PAPER • OPEN ACCESS

# Pulse rate estimation using imaging photoplethysmography: generic framework and comparison of methods on a publicly available dataset

To cite this article: Anton M Unakafov 2018 *Biomed. Phys. Eng. Express* **4** 045001

Biomedical Physics & Engineering Express

PAPER

# Pulse rate estimation using imaging photoplethysmography: generic framework and comparison of methods on a publicly available dataset

Anton M Unakafov[1,2,3]

[1] University of Goettingen, Georg-Elias-Müller-Institute of Psychology, Goßlerstraße 14, 37073, Goettingen, Germany
[2] Max Planck Institute for Dynamics and Self-Organization, Theoretical Neurophysics Group, Am Fassberg 17, 37077, Goettingen, Germany
[3] Leibniz ScienceCampus Primate Cognition, Kellnerweg 4, 37077, Goettingen, Germany

**E-mail:** anton@nld.ds.mpg.de

**Keywords:** imaging photoplethysmography, heart rate, signal processing, benchmark, pulse rate

OPEN ACCESS

RECEIVED
16 January 2018

REVISED
24 March 2018

ACCEPTED FOR PUBLICATION
10 April 2018

PUBLISHED
30 April 2018



## Abstract

*Objective:* to provide an algorithmic framework for comparing methods of pulse rate estimation using imaging photoplethysmography (iPPG), and to investigate performance of several existing methods on a publicly available dataset. *Approach:* first we reveal essential steps of pulse rate estimation from facial video and review methods applied at each of the steps. Then we investigate performance of these methods for DEAP dataset www.eecs.qmul.ac.uk/mmv/datasets/deap/ containing facial videos and reference contact photoplethysmograms. *Main results:* best assessment precision is achieved when pulse rate is estimated using continuous wavelet transform from iPPG extracted by the POS method (overall mean absolute error below 2 heart beats per minute). *Significance:* a framework is provided for theoretical comparison of methods for pulse rate estimation from iPPG; performance of the most popular methods is reported for a publicly available dataset that can be used as a benchmark.

## 1. Introduction

Heart rate is an important indicator of the functional status, psycho-emotional state and health conditions in general. Traditionally, heart rate is either extracted from electrocardiogram or assessed by pulse rate (number of pulse waves transmitted from heart to periphery per unit of time, usually highly correlated with heart rate (Schäfer and Vagedes 2013)) which can be computed from indirect measurements such as photoplethysmogram (PPG, (Allen 2007). However, both techniques require contact sensors, which can be disadvantageous (Perry and Watkins 2011), whereas non-contact heart rate estimation is useful, for example, for detecting driver drowsiness or abnormal state (Sahayadhas *et al* 2012).

For a non-contact assessment of pulse rate, imaging photoplethysmogram (iPPG) has been proposed (Takano and Ohta 2007, Verkruysse *et al* 2008). Similarly to contact PPG, iPPG traces variations in the intensity of light reflected by the skin (see (Allen 2007, Tamura *et al* 2014) for details), but a video camera is used instead of simple photo-detector. Then iPPG is computed from sequence of images, usually acquired from face or palm regions.

Theoretical underpinnings of imaging photoplethysmography are provided in (Hülsbusch 2008, Kamshilin *et al* 2015, Wang *et al* 2017).

Rapid development of iPPG analysis (Tarassenko *et al* 2014, McDuff *et al* 2015, Sun and Thakor 2016) emphasizes importance of comparing algorithms for iPPG-based pulse rate estimation. Theoretical comparison is complicated since algorithms for iPPG acquisition consist of multiple steps that are often non-uniformly described. For empirical comparison a publicly available benchmark dataset is required since pulse rate estimates reported in different studies are not comparable due to the differences in experimental conditions. However, to the best of our knowledge no suitable dataset has been proposed for iPPG benchmarking. Uncompressed video normally used for iPPG acquisition is too large for publishing on-line. Attempts to compare existing algorithms on publicly available MAHNOB (Soleymani *et al* 2012) and MMSE-HR (Zhang *et al* 2016) datasets were made in (Li *et al* 2014, Tulyakov *et al* 2016). However,

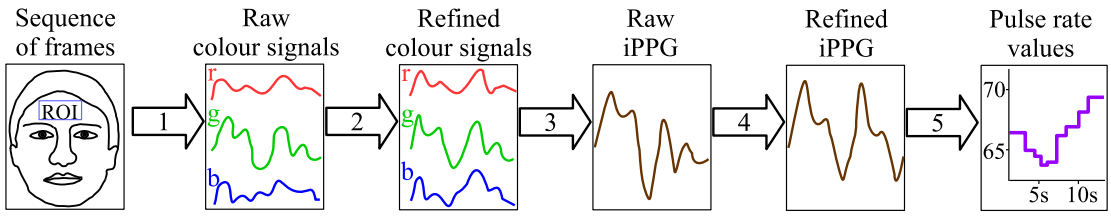


**Figure 1.** Five steps of pulse rate estimation from facial video using iPPG.

MAHNOB dataset does not seem suitable for iPPG benchmarking as the videos underwent strong compression making consistent iPPG extraction impossible (Wang *et al* 2017). The same seems to be the case for MMSE-HR since pulse rate estimation errors reported for this dataset in (Tulyakov *et al* 2016) are very high.

To overcome the problems of comparing algorithms, we suggest a framework describing main steps of iPPG-based pulse rate estimation; we discuss popular methods employed at various steps and compare their performance on a publicly available dataset (Koelstra *et al* 2012), containing facial video and reference contact PPG. We report experimental results demonstrating how the choice of the methods for each step influences the overall quality of pulse rate estimation.

Our framework consists of five steps[4]. Methods used at single steps of pulse rate estimation were previously compared in (Holton *et al* 2013, Cui *et al* 2015, Wang *et al* 2017); here we combine methods used at each of five steps to find their optimal configurations.

## 2. Materials and methods

### 2.1. Dataset description

The Dataset for Emotion Analysis using EEG, Physiological and video signals (DEAP, (Koelstra *et al* 2012)) contains physiological recordings and frontal face videos of 22 human volunteers watching music videos in 40 one-minute trials. We denote trials as P$x$ T$y$, where $x$ is the number of participant in DEAP and $y$ is the number of trial. Altogether, DEAP dataset consists of 861 one-minute trials with facial video and reference contact PPG data (37 trials for P11; 39 for P3, P5 and P14; 40 for other participants). We rejected 13 trials where large part of the face was occluded (P4 T17; P6 T24; P12 T14, T18; P15 T12, T16, T23; P18 T4, T10; P22 T13, T18-T20) since for these videos consistent iPPG extraction was not possible.

Video were acquired under controlled illumination using a SONY DCR-HC27E camcorder. Original videos were recorded in DV PAL format and then transcoded to 50 FPS deinterlaced video using the h264 codec. The resolution of all the videos was 720 × 586.

Contact PPG was acquired from the left thumb using a Biosemi ActiveTwo system (Koelstra *et al* 2012). The signal was recorded at 512 Hz and then downsampled to 256 Hz. We have selected segments of PPG corresponding to the trials, each segment was detrended by subtracting running mean over a window of 1 s, and filtered using 2nd order Butterworth filter as suggested in (Elgendi *et al* 2013). Then we computed reference pulse rate values from PPG as inverse intervals between diastolic minima (Schäfer and Vagedes 2013) detected automatically by the method proposed in (Elgendi *et al* 2013)[5].

### 2.2. Methods

In this section we propose a generic algorithmic framework of iPPG-based pulse rate estimation[6]. It takes as an input a sequence of $T$ RGB frames; $t$-th frame for $t = 1, 2, \ldots, T$ consists of pixels given by vectors $\boldsymbol{c}_{i,j}(t) = (r_{i,j}(t), g_{i,j}(t), b_{i,j}(t))^\top$, where $r_{i,j}(t)$, $g_{i,j}(t)$, $b_{i,j}(t)$ are the red, green and blue channels for the pixel with coordinates $(i, j)$; $\boldsymbol{v}^\top$ stands for the transposed vector $\boldsymbol{v}$. The algorithm consists of five steps schematically shown in figure 1; below we consider them in details.

(i) For every frame $t = 1, 2, \ldots, T$ select the region of interest ROI($t$) as a set of pixels containing

[4] A similar three-step framework was proposed in (Rouast *et al* 2016), but that article gives an overview of iPPG acquisition while here we focus on the algorithmic details of iPPG processing steps.

