak detection algorithms and benchmark the performance of algorithms on a large dataset of PPG and ECG signals collected during home polysomnography (PSG).

2. MESA Dataset

The Multi-Ethnic Study of Atherosclerosis (MESA) sleep database [6,7] contains 16,300 hours of full overnight PSG measured from 2,237 patients aged 54-95 years old. PSG recordings were carried out in-home using the Compumedics Somte System. Measurements include ECG, and finger PPG (Nonin 8000 sensor). ECG and PPG signals are sampled at 256Hz. The first 15 minutes of each patient recording was excluded from this study as this period tends to be particularly noisy, given the patient was settling in.

3. Peak Detection

A number of different approaches for PPG peak detection have been published [2, 3, 5, 8]. We evaluated four open source PPG peak detectors listed in Table 1. PPG signals were prefiltered using a zero-phase Butterworth filter, 4th order, bandpass at 0.5Hz-8Hz [9, 10]. After filtering, signals were windowed into segments of 100s with a 10s overlap between consecutive windows. Windowing was necessary to reduce computational complexity and memory constraints faced by certain peak detectors when processing very long PPG recordings. Each peak detector was considered as a black box and was used with the author suggested or default settings apart from the common prefiltering band (0.5Hz-8Hz). Matlab 2020a and Python 3.7 were used as required. An example of the peaks detected is shown in Figure 1.

Table 1. List of PPG peak detector algorithms included in this study.

#	Detector	Toolbox	Implementing Author	Original Author	Ref	Code
1	Pulses	Ecg-Kit	M. Llamedo Soria	J. Lázaro	[2]	Matlab
2	Heartpy	Heartpy	P. van Gent	P. van Gent	[3]	Python
3	CO_PPG	RRest	P. Charlton	C. Orphanidou	[4]	Matlab
4	AdaptPulseSegment	RRest	M. Pimentel	W. Karlen	[5]	Matlab

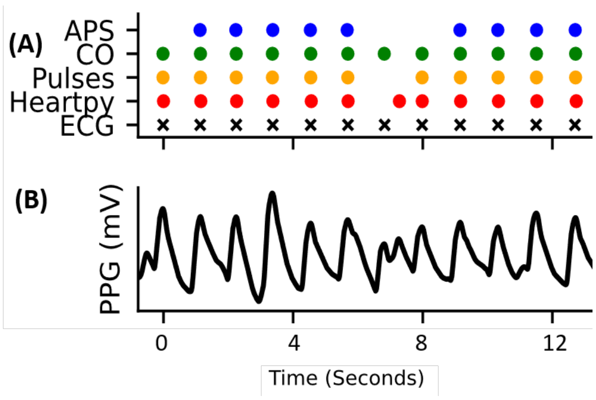

Figure 1. (A) Comparison between PPG-Peaks detected by four open access PPG peak detectors. The reference ECG R-peaks were aligned with the PPG peaks as described in Section 4.1; (B) the input PPG signal.

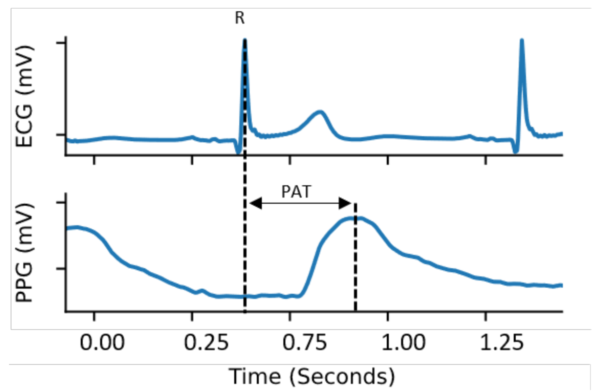

Figure 2. Diagram describing an example the PAT between the ECG and PPG signals.

