



Pulse oximetry based on photoplethysmography imaging with red and green light

Calibratability and challenges

Andreia Moço¹ · Wim Verkrusse²

Received: 12 March 2019 / Accepted: 15 December 2019 / Published online: 1 January 2020
© Springer Nature B.V. 2020

Abstract

Remotely measuring the arterial blood oxygen saturation (SpO_2) in visible light (Vis) involves different probing depths, which may compromise calibratability. This paper assesses the feasibility of calibrating camera-based SpO_2 ($\text{SpO}_{2,\text{cam}}$) using red and green light. Camera-based photoplethysmographic (PPG) signals were measured at 46 healthy adults at center wavelengths of 580 nm (green), 675 nm (red), and 840 nm (near-infrared; NIR). Subjects had their faces recorded during normoxia and hypoxia and under gradual cooling. $\text{SpO}_{2,\text{cam}}$ estimates in Vis were based on the normalized ratio of camera-based PPG amplitudes in red over green light (*RoG*). $\text{SpO}_{2,\text{cam}}$ in Vis was validated against contact SpO_2 (reference) and compared with $\text{SpO}_{2,\text{cam}}$ estimated using red-NIR wavelengths. An *RoG*-based calibration curve for SpO_2 was determined based on data with a SpO_2 range of 85–100%. We found an A_{rms}^* error of 2.9% (higher than the A_{rms}^* for $\text{SpO}_{2,\text{cam}}$ in red-NIR). Additional measurements on normoxic subjects under temperature cooling (from 21 °C to < 15 °C) evidenced a significant bias of -1.7 , CI [$-2.7, -0.7$]%. It was also noted that $\text{SpO}_{2,\text{cam}}$ estimated at the cheeks was significantly biased (-3.6 , CI [$-5.7, -1.5$]%) with respect to forehead estimations. Under controlled conditions, $\text{SpO}_{2,\text{cam}}$ can be calibrated with red and green light but the accuracy is less than that of $\text{SpO}_{2,\text{cam}}$ estimated in the usual red-NIR window.

Keywords PPG imaging · Pulse oximetry · Calibration · Video-health monitoring

1 Introduction

Pulse oximetry is a non-invasive technique that is used for measuring the blood oxygen saturation of arterial blood (SpO_2)¹ [2]. This vital sign is widely monitored in intensive care units, general wards, during surgery, and in ambulant patients [10, 13].

Most commercial SpO_2 devices require a probe to be attached to the finger. SpO_2 devices operate by shining light into the skin and by collecting the photons that travelled through the tissue. A photodetecting element converts the probed light intensity into an electric signal which includes

a photoplethysmographic (PPG) component. Irrespective of data acquisition mode and probing wavelength, PPG signals are modulated by blood volume variations at the microvascular bed of tissue and are recognizable as cardiac-related, temporally-varying and relatively low-amplitude signals [1, 8].

Typically, monitoring SpO_2 involves PPG measurements at red and near-infrared (NIR) wavelengths. This is because PPG-amplitudes acquired using red (e.g., 660–700 nm) are sensitive to the saturation level of the arterial blood whereas the PPG-amplitudes in NIR (e.g. 800–950 nm) enable accounting for individual and temperature variations [3]. The ratio of the normalized PPG-amplitudes in red over NIR (hereafter denoted as *RoIR*) is commonly known as the “ratio-of-ratios” (RRs) method, with the understanding that the PPG amplitude for each wavelength derived from a

✉ Andreia Moço
moco.a.v@gmail.com

¹ Electronic Systems Group, Eindhoven University of Technology, Eindhoven, The Netherlands

² Philips Innovation Group, Philips Research, 5612 AZ Eindhoven, The Netherlands

¹ The saturation of arterial blood is referred to as SaO_2 when measured invasively. The “p” in SpO_2 refers to “peripheral” measurement locations such as the finger, ears or forehead [8, 10].

ratio between a pulsatile component, AC, and non-pulsatile component, DC [22].

While relatively unobtrusive, contact probes are undesirable in patient groups with sensitive skin such as premature infants and the wounded. This motivated the concept of measuring PPG remotely and development of camera-based SpO_2 ($\text{SpO}_{2,\text{cam}}$) systems [21]. One fundamental difference between contact and camera-based/imaging pulse oximetry is the shallower light penetration depth involved in the latter. Verkruysse et al. [20] showed that by using 675 nm (red) and 840 nm (NIR), a single universal calibration curve with acceptable spread between individuals could be achieved. Also van Gastel et al. [18] elaborated on the calibratability of camera-based SpO_2 estimates by showing that motion-robustness improves when more than two wavelengths in the red-NIR range are combined to cancel out motion interference in PPG signals.

Despite the recognized value of using red and NIR wavelengths, there are practical reasons for preferring visible light (Vis) wavelengths. For example, green PPG amplitudes are the strongest across the range of 300–900 nm [19]. Also, modern illumination setups are becoming less reliant on incandescent lamps and the availability of NIR light is no longer a given. Lastly, operating in Vis could enable cost-effective implementations since RGB cameras, including those on smartphones, could be used. As demonstrated by Guazzi et al. [7], the red and blue camera channels of an RGB camera could provide inputs for an adapted RRs-based methodology for estimating SpO_2 . Trends were shown in the logarithmic ratio of the normalized red over normalized green which were associated with SpO_2 changes in five subjects. However, calibrations were subject dependent and temporal consistency was not assessed.

Another concern about the validity of camera-based SpO_2 estimations in Vis is the fact that blue-green and red wavelengths differ in terms of penetration depth. This challenges the usual pulse oximetry assumption of wavelengths interrogating the same vasculature. This aspect is emphasized in Fig. 1 (adapted from [20]) which shows, qualitatively, that the contribution to the camera-based PPG signals acquired using blue and green wavelengths are mostly modulated by arterioles at the upper dermal layers whereas signals acquired using red wavelengths even reach subdermal tissue. The depth-gap between probing wavelengths may increase the inaccuracy of camera SpO_2 estimates because of depth-variations in skin properties and its interference on the RRs, e.g. the blood concentration [16].

The focus of this paper is on the calibration feasibility of $\text{SpO}_{2,\text{cam}}$ using green as a replacement for NIR for normalization, while keeping red as the saturation-sensitive wavelength. We limit our scope to an adapted version of

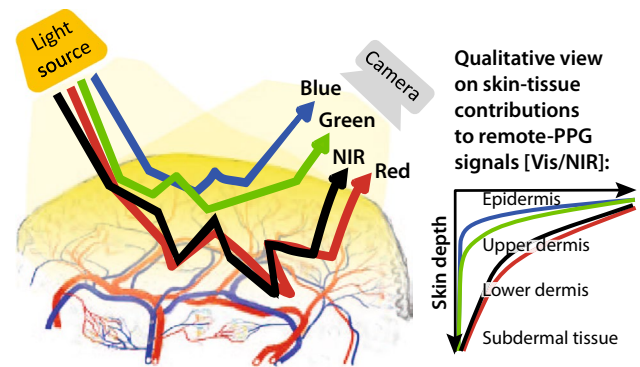


Fig. 1 The penetration depths are wavelength-dependent and PPG signals probed using red-NIR light reach the deep arteriovenous (A-V) plexus and subdermal tissue, whereas signals probed using blue and/or green have depth-origin at arteriolar blood volume variations the upper dermis

the RRs to Vis; i.e., the ratio of normalized red over the normalized green (RoG) as predictor of SpO_2 levels. While this study is ancillary to Verkruysse et al. [20], new data is reported for camera-based PPG signals probed with a camera dedicated to green light. For consistency with Verkruysse’s work, similar statistical, data processing and visualization options were applied to ease study comparisons.

