

# A Practical Noise Stress Test to Assess Performance of Automated Photoplethysmogram Analysis

Reza Firoozabadi, Saeed Babaeizadeh

Advanced Algorithm Research Center, Philips Healthcare, Andover, MA, USA

## Abstract

*Due to a significant spectral overlap between the motion artifact and underlying photoplethysmogram (PPG), reliable automated PPG analysis in real-life environment may be challenging. To evaluate the impact of motion artifact on the accuracy of automated PPG pulse detection, we designed a noise stress test (NST) in which artifact-bearing (noise-added) recordings are assembled from actual recordings by selecting intervals that contain predominantly motion artifact. To assemble the NST database, we analyzed 2000 synchronized electrocardiogram (ECG) and PPG recordings from MIMIC-II database. One-minute segments with the highest and lowest agreement between the ECG beats and the PPG pulses were selected using a semi-automated protocol. The resulting NST database included 52 artifact-free base recordings by visually selecting clean segments with normal pulse rate and rhythm, and 10 pure artifact recordings by selecting segments with negligible spectral content from the base signal. Cross combination of the base and artifact recordings, by calibrating the level of added artifact, generated 520 one-minute PPG signals for each desired signal-to-noise ratio (SNR). For each combined signal, the performance of automatic pulse detection and time-domain pulse rate variability analysis was evaluated by using the annotations from artifact-free base recordings as reference.*

## 1. Introduction

In recent years, using wearable devices for analysing physiological signals has been getting a lot of attention due to their ease of use and low cost. One of the most popular techniques is photoplethysmography (PPG). However, PPG is prone to motion artifact, difficult to protect against, which can corrupt the recorded signal. Some wearable devices use accelerometers to detect periods of movement to discard the noisy intervals, but these additional sensors result in higher complexity, processing requirements, and cost. Furthermore, the knowledge of movement periods, although may help to discard noisy segments, has limited

to no use for cleaning the corrupted recording.

Although the impact of motion artifact on analysis of PPG waveforms has been studied by researchers [1,2], little work has been done to truly test the impact of different levels of motion artifact on the performance of automated algorithms. One approach is to use a technique, usually called noise stress test, which is well established for electrocardiogram (ECG) recorders. However, unlike ECG devices, there are no widely-accepted standards for designing a PPG device. Therefore, different devices may induce different patterns of motion artifact on the PPG waveform. In other words, noise recorded by one PPG device may not be suitable for testing a different PPG device.

In this study, we developed a customizable method to generate an artifact-bearing database with controlled signal-to-noise ratio (SNR) using actual PPG recordings with simultaneously recorded ECG waveforms (to be used as reference). The database can be used to perform a customised noise stress test for any particular device to measure its performance in the presence of artifact. Using this technique, one can evaluate the performance of any automated PPG-based pulse detection algorithm at different levels of SNR. We also measure the interbeat intervals between normal PPG pulses and use them to calculate a few time-domain pulse rate variability (PRV) parameters to study the impact of noise on the reliability of those parameters.

## 2. Methods

We generated a library of SNR-controlled PPG waveforms from actual PPG signals recorded simultaneously with ECG, and used them to study the impact of different levels of noise on our pulse detection algorithm. We also studied the deviation in a few time-domain PRV parameters from their reference values due to added noise.

In the following we describe our proposed technique in details.

### 2.1. Building SNR-controlled database

We analyzed 2000 segments (about 1800 hours) of simultaneously recorded ECG and PPG waveforms in the MIMIC-II database [3]. ECG waveforms were annotated using Philips ST/AR algorithm. Normal ECG beats were counted within 1-min sliding windows shifting at 10-sec steps. The ratio of normal (sinus rhythm) ECG beats ( $N_d$ ) to all beats ( $N_a$ ) in each 1-min window is defined as the quality of that ECG segment.

$$Q_{ECG} = N_d/N_a \quad (1)$$

PPG waveforms were analyzed by our newly designed pulse-detection algorithm where the PPG pulses and their fiducial points (peak, upslope, and trough) were measured for each detected PPG pulse [4]. Detected PPG pulses in each sliding 1-min window with a 10-sec step were matched to their corresponding ECG beats. To perform the matching, any single PPG pulse between two adjacent ECG beats was assigned to the closest beat, provided their distance was smaller than a pre-defined threshold. Otherwise, the detected PPG pulse was ignored and annotated as missing.