4. Benchmarking Methods

Each detector was evaluated for its ability to (1) robustly detect PPG peaks and (2) generate an accurate IPR signal. The first evaluation metric was based on matching the peak of the ECG-R-wave (ECG-Peak) with the systolic peak of the PPG (PPG-Peak) and gives an indication of the detector's ability to detect a peak where one should exist. The second is based on the degree to which the IPR signal matches the IHR reference-signal.

4.1. Peak Matching

This technique uses the positioning of the ECG-Peaks to forecast where a PPG-Peak should be expected on the PPG signal. It provides an indication of the peak detector's ability to detect a PPG-Peak where one should exist. There is a variable delay between the ECG-Peak and the PPG-Peak caused by the time it takes for the heart to eject blood and for the pulse to travel from the heart to the point of measurement [11]. This delay is defined here as the pulse arrival time (PAT) as shown in Figure 2. The PAT is relatively stable over short time periods when the patient is in a steady state [11].

The procedure for Peak Matching consists of the following steps: (1) Extract all ECG-Peaks using epltd [12] a state-of-art ECG peak detection algorithm, and xqrs [13], a weaker ECG peak detection algorithm. (2) Estimate the ECG signal quality index (SQI) for non-overlapping 30s windows using the *bsqi* algorithm from PhysioZoo with default paramaters [14] and epltd as the reference input and xqrs as the test input [14]. *bsqi* computes the F1-score, the harmonic mean of the precision and recall as described in Equation 1, between the ECG-forecast and the PPG-Peaks. (3) Extract all PPG-Peaks from the PPG signal. (4) Calculate the instantaneous PAT as the time elapsed between the each ECG-Peak and the next annotated PPG-Peak. (5) Remove PAT values that are not physiologically possible by limiting the PAT values to a range of 200ms to 540ms, an adjustment from the median PAT value of 480ms reported by Rajala [15]. (6) Smooth the PAT signal with a rolling average filter of 600 points, representing approximately 10 minutes, to reduce the effect of incorrectly annotated PPG-Peaks and other mistakes in the measurement of the PAT. An example of the resulting overnight PAT signal is shown in Figure 3. (7) Following calculation of the rolling-average-PAT, forecast the expected position of the PPG-Peaks from the ECG-Peaks, by adding the instantaneous PAT to the ECG-Peaks. This technique was used in Figure 1 to align the PPG-Peaks with the ECG-Peaks. (8) Compute the relative strength of the PPG-Peak detector using the *bsqi* algorithm from PhysioZoo with parameter agreement_window=0.1 [14]. (9) Remove all windows where the ECG SQI is less than 80% (determined experimentally by visually analyzing a number of PPG waveforms and associated scores) and calculate the mean for all the remaining 30s windows.

$$F_1 = \frac{2 * Se * PPV}{Se + PPV} \quad (1)$$

Where Se is sensitivity and PPV is positive predictive value.

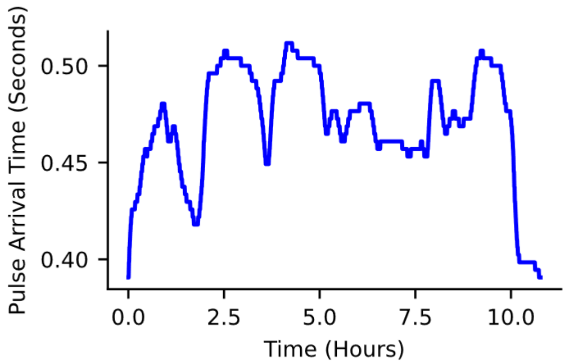

Figure 3. Diagram showing the overnight PAT signal extracted using the ECG-Peaks and PPG-Peaks.

4.2. IHR-IPR Accuracy

In order to test the ability of each peak detector to generate an accurate IPR signal, the IPR is compared to the IHR.

