



# Heart-rate tuned comb filters for processing photoplethysmogram (PPG) signals in pulse oximetry

Ludvik Alkhoury<sup>1</sup> · Ji-won Choi<sup>1</sup> · Chizhong Wang<sup>1</sup> · Arjun Rajasekar<sup>2</sup> · Sayandeep Acharya<sup>2</sup> · Sean Mahoney<sup>3</sup> · Barry S. Shender<sup>4</sup> · Leonid Hrebien<sup>2</sup> · Moshe Kam<sup>1</sup>

Received: 5 September 2019 / Accepted: 26 May 2020 / Published online: 17 June 2020  
© The Author(s) 2020

## Abstract

Calculation of peripheral capillary oxygen saturation ( $SpO_2$ ) levels in humans is often made with a pulse oximeter, using photoplethysmography (PPG) waveforms. However, measurements of PPG waveforms are susceptible to motion noise due to subject and sensor movements. In this study, we compare two  $SpO_2$ -level calculation techniques, and measure the effect of pre-filtering by a heart-rate tuned comb peak filter on their performance. These techniques are: (1) “Red over Infrared,” calculating the ratios of AC and DC components of the red and infrared PPG signals,  $\frac{(AC/DC)_{red}}{(AC/DC)_{infrared}}$ , followed by the use of a calibration curve to determine the  $SpO_2$  level Webster (in: Design of pulse oximeters, CRC Press, Boca Raton, 1997); and (2) a motion-resistant algorithm which uses the Discrete Saturation Transform (DST) (Goldman in J Clin Monit Comput 16:475–83, 2000). The DST algorithm isolates individual “saturation components” in the optical pathway, which allows separation of components corresponding to the  $SpO_2$  level from components corresponding to noise and interference, including motion artifacts. The comparison we provide here (employing the two techniques with and without pre-filtering) addresses two aspects: (1) accuracy of the  $SpO_2$  calculations; and (2) computational complexity. We used both synthetic data and experimental data collected from human subjects. The human subjects were tested at rest and while exercising; while exercising, their measurements were subject to the impacts of motion. Our main conclusion is that if an uninterrupted high-quality heart rate measurement is available, then the “Red over Infrared” approach preceded by a heart-rate tuned comb filter provides the preferred trade-off between  $SpO_2$ -level accuracy and computational complexity. A modest improvement in  $SpO_2$  estimate accuracy at very low SNR environments may be achieved by switching to the pre-filtered DST-based algorithm (up to 6% improvement in  $SpO_2$  level accuracy at  $-10$  dB over unfiltered DST algorithm and the filtered “Red over Infrared” approach). However, this improvement comes at a significant computational cost.

**Keywords** Pulse oximeter · Peripheral capillary oxygen saturation ( $SpO_2$ ) · Comb filter · Electrocardiography (ECG) · Motion artifact · Pulse oximetry · Photoplethysmography (PPG)

## 1 Introduction

Photoplethysmography (PPG) is a noninvasive [1, 2], electro-optic method for detecting the cardiovascular pulse wave generated by the elastic nature of the peripheral vascular arteries excited by the quasi-periodic contractions of the heart [3]. Vital signs such as heart rate, respiratory rate, and blood oxygen saturation are usually extracted from PPG waveforms. Fig. 1 illustrates a clean PPG signal obtained from experimental data in the time domain (Fig. 1a) and in the frequency domain (Fig. 1b). This clean PPG signal was

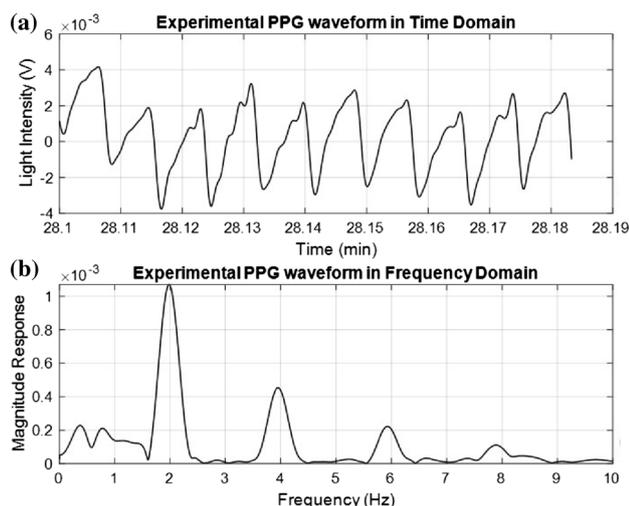
✉ Ludvik Alkhoury  
La256@njit.edu

<sup>1</sup> Department of Electrical and Computer Engineering, Newark College of Engineering, New Jersey Institute of Technology, New Jersey 07102, USA

<sup>2</sup> Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA 19104, USA

<sup>3</sup> Regulatory Affairs Department, Athena GTX, Johnston, IA 501131, USA

<sup>4</sup> Human Systems Department, Naval Air Warfare Center Aircraft Division, Patuxent River, MD 20670, USA



**Fig. 1** **a** Time domain representation of a PPG signal extracted from experimental data—**b** Frequency domain representation of a PPG signal extracted from experimental data

taken from a healthy male human in the course of intensive exercise regime.<sup>1</sup>

A pulse oximeter detects and calculates the absorption of light by functional hemoglobin (oxygenated and deoxygenated hemoglobin) to produce a measurement of the peripheral capillary oxygen saturation ( $\text{SpO}_2$ ).  $\text{SpO}_2$  is an estimate of arterial oxygen saturation ( $\text{SaO}_2$ ) [4]. The use of absorption of light for calculation of  $\text{SpO}_2$  levels makes use of the Lambert–Beer Law. The Law states that if a solute is dissolved in a clear solvent, its concentration can be determined if a light of known wavelength is transmitted through the solution. The incident and transmitted light intensity are logarithmically related to the absorbance of the solution (the absorbance and the concentration of the solution have a linear relationship) [5]. On average, a healthy human has a  $\text{SpO}_2$  value of 95–100% at sea level.  $\text{SpO}_2$  values below 90% are considered low, and are taken as a possible indication for the onset of hypoxia [6].

