Robert Koch, Norman Pfeiffer, Nadine Lang, Björn Eskofier, Oliver Amft, Matthias Struck, Thomas Wittenberg*

Evaluation of HRV estimation algorithms from PPG data using neural networks

Abstract: Heart rate variability (HRV) is a powerful measure to gain information on the activation of the central nervous system and is thus a strong indicator for the overall health and emotional state of a person. Currently, the gold standard for HRV analysis is the examination of R-peaks in electrocardiograms (ECG), which requires a placement of electrodes on the torso. This is often impracticable, especially for the use in daily routines or 24/7 measurements. Photoplethysmograms (PPG) are an alternative to ECG assessment and are easier to acquire, e.g. by using fitness trackers or smart watches. Nevertheless, PPG data is more susceptible to motion artifacts. Hence, goal of this work is to develop and evaluate an artificial neural network (ANN) approach to estimate the R-peak locations in complex PPG signals. Public data collections were used as benchmark to compare our ANN-based approach to state-of-the-art methods. Results show that ANNs can improve HRV estimation during motion. HRV estimations from baseline methods (decision-tree based and automatic multiscale-based peak detection) were compared with the best performing neural network (3L-GRU) using the TROIKA dataset with respect to reference parameters obtained from a manual selection of the peaks in ECG data. In most cases, the neural network based HRV estimation was closer to the reference HRV compared to baseline methods (lower μ and σ) Also, σ is smaller for the best performing ANN approach across most HRV parameters. Inclusion of another PPG or acceleration channel did not affect HRV estimation. Although, the neural network learning approach outperforms conventional methods, the examined PPG-based HRV estimation has still accuracy limitations. Nonetheless, the proposed estimation approach opens up new directions for further improvement.

Keywords: photoplethysmography, PPG, heart-rate variability, HRV, neural networks

***Corresponding author: Th. Wittenberg:** Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany, e-mail: thomas.wittenberg@fraunhofer.de. **R. Koch, N. Pfeiffer, N. Lang, M. Struck:** Fraunhofer IIS, Erlangen; **O. Amft:** University Clinics Erlangen; **B. Eskofier: FAU Erlangen**

https://doi.org/10.1515/cdbme-2020-3130

1 Introduction

Heart rate variability (HRV) is considered as an important parameter when it comes to the estimation of stress level or fitness state [1]. The current gold-standard to determine a person's HRV is the examination of R-peaks in electrocardiograms (ECG), which requires the proper placement of multiple electrodes on the human torso to measure the electrical potential created from the activity of the heart. Photoplethysmograms (PPG) use optical means to acquire information on the heart activity, which can lead to similar results, and is much easier to acquire, e.g. with a smartphone. On the downside, PPG signals are susceptible to motion artifacts. Employing PPGs, the change in blood volume can be detected by sending light using a LED through the skin and measuring the intensity of the light that is either reflected or transmitted over time. As the blood volume increases, more light is absorbed, resulting in a decreased intensity and vice versa. The transmitted or reflected light intensity can thus be measured using a photo diode. Figure 1 depicts a typical PPG signal. One cycle consists of a systolic and a diastolic phase, corresponding to the phases during a cardiac cycle. The diastolic phase may also contain a small dip in the falling flank, denoted as *dicrotic notch*, caused by the closure of the aortic valve generating a small pulse wave by itself [2]. Most commercially available PPG sensors utilize two LEDs operating at wavelengths in the red and NIR spectra around at 660 nm and 800-960 nm. PPG waveforms can be used to observe respiratory rate, blood pressure and oxygen saturation in blood and may give further insight into diagnostic of cardiac diseases along with ECG [3,4]. Theoretically, PPGs can be acquired from any body part. Nevertheless, established locations are the finger and earlobe. At rest, an HRV can be determined considering the peak-to-peak (R-R) intervals in a PPG signal alone. But motion artifacts (yielding electrode displacements) and the complex physio-mechanical processes between the heart and the observed skin act as a strong filter, making the HRV estimation from PPG difficult.

Thus, the goal of this work is to investigate to which extent PPG data can be used as a *surrogate* to ECG data in order to *estimate* the underlying HRV, *without* using the related pulse-rate. As the human body acts as a strong damping filter between the heart and the observed skin, the "shadows" of the ECG R-peaks are somewhere "hidden" in the acquired PPG-signals. Thus, artificial neural networks are trained, and evaluated in order to robustly predict the locations of these "R-peaks" from PPG data. From such estimated R-peaks, the HR and HRV values are approximated and compared to the known HRV.