The procedure for IHR-IPR Accuracy consists of the following steps: (1) Follow Steps 1 and 2 from Section 4.1. (2) From the epltd ECG-Peaks extract the peak-to-peak intervals (RR) and calculate a bidirectional-rolling-average-filtered-RR as described in the PhysioZoo function *filtrr.m* with parameters win_samples=10 and outliers=50 [14]. (3) Convert the filtered-RR signal into an IHR signal using 60/RR-filtered and resample the IHR to 2Hz. (4) Extract all PPG-Peaks from the PPG signal. (5) Extract the peak-to-peak intervals from the PPG-Peaks (PP). (6) Generate an IPR signal from the PP using 60/PP and resample the IPR to to 2Hz. (7) Shift the IPR signal back by 460ms to roughly account for the PAT [15]. (8) Build non-overlapping 30s windows and for each 30s window. (9) Using an IHR upper and lower band of ±5bpm above and below the filtered IHR signal calculate an accuracy-score by counting the ratio of time that the IPR signal is inside the IHR band. (10) Remove all windows where the ECG SQI is less than 80% and calculate the mean for all the remaining 30s windows.

5. Results

One patient had no valid PPG channel and was discarded. Results are presented in Table 2 and Table 3 for

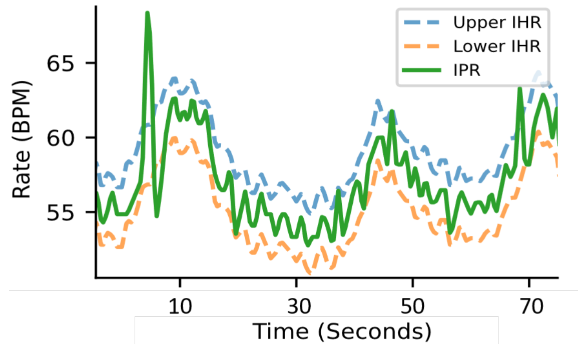

Figure 4. Diagram showing upper and lower IHR bounds with the IPR signal running in the middle.

2,055 out of the total of 2,056 patients in the MESA Sleep database. The "Pulses" detector performed best with a Peak Matching F1-score of 0.940 and an IHR-IPR Accuracy of 89.6%.

Table 2. Results from ECG to PPG Peak Matching

Annotator	PPV	Se	F1	Rank
Pulses	0.944	0.940	0.940	1
CO_PPG	0.951	0.916	0.930	2
Heartpy	0.940	0.919	0.927	3
AdaptPulseSegment	0.939	0.890	0.908	4

Table 3. Results from IHR and IPR Accuracy

Annotator	Accuracy	Rank
Pulses	89.6	1
CO_PPG	87.2	2
Heartpy	86.0	3
AdaptPulseSegment	84.8	4

6. Discussion

As can be seen in Figure 1, not all PPG-Peak detectors perform equally, and each detector has its own strengths and weaknesses. For example, Pulses sometimes fails to detect peaks after a sudden change in amplitude, while AdaptPulseSegment can detect false peaks if the signal is noisy.

All the detectors performed reasonably well when the PPG signals has no noise or anomalies. The results of the Peak Matching technique are presented in Table 2 and show an F1-score value ranging from 0.908 to 0.940. The result of the IHR-IPR Accuracy technique are in the range of 84.8% to 89.6% as shown in Table 3. Both techniques show the same rank ordering, as shown in the *Rank* column. The strict requirement that the IPR signal fall within

±5bpm of the IHR, results in lower than expected IHR-IPR Accuracy. Relaxing this requirement to ±15bpm did not change the rank ordering.

The techniques used here have not been explicitly tested on any irregular heart patterns or beats. For example, we did not look at the impact of arrhythmia (e.g. ectopic beats, atrial fibrillation). Furthermore, findings are for nocturnal recorded data and thus there may be variation of the detectors ranking with data recorded during daytime or specific activities.