In conventional oximetry,  $\text{SpO}_2$  is calculated from the PPG waveforms using the “Red over Infrared” approach, calculating the optical density ratio (ratios of AC and DC components of the red and infrared PPG signals),  $r = \frac{(AC/DC)_{red}}{(AC/DC)_{infrared}}$ , followed by the use of a calibration curve (see [1], as well as a recent variation in [7]). In most systems, weighted moving average filters are commonly used to stabilize the readings [8–10]. However, when the measured subject experiences substantial motion, measurements

become noisier, and the “Red over Infrared” approach sometimes fails, providing false and inaccurate readings [10]. A Discrete Saturation Transform (DST) based algorithm [2] that uses an adaptive noise cancellation filter [11, 12] was proposed to suppress some motion artifact effects on  $\text{SpO}_2$  level calculations, thereby improving pulse oximetry. A 2002 study [13] reviewed the performance of twenty (20) commercial oximeters, and compared  $\text{SpO}_2$  readings from stationary “control hand” of each of the seventy (70) healthy human subjects to readings from the subject’s other hand, which was in motion. In this study, a Masimo SET (Signal Extraction Technology) pulse oximeter, which uses the DST algorithm, exhibited the best performance over all other tested oximeters. Other comparisons of oximeter performances were reported in [14] (from 2016) and [15] (from 2018). The study in [15] also included a DST-based oximeter (Masimo Radical-7). It concluded that in the face of motion artifacts, the DST-based oximeter performed at a similar level to other FDA-cleared pulse oximeters.<sup>2</sup>

In this study we are motivated by the observation that the spectral components of the PPG waveform appear at a fundamental frequency that corresponds to the subject’s heart rate and at its harmonics (e.g., Fig. 1b of this paper and Fig. 7b, d of [16]). A comb filter tuned to these (possibly time-varying) frequencies thus may have the potential to “clean up” the PPG waveform prior to applying the  $\text{SpO}_2$  calculation algorithm. The low computational complexity of a comb filter (when realized in software) may offer a viable alternative to the use of the more computationally complex realization of DST algorithm based systems.

The rest of the paper is organized as follows. Section 2 reviews two popular methods for  $\text{SpO}_2$  calculations from PPG waveforms, namely, the “Red over Infrared” approach and a DST-based algorithm. Section 3 reviews the synthetic and experimentally measured PPG waveforms used to assess  $\text{SpO}_2$  calculations in this study. Section 4 compares the performance (accuracy and computational complexity) of two  $\text{SpO}_2$  calculation methods, namely, the “Red over Infrared” approach and a DST-based algorithm, on synthetic data and on data collected from human subjects. In that section we also quantify the effect of pre-filtering of the PPG signals with a heart-rate tuned comb filter on algorithm performance. The main conclusion is that if an uninterrupted high-quality heart rate measurement is available, then the pre-filtered “Red over Infrared” approach using a heart-rate tuned comb filter provides the preferred trade-off between  $\text{SpO}_2$  level accuracy and computational complexity. While in a very low signal to noise (SNR) environment,

<sup>1</sup> Details about the exercise profile are provided in Sect. 3.2. This waveform was obtained from stage 8 (active recovery stage).

<sup>2</sup> The pulse oximeters reported upon in [15] were the following: Masimo Radical-7, Nihon Kohden OxyPal Neo, Nellcor N-600, and Philips Intellivue MP5.

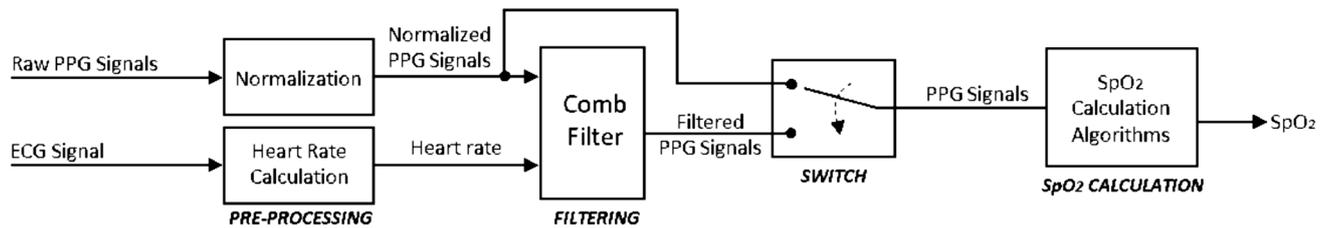


Fig. 2 SpO<sub>2</sub> Calculation procedure

the DST-based algorithm performed somewhat better (up to 6% improvement in accuracy at  $-10$  dB SNR over unfiltered DST algorithm and the filtered “Red over Infrared” approach), its computational complexity was much higher.

## 2 Methods

A block diagram of a processing module for PPG signals towards SpO<sub>2</sub> level calculation is shown in Fig. 2. The module is subdivided into three main stages: (1) pre-processing, (2) filtering, and (3) SpO<sub>2</sub> calculation. The inputs are raw PPG (red and infrared signals) and ECG waveforms, and the outputs are SpO<sub>2</sub> levels. The virtual switch enables comparison of the performance of the SpO<sub>2</sub> calculation module with and without the comb filter.