[5] For the contact PPG from DEAP, this method did not detect some diastolic minima since their amplitudes varied significantly. To alleviate this problem we introduced two modifications. First, to detect minima with varying amplitudes we determined offset level $\alpha$ (Elgendi *et al* 2013, equation (7)) not as the mean of the whole signal, but as a running mean over a window of 7 s. Second, to reject false positives that may arise from the first modification we added a post-processing step: after the method detects diastolic minima $\mathrm{DM}_1$, $\mathrm{DM}_2, \ldots, \mathrm{DM}_N$, we rejected $\mathrm{DM}_i$ if

$$\mathrm{DM}_{i+1} - \mathrm{DM}_{i-1} < \frac{2}{3}\min\left(\mathrm{DM}_{i-1} - \mathrm{DM}_{i-3}, \frac{2.3}{N-1}\sum_{j=2}^{N}(\mathrm{DM}_j - \mathrm{DM}_{j-1})\right),$$

where coefficients $\frac{2}{3}$ and 2.3 were selected empirically.

[6] This framework describes most of the existing algorithms; notable exceptions are the Spatial Subspace Rotation algorithm (Wang *et al* 2016) and Self-Adaptive Matrix Completion algorithm (Tulyakov *et al* 2016) based on the analysis of non-average colour channels for facial pixels or subregions.

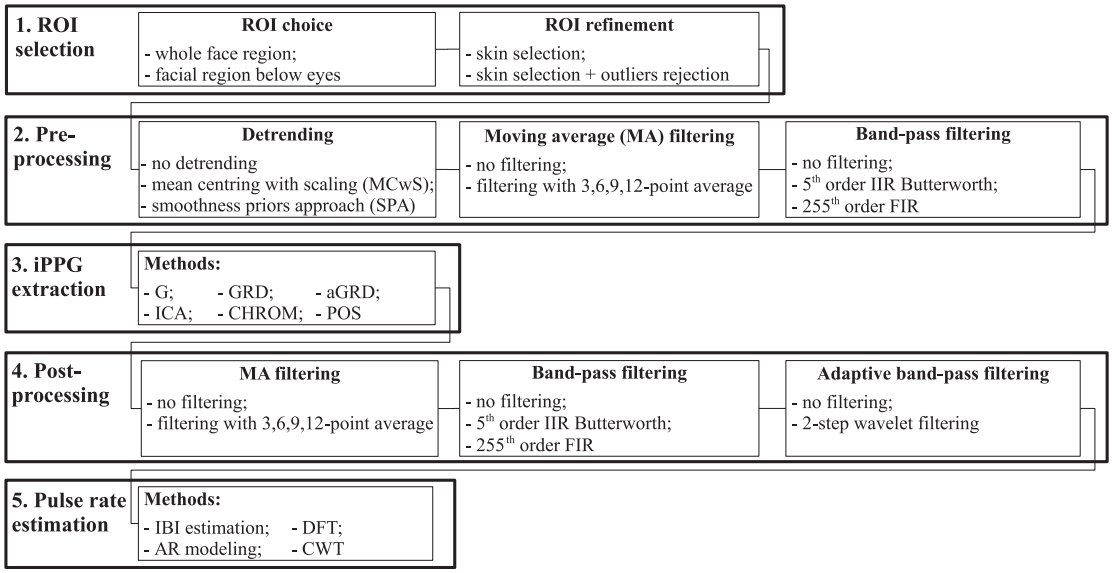


**Figure 2.** Scheme of considered methods. Big blocks represent five steps of pulse rate estimation (see figure 1), each box inside the block represents a sub-step and contains a list of methods used at this sub-step. We tested various combinations of methods, each time taking one method for every sub-step.

PPG-related information, and compute average colour intensities over ROI (colour signals):

$$\boldsymbol{c}^0(t) = (r^0(t),\, g^0(t),\, b^0(t))^\top = \frac{1}{|\mathrm{ROI}(t)|} \sum_{(i,j)\in \mathrm{ROI}(t)} \boldsymbol{c}_{i,j}(t), \qquad (1)$$

where $|\mathrm{ROI}(t)|$ is the number of pixels in $\mathrm{ROI}(t)$ (see subsection 2.2.1 for $\mathrm{ROI}(t)$ selection).

(ii) Compute refined colour signals $\boldsymbol{c}(t) = (r(t),\, g(t),\, b(t))^\top$ by pre-processing $\boldsymbol{c}^0(t)$ (subsection 2.2.2).

(iii) Extract raw iPPG as a combination of refined colour signals:

$$\mathrm{iPPG}^0(t) = w_\mathrm{r}(t)r(t) + w_\mathrm{g}(t)g(t) + w_\mathrm{b}(t)b(t)$$

with weights $w_\mathrm{r}(t),\, w_\mathrm{g}(t),\, w_\mathrm{b}(t) \in \mathbb{R}$ (see subsection 2.2.3 for weights calculation).

(iv) Post-process raw signal $\mathrm{iPPG}^0(t)$ to get refined signal $\mathrm{iPPG}(t)$ (subsection 2.2.4).

(v) Estimate pulse rates from processed iPPG signal (subsection 2.2.5).

We tested several popular methods for every step of the estimation algorithm (figure 2) in order to find out which combinations of methods provide most precise pulse rate estimation.

### 2.2.1. Selecting region of interest

Colour intensities $\boldsymbol{c}_{i,j}(t)$ are averaged over ROI to compute colour signals $\boldsymbol{c}^0(t)$ by (1). As PPG-induced variations of facial colour are weak in comparison with noise and artefacts, the aim of ROI selection is to choose pixels containing maximal pulsatile information, so that averaging reduces noise while preserving the iPPG signal. ROI selection consists of two sub-steps: initial choice of facial region for iPPG acquisition (ROI choice) and excluding irrelevant pixels (ROI refinement).

ROI choice. The most popular approach is to take a rectangular ROI encompassing the whole face (Lewandowska *et al* 2011, Poh *et al* 2011, de Haan and Jeanne 2013, Mannapperuma *et al* 2014). Other popular regions are the whole face excluding eye region (McDuff *et al* 2014b, Li *et al* 2014) and forehead (Verkruysse *et al* 2008).

In DEAP dataset for some participants EEG cap covered most of the forehead, which hindered using the forehead region; therefore we considered the whole face region and the facial region below eyes. In both cases we detected facial rectangle for each frame by the commonly-used cascade classifier (Lienhart 2000) constructed by means of the Viola-Jones algorithm (Viola and Jones 2001). We took the width of ROI equal to 80% of the estimated face width as recommended in (Poh *et al* 2011).

ROI refinement. Even when ROI is selected properly, some pixels may not contain iPPG signal. Examples include non-skin pixels (for instance, hair), over or under-lit areas, damaged pixels in the sensor. To exclude such pixels ROI-refinement methods are used, here we considered two of them.

First, non-skin pixels were discarded. This was an essential part of ROI refinement for DEAP since in many videos cables hung in front of participants' faces. We used simple HSV masking[7]: pixels with hue, saturation or value outside of the ranges [0°, 46°], [23, 132]

[7] HSV colour model is generally considered to be most useful for skin detection (Zarit *et al* 1999, Vezhnevets *et al* 2003), another popular choice is YCbCr model (Bousefsaf *et al* 2013). See also (Wang *et al* 2015) for a more elaborate approach to skin selection for iPPG acquisition.

and [88, 255], respectively, were considered non-skin and discarded (ranges were selected empirically to provide effective skin selection for the entire dataset).

Then we discarded outliers: pixels $(i, j)$ that did not satisfy the following inequality (Tasli *et al* 2014):

$$|\boldsymbol{c}^0(t) - \boldsymbol{c}_{i,j}(t)| < \gamma\boldsymbol{\sigma}_{\mathrm{ROI}}(t), \quad \text{where}$$
$$\boldsymbol{\sigma}_{\mathrm{ROI}}(t) = \sqrt{\frac{1}{|\mathrm{ROI}(t)|}\sum\nolimits_{(i,j)\in\mathrm{ROI}(t)}(\boldsymbol{c}_{i,j}(t) - \boldsymbol{c}^0(t))^2}.$$

In (Tasli *et al* 2014) $\gamma = 3$ was used; since this value did not provide effective outliers rejection for DEAP videos, we used $\gamma = 1.5$.

Another important part of ROI refinement is motion compensation (Kumar *et al* 2015, Wang *et al* 2015). We did not use it here as no prominent head movements was observed for DEAP dataset.

*2.2.2. Pre-processing of colour signals*

At this step refined colour signals $\boldsymbol{c}(t)$ are computed from raw signals $\boldsymbol{c}^0(t)$ for $t = 1, \ldots, T$ by suppressing noise and artefacts. To preserve relevant information, frequency components in human heart rate bandwidth (40–240 beats per minute (BPM), which corresponds to 0.65–4 Hz) should not be suppressed. Typical pre-processing sub-steps are detrending, band-pass and moving average filtering (see figure 2, Step 2). They are often used in combination (Holton *et al* 2013, Li *et al* 2014), but some sub-steps can be omitted or applied at post-processing (Step 4, see subsection 2.2.4).

