Evaluation of HRV from Repeated Measurements of PPG and Arterial Blood Pressure Signals

Andrejs Fedjajevs[1], Willemijn Groenendaal[1], Lars Grieten[1,2], Carlos Agell[1], Pieter M Vandervoort[2,3], Evelien Hermeling[1]

[1]Stichting Imec the Netherlands, Eindhoven, the Netherlands
[2]Future Health Department, Ziekenhuis Oost-Limburg, Genk Belgium
[3]Department of Cardiology, Ziekenhuis Oost-Limburg, Genk, Belgium

Abstract

Heart rate variability (HRV) is a valuable non-invasive indicator of the autonomic nervous system. Guidelines prescribe that HRV is calculated from RR intervals of the ECG signal. With the development of wearables and the availability of unobtrusive photoplethysmography (PPG) sensors inter-beat intervals (IBI) are considered as a viable HRV source. We compare short-term HRV from ECG, PPG and arterial blood pressure (ABP), which is a pulse signal of the best obtainable quality. We evaluate several characteristic points to derive IBI and validate the accuracy of individual HRV metrics for each signal.

We analysed 69 extracts (5 min) of rest data from 25 healthy volunteers recorded three times within two hours. ABP is superior to PPG, but both can be used for pulse rate variability (PRV) of subjects at rest, though not all metrics are equally reliable. The most error-prone features are high frequency power and pnn50 with relative errors of 22% (16%) and 30% (12%) for PPG (ABP). Bland-Altman plots reveal steeper regression trend for PPG-based PRV and tendency to overestimate low RMSSD and SDNN values more. Intra-subject volatility of explored metrics exceeds the errors by at least several times: absolute power features may variate by more than 100%, but their mean errors are up to 20%.

1. Introduction

Heart rate variability (HRV) allows non-invasive monitoring of the autonomic nervous system. Predictive capabilities of HRV have been studied in connection to cardiovascular and mental diseases, sleep problems and exercise [1]. According to the guidelines, HRV analysis requires extracting RR intervals from a continuous ECG signal [2]. However, several studies use inter-beat-intervals (IBI) of the photoplethysmography (PPG) signal to obtain HRV, which is then sometimes called pulse rate variability (PRV). The rationale behind this is the use of PPG technology in ambulatory monitoring, e.g., in smartwatches. It has been shown by comparing PRV from PPG to HRV that the former is a suitable replacement for HRV during resting conditions [3].

The aim of this work is to explore the limits of agreement among short-time HRV and PRV from two sources: PPG and invasive arterial blood pressure (ABP) – the gold reference blood pressure signal, providing the best signal quality. Comparison of the two pulse signals in this context may give new insight about major PRV error sources and the lower limit achievable by PPG modality. We define the optimum characteristic point to use for IBI extraction from pulse signals. We show experimental error ranges for various PRV metrics and compare them to physiological variations across multiple measurements.

2. Methods

2.1. Dataset

The dataset was collected at Ziekenhuis Oost-Limburg in Genk, Belgium (ethical committee approval number 16/039U). 28 male volunteers belonging either to normotensives or diagnosed hypertensives (age 52±7 years, BMI 27±4 kg/m^2) were recruited. Each participant signed a consent form, and the study was conducted according to the Declaration of Helsinki. One subject was excluded from analysis due to multiple arrhythmic events identified in the extracted data.

At the start of the protocol the subject was equipped with three sensing modalities: 1. Arterial line inserted into radial artery on the non-dominant arm, 2. Infra-red PPG at the index finger of the same arm (Biopac PPG100C, 240 Hz) and 3. ECG (Biopac ECG100C, 240Hz). Subjects were placed in supine position and performed a protocol of three repetitive sessions (S1-S3) with 20 minutes breaks in between. Each session included several activities, separated with 3 minutes of dedicated rest in the middle of each session, during which participants closed their eyes and relaxed. In-between activities extra (unlabeled) rest time was allocated.

2.2. Data processing

Data processing and analyses were performed using MATLAB (R2018a, MathWorks, Inc.). Data from the two acquisition systems was synchronized by aligning the respective ECG signals. Next, all signals were resampled to 240 Hz.