- (1) In the pre-processing stage, raw PPG signals are normalized (Sect. 2.1). Concurrently, the heart rate (HR) is calculated from an electrocardiography (ECG) waveform which is assumed to be available (Sect. 2.2).
- (2) In the filtering stage, the normalized PPG waveforms are processed with a heart-rate tuned peak comb filter (Fig. 3b) that uses the calculated HR as a reference signal (Sect. 2.3). The filter presents its lowest attenuation at the heart rate frequency and its principal harmonics, and higher attenuation otherwise.
- (3) The virtual switch (Fig. 2) allows us to compare the SpO<sub>2</sub> estimate that uses the normalized PPG signals to the estimate that uses these signals after comb filtering.
- (4) In the SpO<sub>2</sub> calculation stage, we use one of two different algorithms “Red over Infrared” [10] or a DST-based algorithm [2].

### 2.1 Normalization

Raw PPG (red and infrared) waveforms have two main components, namely: an AC component due to the light absorbed by pulsatile arterial blood, and a DC component due to the light absorbed by non-pulsatile components, such as tissues, venous, and capillary blood [17]. Since the DC component varies from one person to another (depending

on variables such as skin tone and tissue thickness), a normalization process is commonly used. The normalization is done by dividing the signal’s AC component by its constant DC component.

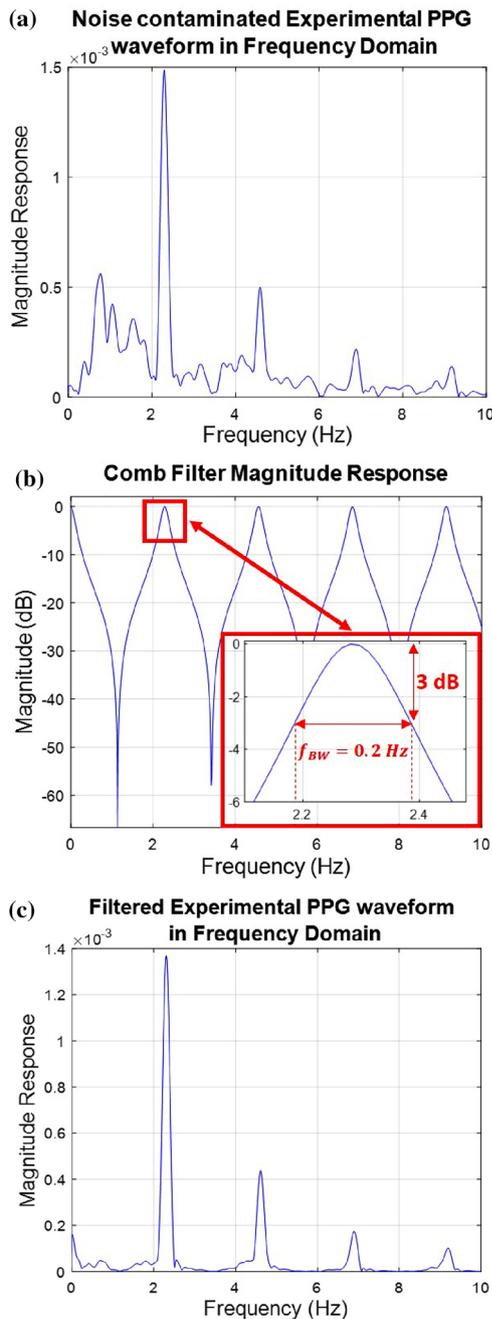
### 2.2 Heart rate calculation

We assume that we have access to the ECG waveform of the subject whose SpO<sub>2</sub> level we measure. The ECG waveform is known to be less susceptible to motion noise than the PPG waveform [18, 19]. The heart rate (beats per minute) was calculated in our study from an ECG signal through the Pan and Tompkins algorithm [20]. Since the fundamental frequency of the PPG signal is the heart rate, we use the heart rate to tune the comb filter. The comb filter discriminates against the portion of the PPG input signals which are not at the heart rate frequency or one of its principal harmonics.

### 2.3 Comb filter

The spectral components of the PPG waveform appear at a fundamental frequency (corresponding to the subject’s heart rate) and its harmonics (Fig. 1b of this paper and Fig. 7b, d of [16]). The use of a comb peak filter tuned to these frequencies may therefore serve to clean up the PPG waveform. The filter exhibits low attenuation at the fundamental frequency and its harmonics, and high attenuation in the intermediate regions between these frequencies (see Fig. 3b). In this manner, the filter reduces noise that resides in the intermediate regions. In order to reject as much noise as possible, we want the ‘peaks’ of the filter to be narrow. On the other hand, overly narrow peaks are likely to miss the PPG harmonics if the tuning is not exact (if the filter is not tuned exactly to the heart rate) (Fig. 4). Therefore, a compromise is needed between tuning accuracy and noise-rejection capability. We employed an IIR comb filter with the transfer function

$$H(z) = \beta \frac{1 + z^{-k}}{1 - \alpha z^{-k}}, \quad (1)$$



**Fig. 3** **a** Frequency domain representation of a noise contaminated experimental PPG waveform of fundamental frequency  $f_0 = 2.29$  Hz—**b** Magnitude response of a tuned comb filter—**c** Frequency domain representation of the comb-filtered PPG waveform

where  $\alpha$  and  $\beta$  are two positive scalars and  $K$  is the comb filter’s order. The design equations are shown in Table 1.