The ratio of the number of matched PPG pulses in a 1-min intervals ( $N_m$ ) to the number of all ECG beats in that interval is used as a measure of PPG quality defined as

$$Q_{PPG} = N_m/N_a \quad (2)$$

### a. Selecting clean segments

To select a clean (artifact-free) PPG segment, the following hybrid high quality measure is defined which mostly takes into account the quality of PPG segment, but also incorporates the quality of corresponding ECG segment.

$$Q_H = Q_{PPG}^2 \cdot Q_{ECG} \quad (3)$$

$Q_{ECG}$  is used to reduce the impact of low-quality ECG segments on the evaluation of our PPG analysis.

The 1-min PPG segments with highest  $Q_H$  are selected as clean segments. After visual examination of all automatically-selected clean PPG segments, 52 were chosen as our base artifact-free waveforms.

Figure 1 shows an example of a clean PPG segment used as a base recording in the SNR-controlled PPG database.

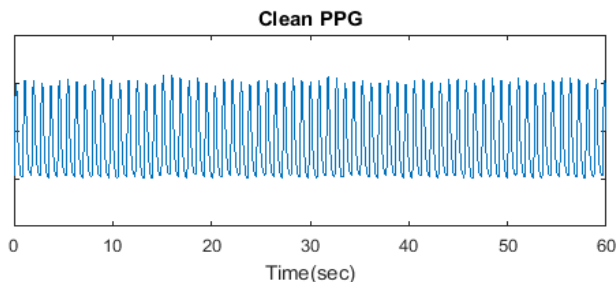


Figure 1. An example of a clean PPG segment used as a base recording in the SNR-controlled PPG database.

### b. Selecting artifact segments

To select the pure artifact segments, we define the PPG low quality measure as

$$Q_L = (1/Q_{PPG}^2) \cdot Q_{ECG} \quad (4)$$

which includes the quality of both the PPG segment (as a major factor) and the ECG reference segment (to some extent). Segments significantly corrupted by artifact are automatically detected by the algorithm. The spectral contents of the 1-min artifact PPG segments and their corresponding ECG segments were visually reviewed to verify that they indeed did not contain any significant level of the underlying PPG spectral contents. A total of 10 1-min artifact segments was chosen as the set of pure artifact waveforms.

Figure 2 shows all selected artifact segments used in generation of the SNR-controlled PPG database.

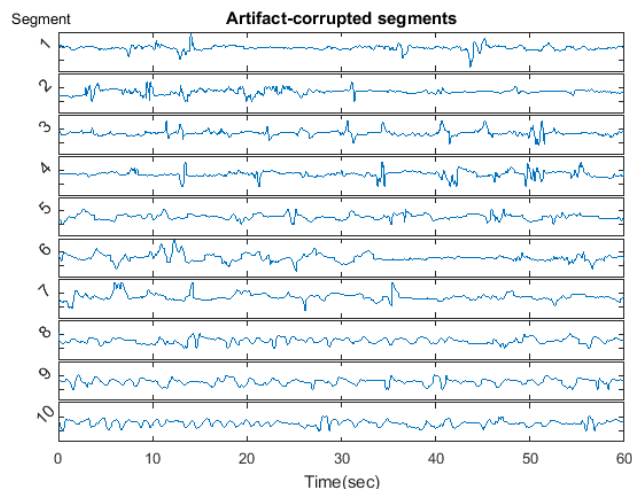


Figure 2. Artifact segments chosen for the SNR-controlled PPG database.