From each session, chunks of 270-300 seconds were extracted, containing the three-minute dedicated rest period and 2 minutes of adjacent unmarked data when subjects were also at rest. The total number of short-time-HRV-enabling extracts was 69 (19 subjects with 3 sessions, 6 subjects with 2). Two subjects did not have long-enough rest data segments.

Resulting waveforms were subject to automated beat detection with subsequent visual inspection and manual correction. This procedure was done using validated beat detectors for ECG and PPG signals coupled with PALMS - a GUI for generic time-series annotation [4]. The PPG algorithm outputs three fiducial points: upstroke (peak of the derivative), foot (last increasing gradient through defined amplitude threshold) and peak (first decreasing gradient through zero). Because of the similarity in shape of the ABP signal and the PPG signal (Figure 1), the same algorithm with changed parameters was used for processing of the ABP.

HRV features were calculated from the extracted beat-to-beat intervals of full segments. Selected features included time- (NNmean, RMSSD, SDNN, pnn50), and frequency-domain metrics related to the power in a band of choice: very low (VLF), low (LF) and high (HF) frequency. Additionally, the normalized aliases (e.g., VLF_{norm}) w.r.t. the total band power [5] were calculated. Frequency information was extracted using the Welch method with Hamming window and 50% overlap [6].

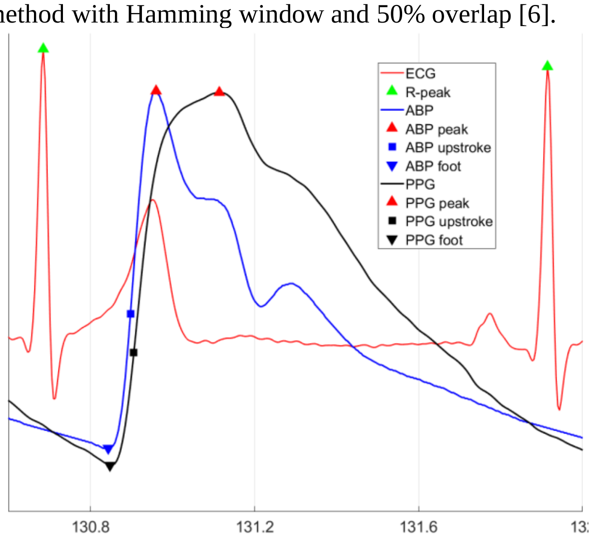

Figure 1. Fiducials used for RR and IBI extraction: R-peak of the ECG (red), and peak\upstroke\foot of the PPG (black) and ABP (blue)

3. Results

A total of 69 segments from 25 subjects were analyzed to compare PPG and ABP signals eligibility for HRV. Typical waveforms and beat-detector outputs are shown in Figure 1. The delay between the PPG and ABP is location dependent: nearly zero at the beginning of the rising edge and growing to tens of ms at the peak, where the PPG profile is also more smoothed. This indicates that the choice of point to extract IBI can have impact on the PRV.

We start by comparing IBI series with ECG RR intervals directly. Table 1 presents the average Pierson correlation, mean absolute error (MAE) and mean percentual error (MPE) of both series when calculated from different characteristic points: peak, upstroke and foot.

Table 1. ABP- and PPG-derived IBI correlation with RR intervals from ECG and corresponding MAE and MPE (mean ± SD), (* denotes statistical difference between modalities, p-value<0.01)

Metric	Fiducial	ABP	PPG
Correlation	Peak*	0.970±0.071	0.883±0.113
	Upstroke	0.995±0.004	0.993±0.007
	Foot*	0.995±0.004	0.988±0.016
ME (ms)	Peak*	5.3±6.8	13.5±8.3
	Upstroke*	3.0±0.8	3.7±1.1
	Foot*	3.2±0.8	4.1±2.1
MPE (%)	Peak*	0.51±0.62	1.33±0.82
	Upstroke*	0.29±0.08	0.36±0.11
	Foot*	0.31±0.08	0.41±0.25

Based on this data the upstroke is a preferred location: it brings the ABP (PPG) IBI error down to 0.29% (0.36%) from 0.51% (1.33%) in case of using the peak. IBI from both sources are very close to ECG RR series (correlation above 0.99 and single-digit MAE), but the Wilcoxon rank sum test confirms ABP errors are consistently lower (p-value<0.01). Following results are reported using upstroke-to-upstroke IBI.