Detrending is important since pulsatile component of iPPG has much lower amplitude than the slowly-varying baseline (Hülsbusch 2008). A simple detrending method consists in mean-centring with scaling (de Haan and Jeanne 2013):

$$\boldsymbol{c}(t) = \frac{\boldsymbol{c}^0(t) - \boldsymbol{m}(t, L)}{\boldsymbol{m}(t, L)},$$

where $\boldsymbol{m}(t, L) = \frac{1}{L}\sum_{k=0}^{L-1}\boldsymbol{c}^0(t-k)$ is an $L$-point running mean of colour vectors $\boldsymbol{c}^0(t)$; we used $L$ corresponding to 1 s. Mean-centring with scaling is used by default with several methods of iPPG extraction (see subsection 2.2.3).

Another popular detrending method is smoothness priors approach (SPA, Tarvainen *et al* 2002) used in (Poh *et al* 2011, Li *et al* 2014). To remove trend without affecting the heart rate bandwidth we employed SPA with control parameter $\lambda = 300$, which suppressed frequencies below 0.55 Hz, see (Tarvainen *et al* 2002) for details.

Moving average (MA) filtering smooths the signal and suppresses high-frequency noise. MA filtering with $M$-point average is provided by the following equation:

$$\boldsymbol{c}(t) = \frac{1}{M}\sum_{k=0}^{M-1}\boldsymbol{c}^0(t-k).$$

When choosing $M$ one should take into account that $M$-point MA filter suppresses frequencies $\frac{n}{M}F_{\mathrm{SR}}$ for $n = 1, 2, \ldots$, where $F_{\mathrm{SR}}$ is the sampling rate of the signal, see (Smith 1997), Chapter 16 for details. Since human pulse rate can reach 4 Hz, we recommend using $M < \frac{1}{4}F_{\mathrm{SR}}$. For instance, $F_{\mathrm{SR}} = 50$ Hz requires $M \leqslant 12$, thus we considered MA filtering with 3-, 6-, 9- or 12-point average.

Band-pass filtering suppresses frequency components outside the heart rate bandwidth. Here we employed two commonly used filters, either the 255th order finite impulse response (FIR) filter with linear phase designed using the Hamming window (Lewandowska *et al* 2011, Poh *et al* 2011, Li *et al* 2014) or the 5th order Butterworth infinite impulse response (IIR) filter (Sun *et al* 2013).

*2.2.3. Extracting photoplethysmogram from colour signals*

This step (figure 2, Step 3) can be represented as:

$$\mathrm{iPPG}^0(t) = \boldsymbol{w}(t)\cdot\boldsymbol{c}(t) = w_{\mathrm{r}}(t)r(t) + w_{\mathrm{g}}(t)g(t) + w_{\mathrm{b}}(t)b(t),$$

where $\boldsymbol{w}(t) = (w_{\mathrm{r}}(t), w_{\mathrm{g}}(t), w_{\mathrm{b}}(t))^\top \in \mathbb{R}^3$ are weights of colour signals. For computing these weights the following methods are often used:

- Estimating iPPG by the green signal (**G** method). This approach is popular (Tarassenko *et al* 2014, Cui *et al* 2015) due to its simplicity, in this case $\boldsymbol{w}(t) = (0, 1, 0)$ that is
$$\mathrm{iPPG}^0(t) = g(t).$$

- Estimating iPPG by the green signal while the red signal is considered as containing artefacts (green-red difference or GRD method). Here $\boldsymbol{w}(t) = (-1, 1, 0)^\top$, thus
$$\mathrm{iPPG}^0(t) = g(t) - r(t).$$
This method was first proposed in (Hülsbusch 2008), Chapter 6 as a robust alternative to G method.

- Adaptive green-red difference (aGRD, Feng *et al* 2015) computes iPPG as
$$\mathrm{iPPG}^0(t) = \|\boldsymbol{c}^0(t)\|\left(\frac{g(t)}{g^0(t)} - \frac{r(t)}{r^0(t)}\right), \qquad (2)$$
where $\|\boldsymbol{c}^0(t)\| = \sqrt{(r^0(t))^2 + (g^0(t))^2 + (b^0(t))^2}$. Pre-processing is essential for this method since otherwise $g(t) \equiv g^0(t)$ and $r(t) \equiv r^0(t)$ result in $\mathrm{iPPG}^0(t) \equiv 0$. Originally, a band-pass filtering was used (Feng *et al* 2015).

- Decomposing colour signals into components by means of blind source separation (BSS) and

choosing the component with the most prominent peak in the heart rate bandwidth. Independent component analysis (ICA) is the most popular BSS technique for iPPG computation (Holton *et al* 2013, Wang *et al* 2017). Here we used JADE algorithm of ICA (Cardoso 1999) as suggested in (Poh *et al* 2010, 2011)[8].

- CHROM method (de Haan and Jeanne 2013) employs a model of PPG-induced variations in colour intensity and defines iPPG signal as

$$\mathrm{iPPG}^0(t) = x_1(t) - \frac{\sigma_1(t, L)}{\sigma_2(t, L)} x_2(t), \tag{3}$$

where $\sigma_1(t, L)$, $\sigma_2(t, L)$ are $L$-point running standard deviations of $x_1(t) = 0.77r(t) - 0.51g(t)$ and $x_2(t) = 0.77r(t) + 0.51g(t) - 0.77b(t)$, respectively:

$$\sigma_i(t, L) = \sqrt{\frac{1}{L-1}\sum_{k=0}^{L-1}(x_i(t-k))^2 - \frac{1}{L(L-1)}\left(\sum_{k=0}^{L-1} x_i(t-k)\right)^2} \tag{4}$$

for $i = 1, 2$. We followed (de Haan and Jeanne 2013) in taking $L$ corresponding to 1.6 s.

- The recently proposed POS method (Wang *et al* 2017) can be considered as an improved and simplified version of CHROM:

$$\mathrm{iPPG}^0(t) = x_1(t) + \frac{\sigma_1(t, L)}{\sigma_2(t, L)} x_2(t),$$

where $\sigma_1(t, L)$ and $\sigma_2(t, L)$ are $L$-point running standard deviations (4) of $x_1(t) = g(t) - b(t)$ and $x_2(t) = g(t) + b(t) - 2r(t)$, respectively. We used $L$ corresponding to 1.6 s as suggested in (Wang *et al* 2017).

In order to make CHROM and POS compliant with our framework, we introduced minor algorithmic changes not affecting the nature of the methods. Namely, we used running means and standard deviations instead of computing iPPG signal in overlapped windows. For the considered dataset the performance of our modified versions was slightly better than of the original methods.

Note that special pre-processing is required for some iPPG extraction methods, namely mean-centring with scaling for GRD, CHROM and POS, band-pass filtering for aGRD. When testing the effect of pre-processing (see figure 2, Step 2) we always used required pre-processing with these iPPG extraction methods.

[8] ICA-based iPPG extraction incorporates specific pre-processing (subtracting the mean and dividing by standard deviation of each channel (McDuff *et al* 2014a); see (de Haan and Van Leest 2014) for the criticism of this approach) and post-processing (inverting iPPG if it was flipped during ICA (McDuff *et al* 2014a)). We used these methods as a part of ICA but did not describe them separately due to their limited applicability for non-ICA-based iPPG extraction.

*2.2.4. Post-processing of imaging photoplethysmogram*

Post-processing (figure 2, Step 4) improves quality of iPPG signal and is especially necessary if noise and artefacts were not removed at pre-processing (Step 2) or if iPPG was extracted at Step 3 in a non-linear fashion (which is the case for aGRD, ICA, CHROM and POS). Here we consider three typical sub-steps of post-processing: band-pass, MA and adaptive band-pass filtering.

Band-pass and MA filtering described in subsection 2.2.2 as pre-processing sub-steps can be also used at post-processing (Poh *et al* 2010, 2011), this results in different iPPG signal for all considered methods of iPPG extraction except for linear G and GRD methods.

Adaptive band-pass filtering assumes that frequency components of iPPG signal pertaining to pulse rate have relatively high power; then weak components correspond to noise and should be suppressed (Hülsbusch 2008, Bousefsaf *et al* 2013, Wang *et al* 2015, Feng *et al* 2015). Here we used a two-step wavelet filtering suggested in (Bousefsaf *et al* 2016). First we performed continuous wavelet transform of iPPG and filtered wavelet coefficients with a wide Gaussian window centred at scale corresponding to the maximum of squared wavelet coefficients averaged over 15 s temporal running window. Then we applied usual Gaussian filter. The filtered signal was reconstructed by performing the inverse continuous wavelet transform. See (Bousefsaf *et al* 2016) for details.

Modifications of iPPG signal provided by MA, band-pass and wavelet filtering are shown in figure 3.

*2.2.5. Estimation of pulse rate*

We consider here four most popular methods of pulse rate estimation (figure 2, Step 5).

