Accelerometry-Guided Inter-Beat-Interval Assessment from Wrist Photoplethysmography

Peter H Charlton[1,2], Joachim A Behar[3], Márton Áron Goda[3], Jonathan Mant[1], Panicos A Kyriacou [2]

[1] University of Cambridge, Cambridge, United Kingdom
[2] City, University of London, London, United Kingdom
[3] The Faculty of Biomedical Engineering, Technion-IIT, Haifa, Israel

Abstract

Wearable photoplethysmography devices such as smartwatches can detect possible arrhythmias from inter-beat intervals (IBIs). However, photoplethysmogram (PPG) signals are highly susceptible to motion artifact. This study investigated using simultaneous accelerometry signals to determine whether IBIs can be reliably measured from PPG signals. The PPG-DaLiA and WESAD datasets were used. These datasets contain wrist accelerometry and PPG signals collected from 15 subjects during activities of daily living and mental stress tasks. IBIs were estimated from PPG signals using the 'MSPTD' beat detection algorithm. PPG-based IBIs were deemed accurate if the resulting instantaneous heart rate (IHR) was within $\pm$ 5 bpm of a reference ECG-derived IHR. The mean absolute deviation (MAD) of the accelerometry signal was able to predict whether PPG-derived IBIs were accurate, with an area under the precision-recall curve (AUPRC) of 0.82 on all data. An optimal MAD threshold of 12.9 milli-gravitational units was identified. However, performance was poorer during stress (AUPRC of 0.54). In conclusion, accelerometry can be used to identify periods when IBIs can be accurately measured from PPG signals during activities associated with movement, but is not reliable during stress.

1. Introduction

Wearable devices equipped with photoplethysmography sensors provide opportunity to monitor inter-beat intervals (IBIs) unobtrusively in daily life. Smartwatches, hearables, and smart rings typically use photoplethysmography for heart rate monitoring. IBIs, the intervals between successive heartbeats, can also be estimated from the photoplethysmogram (PPG) signal. IBIs can be used in a range of applications, from identifying potential arrhythmias to assessing stress levels [1]. However, PPG signals are highly susceptible to noise, particularly during motion, and this can affect the accuracy of IBI estimation [2].

Broadly, there are two approaches to dealing with noisy PPG signals: the first is to remove motion artifact from a signal so it can be used for analysis; the second is to assess signal quality and only retain high quality (*i.e.* low noise) periods for analysis [3]. Motion artifact removal has the advantage of enabling analysis during exercise when it is difficult to obtain high-quality signals, but the disadvantage that only an approximate PPG signal is available for analysis. Conversely, signal quality assessment has the advantage of using the measured signal for analysis, but does not enable analysis in the presence of noise.

An alternative approach could be to use simultaneous accelerometry signals to identify periods when IBIs can be accurately estimated from PPG signals. This is based on the assumption that inaccuracies in IBIs are caused by movement-induced noise (*i.e.* motion artifact). This is a practical approach since many wearables have accelerometers. Indeed, this approach has been used in Apple and FitBit devices to identify periods of PPG for atrial fibrillation detection [4, 5]. However, to our knowledge there has been little research into how to use accelerometry signals to identify periods when IBIs can be estimated accurately.

The aim of this study was to investigate whether accelerometry signals could be used to predict whether or not IBIs could be accurately measured from simultaneous PPG signals at the wrist. The objectives were to: (i) assess the classification performance of such an approach; (ii) identify an optimal classification threshold; and (iii) assess performance across different activities.

2. Methods

2.1. Datasets

The publicly available PPG-DaLiA and WESAD datasets were used [6, 7]. These contain accelerometry and PPG signals recorded at 64 Hz at the wrist from 15 subjects using an Empatica E4, alongside ECG signals recorded at 75 Hz at the chest using a RespiBAN Professional.

ISSN: 2325-887X DOI: 10.22489/CinC.2023.046

The PPG-DaLiA dataset contains data collected during a protocol of activities of daily living. In this study we used data collected during all the available activities, consisting of: sitting, working, walking, taking a lunch break, stair climbing, car driving, table soccer, and cycling. The WESAD dataset contains data collected during a protocol designed to induce mental stress. In this study we used data from baseline, amusement, stress, and meditation phases (the second meditation phase was not used in this study).

2.2. Inter-beat interval estimation

IBIs were estimated from PPG signals as follows. First, PPG signals were segmented into 20 s windows with 5 s overlap. Second, beats were detected using the 'MSPTD' beat detection algorithm [8], which has previously been found to perform well [9]. Third, the middle-amplitude points of each systolic upslope were identified from the onset and peak locations provided by the beat detector. These were used as they have been found to be more suitable for pulse rate variability analysis than onsets or peaks [10]. Fourth, repeated beat detections due to overlapping windows were removed. Finally, IBIs were calculated as the time delays between successive middle-amplitude points.