We have selected  $f_{BW} = 0.2$  Hz (capturing 97.5% of the total power of the signal of interest). For this selection and  $f_0 = 1$  Hz, the null-to-null bandwidth of the filter’s lobes is 1 Hz and the 10 dB bandwidth is 0.49 Hz. The sampling rate was  $f_s = 256$  Hz.

Figure 3 shows a frequency domain plot of a noise contaminated PPG signal measured on a healthy male subject during aerobic exercise (see Sect. 3.2 for more details). The signal is passed through a heart-rate tuned comb filter whose transfer function (magnitude response) is shown in Fig. 3b. The fundamental frequency (frequency of the heart rate) of the subject is  $f_0 = 2.29$  Hz. The parameters of the comb filter are  $K = 112$ ,  $\alpha = 0.7570$ , and  $\beta = 0.1215$ . Fig. 3c shows the clean PPG signal emerging from the comb filter.

### 2.4 SpO<sub>2</sub> level calculations

#### 2.4.1 The “red over infrared” approach

In the “Red over Infrared” approach, two light sources of different wavelengths,  $\lambda_r$  and  $\lambda_{ir}$  (red and infrared light, respectively), are used. The optical density ratio ‘r’ is defined as the ratio of the normalized red to the normalized infrared waveforms. In our project, we have used the Texas Instruments AFE4490 as the analog front-end for the pulse oximetry system—using diodes of wavelength of  $\lambda_r = 660$  nm for the red light source and  $\lambda_{ir} = 900$  nm for the infrared light source. In order to calculate SpO<sub>2</sub> levels, we employed first the calibration curve (2) which was provided by the manufacturer as the standard model [21].

$$SpO_2 = 110 - 25 \times r. \tag{2}$$

To study the sensitivity of our statistical results and main conclusions (Sect. 4.2 and Table 4) to the specification of the calibration curve, we have also employed two alternate calibrations curves in this study (viz., we calculated the statistics separately for each one of three different calibration curves, see Fig. 5).

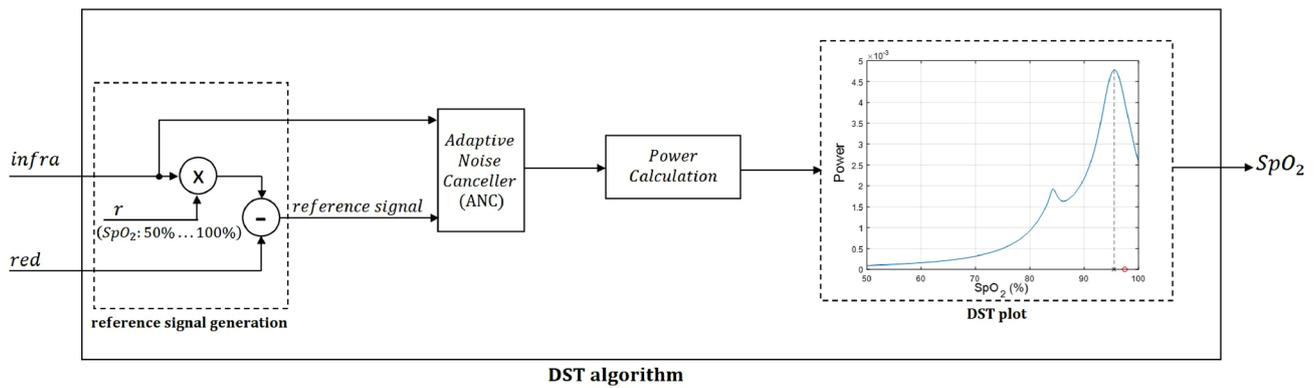
The first alternate curve is provided by the Lambert–Beer method [22], shown in Fig. 5 as a red trace. Notably, the standard model (blue trace) shows a relationship between SpO<sub>2</sub> and ‘r’ which is “to the right and above” Lambert–Beer curve (3). Hence, the standard model overestimates the SpO<sub>2</sub> level when compared to the Lambert–Beer estimate at the same value of ‘r’.

The second alternate curve is “to the left and below” the Lambert–Beer curve, and hence underestimates the SpO<sub>2</sub> level when compared to the Lambert–Beer estimates. We denote this curve (4) “underestimation calibration curve” (it has the same slope (− 25) as the standard model). The equations of the alternate calibration curves are as follows.

Lambert–Beer estimation calibration curve:

$$SpO_2 = \frac{100(r \times Ext(Hb, \lambda_{ir}) - Ext(Hb, \lambda_r))}{r(Ext(Hb, \lambda_{ir}) - Ext(HbO_2, \lambda_{ir})) + Ext(HbO_2, \lambda_r) - Ext(Hb, \lambda_r)}, \tag{3}$$

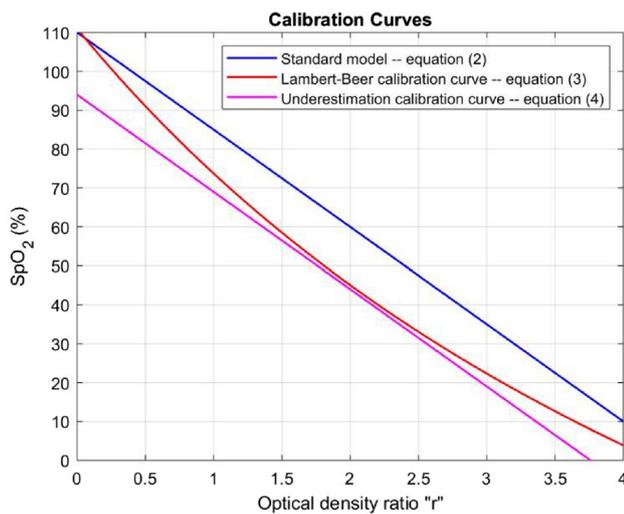
where  $Ext(Hb, \lambda_r)$  and  $Ext(Hb, \lambda_{ir})$  are the extinction coefficients of deoxygenated hemoglobin, and  $Ext(HbO_2, \lambda_r)$  and