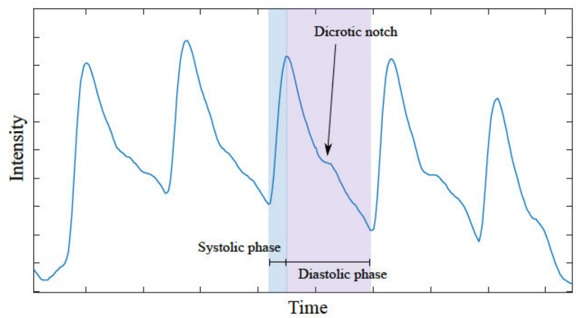

Figure 1: Example time series of a PPG with five periods, each consisting of a systolic and diastolic phase.

2 Related Work

In the past years, various research groups have proposed different approaches for the robust detection of HRV from PPG data. Kageyama et al [2] were one of the first groups investigating HRV from PPG data. Using a finger clip, ECG and PPG data was acquired over 5 minutes, while subjects were lying, sitting or standing at rest. R-Peaks in both data streams were selected manually. Inconsistencies were caused by errors in RR interval measuring, due to movements of the subject and slipping of the PPG sensor. It was concluded that PPG can be used as a substitute to estimate ECG. Murthy et al. [5] compared results of spectral analysis of PPG signals from 5 healthy subjects and 5 patients with cardiovascular diseases, where PPG was acquired in supine position over 15 minutes from the earlobe. Additionally, low frequency components caused by movement were observed. HRV was computed using peaks in PPG, assumed to be selected manually. Significant differences in the power spectra of the HRV could be observed between healthy and cardiovascular patients. A comparison between ECG- and PPG-based HRV was conducted by Giardino et al [6] with 16 sitting subjects measured over 5 minutes and inducing a Stroop color test as stress. Single channel ECG from chest and finger PPG were acquired. PPG peak detection was performed with a maxima search. For HRV detection in ECG and PPG the power-spectrum of the band pass filtered data was used. A high correlation between ECG and PPG ($r = .99$) could be obtained, but with higher periodic variations in PPG. Teng & Zhang [7] investigated PPG and ECG based HRV detection for 16 healthy males sitting, stepping and recovering from exercise, using finger PPG and chest ECG. PPG peaks were selected manually. HRV rates strongly agreed during the resting phase but diverged for exercise and recovery phases. In a similar study (8 healthy males on a treadmill at 9 km/h for 3 min) ECG and PPG were recorded 5 minutes after exercise [8]. PPG peaks were detected using wavelet transform with manual correction, power spectrum was considered for pre- and post-exercise data. It was found that high-frequency analysis is more susceptible to arterial blood conditions and respiratory movement for PPG than for ECG based HRV. Pinheiro et al. [9] compared HRV extraction from PPG using 33 healthy subjects at rest and after exercise on a treadmill and 35 patients with various diseases in supine position over 3 minutes. PPG-based HRV estimation was based on manually selected reference points. HRV parameters were determined from ECG and PPG. PPG-based HRV could only be obtained during rest, as robust R-peaks were hard to extract. To improve HRV estimation for signals corrupted by motion artifacts Alqaraawi et al. [10] proposed a multiscale peak detection algorithm extended by Bayesian learning to provide probabilities for the existence of peaks. 3 subjects were assessed with ECG and wrist PPG with 5 minutes of artifact free signals and 5 minutes with various movement activities. Additionally, the signals were corrupted with artifacts at varying SNRs. Results showed improvements in the HRV detection scheme. Koenig et al. [11] investigated correlations between ECG- and PPG-derived HRV using a smartphone camera serving as photodetector. The green channel was used to extract the PPG signal locations of steepest slopes of pulse waves were used for HRV estimation. PPG and ECG were acquired from 68 subjects for 2 minutes after 3 minutes of exercise, where PPG was extracted from the smartphone video, while the subjects had their finger placed on the lens. Correlation of heart rate (HR) was $r > .99$ and $r = .90$ for HRV. Similar results using a smartphone camera were achieved by Plews et al. [12] with 26 athletes. HRV estimation was based on peak-detection in the PPG signal. Ground truth was obtained from RR intervals of the chest. Correlation was very high ($r \sim 1.00$) for normal and paced breathing, leading to the conclusion that PPG derived HRV analysis is the preferred approach due its ease of use. One of the most complex settings is the estimation of HRV parameters without contact to the subjects. This was investigated by Fukunishi et al. [13] on 3 subjects recorded for 60 seconds with a video camera (distance 4 m) during resting and under cognitive stress. Ground truth was obtained from blood volume pressure. Peaks in PPG data were determined using a decision tree, filtering out candidates too close to each other or below a relative threshold. HRV was estimated from PPG and BVP spectrograms.