Reference IBIs were derived from ECG signals as follows. The PPG-DaLiA dataset contains manual beat annotations which were used in this study. The WESAD dataset does not contain manual annotations, so beats were detected as described in [9]: (i) beats were detected using two separate ECG beat detectors ('jqrs' and 'rpeakdetect' algorithms [11–13]); and (ii) 'correct' beats were identified as those which both beat detectors detected within 150 ms of each other. Any 20 s windows in which the two beat detectors did not agree were excluded from the analysis.

PPG and ECG signals were not necessarily precisely time-aligned, so PPG- and ECG-derived IBIs were time-aligned using a slight modification to our approach described in [9]. Briefly, previously we found the optimal lag between signals as the lag which resulted in the highest proportion of PPG beats aligning with ECG beats (within $\pm$ 150 ms). In this work we improved the approach to re-calculate the optimal lag for every 300 ECG beats, thereby accounting for clocks drifting during a recording.

Any periods of missing ECG or PPG signals were excluded, as identified by a flat line lasting $>$0.2s.

2.3. Assessing level of movement

The level of movement was quantified from accelerometry signals by calculating the mean absolute deviation (MAD) over 5 s windows. The MAD is calculated from tri-axial accelerometry as follows [14]. First, the resultant acceleration (r_i) is calculated from accelerations in x, y and z directions, in milli-gravitational (mg) units, as

$$r_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad . \tag{1}$$

Then, the MAD is calculated as

$$MAD = \frac{1}{n}|r_i - \overline{r}| \quad . \tag{2}$$

This approach has been previously recommended to classify physical activities according to their intensity [14, 15].

2.4. Statistical analysis

IBIs were deemed accurate if the instantaneous heart rate (IHR) calculated from PPG-derived IBIs was within $\pm$ 5 bpm of the ECG-derived IHR. The performance of the MAD for classifying PPG periods according to whether or not IBIs could be accurately estimated was quantified using the areas under the receiver operator curve (AUROC) and the precision-recall curve (AUPRC). The accuracy of IBIs was summarised using mean absolute error (MAE).

3. Results

3.1. Dataset characteristics

Table 1 summarises the datasets. In total, 2,156 mins of data were included in the analysis, and PPG-derived IBIs were correct for 59.1% of this time. The proportion of IBIs which were correct varied greatly between activities, from 93.7% during meditation to 17.4% during stair climbing.

Table 1. Dataset characteristics.

Dataset	No. subjs	Duration per subj (mins)	Prop. IBIs correct (%)
meditation	15	6.3 (6.1-6.3)	93.7
amusement	15	5.8 (5.8-5.8)	80.4
baseline	15	19.1 (18.8-19.3)	76.0
stress	15	10.3 (10.0-10.8)	34.5
sitting	15	9.7 (9.7-9.9)	83.3
working	14	19.8 (19.7-20.4)	72.0
lunch break	14	32.3 (28.7-37.2)	60.2
car driving	15	15.0 (14.1-15.7)	59.9
cycling	15	7.7 (6.7-8.1)	39.5
table soccer	15	4.7 (4.5-5.2)	27.7
walking	14	10.7 (9.4-11.5)	25.9
stair climbing	15	7.4 (6.7-7.7)	17.4

3.2. Classification performance

Table 2 presents the results on the performance of using the MAD to predict whether PPG-derived IBIs were accurate. It produced an AUROC of 0.78 and an AUPRC of 0.82 across the entire dataset.

Table 2. Performance of using MAD to predict whether PPG-derived IBIs were accurate.

Dataset	AUROC	AUPRC
Combined	0.78	0.82
Mental stress	0.82	0.88
Daily living	0.77	0.79

3.3. Identifying an optimal threshold

Optimal thresholds were identified as those which produced IBI MAEs of <5 bpm. Figure 1 shows how the MAE varied with different thresholds (left axis). The higher the threshold, indicating a higher level of movement, the lower the accuracy of IBIs (as indicated by higher MAEs). In the future, another consideration could be the proportion of data which falls below a threshold, and is therefore used for analysis. Here, the optimal threshold may differ according to the level of accuracy required. Applications such as pulse rate variability analysis may require a high level of accuracy and therefore a lower threshold, whereas arrhythmia detection may require a lower level of accuracy, and therefore tolerate a higher threshold.