- Interbeat interval (IBI) estimation is the most direct way to assess pulse rate, however this approach is rarely used for iPPG since precise IBI estimation is often problematic (Schäfer and Vagedes 2013, Elgendi *et al* 2013, Kamshilin *et al* 2016). IBI corresponds to a cardiac cycle; thus momentary pulse rate is equal to the inverse IBI duration. IBI is usually defined for iPPG as time between successive systolic peaks (Schäfer and Vagedes 2013) using some method of peak detection; here we employed method of (Elgendi *et al* 2013) with modifications described in subsection 2.1, see figure 4 for an illustration. For accurate IBI estimation we increased sampling rate of iPPG signal from 50 to 250 Hz using cubic spline interpolation as suggested in (Takano and Ohta 2007).
- Another approach is to assess average pulse rate as frequency corresponding to maximal power spectral density (PSD). By computing PSD over $N$ points one estimates average pulse rate value over time interval $\tau = N/F_{SR}$, where $F_{SR}$ is the sampling rate

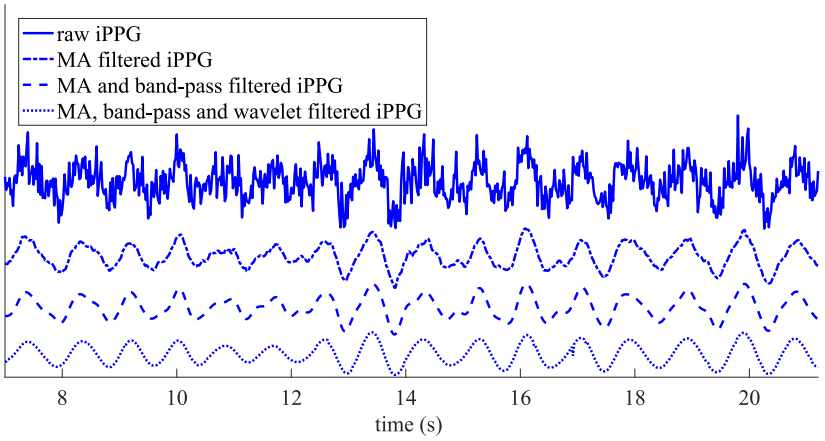


**Figure 3.** Effect of various post-processing methods on iPPG extracted from P1 T24 data using POS.

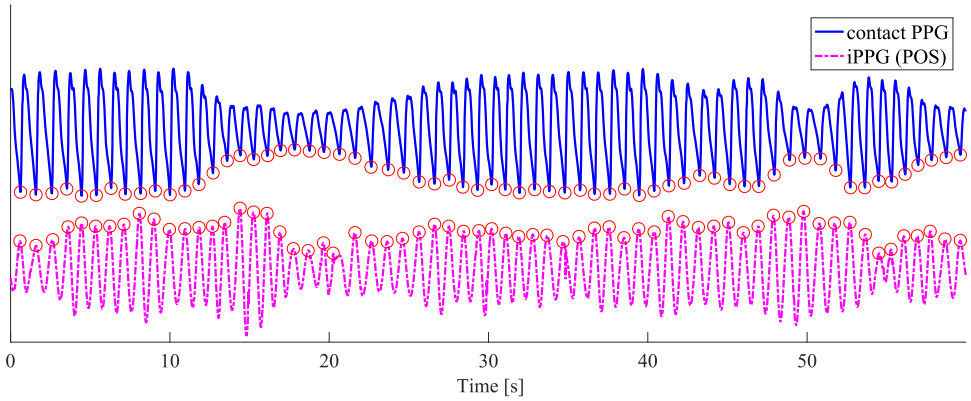


**Figure 4.** Contact PPG and iPPG signal (extracted using POS and post-processed by MA, band-pass and wavelet filtering) for P1 T24, red circles indicate diastolic minima for PPG and systolic peaks for iPPG detected using algorithm from (Elgendi *et al* 2013) with modifications described in subsection 2.1. Note that for contact PPG signal interbeat intervals are estimated from diastolic minima since they are more clear and prominent than peaks.

of the iPPG signal ($F_{SR} = 50$ Hz for DEAP). PSD is usually estimated by Discrete Fourier Transform (DFT) or by autoregressive (AR) modelling.

DFT is a direct way to estimate PSD (Poh *et al* 2011, de Haan and Jeanne 2013). Yet, DFT is often criticized (Hülsbusch and Blazek 2002, Holton *et al* 2013) since its frequency resolution is $60/\tau$ BPM ($1/\tau$ Hz) and leads to a crude estimation of pulse rate for $\tau < 20$ s, while taking $\tau > 20$ s hinders tracking of pulse rate variations. Here we used $N = 1024$, which resulted in averaging pulse rate over $\tau = 20.48$ s.

AR modelling considers iPPG as an output of linear system with added white noise (Takano and Ohta 2007, Tarassenko *et al* 2014); parameters of this system are estimated to compute PSD. In comparison with DFT, AR modelling yields improved resolution for short samples. We have implemented AR modelling using Burg's method (Matlab function `pburg`) and employed models either of 23rd order (for iPPG signal with wavelet filtering at Step 4) or 34th order (without wavelet filtering) as these settings provided best pulse rate estimation (we have tested orders 5, …, 80).

- Continuous Wavelet Transform (CWT) provides a promising alternative to DFT and AR modelling (Hülsbusch and Blazek 2002). We have implemented CWT using Matlab function `cwtft`; we utilized Morlet wavelet (Hülsbusch 2008, Bousefsaf *et al* 2013) and scales corresponding to 0.325–25 Hz with factor $2^{0.03125}$. We have chosen these scales to have a sufficiently good coverage of human heart rate bandwidth (0.65–4Hz): 0.325 Hz is twice lower than minimal pulse rate, while 25 Hz is the half of iPPG signal sampling rate. Factor $2^{0.03125}$ provides 32 scales per octave.

Since DFT and AR modelling estimate only average pulse rate, in order to make all estimates comparable, we averaged pulse rate estimates for IBI and CWT in windows of $\tau = 20.48$ s.

**Table 1.** Quality metrics (averaged over all epochs and participants) for pulse rate estimates computed from iPPG with best pre- and post-processing (table 2); best values of metrics for each iPPG extraction method are shown in **bold**.

| iPPG extraction | MAE, BPM | | | | RMSE, BPM | | | | $PE_{3.5}$, % | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | IBI | DFT | AR | CWT | IBI | DFT | AR | CWT | IBI | DFT | AR | CWT |
| G | 5.91 | 6.54 | 5.62 | **5.35** | 7.42 | 9.62 | **6.78** | 7.62 | 44 | 54 | 44 | **58** |
| GRD | 4.05 | 3.82 | 3.99 | **3.07** | 5.30 | 6.36 | 5.11 | **4.96** | 60 | 73 | 60 | **78** |
| aGRD | 4.41 | 4.41 | 4.28 | **3.59** | 5.70 | 7.07 | **5.45** | 5.55 | 57 | 70 | 58 | **74** |
| ICA | 3.91 | 3.61 | 3.71 | **2.94** | 5.41 | 6.03 | 4.87 | **4.77** | 64 | 75 | 63 | **79** |
| CHROM | 3.46 | 2.70 | 3.05 | **2.08** | 4.64 | 4.70 | 4.04 | **3.46** | 65 | 81 | 69 | **86** |
| POS | 3.13 | 2.61 | 2.91 | **1.99** | 4.30 | 4.50 | 3.80 | **3.25** | 70 | 81 | 71 | **87** |

Note that methods of pulse rate estimation have been recently compared in (Cui *et al* 2015) for iPPG extracted using G method, but in that study CWT and AR modelling were not considered, while DFT was used either with long windows of 30 s resulting in low time resolution or with short windows of 2 s providing very low frequency resolution.

### 2.3. Metrics

To investigate the quality of pulse rate estimation, we split each trial (see section 2.1) into epochs of 20.48 s with 9.88 s (approximately 50%) overlap and got five epochs per trial. For each epoch $i$ we compared estimated average pulse rate $\mathrm{PR}_i$ with the averaged reference value $\mathrm{PR}_i^{\mathrm{ref}}$. The following quantities were used to assess estimation performance for the epochs of each participant.

Mean absolute error (MAE) is given by

$$\mathrm{MAE} = \frac{1}{N}\sum_i |\mathrm{PR}_i - \mathrm{PR}_i^{\mathrm{ref}}|, \tag{5}$$

where $N = 5$ epochs *per trial* $\times$ amount of trials and $i$ is the number of epoch. MAE $\approx$ 3 BPM was observed in (Tarassenko *et al* 2014) for epochs comprising 4 heart beats (approximately 4 s) and MAE $\approx$ 2.5 BPM on average in (Lewandowska *et al* 2011) for 30 s epochs.