**Fig. 4** DST algorithm block diagram and DST plot for a noise-contaminated synthetic PPG signal of SNR of 0 dB. The SpO<sub>2</sub> level calculated by the algorithm corresponds to the right-most peak in the DST plot (output power vs. SpO<sub>2</sub> level)

**Table 1** Comb filter design equations

| Comb filter design equations  |  |
|---|--|
| $K = \frac{f_s}{f_0}$ (dimensionless)   | $f_0$ is the fundamental frequency (heart rate) in Hz. $f_s$ is the sampling frequency |
| $\beta = \frac{1-\alpha}{2}$ (dimensionless)  | Gain at fundamental frequency and its harmonics set to 1                               |
| $f_{BW} = \cos^{-1}\left(\frac{2\alpha}{\alpha^2+1}\right) \times \frac{f_0}{180}$ (Hz) | 3 dB bandwidth set   |



**Fig. 5** Calibration curves used for sensitivity study

$Ext(\text{HbO}_2, \lambda_{ir})$  are the extinction coefficients of oxygenated hemoglobin, for wavelengths  $\lambda_r$  and  $\lambda_{ir}$ , respectively. In this study, the following extinction coefficients were taken from [23]:

$$\begin{aligned} Ext(\text{Hb}, \lambda_r) &= 3226.6 \text{ cm}^{-1}(\text{moles/l})^{-1}, \\ Ext(\text{Hb}, \lambda_{ir}) &= 761.84 \text{ cm}^{-1}(\text{moles/l})^{-1}, \\ Ext(\text{HbO}_2, \lambda_r) &= 319.6 \text{ cm}^{-1}(\text{moles/l})^{-1}, \\ Ext(\text{HbO}_2, \lambda_{ir}) &= 1198 \text{ cm}^{-1}(\text{moles/l})^{-1}. \end{aligned}$$

Underestimation calibration curve:

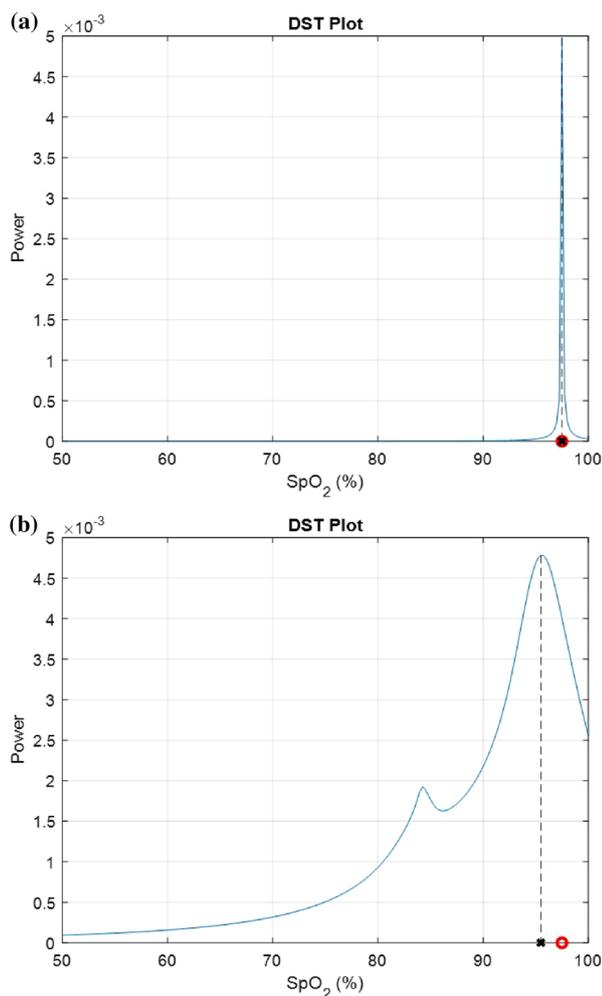
$$SpO_2 = 94 - 25 \times r. \tag{4}$$

### 2.4.2 The DST algorithm

The DST algorithm [2] was derived to measure SpO<sub>2</sub> levels in the face of motion noise. In developing the algorithm, it was assumed that the clean PPG signal of interest is contaminated by additive noise, uncorrelated with the signal. Figure 4 is the block diagram of the DST algorithm. The red and infrared PPG signals are the inputs and the SpO<sub>2</sub> level is the output. A family of reference signals is generated for each optical density ratio corresponding to SpO<sub>2</sub> values ranging from 50 to 100% at a resolution of 0.5%. The reference signal is defined as

$$reference\ signal(t) = infra(t) \times r - red(t). \tag{5}$$

Here, ‘r’ is an arbitrary optical density ratio value that corresponds to SpO<sub>2</sub> levels ranging from 50% to 100% (we use the calibration curve Eq. (5), which gives the corresponding values of ‘r’ of 0.4 to 2.4). “red(t)” and “infra(t)” are the time-dependent red and infrared PPG signals collected on a range of  $t \in [0, T]$  (T is typically 10 s); they serve as the two inputs of the DST algorithm.