### c. Generate artifact-bearing segments

To generate the SNR-controlled artifact-bearing database, properly-scaled artifact segments at desired SNR values were combined with the clean PPG segments. Initially, the baseline of both the clean PPG and artifact-corrupted PPG segments were removed. A single root-mean-square (rms) amplitude ( $R_c$ ) was calculated for each 1-min clean PPG segment. For the artifact segments, rms amplitude at any sample  $s$  was calculated in 1-sec sliding windows around the sample ( $R_a(s)$ ).

The artifact segment scaling factor was determined by

$$\alpha(s) = 10^{(-SNR/20)} \frac{R_c}{R_a(s)} \quad (5)$$

The artifact-added PPG segment was then defined as

$$PPG_{\text{artifact-added}}(s) = PPG_{\text{clean}} + \alpha(s) \cdot PPG_{\text{artifact}} \quad (6)$$

Figure 3 shows an example of the SNR-controlled PPG segments generated from one clean and one artifact segment scaled at different levels of SNR.

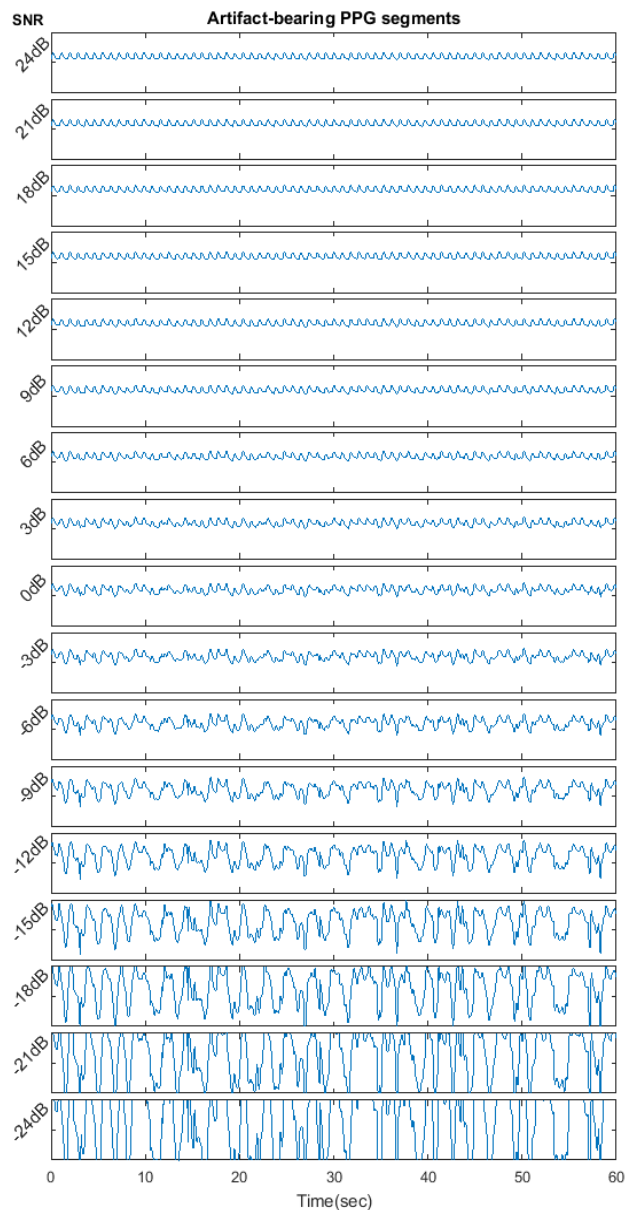


Figure 3. An example of the SNR-controlled PPG segments generated from one clean and one artifact segment by lowering SNR from 24dB to -24dB.