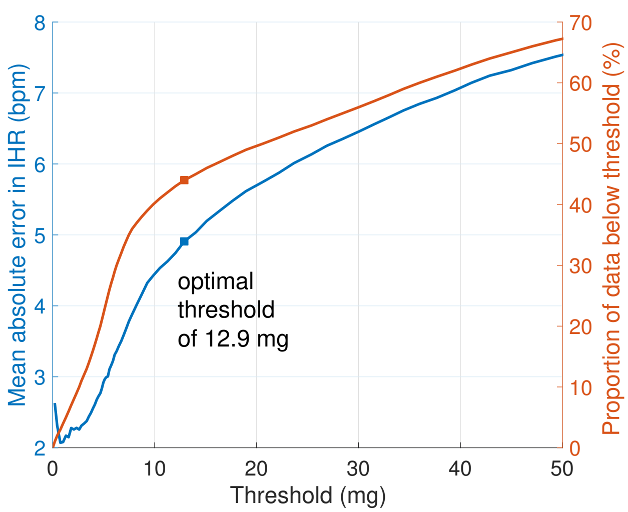

Figure 1. Performance at different MAD thresholds: (left axis) the mean absolute error (MAE) in instantaneous heart rate (IHR); (right axis) proportion of data below threshold.

Table 3 presents the optimal thresholds which produce a MAE in IHRs of <5 bpm, for the combined dataset and separately for mental stress tasks and physical activities. The optimal threshold for the combined dataset was 12.9 mg, and 44.0 % of the data fell below this threshold. This threshold approximately corresponds to cutoffs between sedentary behaviour and low-intensity activities [14, 15].

Table 3. Optimal MAD thresholds to predict whether PPG-derived IBIs were accurate.

Dataset	MAD threshold (mg)	Proportion below threshold (%)
Combined	12.9	44.0
Mental stress	10.9	72.0
Daily living	17.1	35.0

3.4. Performance across different activities

Figure 2 shows the MAE in IHRs for data below the optimal MAD threshold across different activities. The MAE was <5 bpm as expected for all activities except stress, table soccer, and walking.

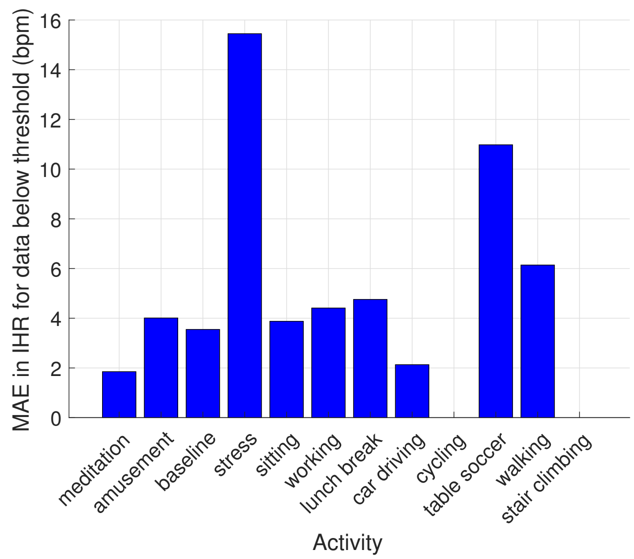

Figure 2. The MAE in IHR for data below the optimal MAD threshold for different activities.

Figure 3 helps understand the reasons for, and consequences of, the poorer performance during these activities. The figure shows, for each activity, the proportion of data for which: (i) the MAD was below the threshold and IBIs were correct (blue); (ii) the MAD was below the threshold and IBIs were incorrect (red); and (iii) the MAD was above the threshold (and therefore data would not be analysed) (yellow). Considering walking and table soccer: only a very small proportion of data fell below the threshold in these activities (5.8 % and 0.3% respectively), so the impact of high MAEs would be minimal during these activities as most data would be excluded from analysis. In contrast, a relatively high proportion of data collected during stress fell below the threshold (53.0%), so the impact of high MAEs would be substantial during stress. The poorer performance during stress may be because: (i) stress induces noise in the PPG which is not entirely attributable to movement; and/or (ii) movement during stress does induce noise, but the movement levels fall below the threshold.