Table 2. Errors of ABP- and PPG-derived PRV (mean ± SD). Features with statistically different MPE distributions denoted by * (p-value<0.01).

Metric	ME		MPE (%)	
	ABP	PPG	ABP	PPG
NNmean	0.02±0.0	-0.01±0.06	0.002±0.005	-0.001±0.006
RMSSD*	1±0.8	1.6±1.4	4.5±4	7.8±10.6
SDNN*	0.5±0.3	0.7±0.4	1.2±1.1	1.7±1.4
pnn50	0.1±1.1	0.3±1.2	11.9±68.8	29.8±88.5
VLF*	4.4±8	8.5±12.5	0.4±0.8	0.9±1.2
LF*	12±24.7	27.7±35.4	1.6±5.2	4.2±6.2
HF	27.3±15.8	31.8±21.7	15.8±11.3	21.5±31
VLF_{norm}	-1.1±1.5	-1.5±1.7	-2.2±2.9	-2.9±3.3
LF_{norm}*	-0.4±1.3	0±1.3	-1.1±4.5	0.2±4.8
HF_{norm}	1.4±1.4	1.5±1.5	12.5±9.2	16.4±25.9

To explore both proxy options further, we evaluated the error of individual PRV metrics (Table 2). Similar to IBI errors, most features' mean errors (ME) of ABP are superior to PPG, but statistical tests only confirm significance (p-value<0.01) for RMSSD, SDNN, VLF, LF and LF_{norm} because of wide confidence intervals. The MPE errors show differences in both signals' accuracy expressed as % of the reference HRV values. As such, pnn50 ME difference is 0.1±1.1 vs 0.3±1.2%, but because of the low reference values, after converting to MPE, it is a 20% downgrade between ABP and PPG (11.9±68.8 vs 29.8±88.5%). Other time-domain metrics exhibit less divergence between the two modalities: more mathematically advanced RMSSD MPE is 4.5% and 7.8%, whereas SDNN MPE is 1.2% and 1.7% for ABP and PPG, accordingly. Among the frequency domain features, VLF and LF have near 0% offset, but HF is overestimated by 16% (ABP) and 22% (PPG). This can be expected because HF is foremost susceptible to noise and tiniest beat detection inconsistencies. Therefore, also power normalization is more effective for HF_{norm} (error reduced by approx. 5%) and has no or light negative impact on VLF_{norm} and LF_{norm}.

Additionally, Figure 2 shows Bland-Altman plots of selected metrics. The confidence intervals (dashed lines) for ABP PRV are fully included within PPG confidence intervals. We observed minimally steeper slopes of PPG regression lines (solid red) in comparison with those of ABP (solid blue), but for all metrics the slope coefficient is of the same sign. For instance, RMSSD slope values are -0.024 and -0.05, accordingly.

Table 3. Reference HRV values (mean ± SD) across all sessions and the maximum change between any two sessions

Metric	Ref. value	Max change	Max change (%)
NNmean, ms	1034±106	34±30	3±3
RMSSD, ms	29±13	6±11	24±29
SDNN, ms	52±18	12±11	29±25
pnn50, %	8±8	3±6	Inf±Inf
VLF, ms^2	1426±1219	901±1121	115±117
LF, ms^2	747±527	341±421	66±68
HF, ms^2	287±529	221±730	73±132
VLF_{norm}, %	55±14	15±12	38±40
LF_{norm}, %	32±10	11±9	52±61
HF_{norm}, %	13±9	6±6	58±50

In order to better understand the impact of using a proxy PRV we study reference HRV values' variability through repeated measurements. The measurements were done in similar conditions within 2 hours timeframe, meaning that observed changes can be considered as physiological. Table 3 gives an overview of average values and standard deviation for each HRV metric across the population. The last two columns present the maximum change recorded between any two sessions (averaged across subjects). In general, the mean values match previously reported short-term HRV norms [5]. Most metrics exhibit substantial standard deviation, which is due to the differences in the population. On top of this, also intra-subject changes (between sessions) increase as metric's complexity increases: time-domain RMSSD and SDNN may change on average 25-30% from session to session. Absolute power metrics VLF, LF and HF are more volatile, and their changes can exceed 100%. Power normalization helps keeping them more stable: VLF_{norm}, LF_{norm} and HF_{norm} take 55%, 32% and 13% of the total spectra.