## 2.2. Pulse detection algorithm

At each SNR, the timestamp of PPG pulses detected by the algorithm are compared with reference pulses in corresponding clean PPG segment; the closest pulse to each reference pulse within a pre-defined interval is labeled as normal and the other detections are labeled as abnormal.

## a. Performance evaluation

At each SNR, sensitivity (Se) and positive predictive value (PPV) of the pulse detection algorithm were calculated by counting the number of true positives (TP – PPG pulses correctly detected), false positives (FP – detections where no reference pulse existed), and false negatives (FN – undetected PPG pulses) as follows

$$Se = TP / (TP+FN) \quad (7)$$

$$PPV = TP / (TP+FP) \quad (8)$$

An example of true and false positives, as well as false negative PPG pulses, is depicted in figure 4.

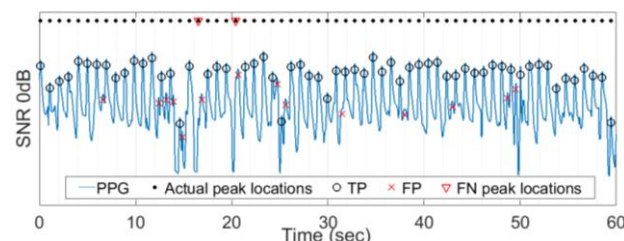


Figure 4. An example of true positive (TP), false positive (FP), and false negative (FN) PPG pulse detection for an artifact-bearing PPG signal at 0dB. The black dots at the top of panel show the location of reference pulses in the corresponding clean PPG.

## b. PRV measurements

At each SNR level, the interbeat intervals between the peaks of adjacent normal PPG pulses were measured to calculate instantaneous pulse rate and a number of time-domain PRV parameters (including pNN50, SDNN, and RMSSD [5]).

## 3. Results

We generated 8,840 one-minute artifact-bearing signals (52 clean segments cross-combined with 10 artifact segments, each at 17 different SNR levels from -24dB to 24dB) to use as the evaluation database.

At each SNR, the performance parameters of sensitivity and PPV were calculated and averaged over 520 PPG segments. As shown in Figure 5, reduction of SNR from 24dB to -24dB decreases the sensitivity and PPV from 100% to 35% and 52%, respectively. At 0dB, these values are 97% and 88%, respectively.

Figure 6 shows the mean pulse rate (meanPR) and a number of time-domain PRV parameters versus SNR. The 52 fine plots in each panel are the averages of that parameter for 10 artifact-bearing segments based on each clean segment. The bold plots are the medians of these averages. As seen, meanPR is not sensitive to artifact and can be calculated accurately even at very low SNR.

However, as is intuitively expected, the PRV parameters (pNN50, SDNN, and RMSSD) significantly increase by reducing SNR. For example, pNN50 increases by 7% at 12 dB and 36% at 0dB from its reference artifact-free value of 0.3%. SDNN increases by 5 milliseconds at 12dB and 28 milliseconds at 0dB.

#### 4. Discussion and conclusions

Unlike ECG devices, there are no widely-accepted standards for designing a PPG device. Therefore, different devices may react differently to motion artifact. Using our method, one can perform a customizable noise stress test on their PPG analysis algorithm in order to assess its performance expected in a real-life environment.

As hypothesized, motion artifact was shown to significantly impact the accuracy of automatic pulse detection and the validity of the PRV parameters.

Although we detected a series of independent artifact segments and cross-combined them with clean data and verified the results statistically, the study resources can be extended in order to achieve more reliable results. The number and length of segments of pure artifact and clean data in any available database (e.g. MIMIC II) is limited. Hence, to expand the study new experiments and customized data collection may be needed.

#### References

- [1] Kim BS, Yoo SK. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans Biomed Eng* 2006;53:566-8.
- [2] López-Silva S, Giannetti R, Dotor ML, et al. Heuristic algorithm for photoplethysmographic heart rate tracking during maximal exercise test. *J of Med and Biol Eng* 2008;32:181-8.
- [3] Goldberger AL, Amaral LAN, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 2000;101:215-20.
- [4] Firoozabadi R, Helfenbein ED, Babaeizadeh S. Efficient noise-tolerant estimation of heart rate variability using single-channel photoplethysmography. *J of Electrocardiol* 2017;Published online.
- [5] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology: Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur Heart J* 1996;17:354-81.