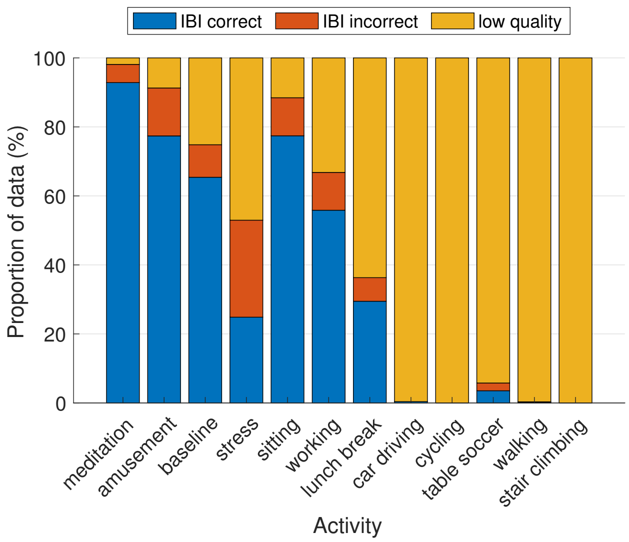

Figure 3. The proportion of data for each outcome for different activities.

4. Conclusion

Accelerometry can be used to identify periods when IBIs can be accurately measured from PPG signals during physical activities with reasonable performance. However, performance was poorer during mental stress. A potential application of the accelerometry-based approach is as an adjunct to existing PPG-based approaches, and future work should investigate whether a combined approach provides improved performance over either approach on its own.

Acknowledgments

This work was supported by the British Heart Foundation (grant number FS/20/20/34626). This article is based upon work from COST Action CA18216 VascAgeNet, supported by COST (European Cooperation in Science and Technology, www.cost.eu)).

References

[1] P. H. Charlton *et al.*, "The 2023 wearable photoplethysmography roadmap," *Physiological Measurement*, vol. [in press], 2023.

[2] H. J. Baek and J. Cho, "Novel heart rate variability index for wrist-worn wearable devices subject to motion artifacts that complicate measurement of the continuous pulse interval," *Physiological Measurement*, vol. 40, no. 10, p. 105010, 2019.

[3] J. Park *et al.*, "Photoplethysmogram Analysis and Applications: An Integrative Review," *Frontiers in Physiology*, vol. 12, p. 808451, 2022.

[4] M. V. Perez *et al.*, "Large-scale assessment of a smartwatch to identify atrial fibrillation," *New England Journal of Medicine*, vol. 381, no. 20, pp. 1909–1917, 2019.

[5] S. A. Lubitz *et al.*, "Detection of atrial fibrillation in a large population using wearable devices: The Fitbit Heart Study," *Circulation*, vol. 146, pp. 1415–1424, 2022.

[6] A. Reiss *et al.*, "Deep PPG: large-scale heart rate estimation with convolutional neural networks," *Sensors*, vol. 19, no. 14, p. 3079, 2019.

[7] P. Schmidt *et al.*, "Introducing WeSAD, a multimodal dataset for wearable stress and affect detection," in *Proc. ICMI*, 2018, pp. 400–408.

[8] S. M. Bishop and A. Ercole, "Multi-scale peak and trough detection optimised for periodic and quasi-periodic neuroscience data," in *Intracranial Pressure and Neuromonitoring XVI. Acta Neurochirurgica Supplement*, T. Heldt, Ed. Springer, 2018, vol. 126, pp. 189–195.

[9] P. H. Charlton *et al.*, "Detecting beats in the photoplethysmogram: benchmarking open-source algorithms," *Physiological Measurement*, vol. 43, no. 8, p. 085007, 2022.

[10] E. Peralta *et al.*, "Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals," *Physiological Measurement*, vol. 40, no. 2, p. 025007, 2019.

[11] J. Behar *et al.*, "A comparison of single channel fetal ecg extraction methods," *Annals of Biomedical Engineering*, vol. 42, no. 6, pp. 1340–1353, 2014, iSBN: 1043901409939.

[12] A. E. Johnson *et al.*, "R-peak estimation using multimodal lead switching," in *Proc CinC*, vol. 41. IEEE, 2014, pp. 281–284.

[13] G. Clifford, "rpeakdetect.m," place: Available at: http://www.mit.edu/˜gari/CODE/ECGtools/ecgBag/rpeakdetect.m.

[14] H. Vh-Ypy *et al.*, "A universal, accurate intensity-based classification of different physical activities using raw data of accelerometer," *Clinical Physiology and Functional Imaging*, vol. 35, no. 1, pp. 64–70, 2015.

[15] K. Bakrania *et al.*, "Intensity thresholds on raw acceleration data: Euclidean norm minus one (ENMO) and mean amplitude deviation (MAD) approaches," *PLoS ONE*, vol. 11, no. 10, p. e0164045, 2016.

Address for correspondence:

Dr Peter H. Charlton
Strangeways Research Laboratory, 2 Worts' Causeway, Cambridge, CB1 8RN, United Kingdom
pc657@cam.ac.uk