Extension of the PhysioNet Brno University of Technology Smartphone PPG Database

Andrea Nemcova[1], Radovan Smisek[1,2], Martin Vitek[1], Lucie Saclova[1], Marina Filipenska[1], Eniko Vargova[1], Pavlina Sikorova[1], Pavel Galik[1], Oto Janousek[1], Lukas Smital[1]

[1]Brno University of Technology, Department of Biomedical Engineering, Brno, Czech Republic
[2]The Czech Academy of Sciences, Institute of Scientific Instruments, Brno, Czech Republic

Abstract

We introduce an extension of the PhysioNet Brno University of Technology Smartphone PPG Database (BUT PPG) suitable for health monitoring.

BUT PPG is extended by PPG, ECG, and ACC signals recorded simultaneously in healthy subjects. RGB PPG signals were measured from the ear and finger by smartphone camera. The signals were measured either at rest or under various conditions causing artifacts. PPG signals were annotated in terms of quality into 2 classes (good/poor) by 3 experts and their consensus was provided. In ECGs, QRS complexes were detected and manually verified. The data are supplemented by glycaemia, blood pressure, and saturation values.

The extension of the BUT PPG database includes 3,840 new 10s PPG signals from 38 subjects (19 males, 19 females, age 19-76). In the extension of the BUT PPG database, 21 % of signals are of good quality. Signals from the ear (n = 1,962) are of lower quality (11 % of good quality) than signals from the finger (n = 1,878) where 31 % of signals are of good quality. The dataset includes 49,540 manually verified heartbeat positions.

The complete BUT PPG database includes 3,888 signals from 50 subjects. The database is primarily intended for the development of smartphone PPG quality assessment and heart rate estimation algorithms.

1. Introduction

This paper introduces an extension of the Brno University of Technology Smartphone PPG Database (BUT PPG) [1], [2] which is freely available on PhysioNet [3]. Original database includes photoplethysmograms (PPG) measured by smartphone camera and reference electrocardiograms (ECG) simultaneously measured by Bittium Faros 180 device. The database is provided by annotations of PPG quality and reference heart rates (HR). It consists of 48 10 s recordings from 12 subjects.

State-of-the-art among smartphone PPG datasets follows. The Welltory PPG dataset [4] includes 21 recordings (duration 68-112 s) from 13 subjects supplemented by RR intervals from Polar H10 chest strap. 1KPPG dataset [5] includes smartphone RGB PPG and ACC signals, some patient details and patient vitals – single value per signal of weight, height, HR, respiratory rate, blood oxygen saturation (SpO2), and blood pressure (BP). The dataset includes 1,000 signals from 1,000 subjects. Fingertip Videos for Heart Rate estimation database [6] includes 51 PPG recordings from 24 participants with single-value HR as a reference. Blood dataset HH [7] is primarily intended for hemoglobin level determination. However, it includes smartphone PPGs from 180 subjects. AnxiECG-PPG Database [8] includes PPGs and ECGs from 47 subjects and is intended to anxiety assessment. The two latter mentioned databases are not freely available.

While nowadays artificial intelligence (AI) coming to the fore to develop more robust and accurate algorithms, large data volumes are necessary. Moreover, verified annotations are more than welcome.

The role of smartphones in healthcare is poised to expand, paving the way for more innovative and effective cardiovascular health solutions. Nowadays, there already exist a few CE (class II) and/or FDA marked smartphone applications for atrial fibrillation detection, e.g. Preventicus Heartbeats, FibriCheck, Happitech. In Germany, there are apps on prescription (DiGA = Digital Health Applications) [9].

The extension of BUT PPG database adds 3,840 new 10s smartphone RGB PPG signals from 38 subjects along with reference ECGs, accelerometric (ACC) data (3-axes), manually verified heartbeat (QRS complex) positions, SpO2, BP, glycaemia, and annotations of PPG signal quality and reference HR. The data were measured under various conditions to include common real artifacts, which enhances the potential of their use (e.g. for algorithms including quality-regarding user feedback during measurement).

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.055

2. Methods

2.1. Data acquisition

The data were measured on 38 volunteers, of which 19 were males and 19 were females. The age range was 19-76 years (33±17 years), height range was 154-194 cm (174±10 cm), weight range was 47-110 kg (73±15 kg). Body mass index (BMI) is in the range of 18-33. The Fitzpatrick skin types were II-IV.