7. Conclusion

Four PPG peak detection algorithms were benchmarked on their performance for robust peak detection on PPG signals collected during full overnight polysomnography. The PPG peak detectors were compared using Peak Matching and IHR-IPR. Peak Matching forecast the expected position of the PPG-Peak from the ECG-Peak and PAT and then looked at the agreement. IHR-IPR Accuracy compared the IHR signal from the ECG-Peaks and the IPR signal from the PPG-Peaks and calculated the time ratio of time that IPR is inside vs outside a given boundary of the IHR. The Pulses peak detector performed best on our nocturnal PPG recordings with an F1-score score of 0.940 for Peak Matching and an IHR-IPR Accuracy of 89.6%. It is suggested that the Pulses algorithms be used as the peak detector of choice on overnight PPG recordings.

Acknowledgements

KK and JB acknowledge an anonymous Technion UK donor. PC acknowledges the British Heart Foundation [FS/20/20/34626]

References

[1] Dehkordi P, Garde A, Karlen W, Wensley D, Ansermino JM, Dumont GA. Sleep stage classification in children using photoplethysmogram pulse rate variability. In Computing in Cardiology 2014. IEEE. ISBN 1479943479, 2014; .

[2] Lázaro J, Gil E, Vergara JM, Laguna P. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. IEEE Journal of Biomedical and Health Informatics 1 2014;18(1):240–246. .

[3] van Gent P, Farah H, van Nes N, van Arem B. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. Transportation Research Part F Traffic Psychology and Behaviour 10 2019;66:368–378. .

[4] Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L. Signal Quality Indices for the Electrocardiogram and Photoplethysmogram: Derivation and Applications to Wireless Monitoring. IEEE Journal of Biomedical and Health Informatics 2014;1–1. .

[5] Karlen W, Ansermino JM, Dumont G. Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. ISBN 978-1-4577-1787-1. ISSN 1557170X, 8 2012; .

[6] Dean DAA, Goldberger ALL, Mueller R, Kim M, Rueschman M, Mobley D, Sahoo SSS, Jayapandian CPP, Cui L, Morrical MGG, Surovec S, Zhang GQ, Redline S. Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource. Sleep 5 2016; 39(5):1151–1164. .

[7] Zhang GQ, Cui L, Mueller R, Tao S, Kim M, Rueschman M, Mariani S, Mobley D, Redline S. The National Sleep Research Resource: towards a sleep data commons. Journal of the American Medical Informatics Association 10 2018; 25(10):1351–1358. .

[8] Scholkmann F, Boss J, Wolf M. An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals. Algorithms 11 2012;5(4):588–603. .

[9] Park C, Shin H, Lee B. Blockwise PPG Enhancement Based on Time-Variant Zero-Phase Harmonic Notch Filtering. Sensors 4 2017;17(4):860. .

[10] Tanweer KT, Hasan SR, Kamboh AM. Motion artifact reduction from PPG signals during intense exercise using filtered X-LMS. In Proceedings - IEEE International Symposium on Circuits and Systems. Institute of Electrical and Electronics Engineers Inc. ISBN 9781467368520. ISSN 02714310, .

[11] Jago JR, Murray A. Repeatability of peripheral pulse measurements on ears, fingers and toes using photoelectric plethysmography. Clinical Physics and Physiological Measurement 1988;9(4):319–329. .

[12] Hamilton P. Open source ECG analysis. In Computers in Cardiology. IEEE. ISBN 0-7803-7735-4; .

[13] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 6 2000;101(23). .

[14] Behar JA, Rosenberg AA, Weiser-Bitoun I, Shemla O, Alexandrovich A, Konyukhov E, Yaniv Y. PhysioZoo: a novel open access platform for heart rate variability analysis of mammalian electrocardiographic data. Frontiers in physiology 2018;9:1390. .

[15] Rajala S, Ahmaniemi T, Lindholm H, Taipalus T. Pulse arrival time (PAT) measurement based on arm ECG and finger PPG signals - Comparison of PPG feature detection methods for PAT calculation. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. Institute of Electrical and Electronics Engineers Inc. ISBN 9781509028092. ISSN 1557170X, 9 2017; .

Address for correspondence:

Email: kkotzen@campus.technion.ac.il