Root-mean-square error (RMSE) is given by

$$\mathrm{RMSE} = \frac{1}{N}\sqrt{\sum_i (\mathrm{PR}_i - \mathrm{PR}_i^{\mathrm{ref}})^2}. \tag{6}$$

RMSE is more sensitive to large estimation errors than mean absolute error, and a small number of large errors results in high RMSE and low MAE values. Pulse rate estimates from uncompressed video of stationary subjects usually have RMSE in range of 1–2 BPM for epochs of 30–60 s (Poh *et al* 2011, Li *et al* 2014, Bousefsaf *et al* 2016). RMSE $<$ 1 BPM observed in (de Haan and Jeanne 2013) is obtained for video recorded under dedicated professional illumination, which makes results incomparable with those for DEAP. On the other hand, RMSE $>$ 6 BPM reported in (Li *et al* 2014, Tulyakov *et al* 2016) for the MAHNOB dataset and RMSE $>$ 10 BPM obtained in (Tulyakov *et al* 2016) for the MMSE-HR dataset are too high and indicate limited usefulness of these datasets for iPPG-based pulse rate estimation.

Percentage of epochs (PE) for those pulse rate estimation error is below 3.5 BPM[9] is given by

$$\mathrm{PE}_{3.5} = \frac{1}{N}\{i\colon |\mathrm{PR}_i - \mathrm{PR}_i^{\mathrm{ref}}| < 3.5\ \mathrm{BPM}\}. \tag{7}$$

We also assess quality of iPPG signal by signal-to-noise ratio (SNR) defined as (de Haan and Jeanne 2013):

$$\mathrm{SNR} = \frac{1}{N}\sum_i 10\log_{10}\frac{\sum_{f=40\ \mathrm{BPM}}^{240\ \mathrm{BPM}} (\hat{S}_i(f))^2 U_i(f)}{\sum_{f=40\ \mathrm{BPM}}^{240\ \mathrm{BPM}} (\hat{S}_i(f))^2 (1 - U_i(f))}, \tag{8}$$

where $\hat{S}_i(f)$ is the spectrum of the $i$-th iPPG epoch computed by using DFT and $U_i(f)$ indicates whether frequency component $f$ is attributed to the signal ($U_i(f) = 1$) or to noise ($U_i(f) = 0$):

$$U_i(f) = \begin{cases} 1 & \text{if } |f - \mathrm{PR}_i^{\mathrm{ref}}| \leqslant \Delta f \text{ or } |f - 2\mathrm{PR}_i^{\mathrm{ref}}| \leqslant 2\Delta f, \\ 0 & \text{otherwise.} \end{cases}$$

In order to make results comparable with those in (Wang *et al* 2017) we take $\Delta f = \frac{50 \cdot 60}{1024} \approx 2.93$ BPM.

## 3. Results and discussion

### 3.1. Overview

In figure 5 and in table 1 we present values of mean absolute error and other quality metrics for pulse rate estimates and iPPG extraction methods under best pre- and post-processing (see table 2). In all cases best results are obtained for the whole face ROI with skin selection and outliers rejection (see figure 2, Step 1).

The lowest estimation errors were achieved when using POS for iPPG extraction and CWT for pulse rate estimation. Altogether, values of quality metrics are comparable with those reported in the literature for pulse rate estimation from uncompressed video (see subsection 2.3).

Below we discuss influence of various steps on the pulse rate estimation quality. We begin with methods for

[9] In (Holton *et al* 2013) best method estimates pulse rate with error below 6 BPM for $\mathrm{PE}_6 = 87\%$ of epochs. Here we are interested in percentage of epochs for those pulse rate is estimated well; precision of 6 BPM seems insufficient for this, so we bound error by 3.5 BPM (5% of average human pulse rate 70 BPM).

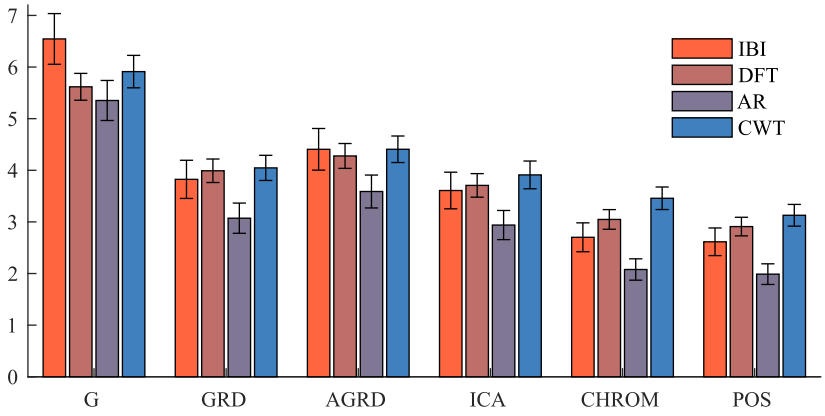


**Figure 5.** Mean values of absolute error with 95% confidence intervals (estimated as 1.96 times standard error) for pulse rate estimates computed from iPPG with best pre- and post-processing (table 2).

**Table 2.** Pre- and post-processing providing most precise pulse rate estimates (processing for IBI and DFT also results in highest SNR); IIR and FIR stand for corresponding band-pass filtering, MA*x*—for moving average filtering with length *x*, WF—for wavelet filtering. Mean-centring with scaling is beneficial for all and is not included here.

| iPPG extraction | Pre-processing | | Post-processing | | |
|---|---|---|---|---|---|
| | IBI, DFT, CWT | AR | IBI, DFT | CWT | AR |
| G | – | | MA12, FIR, WF | | |
| GRD | – | | MA12, FIR, WF | | MA12, IIR, WF |
| aGRD | FIR | IIR | MA12, WF | | |
| ICA | FIR, MA12 | IIR, MA12 | WF | | |
| CHROM | – | | MA12, FIR, WF | MA9, FIR, WF | MA12, WF |
| POS | – | | MA12, FIR, WF | MA9, FIR | MA12, WF |

**Table 3.** Average SNR and quality metrics for CWT pulse rate estimates from iPPG extracted using pre- and post-processing settings providing best SNR. For comparison we include overall SNR values from (Wang *et al* 2017) (except for aGRD, which was not tested in that study). Best values for each quality metrics are shown in **bold**.

| iPPG extraction | With wavelet filtering | | | | Without wavelet filtering | | | | Previously reported SNR, dB |
|---|---|---|---|---|---|---|---|---|---|
| | MAE, BPM | RMSE, BPM | $PE_{3.5}$, % | SNR, dB | MAE, BPM | RMSE, BPM | $PE_{3.5}$, % | SNR, dB | |
| G | 5.40 | 7.70 | 58 | −3.24 | 5.60 | 8.20 | 58 | −4.16 | −1.90 |
| GRD | 3.15 | 5.00 | 77 | −0.89 | 3.11 | 5.08 | 78 | −2.14 | 3.67 |
| aGRD | 3.59 | 5.55 | 74 | −1.26 | 3.63 | 5.75 | 74 | −2.45 | — |
| ICA | 2.98 | 4.88 | 79 | −0.60 | 3.02 | 5.11 | 80 | −1.93 | 1.92 |
| CHROM | 2.10 | 3.39 | 85 | −0.20 | 2.12 | 3.56 | 86 | −1.67 | 3.86 |
| POS | 2.04 | **3.16** | **87** | **0.30** | **2.01** | 3.17 | **87** | −1.19 | **5.16** |

iPPG extraction (see figure 2, Step 3), since results for different methods vary considerably (subsection 3.2). We proceed with ROI selection (Step 1), pre- and post-processings (Steps 2, 4) and finish with pulse rate estimation (Step 5) in subsection 3.3–3.5, respectively.

### 3.2. Step 3: Imaging photoplethysmogram extraction

In table 3 we present performance metrics for all considered iPPG extraction methods. We use pre- and post-processing (Steps 2, 4) ensuing best SNR (see table 2) and estimate pulse rate by CWT since this method provides best results at Step 5.

As you can see from table 3, POS had the highest signal-to-noise ratio and provided most precise pulse rate estimation. The ranking of the methods is in line with the results reported in (Wang *et al* 2017) for uncompressed videos representing various types of motion and several illumination settings. The only difference is that in (Wang *et al* 2017) GRD performs better than ICA. We explain this difference by sensitivity of ICA to the number of source components in