Observed short-term physiological changes in Table 3 are higher than PRV errors of corresponding features. For instance, pnn50 as one of the least reliable features has ME 0.3±1.2% for PPG and intra-subject changes are on average 3±6%. The highest calculated error 21.5% (PPG HF) constitutes less than one third of the observed intra-subject changes of 73%.

3. Discussion

Aiming to evaluate PPG signal as potential ECG substitute for HRV analysis, in this study we compared PPG to another reference (for pulse signals) – invasive ABP.

IBI series themselves do not exhibit significant discrepancies between the two modalities, but ABP is marginally better than PPG by correlation and mean errors when comparing both to ECG. Out of three options for R-peak substitute, the upstroke is preferred for both pulses, because of lower mean IBI errors in comparison to peak and foot. Other researchers came to similar results regarding the optimal identification point on a PPG waveform [7]. The reason for that can be that the extrema of a pulse are normally noisier and less well defined. The upstroke makes PRV accuracy less dependent on the noise level and the waveform.

As the complexity of PRV metrics increases so does the estimation error, but ABP appears statistically more accurate than PPG for RMSSD, SDNN and LF. It is best seen by comparing their MPE w.r.t. the reference. Mostly the difference is a few percent, but pnn50 has shown the largest spread between ABP and PPG (20%). Discrete nature of the feature (intervals occurrences are counted), low reference value, a sampling period of >4 ms and noisier PPG signal – these factors together contribute to such an effect. Previous study also reported pnn50 errors up to 30% [8].

All metrics' errors are at least several times less than intra-subject variability between repetitive measurements: pnn50 and HF also show the highest volatility across sessions. Therefore, even the most error-prone features still are eligible for certain PRV use-cases such as tracking changes.

4. Conclusions

All reported PRV metrics may be used instead of HRV since the errors do not surpass potential intra-subject changes. ABP- and PPG-derived PRV both have high correlation with conventional HRV, though, the ABP performance is slightly better and, presumably, defines the lowest error boundary for pulse signals. The results, however, only relate to the case when no motion artefacts are present in the data.