In summary, PPG can substitute ECG for HRV estimation for individuals at rest, whereas psychological and physiological stress reduces the agreement between PPG and ECG derived HRV. As most studies rely on peaks in PPG as reference points for inter-beat intervals, novel approaches are necessary to robustly estimate such peak locations in complex and noisy PPG data. In this work, we evaluate complex Artificial Neural Networks (ANNs) for R-peak estimation and HRV approximation, thus explicitly excluding any pre-processing steps.

3 Materials

Two publicly available PPG data sets were used, cf. Table 1. The *TMBE* data set [14] consists of synchronized raw PPG data acquired from finger pulse oximeter and single-lead ECG signals with 100 Hz, and up-sampled to 300 Hz obtained from 29 children and 13 adults over 8 minutes. Peaks in PPG as well as R-peaks in ECG were annotated. The *TROIKA* data set [15] was used as benchmark for the 2015 signal processing cup to estimate heart rate from PPG acquired from wrist devices during physical exercise and thus includes moving artifacts. The dataset consists of 23 datasets from subjects running on a treadmill. The dataset does not include R-peaks annotations, which were manually added for this work.

Table 1: Overview of the publicly available data collections used

	TBME	TROIKA
Acquisition rate [Hz]	300	125
Annotated ECG	Yes	Yes
Annotated PPG	Yes	No
PPG lambda [nm]	660/940	609
PPG type	Trans.	Refl.
Duration [min]	8	5
Activities	---	Running

4 Methods

To investigate if deep learning methods can improve HRV estimation from PPG data, 5 different ANN architectures were considered, a 1-layer CNN [16], a 1-layer and a 3-layer gated recurrent unit (GRU) [17] using PPG as input. Furthermore two 3-layer GRUs using PPG and additional input channels (PPG2 and acceleration). PPG signals were sliced using a sliding window, whereas each slice was labeled as '1' if the first sample in the slice corresponds to a peak pulse in the reference ECG, otherwise as '0'. Varying positions of the peak pulse and different window widths were considered.

From the annotated data, R-peak locations were predicted from PPG slices using the benchmark datasets. Afterwards HRV parameters based on the RR intervals (μ_{RR}, $\boldsymbol{\sigma}_{RR}$, RMSSD, pNN50, SD2, SD1/SD2, μ_{HR}, HF) were computed from predictions over 1 min segments with 30s overlap. The extracted HRV parameters were compared with corresponding values estimated using conventional methods: *manual* peak selection, *decision tree*-based peak detection (DTPD) and *automatic multiscale*-based peak detection (AMPD). Additionally, as ground truth, the same HRV parameters were computed from annotations of reference ECG.

Records from the TROIKA and TBME datasets were used for training and validation (40%-60% split). As ANNs are known to compensate noisy input data, no preprocessing (except data normalization) was applied to the PPG data. Input layers of the ANNs were adjusted to accommodate different sampling rates in the data sets.

Evaluation was done using a framework implemented in Python on a desktop computer (i7 3,4 GHz, x64, 16GB RAM) with a single GPU (NVIDIA GeForce GTX 1060). Algorithms were implemented using Tensorflow 2.0.

5 Results

Table 2 shows quantitative results of HRV parameter estimations from baseline methods (AMPD, DTPD) and the best performing neural network (3L-GRU) using the TROIKA dataset with respect to the *reference* parameters obtained from a manual selection of the peaks in ECG data. Values μ and $\boldsymbol{\sigma}$ indicate mean errors and standard deviation to obtained reference data. In most cases, the neural network based HRV estimation was closer to the reference HRV compared to baseline methods (lower μ and $\boldsymbol{\sigma}$). Additionally, $\boldsymbol{\sigma}$ is smaller for the best performing neural network approach across most HRV parameters. Inclusion of another PPG or acceleration channel did not affect HRV estimation. The optimal input sliding window length was found to be 1 second.

Table 2: Average errors and standard deviations for HRV parameters over all subjects estimated by reference approaches (AMPD, DTPD) and best ANN approach (3L-GRU) using PPG compared to the ECG-based reference. Analysis for the TROIKA dataset.

Metric	AMPD μ ($\boldsymbol{\sigma}$)	DTPD μ ($\boldsymbol{\sigma}$)	3L-GRU μ ($\boldsymbol{\sigma}$)
μ_{RR} [ms]	**-7.53** (223.05)	35.58 (146.27)	8.24 (**81.79**)
$\boldsymbol{\sigma}_{RR}$ [ms]	-66.83 (71.18)	-58.06 (69.37)	**-37.29** (**59.67**)
μ_{HR} [ms]	-117.10 (99.13)	-117.36 (102.98)	**-63.78** (**65.13**)
pNN50 [%]	-44.87 (32.23)	-49.16 (37.72)	**-32.61** (**30.50**)

SD2 [ms]	-60.33 (78.38)	-43.24 (**78.03**)	**-41.34** (78.43)
SD1/SD2	-0.65 (0.50)	-0.82 (0.51)	**-0.50** (**0.46**)
μHR [bpm]	2.14 (52.07)	-8.13 (35.22)	**-0.46** (**20.42**)
HF [ms^2]	-4.64 (3.97)	-5.87 (5.33)	**-3.06** (**4.21**)

Table 3 exemplarily shows the true- and false positive rates of the 3-Layer GRU network on the TROIKA and TBME datasets. Specifically, the TROIKA data includes a lot of motion artefacts from the physical exercise.