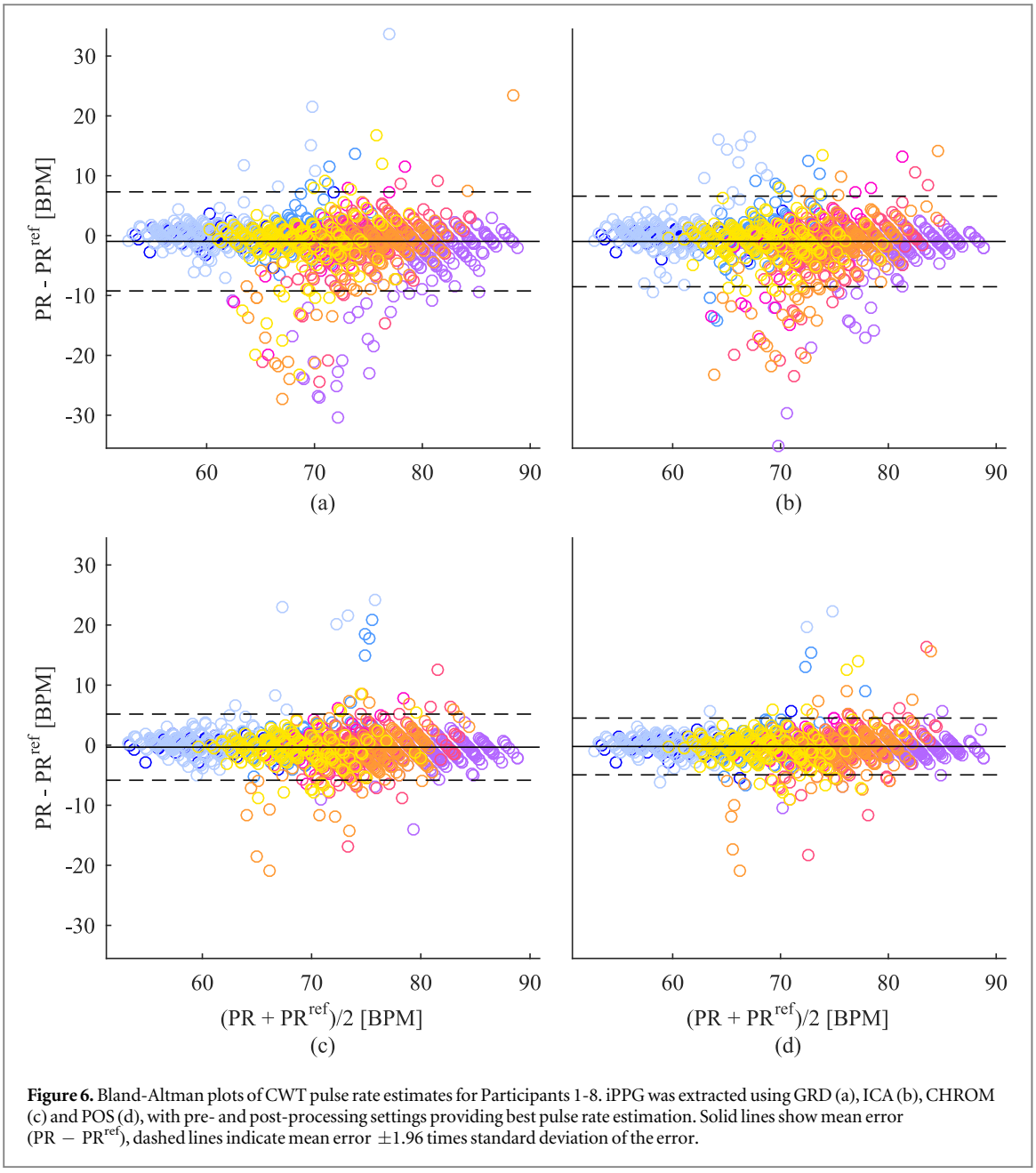


**Figure 6.** Bland-Altman plots of CWT pulse rate estimates for Participants 1-8. iPPG was extracted using GRD (a), ICA (b), CHROM (c) and POS (d), with pre- and post-processing settings providing best pulse rate estimation. Solid lines show mean error ($PR - PR^{ref}$), dashed lines indicate mean error $\pm 1.96$ times standard deviation of the error.

the signal; light variation and motion in (Wang *et al* 2017) introduce additional components to the colour signals and may complicate extraction of pure iPPG by means of ICA.

Note that average SNR for DEAP dataset is worse than values reported in (Wang *et al* 2017). It might be due to the compression of videos in DEAP and using professional dedicated lighting for video acquisition in (Wang *et al* 2017).

Figure 6 shows the Bland-Altman plots for data of eight participants for the four best iPPG extraction methods (POS, CHROM, ICA and GRD). One can see that for the POS method estimation errors are consistently lower than for other methods.

Figure 7 shows average values of SNR and mean absolute error for every participant for the three best iPPG extraction methods (POS, CHROM and ICA). In most cases high SNR corresponds to low mean absolute error, which (as expected) indicates that good quality of iPPG ensures precise pulse rate estimation.

### 3.3. Step 1: ROI selection

ROI choice. Results for the whole face region are better than for the region below eyes, both in terms of iPPG quality and pulse rate estimation. Namely, SNR for signal acquired from the whole face region is at least 0.2–0.3 dB higher (for GRD and aGRD; for other methods the difference is 1.1–1.9 dB), while mean absolute error of pulse rate estimation is lower (contribution varies from 1% for GRD and aGRD to more than 10% for other methods).

ROI refinement. Outliers rejection is always beneficial, as it increases SNR of iPPG signal (0.25–0.5 dB)

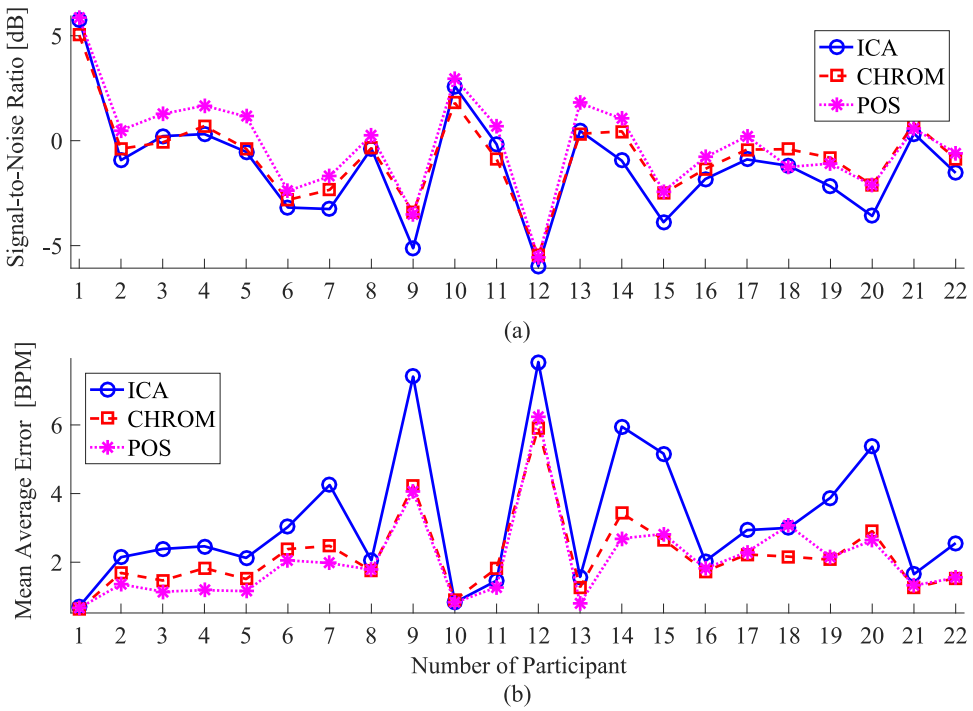


**Figure 7.** SNR (a) and mean absolute error (b) obtained by ICA, CHROM and POS methods for iPPG extraction in combination with best pre- and post-processing settings (table 2) and CWT pulse rate estimation.

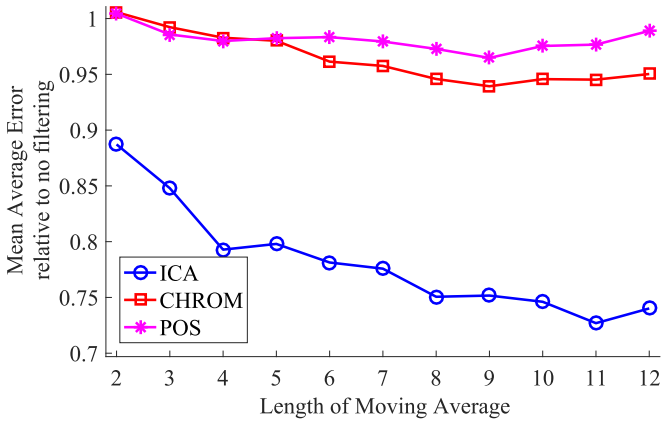


**Figure 8.** Mean absolute error values reflecting the influence of $M$-point MA filtering on pulse rate estimation using CWT (pre- and post-processing according to table 2).

and decreases estimation error (10%–15% when CWT or DFT pulse rate estimation is used in combination with GRD, aGRD, CHROM or POS methods and 5%–10% otherwise).

### 3.4. Steps 2 and 4: Pre-processing and post-processing

Detrending. Among two detrending methods, mean-centring with scaling (MCwS) always improved pulse rate estimation while SPA did not provide any positive effect (probably raw colour signals in our study were too noisy for successful application of this technique). For G, aGRD and ICA methods of iPPG extraction (Step 2), using MCwS increases SNR (average increase is 1.6, 0.7 and 3 dB, respectively) and improves pulse rate estimation (decrease of mean absolute error is above 12%). CHROM, POS and GRD methods use MCwS for pre-processing by default, therefore performance without MCwS was not tested.