Address for correspondence:

Reza Firoozabadi  
 Philips Healthcare  
 3000 Minuteman Rd, Andover, MA 01810  
[reza.firoozabadi@philips.com](mailto:reza.firoozabadi@philips.com)

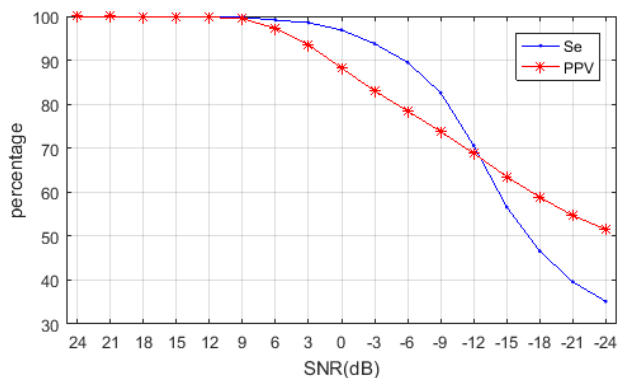


Figure 5. Sensitivity and PPV values averaged over 520 artifact-bearing segments at each SNR.

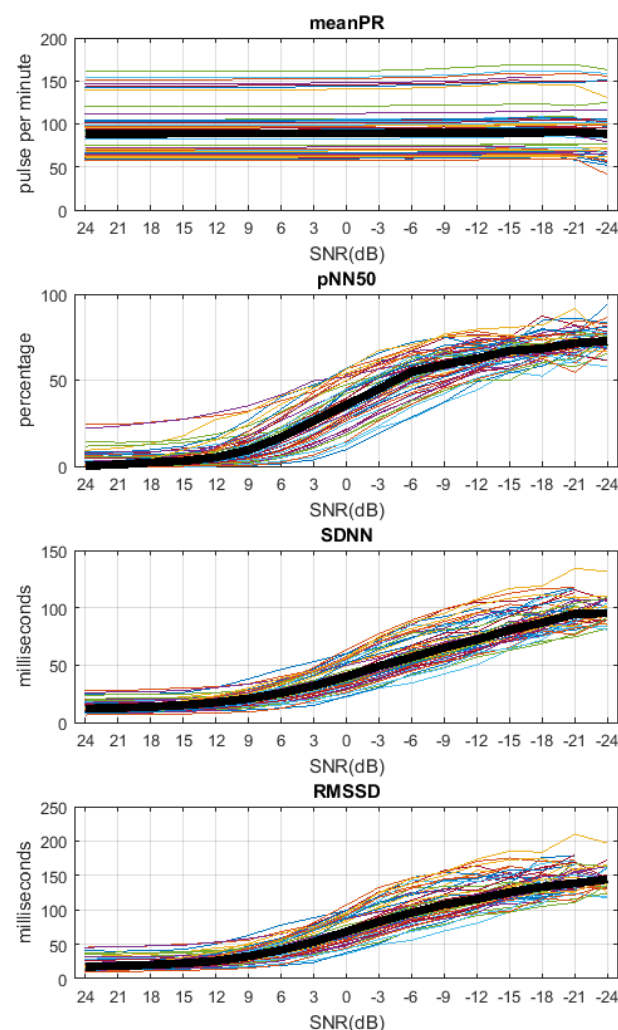


Figure 6. Average pulse rate and time-domain PRV parameters (pNN50, SDNN, and RMSSD) versus SNR. Fine plots are the averages of the parameters over 10 artifact-bearing segments corresponding to each of 52 clean segments. The bold plots are the medians of these averages.