Note that the nomenclature used in this paper is consistent with the recommendation for naming convention of Leonhardt et al. [14]. PPG imaging is used in reference to frameworks or algorithms that are aimed at vital signs extracted remotely from 1-dimensional PPG signals. PPG imaging and camera-based PPG are used interchangeably. Given the nomenclature redundancy in the literature, we clarify that the insights of this paper are applicable to “remote PPG”, “DistancePPG”, “non-contact PPG”, “video-PPG” and “imaging PPG”.

2 Methods

Section 2.1 refers to compliance with ethical standards. Section 2.2 describes the experimental setup and protocols leading to the paired data (camera-based RoG , reference SpO_2). Section 2.3 is dedicated to signal processing and details the determination of the calibration curve (i.e., linear regression) parameters for the SpO_2 level (%) as function of RoG .

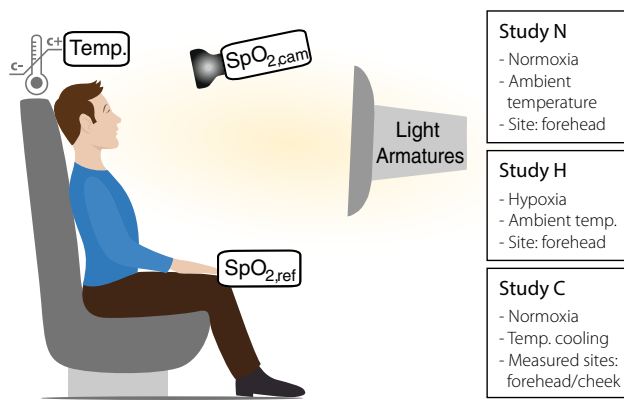


Fig. 2 Setup schematics and protocols/studies. Sitting subjects were measured by monochrome cameras coupled with bandpass filters for R, G and NIR, as well as by finger pulse oximetry ($SpO_{2,ref}$). Two light armatures with incandescent bulbs were used. Gradual temperature cooling or hypoxia were induced in some of the recordings

2.1 Ethical statement

The study was not a clinical trial and was conducted on healthy volunteers [20]. The experimental protocols that were applied in this investigation (studies N, H, and C) were approved by the Philips IRB (Internal Committee on Biomedical Experiments)² and a written informed consent was obtained from each subject. All experiments were carried out in accordance with the ethical standards laid down in the 1964 Helsinki declaration and its later amendments.

2.2 Setup and dataset construction

In a basic set-up protocol (study N), 25 healthy subjects were involved. The subjects foreheads were recorded synchronously by three monochrome cameras (Stingray F046-B FireWire; Allied Vision Technologies, Stadtroda, Germany). These were equipped with bandpass filters with center / bandwidth wavelengths of 560/32 (green), 675/67 (red) and 842/56 (NIR) nm, respectively (Semrock FF01-579/34-25, FF02-675/67-24-D and FF01-842/56-24D; IDEX Corp., Lake Forrest, IL) and sampled at 15 frames per second. Subjects were recorded in the sitting position and motion was minimized by a head support (see Fig. 2).

Stable illumination was provided by two armatures (Falcon Eyes, Hong Kong, China), each equipped with 9 incandescent lamps (Philips, 40 W; DC-powered). Both were placed symmetrically and at a distance of about 1 m from the subjects face. Assuming an isotropic light distribution and a luminous efficacy of 15 lumens per watt, the total luminous

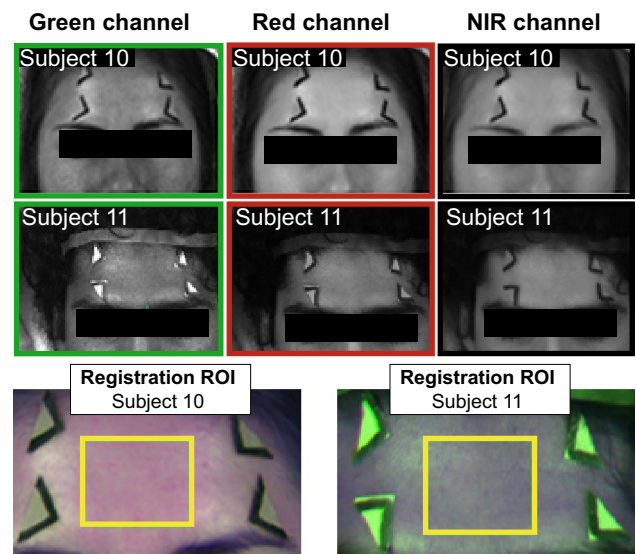


Fig. 3 Representative forehead frames and skin selections (in yellow) for probing camera-based PPG signals in the green, red and NIR camera channels

flux and intensity is estimated as 10.8 kilo-lumens and 859.4 candela, respectively.

To minimize the contamination of specular reflections in facial images, the light sources were placed at an angle of 45° with respect to the camera. However, including polarising filters could have been beneficial for normalized PPG-amplitude measurements in green wavelengths, in particular in subjects with oily skin and/or darker skin tones. Further considerations on specular reflections are included in the discussion section of this paper.

Figure 3 shows representative frames of forehead recordings performed using the green, red and NIR camera channels. To ease the registration of camera video-recordings, markers were placed on the subjects forehead. The markers (consisting of “L”-shaped pieces of paper) were gently attached skin using double-sided adhesive tape. After registration, rectangular skin regions of interest (sROIs) were manually selected in a subset of uniformly illuminated skin pixels within the forehead (see yellow rectangles in Fig. 3).

For arriving at a SpO_2 reference signal, the sample-wise (1 Hz) median of measurements from four commercially available contact probes was taken. The reason for combining commercial probes was improving the precision of the reference signal.

Two protocol adaptations to study N were implemented; one to broaden the range of SpO_2 variations (study H) and another to explore the possible effect of patient centralization (as mimicked by temperature cooling; study C). Study H involved 21 healthy subjects and consisted of measurements performed in a [normobaric] hypoxic tent. Study C was performed in a climate-controlled room with an initial

² IRB contact: PHM Keizer, Senior Ethical & Biomedical Officer, High Tech Campus 34, 5656 AE Eindhoven, The Netherlands.

temperature of 21.4 ± 2.0 °C (average \pm standard deviations; $N = 21$) that was gradually reduced in steps of about 5 °C until a final temperature of 8.4 ± 1.2 °C was reached. Video recordings of the subjects foreheads and cheeks were made before, in the middle of the cooling intervention and when about 5–8 °C was reached. Study C involved ten healthy subjects and three recordings per subject.

Conventional (contact-based) pulse oximetry was obtained in all subjects with the aim of becoming a SpO_2 reference. Four commercially available sensors, coupled to a Philips MP2 patient monitor, were used: a Philips finger and ear sensor (M1191B and M1194A, Philips Medizin-Systeme, Böblingen, Germany), a Masimo finger sensor (LNCS DC-I, Masimo Corporation, Irvine) and a Nellcor finger sensor (DS-100A Medtronic, Dublin, Ireland). The pulse oximetry sensors were synchronized between each other and acquired 1 $SpO_{2,ref}$ (%) sample per second. Note that the camera sensors were synchronized between each other by using LabVIEW software (National Instruments, Austin, TX) while video recordings were synchronized manually with $SpO_{2,ref}$ readings manually. Further details on the dataset preparation are found elsewhere [20].