MA filtering. Quality of iPPG signal and of pulse rate estimation enhanced with increase of MA filter length $M$ and reached maximum for $M = 12$ (MA filtering with $M > 12$ affects heart rate bandwidth and was not tested). The only exception was pulse rate estimation by CWT (Step 5) from iPPG obtained by CHROM and POS methods: in this case best results were observed for $M = 9$. Figure 8 illustrates this effect for estimation error, effect on other quality metrics is similar.

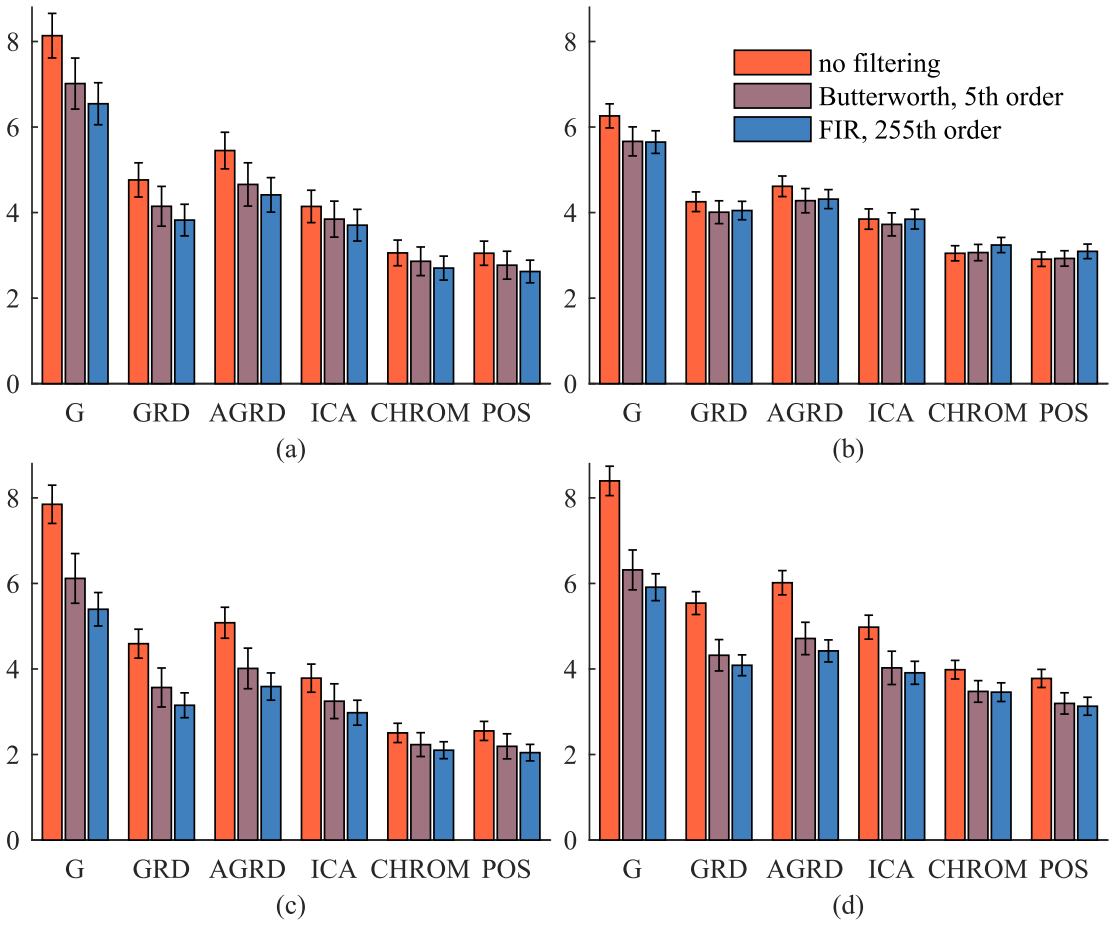


**Figure 9.** Effect of band-pass filtering on mean absolute error estimated by IBI (a), DFT (b), AR modelling (c) and CWT (d) from iPPG with pre- and post-processing according to table 2.

Band-pass filtering in most cases improved pulse rate estimation (see figure 9), but had little positive or even negative effect on pulse rate estimation by AR modelling; we cannot explain this result. The 255th order FIR filter performed slightly better than the 5th order IIR Butterworth filter; this was expected since frequency response of the latter is slightly worse.

Wavelet filtering eliminated of large errors in pulse rate estimation which is reflected by prominent decrease of RMSE (see table 3). However, wavelet filtering only slightly improved estimation error and almost did not change fraction of correctly estimated epochs $PE_{3.5}$. Parameter choice for the wavelet filter deserves a separate study: preserving several harmonics of pulse rate is of interest to keep the shape of iPPG signal.

Pre- versus post- processing. Band-pass and MA filtering was preferable at pre-processing (Step 2) when ICA or aGRD are used at Step 3 and at post-processing (Step 4) for other methods of iPPG extraction, see table 2. For ICA, POS and CHROM using filtering at a different step considerably decreased quality of iPPG and of pulse rate estimation. This is quite unexpected since originally POS and CHROM were proposed with band-pass filtering as pre-processing (de Haan and Jeanne 2013, Wang *et al* 2017), while for ICA post-processing was recommended (McDuff *et al* 2014a). For aGRD band-pass filter is essentially a pre-processing sub-step and we did not observe any difference between using MA filter as pre- and post-processing.

### 3.5. Step 5: Pulse rate estimation

The best results in terms of all metrics were provided by CWT. This method is especially useful since it allows to estimate not only average but also momentary pulse rate (see figure 10(a)).

Other tested methods for pulse rate estimation have certain drawbacks. DFT provided the second best result in terms of estimation error and fraction of correctly estimated epochs $PE_{3.5}$ (table 1), but it had low frequency resolution (see figure 10(b)) and highest RMSE. IBI estimation (figure 10(c)) had the lowest overall performance that can be explained by the insufficient quality of iPPG extracted from compressed DEAP videos. IBI filtering techniques (McDuff *et al* 2014a) may improve precision of

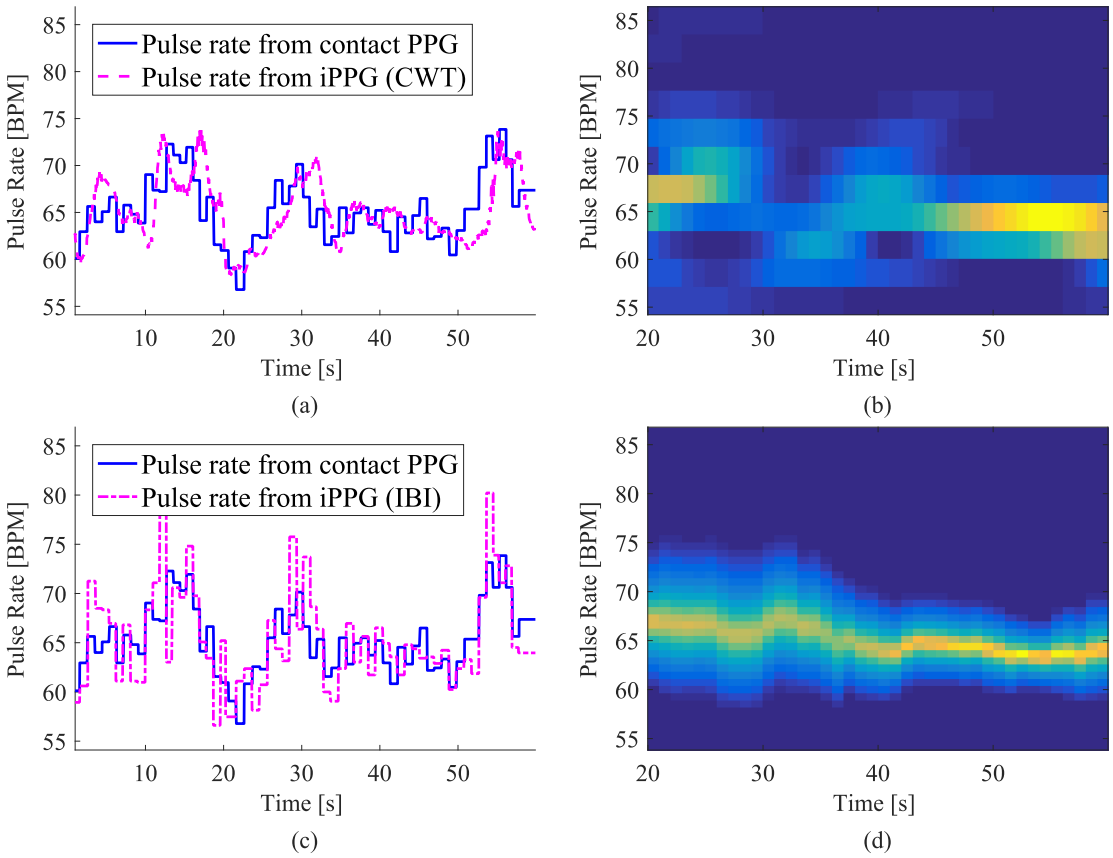


**Figure 10.** Performance of pulse rate estimation methods for iPPG signal extracted using POS method from data P1 T24: momentary pulse rate estimated by CWT (a, smoothed by 1-s moving average) and IBI (c), spectrograms of iPPG signal estimated using DFT (b) and by AR modelling (d).