**Fig. 6** **a** DST plot on a clean synthetic PPG signal—**b** DST plot on noise-contaminated synthetic PPG signals (SNR = 0 dB). The red circle is the SpO<sub>2</sub> ground truth and the black ‘x’ is the SpO<sub>2</sub> level that the DST algorithm calculates. The SpO<sub>2</sub> ground truth for both subplots was 97.5%

The DST algorithm employs Adaptive Noise Cancellation (ANC) filters [11, 12] to remove noise, and provide a “clean” SpO<sub>2</sub>. For each arbitrary ‘r’ (corresponding to an SpO<sub>2</sub> level between 50 and 100%), the reference signal and the infrared signal are fed into an ANC filter which identifies and removes frequency components which are in common between the two signals [2]. The power of the signal collected at the output of the ANC is calculated for each reference signal. A “DST plot” is generated, with the SpO<sub>2</sub> values used to generate the reference signals on the abscissa, and the power at the output of the ANC for each reference signal on the ordinate. Figure 6 shows the DST plots for clean (Fig. 6a) and noise-contaminated (Fig. 6b) synthetic PPG signals (for the way synthetic PPG signals were generated, see Sect. 3.1). In the case of a clean PPG signal, the DST plot shows only one peak. Its location corresponds to

the SpO<sub>2</sub> level estimate (Fig. 6a). The SpO<sub>2</sub> level calculated by the DST algorithm for this synthetic PPG signals (the black ‘x’ on Fig. 6a) matches the SpO<sub>2</sub> level for which it was created (red circle on Fig. 6a). For the noisy signal, two distinct peaks will typically appear, as shown in Fig. 6b. One peak corresponds to the true SpO<sub>2</sub> level and the other peak is generated by noise. The right-most peak is considered to correspond the true SpO<sub>2</sub> level. The SpO<sub>2</sub> level calculated by the DST algorithm in the example used for Fig. 6b (the black ‘x’) deviates slightly from the ground truth (red circle). This difference is attributed to noise. The signal to noise ratio (defined in Eq. (10)) in this example was 0 dB, which is quite low.

### 3 Generation of PPG and ECG signals

#### 3.1 Synthetic data generation

In order to study behavior, performance, and tradeoffs in the design of SpO<sub>2</sub> estimators, we developed a synthetic PPG signal generator. We modeled the PPG red (Eq. (6a)) and infrared (Eq. (6b)) waveforms as the sum of a constant DC component and an AC component. The AC component is the sum of four sinusoids of different amplitude<sup>3</sup> A<sub>j</sub> (j = 1, 2, 3, 4) (Eq. (7)). The first is at a frequency ranging from 0.5 to 3.5 Hz (corresponding to the subject’s heart rate and serving as the fundamental frequency). The three other sinusoids are its second, third, and fourth harmonics.

$$Red(t) = DC_{red} + AC_{red} \tag{6a}$$

$$Infrared(t) = DC_{infrared} + AC_{infrared} \tag{6b}$$

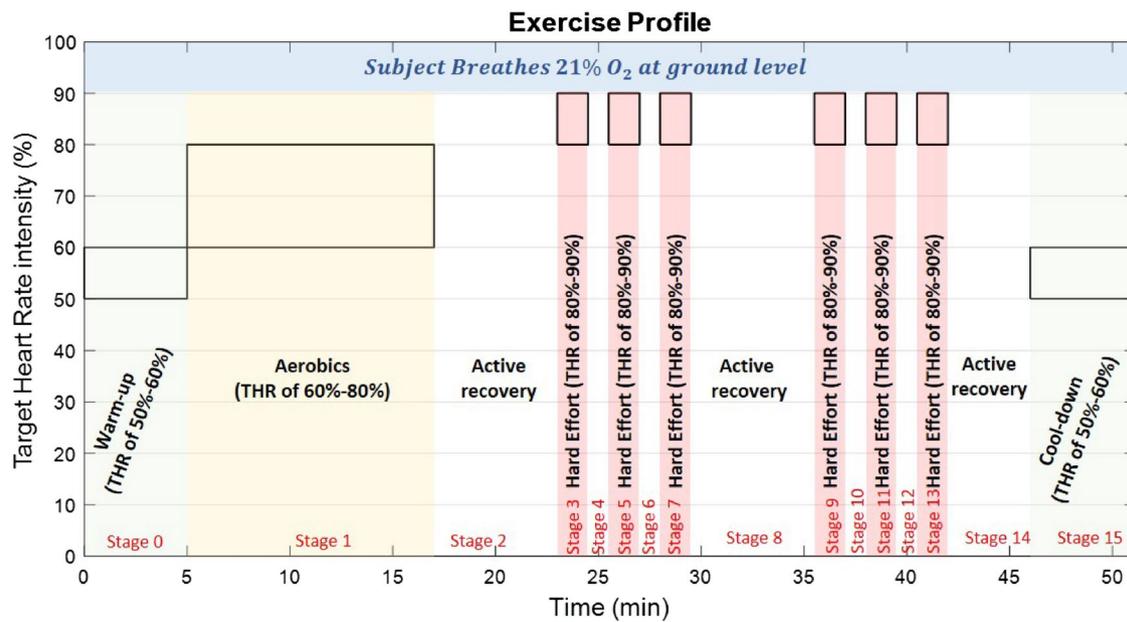
$$AC_{red}(t) = \sum_{j=1}^4 -A_j \sin(2\pi \times j \times f_0 \times t) \tag{7}$$

where A<sub>1</sub> = 1.242 × 10<sup>-3</sup>, (A<sub>2</sub> = 0.835 × 10<sup>-3</sup>, A<sub>3</sub> = 1.899 × 10<sup>-4</sup>, and A<sub>4</sub> = 0.786 × 10<sup>-4</sup>.