Approximately half of the PPGs were measured from ear using front smartphone camera with smartphone in telephoning position. The second half was measured from finger using the rear smartphone camera.

PPGs were acquired by smartphone Huawei P20 Pro. Default camera app was used with sampling frequency (fs) of 30 Hz. As a reference, one-lead ECG was recorded using Bittium Faros 180 with fs = 1,000 Hz. By the same device, 3-axes ACC data were measured with fs = 100 Hz. SpO2 was continuously measured with fs = 1 Hz using pulse oximeter Gima. Before each session, BP and glycaemia were measured by Fora D40 and Fora Diamond Mini, respectively.

There were two measurement sessions for each subject consisting of rest and activity (artifact) parts. The measurement protocol started with initial 15 s for stabilization of signals (not in the database), continuing with 140 s rest, 20 s higher pressure of finger/ear on camera, 20 s rest, 20 s moving the finger/ear on the lens, 20 s rest, 20 s walking, 20 s rest, 20 s coughing (at least 3×), 20 s rest, 20 s laughing, 20 s rest, 40 s changing the light, 20 s rest, 20 s talking, ended by 120 s rest.

This human study was approved by the Institutional Review Board of DBME Faculty of Electrical Engineering and Communication, Brno University of Technology, on September 30, 2021 (IRB Protocol EC:EK:06b/2021). Written informed consent was obtained from all subjects prior to the study.

2.2. Preprocessing

The smartphone videos were processed to extract red, green, and blue PPG from RGB parts of the video. Each frame of the R, G, and B channels of the video was averaged to obtain one sample of relevant PPG for each color channel. The resultant PPG signals were inverted.

Data were manually synchronized and divided into 10 s segments. Altogether, the new part of the database consists of 3,840 PPG segments. There are 1,962 PPG segments from ear and 1,878 from finger.

2.3. Beat detection and verification

The QRS complexes were automatically detected in ECG using QRS detector combining three independent methods [10]. Then the detections were manually verified and corrected. Median HR was calculated for each 10 s segment. The new part of the database includes altogether 49,540 beats.

2.4. Annotating

Annotating of red PPG signal quality was done by six experts (biomedical engineers). Each signal was annotated by three experts. The experts should assess the HR of PPG signals using tailor-made annotating software, described in our previous study [1] in detail. They can set the annotation to 0, if it was not possible to estimate HR. They have access to PPG signals only (ECG signals were not available to them since these serve as a reference). The PPG segments appeared in a random manner and the annotators did not know whether the segment was sensed at rest or in motion.

Then the PPG signals were classified into two classes: 1 – good quality, 0 – poor quality. HR from experts were compared to reference HR calculated from ECG. If the absolute difference was ≤ 5 beats per minute (bpm), the signal was considered as good quality from the perspective of one annotator. The maximum absolute HR difference of 5 bpm is based on the international standard IEC 60601-2-27 [11]. Detailed explanation is in [1]. For the final consensus, at least two out of three annotators must have agreed that the signal is of good quality. Otherwise, the signal was considered as poor quality.

3. Results

3.1. Analysis of annotations

Various activities negatively influence the quality of PPGs, especially in case of ear PPGs. The percentage of quality signals separately from ear and finger during rest and various activities is shown in Figure 1.

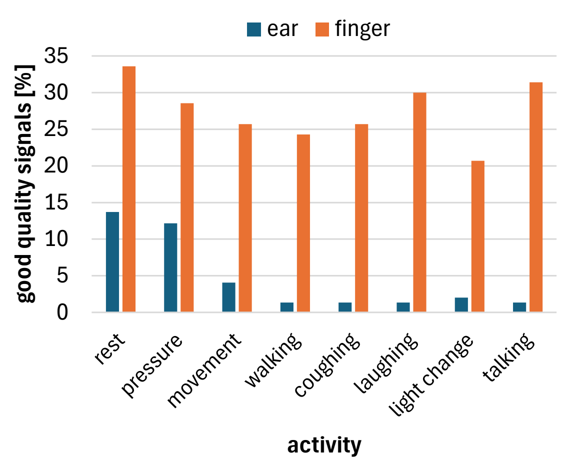

Figure 1. Percentage of good quality signals during rest and various activities.

The extension of the database includes 795 (20.7 %) good quality PPGs and 3,045 (79.3 %) poor quality PPGs. 10.6 % of PPGs from ear and 31.3 % PPGs from finger are of good quality.