Table 3: True-positive (TP) and false positive (FP) rates of the best-performing ANN-based approach (3-Layer GRU) across TROIKA and TBME datasets. Reference R-peak counts were obtained by manual ECG annotation.

	#R-Peaks	TP	FP
TROIKA	8.178	7.114	1.082
TBME	25.833	17.518	8.613

6 Discussion and Conclusion

The pulse-transit-time (PTT) is the time between the occurrence of the R-peak in an ECG and the corresponding peak in the PPG signal. PTTs are not constant over time and are strongly dependent on the subject's activity. Hence, PTTs can be affected by many overlapping factors, which obfuscate R-peak locations in a PPG signal. Thus, the estimation of HRV based on PPG peaks is usually inaccurate and can additionally be distorted by the individuum's blood pressure, heart activity, state of the vessel walls, motion or respiration.

The present work showed that ANNs can improve PPG-based HRV estimation accuracy compared to conventional peak detection methods. However, the discrepancy to ECG-based HRV methods remains large, which may prevent the PPG-based approach from getting applied. Nonetheless, accuracy may get further improved by incorporating additional information channels that affect PPT, e.g. respiration. Also using deeper neural networks and investigating the certainty of predictions using e.g. Bayesian learning could provide further insights and provide a tool to reduce false positives.

This work is based on the master's thesis of RK at the FAU Erlangen. Parts of the research has been funded by the German Federal Ministry of Education and Research (BMBF) under the project **TraMeExCo** (011S18056A).

References

[1] Shaffer et al.: A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Frontiers Psych. 5, 1040, 2014.

[2] Kageyama et al.: Accuracy of pulse rate variability parameters obtained from finger plethysmogram: A comparison with heart rate variability parameters obtained from ECG. J Occup. Health 39(2):154-155,1997.

[3] Lindberg & Öberg: Photoplethysmography. Medical & Biological Engineering & Computing 29(1):48-54,1991.

[4] Karlen et al.: Multiparameter respiratory rate estimation from the photoplethysmogram. Trans Biom Eng 60(7):1946-53, 2013.

[5] Murthy et al.: Analysis of photoplethysmographic signals of cardiovascular patients. Proc's EMBC pp. 2204-2207, 2001.

[6] Giardino et al.: Comparison of finger plethysmograph to ECG in the measurement of heart rate variability. Psychophysiology 39(2):246-53,2002.

[7] Teng & Zhang: Study on the peak interval variability of photoplethysmographic signals. Proc's EMBS pp. 140-141, 2003.

[8] Lin et al: Comparison of heart rate variability from PPG with that from ECG. Proc's IFMBE, pp. 213-15. 2014.

[9] Pinheiro et al.: Can PPG be used for HRV analysis? Proc's EMBC,pp. 2945-9, 2016.

[10] Alqaraawi et al.:Heart rate variability estimation in photoplethysmography signals using Bayesian learning approach. Healthcare Tech. Letters 3(2):136-42, 2016.

[11] Koenig et al.: Validation of a new heart rate measurement algorithm for fingertip recording of video signals with smartphones. Telemedicine & E-health 22(8):631-6, 2016.

[12] Plews et al.: Comparison of heart-rate-variability recording with smartphone photoplethysmography, Polar H7 chest strap, and electrocardiography. Int J Sports, Physiology & Performance 12(10):1324-28, 2017.

[13] Fukunishi et al.: Improvements in remote video based estimation of heart rate variability using the Welch FFT method. Artif. Life & Robotics 23(1):15-22,2017.

[14] Karlen et al.: Multiparameter respiratory rate estimation from photoplethysmogram. Trans Biom Eng. 60(7):1946-53, 2013.

[15] Zhang et al.: TROIKA: A general framework for heart rate monitoring using wrist-type PPG signals during intensive physical exercise. Trans. Biom. Eng. 62 (2):522-531. 2015.

[16] Chandra et al :Robust heartbeat detection from multimodal data via CNN-based generalizable information fusion. IEEE Trans. Biomed. Eng. 66(3)710-7, 2019.

[17] Saadatneja et al.: LSTM-based ECG classification for continuous monitoring on personal wearable devices. IEEE J Biomed & Health Inf, 2019.