SpO<sub>2</sub> levels and the optical density ratio ‘r’ are related by the calibration curve (see Eq. (2)). In simulations on synthetic PPG signals, we varied the ratio of red-to-infrared signals in order to get the desired SpO<sub>2</sub> level. The optical density ratio ‘r’ is defined as

$$r = \frac{\left(\frac{AC}{DC}\right)_{red}}{\left(\frac{AC}{DC}\right)_{infrared}} = \frac{Normalized_{red}}{Normalized_{infrared}}, \tag{8}$$

<sup>3</sup> The values of A<sub>j</sub>s were empirically derived from clean PPG waveforms taken from human subjects at rest.



**Fig. 7** Illustration of the phases of the exercise experimental profile along with Target Heart Rate (THR) intensity. Black boxes indicate the range of THR in each exercise phase

Hence,

$$AC_{infrared} = \frac{DC_{infrared}}{DC_{red} \times r} \times AC_{red}. \quad (9)$$

The ECG signal is generally composed of P, QRS, T, and U waves [24]. Islam et al. [25] modeled a synthetic clean ECG waveforms by assuming that the QRS, Q, and S portions of the ECG signal can be represented by triangular waveforms, and the P, T, and U portions can be represented by the positive half period of a sinusoidal waveform. Their model was used in this study to generate the synthetic ECG waveforms [26].

### 3.2 Human subject data (Real PPG)

We have used human subject data from an experiment conducted in 2018–2019 by the United States Naval Air Warfare Center Aircraft Division (NAWCAD). Data were collected from fourteen (14) test subjects who were briefed and provided informed consent. All subjects were volunteers, military or civil-service personnel in the employ of the US federal government. All subjects were non-smokers and were reviewed by a medical monitor to ensure that they were physically fit to participate.<sup>4</sup>

Each subject was exposed to the following profile of activities (Fig. 7).

1. Warm-up—a warm-up and stretching session, designed to raise and maintain Target Heart Rate (THR)<sup>5</sup> intensity of 50–60% (approximately 5 min).
2. Aerobic—run on treadmill/bike session to raise and maintain THR of 60–80% (approximately 12 min), followed by active recovery (approximately 6 min), designed to decrease the heart rate by a reduced-intensity jog/walk or cycling activity.
3. Anaerobic—two sessions, each consisting of three 90-s periods of strenuous effort (treadmill/bike) designed to raise and maintain THR of 80–90%, separated from each other by one (1) minute of active recovery, designed to decrease heart rate by a reduced-intensity jog/walk or cycling; the two sessions were separated from each other by a period of active recovery (approximately 6 min). Next, the subject undertook a period of active recovery (around 4 min).
4. Cool-down—a cool-down and stretching session to maintain a THR of 50–60% (about 5 min).

Figure 7 shows all stages of this regime, marked stage 0 to stage 15. During the experimental runs, signals were recorded from a prototype dual pulse oximeter system with an accelerometer worn on the arm developed by Athena GTX (Holistic Modular Aircrew Physiologic Status

<sup>4</sup> Study approved by Naval Air Warfare Center Aircraft Division IRB, protocol FWR21070114H, original approval date: 12 June 2017. Air Force Research Lab (AFRL) IRB protocols comply with DoD Directive 3216.02, Title 25, CFR 46, and are in compliance with the Declaration of Helsinki Revision 6, 2008.

<sup>5</sup> The target Heart Rate (THR) is determined using the Karvonen formula [27]  $THR = ((HR_{max} - HR_{rest}) \times (\%intensity)) + HR_{rest}$ , where  $HR_{max} = 208 - 0.7 \times age$  [28].

(HMAPS) Monitoring System), and ECG electrodes affixed to the chest. Additionally, a Nonin 8000R reflectance pulse oximetry sensor was mounted on the subject's temple and connected by wires to the Nonin Wrist-Ox 3150 for data processing. The 8000R sensor was mounted on the subject's temple in a solid and stable manner that minimized sensor movements or dislocations on the temple surface even during strenuous exercise. SpO<sub>2</sub> levels calculated by the temple Nonin 8000R sensor were considered the "ground truth" due to the high accuracy of this sensor.<sup>6</sup> Performance of the arm-mounted oximeters was assessed with respect to the readings of the temple-mounted 8000R sensor (Sect. 4.2).

## 4 Impact of a heart-rate tuned comb filter on SpO<sub>2</sub> measurement performance

### 4.1 SpO<sub>2</sub> level calculations on synthetic data

We set out to assess the effect of the heart-rate tuned comb filter on accuracy of calculated SpO<sub>2</sub> levels.<sup>7</sup> To this end, we employed the "Red over Infrared" approach and the DST-based algorithm on a 10-s long synthetic PPG waveforms, with and without pre-filtering of the PPG signal by a comb filter. SpO<sub>2</sub> level and heart rate were set to 95% and 60 bpm, respectively. We modeled the motion artifact noise,  $N$ , as a bandpass (0.5–5 Hz) filtered AWGN (as was done in [28]), added to the clean PPG signal,  $S$ . The signal-to-noise ratio (SNR) is

$$SNR = \frac{Var(S)}{Var(N)}. \quad (10)$$

On Fig. 8 we show the Root Mean Square Error (RMSE) of SpO<sub>2</sub> levels calculated from synthetic PPG signals for an SNR ranging from –10 to +10 dB. We used the "Red over Infrared" approach and the DST-based algorithm, both before and after processing the synthetic PPG signals with the heart-rate tuned comb filter. The RMSE is defined as

<sup>6</sup> In general, reflectance pulse oximetry, such as the method used by Nonin 8000R is known to be much less vulnerable to artifacts (including motion artifacts). The manufacturer reports that SpO<sub>2</sub> accuracy of the Model 8000R sensor was determined through an induced hypoxia study on healthy subjects over the range of 70% to 100% [29]. The resulting SpO<sub>2</sub> accuracy was  $\pm 2A_{rms}$  in the range 80–100% and  $\pm 3A_{rms}$  in the range 70–80%.  $\pm 1A_{rms}$  encompasses 68% of the population at zero bias.