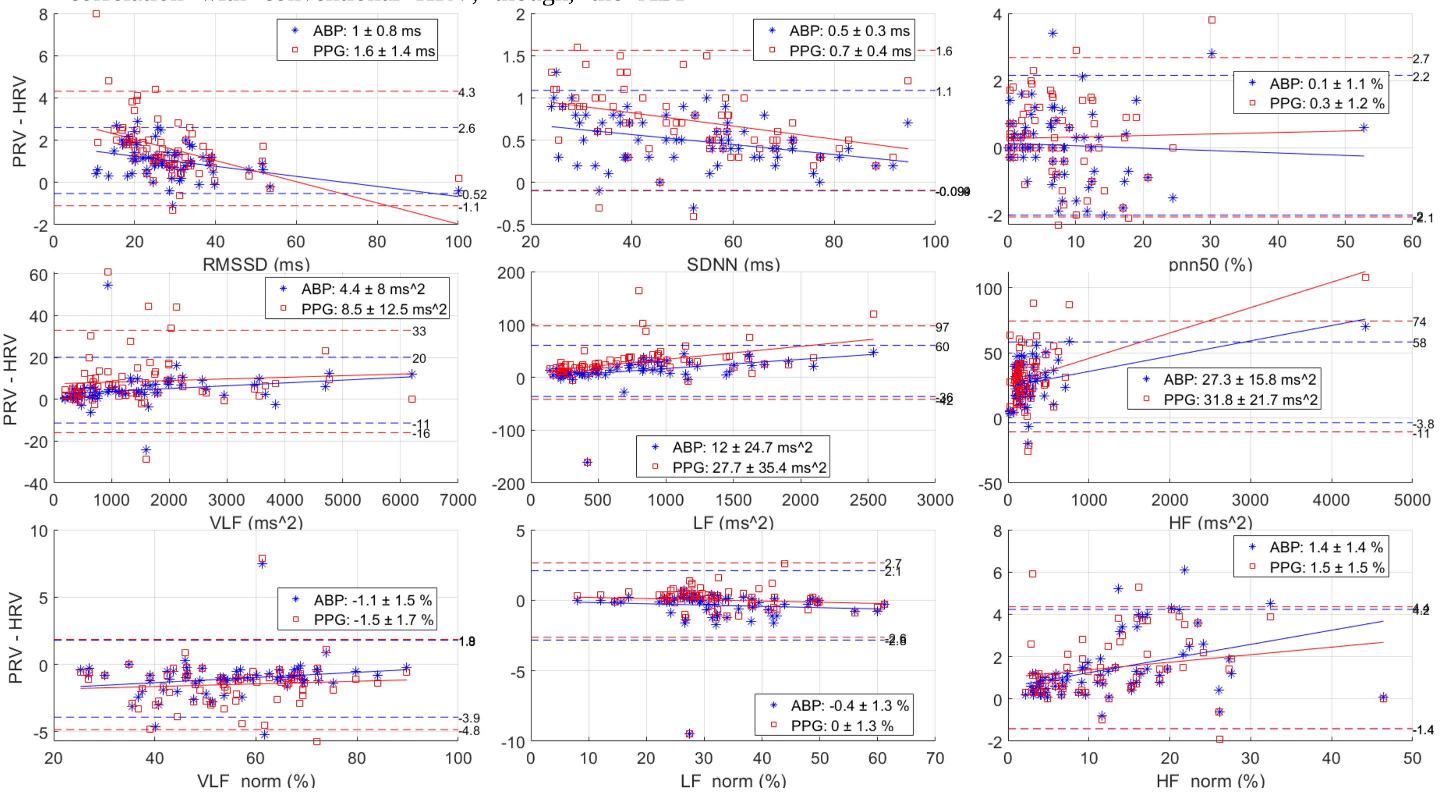

Figure 2. PRV errors vs HRV reference for selected features and two pulse signals. Dashed lines mark 95% confidence interval boundaries, solid lines denote regression trends (blue: ABP fit, red: PPG fit).

Acknowledgments

This research was funded by ITEA INNO4HEALTH 19008 project.

References

[1] Thayer J, Åhs F, Fredrikson M, Sollers J, Wager T. A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. Neuroscience & Biobehavioral Reviews. 2012;36(2):747-756.

[2] Task Force of the European Society Electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Circulation. 1996;93(5):1043–65.

[3] Georgiou K, Larentzakis A, Khamis N, Alsuhaibani G, Alaska Y, Giallafos E. Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. Folia Medica. 2018;60(1).

[4] Fedjajevs A, Groenendaal W, Agell C, Hermeling E. Platform for Analysis and Labeling of Medical Time Series. Sensors [Internet]. MDPI AG; 2020 Dec 19;20(24):7302.

[5] Shaffer F, Ginsberg J. An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health. 2017;5.

[6] Welch P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE trans audio electroacoust. 1967;15(2):70–3.

[7] Pinheiro N, Couceiro R, Henriques J, Muehlsteff J, Quintal I, Goncalves L, et al. Can PPG be used for HRV analysis? Annu Int Conf IEEE Eng Med Biol Soc. 2016;2016:2945–9.

[8] Jeyhani V, Mahdiani S, Peltokangas M, Vehkaja A. Comparison of HRV parameters derived from photoplethysmography and electrocardiography signals. Conf Proc IEEE Eng Med Biol Soc (2015) 2015:5952–5.

Address for correspondence:

Andrejs Fedjajevs.
Stichting Imec the Netherlands, 5656 AE Eindhoven, The Netherlands.
Andrejs.Fedjajevs@imec.nl