The total match of all annotators with consensus in terms of PPGs quality is 93.3 % for all signals, 94.1 % for signals from ear, and 92.4 % for signals from finger. Similar comparison is done with automatized methos for HR and subsequent quality assessment. Three methods were tested:

1) Median HR (77 bpm): Calculated from ECG for the whole database. The quality of each signal was calculated in the same manner as in case of annotators (absolute difference between median HR and reference HR should be ≤ 5 bpm). The match with consensus is 56.6 %.

2) Literature-based mean HR (65 bpm): This HR is based on study of Quer et al. [12] where mean HR was calculated from 92,457 subjects. The quality was assessed in the same manner as previously described. The match with consensus is 67.1 %.

3) Stationary wavelet transform (SWT) based estimation of HR (HR is calculated for each PPG segment): The PPG signal was decomposed by SWT. In 4th frequency band (approx. 0.9–1.9 Hz), the peaks were automatically searched and median HR calculated. The quality was assessed in the same manner as previously described. The match with consensus is 74.1 %.

These numbers are indicative only as these methods did not contribute to consensus.

3.2. PhysioNet

The extension of BUT PPG database was recently published on PhysioNet. The data are in WaveForm Database (WFDB) format, standard for PhysioNet. Each record is in a separate folder with unique six-digit ID which consists of subject identifier (the first free digits) and measurement number for the subject (the last free digits). Each record consists of:

- PPG signal (red channel for records 100001-111004; red, green and blue channels for signals 112001 onwards)
  ID_PPG.dat and ID_PPG.hea
- ECG signal
  ID_ECG.dat and ID_ECG.hea
- ACC signal (for signals 112001 onwards)
  ID_ACC.dat and ID_ACC.hea
- QRS complex positions
  ID.qrs

The annotations (*quality-hr-ann.csv*) are recorded in a CSV file with three columns. The first column contains signal IDs. The second column contains a binary quality indicator: 1 (good quality), 0 (poor quality). The third column contains the reference HR.

File subject-info.csv provides the following information:

- patient demographics (gender, age, height, weight)
- measurement spot: 0 (ear) or 1 (finger)
- motion information: 0 (rest), 1 (higher pressure of finger/ear on camera), 2 (moving the finger/ear on the lens), 3 (walking), 4 (coughing), 5 (laughing), 6 (changing the light), 7 (talking)
- physiological references: BP, glycaemia, SpO2.

4. Discussion

The whole process of BUT PPG database extension is based on previously applied principles, methods and tools [1]. All the QRS complexes were manually verified, thus the reference HR can be considered accurate. The annotators had experience with such annotating, which enhances credibility of final annotations.

BUT PPG database is freely available on PhysioNet. This is still not the norm for the data to be freely available as follows from state-of-the-art part in Introduction. Moreover, majority of databases missed annotations, which is another big advantage of our BUT PPG database.

4.1. Limitations of the study

The limitation of this study can be synchronization as it was done manually. ECG and PPG signals can be thus slightly (within one beat) shifted towards each other. However, for the annotations, only the median HR for the whole segment was taken into consideration, which mitigates this limitation. In PPG segments there can be transition ones – when one activity changes into another.

4.2. Intended use of BUT PPG database

The BUT PPG database is primarily intended to (smartphone) PPG quality assessment (two classes) and HR estimation. The variability of HRs is shown in histogram in Figure 2. It is mostly resting HR as the activities are physically undemanding, thus the variability is quite high.

Secondary, the database can be used for developing and/or testing heartbeat detection algorithms. Thanks to the annotations of BP, glycaemia and SpO2, it can be used for estimation of these parameters from PPG. Moreover, the ACC signals can be used for ECG/PPG quality estimation.