For the current investigation, 36 green, red and NIR recordings from studies N and C were used to assess the calibratability of $SpO_{2,cam}$ from RoG and from $RoIR$. 40 additional recordings from 5 healthy subjects were used from study C. Study C contains 10 forehead measurements and 10 cheek measurements at room temperature conditions, 10 forehead measurements at medium temperature and 10 forehead measurements at low temperature. Each subject was measured twice for each condition and skin site. The duration of the recordings ranged from 3 to 20 min (median, 5 min). The camera video recordings and the conventional pulse oximetry (reference) signals were acquired by LabVIEW and stored in an uncompressed format for offline processing in Matlab (version 2018a, the Mathworks, NA).

2.3 Data analysis and statistics

2.3.1 Signal processing

Reference SpO_2 After manually synchronizing the SpO_2 traces with video recordings, a reference SpO_2 signal was determined for each recording from the median of 4 finger pulse oximetry traces. One reference SpO_2 value, $SpO_{2,ref}^{(i)}$, was stored per recording session i .

PPG imaging: signal extraction After registration of the red, green and NIR images, rectangular ROIs were manually defined such that they only contained skin. For

all subjects, the average ROI area was $17 \pm 4\%$ of the total image area (19k ROI pixels out of 45k total image pixels). For subjects 10 and 11 (exemplified in Fig. 3), the ROI area were 86,868 and 72,927 pixels, respectively.

Raw camera-based PPG signals, $G(t)$, $R(t)$ and $IR(t)$, were obtained as the spatially averaged pixel intensity in a manually selected skin region of interest (sROI) on the subjects forehead. The demarcation of the sROIs was aimed at maximizing skin area while avoiding specular locations and body hair. Because the subjects face was stabilized by a supporting structure, it was sufficient to apply static sROIs over each video recording.

AC/DC-normalization The signals $G(t)$, $R(t)$ and $IR(t)$ were normalized by division by its low-pass filtered component (LPF): $x_{AC/DC} = (x - x_{LPF})/x_{LPF}$. The procedure is known as AC/DC-normalization and is employed as a means to achieve invariance to external lighting fluctuations or slow body motions [5, 19].

To generate low-pass filter coefficients, the Matlab function `butter` was used to generate an 11th order Butterworth filter (normalized cutoff frequency, $0.667/(f_s/2)$, where the parameter $f_s = 15$ Hz is the sampling frequency). Filtering was applied to $G(t)$, $R(t)$ and $IR(t)$ with the function `filtfilt`. The normalized signals were denoted as $G_n(t)$, $R_n(t)$ and $IR_n(t)$.

Normalized signals can be seen as the additive contributions of a source signal, $s(t)$ or $s'(t)$, a common motion signal, $m(t)$, and additive white Gaussian noise, $n(t)$, to account for sensor noise:

$$G_n(t) = \beta_G s'(t) + m(t) + n_G(t) \quad (1)$$

$$R_n(t) = \beta_R s(t) + m(t) + n_R(t) \quad (2)$$

$$IR_n(t) = \beta_{IR} s(t) + m(t) + n_{IR}(t) \quad (3)$$

The β terms translate the wavelength-dependency of the PPG imaging amplitudes and the following holds by definition: $RoG \triangleq \beta_R/\beta_G$ and $RoIR \triangleq \beta_{IR}/\beta_{IR}$. Note that the notation used in Eq. 1 for the PPG source of $G_n(t)$, $s'(t)$, is distinguished from the notation used for $R_n(t)$ and $IR_n(t)$. This translates morphological differences between green and red-NIR signals. The motion $m(t)$ is common for all signals. While there is a line of camera-based PPG research aimed at motion-robustness, e.g. [5, 18], in a datasets comprising recordings with only sporadic and non-severe skin motions, the approaches within the next paragraphs are sufficient for obtaining reliable measurements of RoG and $RoIR$ (see Appendix A).

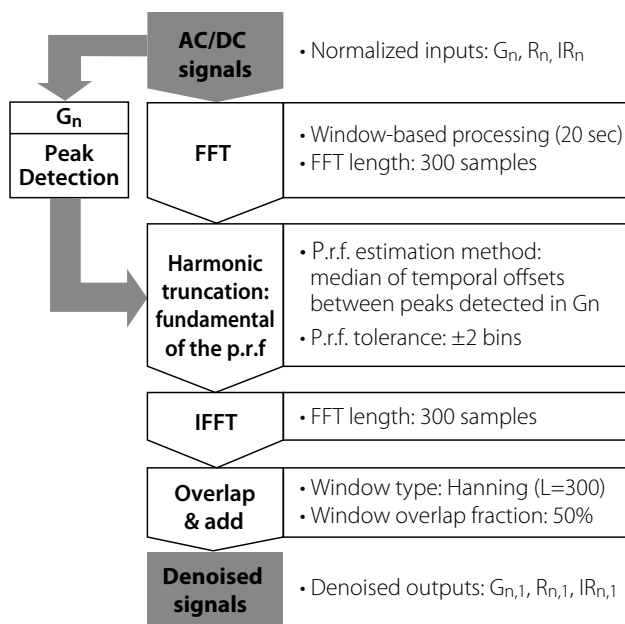


Fig. 4 Approach to denoise camera-based PPG signals

Denoising In a second step, $G_n(t)$, $R_n(t)$ and $IR_n(t)$ were temporally filtered in the frequency domain to the fundamental of the pulse-rate frequency (p.r.f.). This strategy relies on harmonic truncation and is aimed at achieving higher signal sensitivity [12]. Figure 4 depicts the denoising approach used in this study, featuring window-based processing with harmonic truncation guided by the p.r.f. estimated at the signal $G_n(t)$.

The systolic peaks of $G_n(t)$ were first detected by peak detection³. Then, the camera-based PPG signals were processed in an overlap-and-add manner with temporal windows of 300 digital samples (i.e., a duration of 20 s each) and 50% overlap (window type: Hanning). In each window, signals were Fourier-transformed (FFT) and truncated to the frequency bin that corresponds to the p.r.f. (tolerance: ± 2 bins). The denoised signals were denoted as $G_{n,1}(t)$, $R_{n,1}(t)$ and $IR_{n,1}(t)$.

Calculating RRs PPG-amplitudes were determined as the difference between peaks and valleys in denoised PPG imaging signals. By defining the operator $E_{peaks}[\cdot]$ as the median of these cardiac-related systolic peaks and diastolic valleys in signals, the amplitudes of $G_{n,1}(t)$, $R_{n,1}(t)$ and $IR_{n,1}(t)$ were determined as follows:

$$|G_n| = E_{peaks}[G_{n_1}(t)] + E_{peaks}[-G_{n_1}(t)], \tag{4}$$

³ Note that, while any camera channel could have been used to detect peaks in camera-based PPG signals, normalized amplitudes are highest in green wavelengths and this advantage translates to increased reliability of the detected peaks.