IBI estimation. Finally, AR modelling requires an elaborate choice of model order as using orders different from those selected in subsection 2.2.5 resulted in considerably worse pulse rate estimation.

## 4. Conclusion

In this paper we have compared methods for pulse rate estimation using imaging photoplethysmogram (iPPG) extracted from compressed facial videos of DEAP dataset (Koelstra *et al* 2012). These videos have sufficiently good quality and allow more precise pulse rate estimation than MAHNOB (Soleymani *et al* 2012) and MMSE-HR (Zhang *et al* 2016) datasets studied in (Li *et al* 2014, Tulyakov *et al* 2016). DEAP dataset can also be used to investigate estimation of pulse rate variability (McDuff *et al* 2014b) and respiratory rate (Tarassenko *et al* 2014) from iPPG. However, as a benchmark for iPPG extraction this dataset has two potential shortcomings: first, only contact photoplethysmogram is available as a reference while more precise ECG would be preferable; second, a limited number of movements in DEAP videos reduces usefulness of this dataset for testing methods of face tracking and motion artefacts suppression. Therefore a new dataset containing compressed videos and created with a specific aim to provide a benchmark for iPPG extraction would be of interest.

Let us summarize the main results of this work. We have proposed a framework for iPPG-based pulse rate estimation. Using this framework we have compared various methods of iPPG analysis for compressed video from DEAP dataset; best pulse rate estimation was achieved when using following methods.

**Step 1, ROI selection:** whole face ROI with skin selection and outliers rejection.

**Step 2, pre-processing:** mean-centring with scaling; moving average filtering (for ICA) with filter length $M$ close to $\frac{1}{4}F_{SR}$, where $F_{SR}$ is sampling rate in Hz; band-pass filtering (for ICA and aGRD) with 255th FIR filter.

**Step 3, iPPG extraction:** POS; result for CHROM and ICA are also relatively good.

**Step 4, post-processing:** moving average and band-pass filtering (if not used at pre-processing), wavelet filtering.

**Step 5, pulse rate estimation:** continuous wavelet transform.

## Acknowledgments

The author acknowledges funding from the Ministry for Science and Education of Lower Saxony and the Volkswagen Foundation through the program "Niedersächsisches Vorab". Additional support was provided by the Leibniz Association through funding for the Leibniz ScienceCampus Primate Cognition.

The author is grateful to MPI for Dynamics and Self-Organization for the support during the work at the manuscript. The author thanks Dr S. Möller, Dr I. Kagan and Prof. F. Wolf for useful discussions and remarks on the manuscript.

## ORCID iDs

Anton M Unakafov https://orcid.org/0000-0001-6782-5099

## References

Allen J 2007 *Physiol. Meas.* **28** R1
Bousefsaf F, Maaoui C and Pruski A 2013 *Biomed. Signal Process. Control* **8** 568–74
Bousefsaf F, Maaoui C and Pruski A 2016 *Bio-Medical Materials and Engineering* **27** 527–38
Cardoso J F 1999 *Neural Comput.* **11** 157–92
Cui Y, Fu C H, Hong H, Zhang Y and Shu F 2015 *IEEE International Conference on Wireless Communications & Signal Processing (WCSP)* pp 1–5
de Haan G and Jeanne V 2013 *IEEE Trans. Biomed. Eng.* **60** 2878–86
de Haan G and Van Leest A 2014 *Physiol. Meas.* **35** 1913
Elgendi M, Norton I, Brearley M, Abbott D and Schuurmans D 2013 *PLoS One* **8** e76585
Feng L, Po L M, Xu X, Li Y and Ma R 2015 *IEEE Trans. Circuits Syst. Video Technol.* **25** 879–91
Holton B, Mannapperuma K, Lesniewski P and Thomas J 2013 *Physiol. Meas.* **34** 1499
Hülsbusch M 2008 *An Image-Based Functional Method for Opto-Electronic Detection of Skin-Perfusion* (German: RWTH Aachen)
Hülsbusch M and Blazek V 2002 *Medical Imaging* Physiology and Function from Multidimensional Images **4683** 110–8
Kamshilin A, Nippolainen E, Sidorov I, Vasilev P, Erofeev N, Podolian N and Romashko R 2015 *Sci.Rep.* **5** 10494
Kamshilin A, Sidorov I, Babayan L, Volynsky M, Giniatullin R and Mamontov O 2016 *Biomedical Optics Express* **7** 5138–47
Koelstra S, Muhl C, Soleymani M, Lee J S, Yazdani A, Ebrahimi T, Pun T, Nijholt A and Patras I 2012 *IEEE Transactions on Affective Computing* **3** 18–31
Kumar M, Veeraraghavan A and Sabharwal A 2015 *Biomedical Optics Express* **6** 1565–88
Lewandowska M, Rumiński J, Kocejko T and Nowak J 2011 *Federated Conference on Computer Science and Information Systems (FedCSIS)* 405–10
Li X, Chen J, Zhao G and Pietikainen M 2014 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4264–71
Lienhart R 2000 Stump-based 20 × 20 gentle adaboost frontal face detector Accessed 17-04-2018 https://github.com/opencv/opencv/blob/master/data/haarcascades/haarcascade_frontalface_alt.xml
Mannapperuma K, Holton B, Lesniewski P and Thomas J 2014 *Physiol. Meas.* **36** 67
McDuff D, Estepp J, Piasecki A and Blackford E 2015 *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Piscataway, NJ: IEEE) pp 6398–404
McDuff D, Gontarek S and Picard R 2014a *IEEE Trans. Biomed. Eng.* **61** 2593–601
McDuff D, Gontarek S and Picard R 2014b *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Piscataway, NJ: IEEE) pp 2957–60
Perry C and Watkins S 2011 Non-contact vital sign monitoring via ultra-wideband radar, infrared video, and remote photoplethysmography: viable options for space exploration missions *Technical Report* TM-2011-216145 NASA
Poh M Z, McDuff D and Picard R 2010 *Opt. Express* **18** 10762–74
Poh M Z, McDuff D and Picard R 2011 *IEEE Trans. Biomed. Eng.* **58** 7–11
Rouast P, Adam M, Chiong R, Cornforth D and Lux E 2016 *Fronties of Computer Science* **10** 1-15
Sahayadhas A, Sundaraj K and Murugappan M 2012 *Sensors* **12** 16937–53
Schäfer A and Vagedes J 2013 *Int. J. Cardiol.* **166** 15–29
Smith S 1997 *The Scientist and Engineeras Guide to Digital Signal Processing* (San Diego: California Technical Pub.)
Soleymani M, Lichtenauer J, Pun T and Pantic M 2012 *IEEE Transactions on Affective Computing* **3** 42–55
Sun Y, Hu S, Azorin-Peris V, Kalawsky R and Greenwald S 2013 *J. Biomed. Opt.* **18** 061205–061205
Sun Y and Thakor N 2016 *IEEE Trans. Biomed. Eng.* **63** 463–77
Takano C and Ohta Y 2007 *Med. Eng. Phys.* **29** 853–7
Tamura T, Maeda Y, Sekine M and Yoshida M 2014 *Electronics* **3** 282–302
Tarassenko L, Villarroel M, Guazzi A, Jorge J, Clifton D and Pugh C 2014 *Physiol. Meas.* **35** 807
Tarvainen M, Ranta-Aho P and Karjalainen P 2002 *IEEE Trans. Biomed. Eng.* **49** 172–5
Tasli H, Gudi A and den Uyl M 2014 *IEEE International Conference on Image Processing (ICIP)* pp 1410–4
Tulyakov S, Alameda-Pineda X, Ricci E, Yin L, Cohn J and Sebe N 2016 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp 2396–404
Verkruysse W, Svaasand L and Nelson J S 2008 *Opt. Express* **16** 21434–45
Vezhnevets V, Sazonov V and Andreeva A 2003 *Proceedings Graphicon vol 3 (Moscow, Russia)* pp 85–92
Viola P and Jones M 2001 *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (Piscataway, NJ: IEEE) I–511
Wang W, den Brinker A, Stuijk S and de Haan G 2017 *IEEE Trans. Biomed. Eng.* **64** 1479–91
Wang W, Stuijk S and de Haan G 2015 *IEEE Trans. Biomed. Eng.* **62** 415–25
Wang W, Stuijk S and de Haan G 2016 *IEEE Trans. Biomed. Eng.* **63** 1974–84
Zarit B, Super B, Quek F *et al* 1999 *Proceedings of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems* (Piscataway, NJ: IEEE) 58–63
Zhang Z *et al* 2016 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp 3438–46