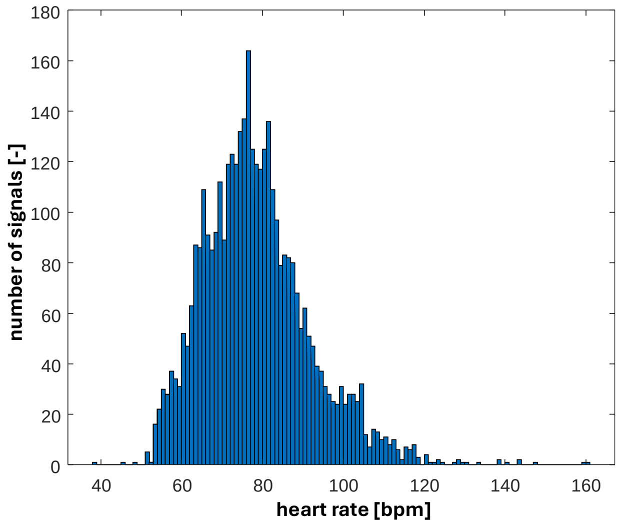

Figure 2. Histogram of HR in the BUT PPG database.

4.3. Challenge

It is possible to extend the database in many ways, thus we encourage other interested researchers and teams to help us with this and together create huge and trustworthy smartphone PPG database. We plan to measure and gradually publish new signals, extend the annotations to green and blue PPGs, annotate heart rhythm disorders in PPG and many more.

5. Conclusion

The whole BUT PPG database which is freely available on PhysioNet includes 3,888 10 s signals from 50 subjects. It consists of PPG and ECG signals and annotations of PPG quality and HR and verified positions of QRS complexes. The data are supplemented by demographic information, BP, glycaemia and SpO2 reference values. The database is primarily intended for the development of smartphone PPG quality assessment and HR estimation algorithms.

Acknowledgments

This work has been funded by the United States Office of Naval Research (ONR) Global, award number N62909-23-1-2050. The authors wish to thank LCDR Geirid Morgan from ONR Code 342 and Dr. Martina Siwek from ONR Global Prague Office for their support.

References

[1] A. Nemcova, E. Vargova, R. Smisek, et al., "Brno University of Technology Smartphone PPG Database (BUT PPG): Annotated Dataset for PPG Quality Assessment and Heart Rate Estimation", *BioMed Research International*, vol. 2021, Article ID 3453007, 2021.

[2] A. Nemcova, R. Smisek, E. Vargova, et al., "Brno University of Technology Smartphone PPG Database (BUT PPG) (version 1.0.0)", *PhysioNet*, 2021.

[3] A. Goldberger, L. Amaral, L. Glass, et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals", *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[4] A. Neshitov, K. Tyapochkin,E. Smorodnikova, et al., "Wavelet Analysis and Self-Similarity of Photoplethysmography Signals for HRV Estimation and Quality Assessment", *Sensors*, vol. 21, no. 20, 2021.

[5] A. Shetty, S. S. Narasimhamurthy, K. S. Nataraj, et al., "1KPPG dataset", *figshare*, 2023, https://figshare.com/articles/dataset/Smartphone_PPG/24511312/2.

[6] A. H. Ayesha, "Fingertip Videos for Heart Rate estimation", *figshare*, 2021, https://figshare.com/articles/dataset/Fingertip_Videos_for_Heart_Rate_estimation/17052410/2.

[7] M. Mahmuda, "Blood_Dataset(HH)", *IEEE Dataport*, 2024, https://dx.doi.org/10.21227/v2xc-4626.

[8] A. C. Almeida, F. Canais, R. Maçorano, et al., "AnxiECG-PPG Database: Synchronized ECG and mobile PPG Signals during Baseline, Physical Activated, and Psychological Activated Conditions", *IEEE Dataport*, 2024, https://ieee-dataport.org/documents/anxiecg-ppg-database-synchronized-ecg-and-mobile-ppg-signals-during-baseline-physical.

[9] Bundesinstitut für Arzneimittel und Midizinprodukte, "DiGA und DiPA", https://www.bfarm.de/DE/Medizinprodukte/Portale/DiGA-DiPA/_node.html.

[10] L. Smital, L. Marsanova, R. Smisek, et al., "Robust QRS Detection Using Combination of Three Independent Methods", *Computing in Cardiology*, vol. 47, 2020.

[11] International Electrotechnical Commission, "Medical Electrical Equipment, Particular Requirements for the Basic Safety and Essential Performance of Electrocardiographic Monitoring Equipment (IEC 60601-2-27)", 2015.

[12] G. Quer, P. Gouda, M. Galarnyk, et al., "Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: Retrospective, longitudinal cohort study of 92,457 adults", *PLOS ONE*, vol. 15, no. 2, 2020.

Address for correspondence:

Andrea Nemcova
Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 3082/12, 616 00 Brno, Czech Republic
nemcovaa@vut.cz