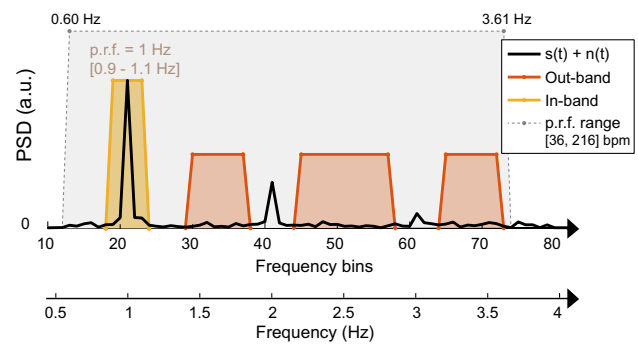


Fig. 5 In and out frequency bands for estimating the SNR of $R_n(t)$. Plotted in black is the power spectral density (PSD) of a sinusoidal signal, s , with a frequency of 1 Hz and 2 harmonics plus sensor noise, n (additive white Gaussian noise). Acronyms: a.u., arbitrary units; p.r.f., pulse-rate frequency

$$|R_n| = E_{peaks}[R_{n_1}(t)] + E_{peaks}[-R_{n_1}(t)], \tag{5}$$

$$|IR_n| = E_{peaks}[IR_{n_1}(t)] + E_{peaks}[-IR_{n_1}(t)]. \tag{6}$$

The RRs between the normalized red over green amplitudes, RoG , and red over NIR, $RoIR$, were finally determined for each subject as $RoG = |R_n|/|G_n|$ and $RoIR = |R_n|/|IR_n|$, respectively. Each recording session provided two data pairs: $(RoG^{(i)}, SpO_{2,ref}^{(i)})$ and $(RoIR^{(i)}, SpO_{2,ref}^{(i)})$. The group descriptives of $SpO_{2,ref}$ and $SpO_{2,cam}$ by using $|R_n|$, $|G_n|$, $RoIR$, RoG were determined as means \pm standard deviations (std) or as 95% confidence intervals (CI).

2.3.2 Calibratability assessment

A metric of signal quality Noise and sporadic motions occur even under controlled measurement conditions. As a metric of signal quality, the signal-to-noise ratio (SNR) of $R_n(t)$ was proposed [20]. This choice was motivated by the fact that $R_n(t)$ has typically the weakest PPG-amplitudes and lower-bounds SNR in the Vis-NIR range. For each video recording i , we obtained $SNR_{R_n}^{(i)}$ as the mean of estimates computed at 20-sec long, non-overlapping windows over $R_n(t)$.

In each window, the transient SNR was computed as the logarithmic ratio of the power of $R_n(t)$, P_{R_n} , over noise, N_{AWGN} . P_{R_n} was estimated in a frequency band containing the fundamental and two harmonic(s) of the pulse-rate frequency. Figure 5 illustrates possible in- and out-bands for estimating P_{R_n} and N_{AWGN} in the frequency domain (FFT length, 300 samples).

The power of $R_n(t)$ was divided by the noise floor estimated in non-pulse bins surrounding the harmonics of the p.r.f. (see out-bands in Fig. 5). Diverging from Verkruyse et al., the signal quality metric is not aimed at excluding subjects from the dataset.

Regression analysis Linear coefficients (C_1, C_2) were fitted to the data pairs $(RoG^{(i)}, SpO_{2,ref}^{(i)})$ as

$$SpO_{2,ref}^{(i)} = C_1 RoG^{(i)} + C_2. \tag{7}$$

A robust algorithm for determining (C_1, C_2) was used instead of the regular least squares method (Matlab function `fitlm` with robust options enabled). Robustness was achieved by using a bisquare function to penalize outliers, thus mitigating the influence of possible outliers or noisy points in the regression coefficients [6, 9]. We computed the significance of the regression, its Pearson’s R-squared (R^2) and the root-mean-squared error metric A_{rms}^* defined as

$$A_{rms}^* = \sqrt{\frac{1}{n} \sum_{i=1}^n |SpO_{2,ref}^{(i)} - SpO_{2,cam}^{(i)}|^2}. \tag{8}$$

The asterisk in A_{rms}^* is used to emphasize that this metric excludes short-time errors (e.g., due to sporadic head movements). For Eq. 7 with coefficients (C_1, C_2) , plots were provided with 99% confidence intervals (CIs; based on the χ^2 distribution for the root-mean squared error of the regression) and the ISO 80601-2-61:2011 requirement of $A_{rms}^* < 4\%$. The same procedure was followed for calibrating $SpO_{2,cam}$ based on $RoIR$.

2.3.3 Challenges to calibrability

Challenge 1 The data obtained during temperature cooling was explored by testing the significance of the mean difference between RoG measured at room, medium and low temperature conditions. Because RoG data was Gaussian-distributed (as assessed by the Shapiro-Wilk test), one way ANOVA tests were used to perform comparisons. Also, we assessed whether temperature-perturbed data points were within the error margins of the calibration curve obtained for $SpO_{2,cam}$.

Challenge 2 Also a switch in the measured skin-site can increase the estimation errors by (7). Skin-site variations were explored by testing the significance of the difference of the means between $RoGs$ obtained at room temperature between paired samples of RoG computed at forehead versus cheeks. The RoG for the two skin sites were also Gaussian distributed and paired-sample t tests were used to test the significance of the mean differences $SpO_{2,cam}$ estimates from forehead vs. cheeks. The significance level p was set at 0.05.

3 Results

3.1 Calibration Constants

Figure 6a shows the scattered data from studies N and H for the pairs $(RoG^{(i)}, SpO_{2,ref}^{(i)})$. A fit to these points enabled 99%

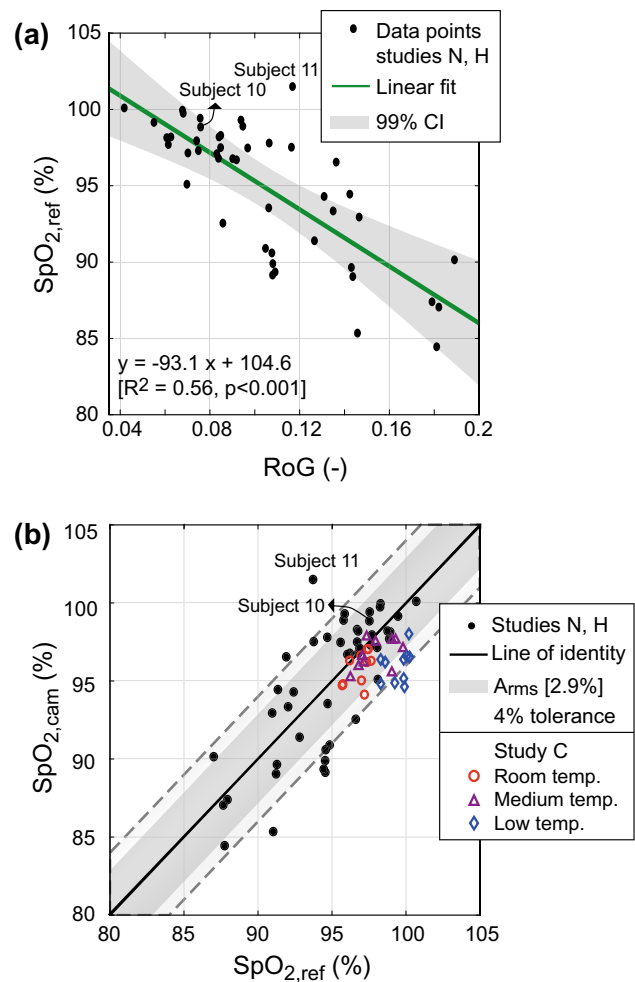


Fig. 6 **a** Scatter plot and regression line along with its corresponding 99% confidence interval ($N = 46$); **b** Calibration data from studies N and H, shown along with data from study C (cooling intervention). The labels “Low”, “Medium”, and “Room” refer to temperatures of approximately 8 °C, 14 °C and 22 °C, respectively

margins of agreement within 4% and an A_{rms}^* of 2.9% (see Fig. 6b; $N = 46$). Comparatively, the pairs $(RoIR^{(i)}, SpO_{2,ref}^{(i)})$ would also enable 99% margins of agreement within 4% and an A_{rms}^* of 1.7%.

Similar to conventional pulse oximetry, RoG and $RoIR$ were negatively associated with SpO_2 . The obtained calibration curves are as follows:

$$SpO_{2,cam}(RoG) = -98.0RoG + 105.0, \tag{9}$$

$$SpO_{2,cam}(RoIR) = -45.2RoIR + 116.8. \tag{10}$$

The fact that the coefficient C_1 is twice as large for $SpO_{2,cam}(RoG)$ than for $SpO_{2,cam}(RoIR)$ suggests a disadvantage of green as an alternative to NIR wavelengths. Still, Eq. 9 was significant and evidenced a moderate strength of association between $SpO_{2,ref}$ and $RoIR$ ($R^2 = 0.58, p < 0.001$; see

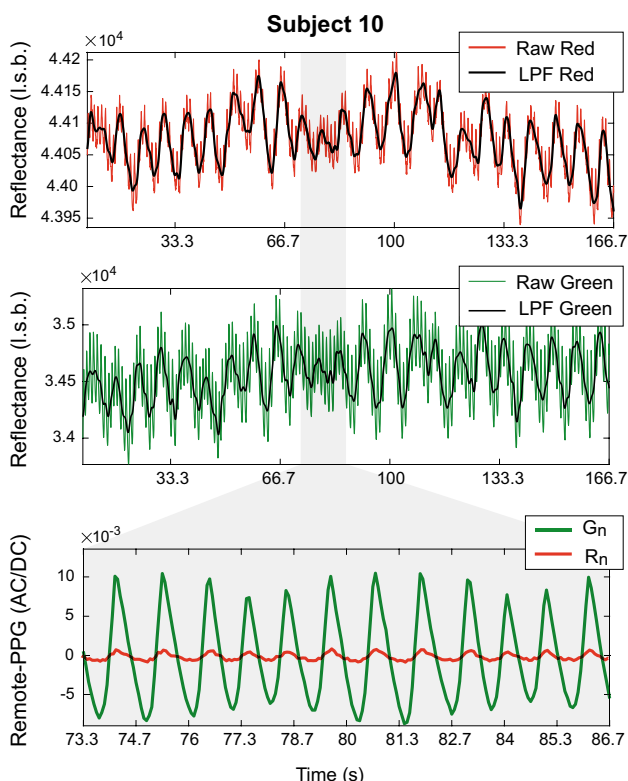


Fig. 7 Raw and normalized camera-based PPG signals for an inlier of the calibration study ($SpO_{2,cam}(RoG)$, 97.6%; $SpO_{2,ref}$, 98.9%; Subject 10, study N). The outcomes of low-pass-filtering (LPF) raw camera-based PPG signals are included

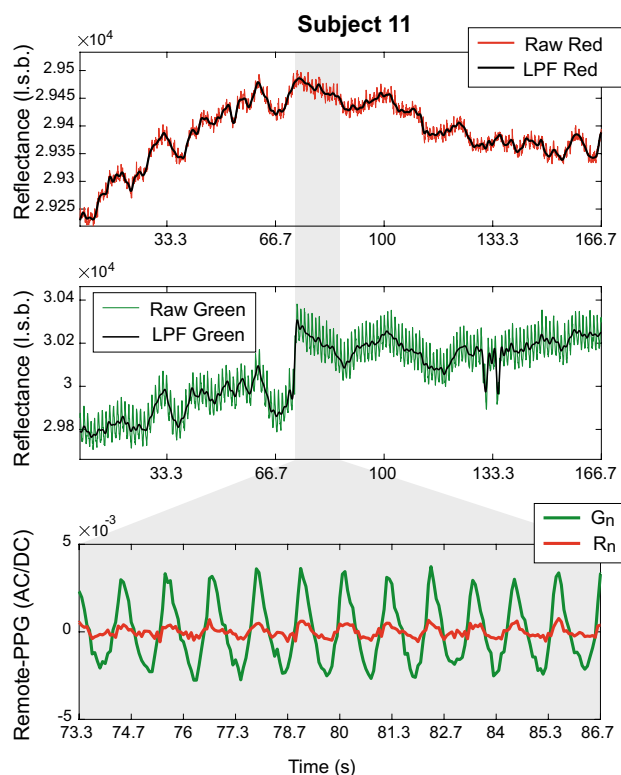


Fig. 8 Raw and AC/DC-normalized camera-based PPG signals for an outlier of the calibration study ($SpO_{2,cam}(RoG)$, 93.7%; $SpO_{2,ref}$, 101.5%; Subject 11, study N). Acronym: LPF, low-pass-filtering

Fig. 6a). We remark that this outcome is below than the attainable by using red-NIR (Eq. 10: $R^2 = 0.86, p < 0.001; A_{rms}^* = 1.7\%$).

To explore possible causes for increased individual errors, the PPG data from an outlier in Fig. 6 (Subject 11), was compared with that of an inlier point (Subject 10). Accordingly, the raw and AC/DC-normalized camera-based PPG signals for subjects 10 and 11 are provided in Fig. 7 and in Fig. 8, respectively. The comparison of this data evidences higher noise levels (relative to PPG-amplitudes) in the outlier situation.

Note that, for subject 11, heavy modulations were observed in raw signals in addition to the cardiac-related component. The frequency of these modulations is about 0.14 Hz, which is lower than the normal range of resting respiratory frequencies (~ 0.3 Hz). It is unclear from the data whether these modulations are due to Mayer waves (~ 0.1 Hz) [11] and/or motions. Still, the pattern was not observed in other subjects.

Realizing that the magnitudes of camera-based PPG signal can differ considerably among populations, in particular among varying skin tones, we use SNR_{Rn} as a

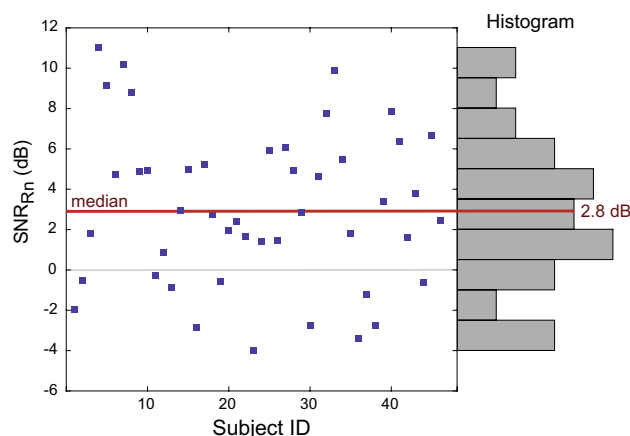


Fig. 9 SNR estimations for the recordings in the calibration dataset ($N = 46$)

describing metric of our dataset. The estimated SNR for the PPG signals in the data from studies N and H is shown in Fig. 9. The median SNR was 2.8 dB and the minimum -4.0 dB. Thus, the calibration constants developed in this study for $SpO_{2,cam}(RoG)$ are independent of the SNR levels, but an $A_{rms}^* < 3\%$ only holds for populations and

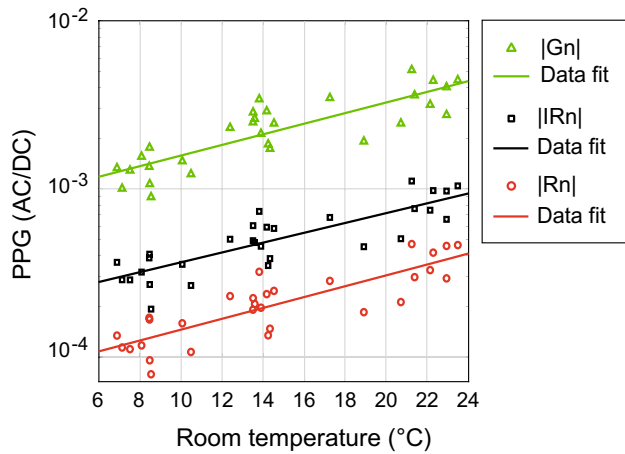


Fig. 10 The amplitudes of the camera-based PPG signals are positively associated with room temperature. The temperature-induced changes in red, NIR and green are proportional

errors for the point subset measured at temperatures > 15 °C was - 1% (CI [- 1.7, - 0.3]%, $p < 0.05, N = 10$; $SpO_{2,cam}, 95.8 \pm 1.0\%$; $SpO_{2,ref}, 96.8 \pm 0.7\%$). For the point subset measured below 15 °C, the mean underestimation error was - 2.4% (CI [- 3.1, - 1.6]%, $p < 0.05, N = 20$; $SpO_{2,cam}, 96.3 \pm 1.1\%$; $SpO_{2,ref}, 98.7 \pm 1.1\%$).

Therefore, while most bias errors fit within the margin of the calibration curve, errors do become larger with temperature cooling.

3.3 Challenge 2: a change in measurement site

Five subjects had their foreheads and cheeks measured successively under normal room temperature (22.1 ± 2.0 °C) and normoxic conditions. Table 1 shows the resulting $SpO_{2,cam}$ estimates (means±std) from the *RoG* and *RoIR* measurements performed at the forehead and cheek (study C). Table 1 indicates that *RoG*-based SpO_2 estimates are more affected by a

Table 1 Group statistics (means±std) for $SpO_{2,cam}$ estimated by *RoG* or by *RoIR* for forehead versus cheek, $SpO_{2,ref}$ and bias errors

$SpO_{2,cam}$	Forehead	$SpO_{2,ref}$	Bias [CI 95%]
by <i>RoG</i>	$95.8 \pm 1.2^{*\dagger}$	$96.8 \pm 0.7^*$	- 1.0 [- 1.8, - 0.2]*
by <i>RoIR</i>	98.6 ± 1.3	96.8 ± 0.7	1.0 [0.0, 2.0]
$SpO_{2,cam}$	Cheek	Reference	Bias [CI 95%]
by <i>RoG</i>	$93.2 \pm 2.8^{*\dagger}$	$96.8 \pm 0.6^*$	- 3.6 [- 5.7, - 1.5]*
by <i>RoIR</i>	$99.2 \pm 1.7^*$	$96.8 \pm 0.6^*$	1.8 [0.2, 3.2]*

Legend: *Significant bias between $SpO_{2,cam}$ (%) and $SpO_{2,ref}$ (%); †Significant bias between *RoG*-based estimates acquired at the forehead and cheek

recording conditions for which the median SNR_{Rn} exceeds 3 dB.

3.2 Challenge 1: temperature cooling

Cooling reduced PPG-amplitudes at all wavelengths. Figure 10 depicts separate regressions that were calculated for the logarithm of the camera-based PPG amplitudes (*IGn*), *IRn*, and *IRn*) as function of temperature.

The correlations for associations of temperature with PPG-amplitudes were significant ($p < 0.001$) and resulted in R^2 values of about 0.6. Based on the slopes of the regressions, for a temperature decrease of 18 °C, PPG-amplitudes decrease by a factor of seven.

To assess whether cooling also influences $SpO_{2,cam}$, it is useful to observe Fig. 6b, which refers to the cooling intervention (study C; the $A_{rms}^* = 2.9$ is the error margin for estimates based on $SpO_{2,cam}(RoG) = -98.0RoG + 105.0$; Eq. 9). The scattering of the data for medium and low temperatures indicates that SpO_2 was underestimated. The cooling-related bias was confirmed by *t* test comparisons. The average underestimation

skin-site change between cheek and forehead than *RoIR*-based estimates.

$SpO_{2,cam}$ estimates obtained at the forehead and cheek were compared by means of paired sample *t* tests. As shown in Table 1, the bias errors in $SpO_{2,cam}$ obtained by *RoG* or *RoIR* at the forehead are insignificant or within the error margin of the calibration curves. Conversely, the bias errors for estimations at the cheek are significant and exceed the A_{rms}^* of its respective calibration curves. Table 1 evidences that a change in skin-site measurement is only significant for *RoG*-based $SpO_{2,cam}$. Changing the measurement site from forehead to cheeks resulted in an average underestimation error for $SpO_{2,cam}$ by *RoG* of - 3.6% CI [- 5.7, - 1.5]% (i.e., above the A_{rms}^* of the calibration curve).

4 Discussion

This study assessed the calibratability of SpO_2 with center wavelengths at 560 nm (green) and at 675 nm (red). The study was performed on the basis of data from 46 healthy adult individuals recruited for [20]. Stationary sitting subjects had their forehead and/or cheeks measured under ambient room temperature during normoxic conditions and in an hypoxic tent. A calibration curve for SpO_2 as function of RoG was obtained at the forehead. The goodness of this fit was assessed in terms of the strength of correlation and by the A_{rms}^* metric (discarding short-term errors).

The presented data evidence that a single calibration curve can be used to estimate SpO_2 with cameras operating in Vis with an A_{rms}^* error $< 3.0\%$. Concluding upon the feasibility of RoG -based $\text{SpO}_{2,cam}$ is not trivial because of the considerable penetration depth-gap when compared to camera-based pulse oximetry using red and NIR. It is possible that the penetration depth-gap between red and green wavelengths makes the calibrations susceptible to physiological variations of the skin properties [15]. We minimized possible errors by constraining the measurement conditions, namely by defining a setup and protocol with static, straight-up sitting subjects, a defined measurement site (forehead) and by controlling the room temperature.

The A_{rms}^* error obtained for $\text{SpO}_{2,cam}$ using red and green was higher than the A_{rms}^* error obtained by using red and NIR but acceptable in view of typical A_{rms}^* values of 2% or 3% for commercial pulse oximeters. Still, coping with sensor noise under the requirement to periodically update SpO_2 estimates (e.g., of about just 10 s) is more difficult for $\text{SpO}_{2,cam}$ than for their contact-based counterparts which typically achieve a higher SNR and sampling rates. These technical issues can add 2% extra error, which means that technical developments are important to leverage the potential advantage of camera-based pulse oximetry.

The assumption of universal calibratability was challenged by a cooling intervention which mimics patient centralization. From the cold-induced vasoconstriction, we found that the $\text{SpO}_{2,cam}$ method by RoG tends to underestimate SpO_2 . The A_{rms}^* for estimations performed under medium and low ambient temperatures was $< 2\%$. The seemingly parallel lines suggest that calibrations could hold in the investigated temperature range. Although this result is encouraging, we remark that the temperatures achieved during the cooling intervention comprised 8–15 °C. Given the extent of this range, future work is necessary to challenge $\text{SpO}_{2,cam}$ under more extreme cooling conditions and to assess actual centralized patients.

In the second challenge of this study, it was seen that a change in skin-site measurement leads to considerable

measurements errors. This is reasonable given the depth-gap of the PPG-signals acquired using green versus red-NIR wavelengths [16]. As an implication, RoG -based $\text{SpO}_{2,cam}$ estimated obtained at cheeks require dedicated calibration coefficients to avoid bias errors of 2–6% (see Table 1). The effect of a change in skin site was also observed for $RoIR$ -based $\text{SpO}_{2,cam}$, although in a lesser extent. Thus, one may assume that the calibration curves for $\text{SpO}_{2,cam}$ in Vis are specific to the skin-site at which they were developed.

When it comes to continuous monitoring, we see an opportunity to achieve $\text{SpO}_{2,cam}$ by using green wavelengths, or possibly blue, as an alternative for NIR. However, the relatively lower performance of Vis measurements with respect to red-NIR is not yet fully understood. We believe that contributions which enable a knowledge transfer from contact-PPG are an efficient way to increase our understanding of PPG imaging. Multi-channel photoplethysmographic research systems are enablers of this promising research direction [4, 17]. In this regard, replacing commercial SpO_2 probes in our own setup by raw separate PPG probes for green and red light, synchronized with the camera-recordings, could have provided invaluable insights into the PPG signals under investigation.

We hope that future technical improvements in camera sensors, data acquisition practices and methods will lead to technical solutions for coping with the penetration depth gap(s) between probing wavelengths, motion interference, low SNR conditions and specular reflections. Short-term errors may be mitigated in future work by adapting motion-robust algorithms, e.g., [18]. These factors, combined with the careful standardization of measurement conditions will be key to improve $\text{SpO}_{2,cam}$ further.

Because of its relevance to the $\text{SpO}_{2,cam}$ methodology, we proceed this discussion with considerations on specular reflections and motion interference.

Specular reflections Polarizing filters are useful for optical skin measurements and are also seen in camera-based PPG works aimed at accurate amplitude measurements, e.g. [16]. When placed in front of the cameras and light sources in a cross-polarizing scheme, specular reflections are cancelled out, which ensures that the normalized PPG-amplitudes are not underestimated.

In study for which the video recordings were originally intended [20], polarizing filters were not essential because red and NIR light penetrate effectively into the skin, which makes the fraction of specularly reflected light small in comparison with the total scattered light.

However, the absence of polarizing filters in this particular study implies that the relatively lower performance of the $\text{SpO}_{2,cam}$ method by RoG (in comparison with its $RoIR$ -based counterpart) can be partly attributed to contamination by specular reflections. Indeed, the inclusion of polarizers

could have improved measurements in green wavelengths, in particular in subjects with oily skin and/or darker skin tones. Future work is valuable to quantify the performance gains enabled by polarizing filters in Vis-based calibrations.

Motion interference The importance of motion robustness is well emphasized in the camera-based PPG literature, e.g., [5, 16, 18]. Therefore, the cautious reader may wish reassurance that motion artifacts did not affect the calibration results provided in this paper.

Profiting from the IR(t) signal, which can be regarded as a third reference signal for algorithmic purposes, we have developed a means to estimate and cancel out a motion interference component in normalized camera-based PPG signals (see Appendix A for details). This enabled the calibratability analysis to be repeated based on “uncontaminated” signals. Interestingly, for SpO_{2,cam} by RoG, the calibration coefficients were almost unperturbed ($C_1 = -96.4, C_2 = 104.8$), with an A_{rms}^* of 3.0% and $R^2 = 0.57$ ($p < 0.001$). Similar observations held for SpO_{2,cam} by RoIR ($C_1 = -43.2, C_2 = 115.9; A_{rms}^* = 1.7%; R^2 = 0.86; p < 0.001$). Thus, as far as it pertains to coping with motion interference, the RRs-methodology can be used in SpO₂ estimations based on recordings with no more than sporadic motions. We hope that our study will be followed by algorithmic developments aimed at coping with higher extents of motion interference.

5 Conclusion

The calibratability of camera-based pulse oximetry using red and green wavelengths was demonstrated by a robust linear regression on data from 46 healthy individuals measured under normoxic and hypoxic conditions (SpO₂ range, 85% – 100%). This represents an extension of the earlier work in which the calibratability of SpO₂ measurements using NIR and red was demonstrated. The results obtained in this paper indicate an A_{rms}^* error for SpO₂ based RoG measurements of 3.0%, which is compatible with ISO 80601-2-61:2011 but larger than its counterpart measurements performed in the red–NIR window ($A_{rms}^*, 1.7%$).

Errors increased when we measured at lower temperatures or at a slightly different skin site (cheek instead of forehead). These sensitivities result from probing the vasculature at different depths for red and green. Our results support the statement that a single calibration curve for estimating SpO_{2,cam} using red and green is feasible, though estimation errors are higher than by using red and NIR.

Acknowledgements For planning and conducting the acquisition of the data that enabled this study, we acknowledge the contributions of Mukul Roque, Mohammed Meftah, Dr. Ihor Kirenko, Dr. Erik Bresch, and Marek Bartula. For helpful discussions, the authors thank Prof. Gerard de Haan, Dr. Mark van Gastel, and Michel Asselman, from

Philips Research, Eindhoven. For revising the manuscript, we thank Prof. Gerard de Haan and Benoit Balmaekers.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest.

Appendix A: Motion interference cancellation

A core task in cancelling a motion interference signal is its estimation from normalized camera-based PPG signals. To do this, the following principles were explored: 1. the periodicity of the camera-based PPG signal source, $s(t)$; 2. the randomness of motion signals; and, most importantly, 3. the assumption that camera-based PPG and motion artifacts are uncorrelated.

In particular, recognizing that $R_n(t)$ and $IR_n(t)$ are suitable candidates for the purpose of determining a motion estimate $m_e(t) \approx m(t)$ (see Eqs. 2, 3), we began by coarsely estimating RoIR from the available signals as $RR_0 = |R_n|/|IR_n|$.

The RR_0 estimate was refined by finding δ_{opt} such that $RR(\delta_{opt}) = RR_0 + \delta_{opt}$. To enable this, the principle (3) was translated into to the formulation given by Eq. 11.

$$\begin{aligned} & \underset{\delta}{\text{minimize}} && \sum_k |(G_n[k] - m_e[k])m_e[k]| \\ & \text{subject to} && m_e[k] = \frac{R_n[k] - IR_n[k]RR(\delta)}{1 - RR(\delta)}, \quad (11) \\ & && RR(\delta) = RR_0 + \delta; \delta \in [-0.15, 0.15]. \end{aligned}$$

with index k being the temporal index. Once δ_{opt} was available, we estimated the optimal m_e as $\frac{R_n - IR_n RR(\delta_{opt})}{1 - RR(\delta_{opt})}$. Minimizing the correlation between the motion-corrected camera-based PPG signal in green, $y[k] = G_n[k] - m_e[k]$, and $m_e[k]$, has resulted in unique solutions for all subjects in the dataset.

However, estimation errors were noticeable in subjects with low SNR. To ameliorate for the issue, signals were processed in consecutive windows of 1500 digital samples. Whenever artifact contamination in a particular window was practically irrelevant and/or indistinguishable from the sensor noise level, no correction was performed. Otherwise, wavelet signal denoising was applied to the final m_e using the Matlab function `wdenoise` (settings: wavelet type “sym4”, level 4).

Motion interference cancellation was applicable to windows in 19 recordings out of 46 (41% of the data from studies N and H). Figure 11a–c exemplifies $R_n(t)$ with its version after motion interference cancellation, $R_n(t) - m_e(t)$, in representative subjects at which motion cancellation

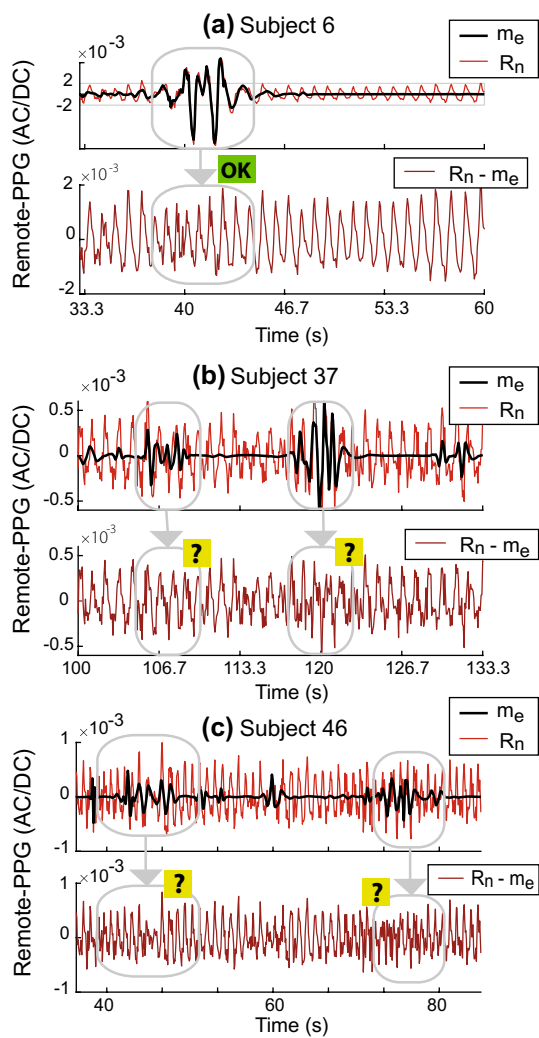


Fig. 11 Examination of outcomes of the motion interference cancellation in $R_n(t)$ signals. The label “OK” indicates intervals where the signal reconstruction was satisfactory. Conversely, “?” evidences doubtful need for motion interference cancellation or imperfect signal reconstruction

was performed. Overall, results suggest an advantage of cancelling strong motion artifacts but minor gains for the remainder.

References

- Allen J. Photoplethysmography and its application in clinical physiological measurements. *Physiol Meas*. 2007;28(3):1–40.
- Aoyagi T. Pulse oximetry: its invention, theory, and future. *J Anesth*. 2003;17(4):259–66.

- Bosschaart N, Edelman GJ, Aalders MCG, van Leeuwen TG, Faber DJ. A literature review and novel theoretical approach on the optical properties of whole blood. *Lasers Med Sci*. 2014;29(2):453–79.
- Budidha K, Rybynok V, Kyriacou PA. Design and development of a modular, multichannel photoplethysmography system. *IEEE Trans Instrum Meas*. 2018;67(8):1954–65.
- de Haan G, Jeanne V. Robust pulse-rate from chrominance-based rPPG. *IEEE Trans Biom Eng*. 2014;60(10):2878–86.
- Dumouchel W, O’Brien F. Integrating a robust option into a multiple regression computing environment. New York: Springer; 1991. p. 41–8.
- Guazzi AR, Villarroel M, Jorge J, Daly J, Frise MC, Robbins PA, Tarassenko L. Non-contact measurement of oxygen saturation with an rgb camera. *Biomed Opt Express*. 2015;6(9):3320–38.
- Hertzman AB. Photoelectric plethysmography of the fingers and toes in man. *Exp Biol Med*. 1937;37(3):529–34.
- Holland PW, Welsch RE. Robust regression using iteratively reweighted least-squares. *Commun Stat*. 1977;6(9):813–27.
- Hu S, Peris VA, Echiadis A, Zheng J, Shi P. Development of effective photoplethysmographic measurement techniques: from contact to non-contact and from point to imaging. *Proc IEEE-EMBS*. 2009;2009:6550–3.
- Julien C. The enigma of Mayer waves: facts and models. *Cardiovasc Res*. 2006;70(1):12–21.
- Kamshilin AA, Miridonov S, Teplov V, Saarenheimo R, Nippolainen E. Photoplethysmographic imaging of high spatial resolution. *Biomed Opt Express*. 2011;2(4):996–1006.
- Kim SJ, Port AD, Swan R, Campbell JP, Chan RVP, Chiang MF. Retinopathy of prematurity: a review of risk factors and their clinical significance. *Surv Ophthalmol*. 2018;63(5):618–37.
- Leonhardt S, Leicht L, Teichmann D. Unobtrusive vital sign monitoring in automotive environments—a review. *Sensors*. 2018;18(9):3080.
- Moço A, Stuijk S, de Haan G. New insights into the origin of remote PPG signals in visible light and infrared. *Sci Rep*. 2018;8(1):8501.
- Moço A, Stuijk S, de Haan G. Posture effects on the calibrability of remote pulse oximetry in visible light. *Physiol Meas*. 2018;40(3):035005.
- Rybynok V, May, JM, Budidha K, Kyriacou PA. Design and development of a novel multi-channel photoplethysmographic research system. In: *Proceedings of the IEEE-PHT*. 2013;267–270.
- van Gastel M, Stuijk S, de Haan G. New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring. *Sci Rep*. 2016;6(38):609.
- Verkruysse W, Svaasand LO, Nelson JS. Remote plethysmographic imaging using ambient light. *Opt Express*. 2008;16(26):434–45.
- Verkruysse W, Bartula M, Bresch E, Rocque M, Meftah M, Kirenko I. Calibration of contactless pulse oximetry. *Anesth Analg*. 2017;124(1):136–45.
- Wieringa FP, Mastik F, van der Steen AFW. Contactless multiple wavelength photoplethysmographic imaging: a first step toward SpO_2 camera technology. *Annals Biom Eng*. 2005;33(8):1034–41.
- Wukitsch MW, Petterson MT, Tobler DR, Pologe JA. Pulse oximetry: analysis of theory, technology, and practice. *J Clin Monit*. 1988;4(4):290–301.