<sup>7</sup> In this section we have used calibration curve (2). We have repeated the calculation for calibration curves (3) and (4) and the trends and conclusion remain the same.

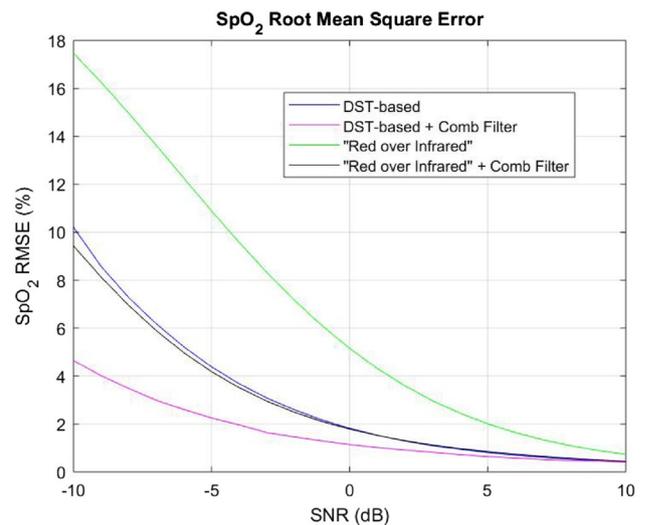


Fig. 8 SpO<sub>2</sub> Root Mean Square Error using "Red over Infrared" approach and the DST-based algorithm, with and without a comb filter

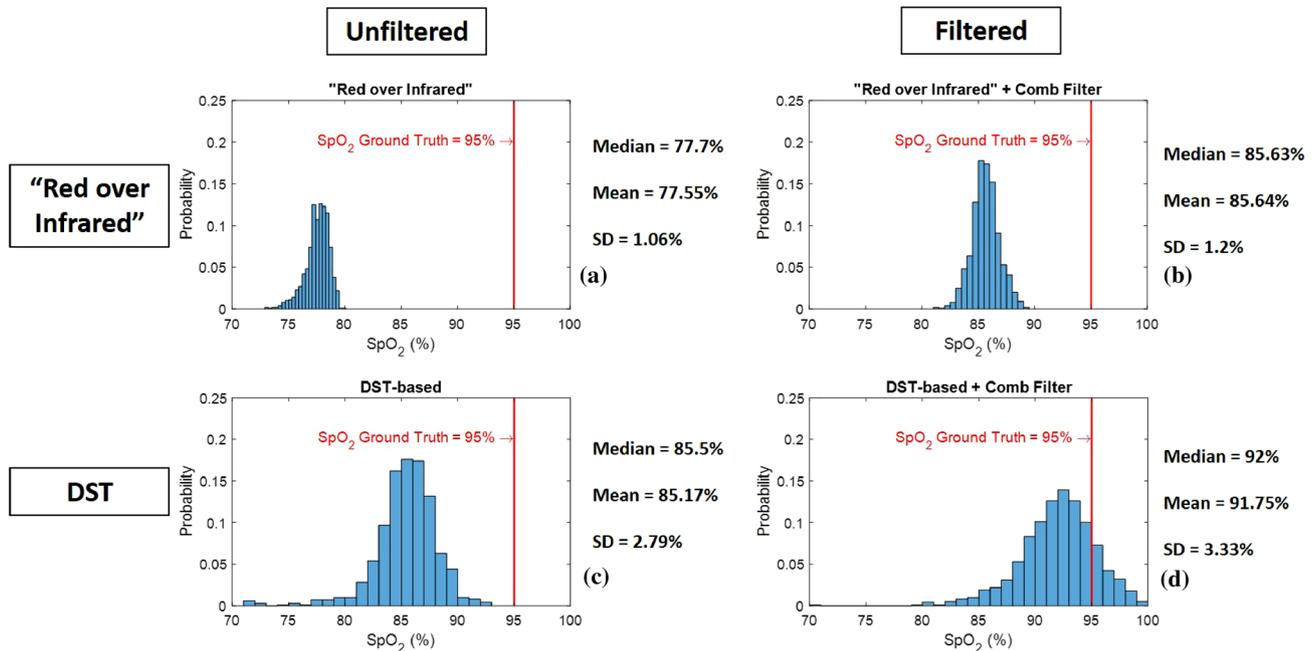
$$RMSE(i) = \sqrt{\frac{1}{P} \sum_{p=1}^P (\text{SpO}_{2,cal.}(p) - \text{SpO}_{2,g.t.})^2}, \quad (11)$$

where 'SpO<sub>2,cal.</sub>( $p$ )' is one value (out of the 1000 realizations) of the SpO<sub>2</sub> level calculated using one of the tested methods ("Red over Infrared" or DST-based algorithm, with and without comb filtering); 'SpO<sub>2,g.t.</sub>' is the ground truth SpO<sub>2</sub> value; ' $P$ ' is the number of SpO<sub>2</sub> levels calculated for each SNR, and ' $i$ ' the SNR value at which the RMSE is calculated. In our study, SpO<sub>2,g.t.</sub> = 95%,  $P = 1000$  realizations, and  $i = -10, -9, \dots, -1, 0, 1, \dots, 9, 10$  (dB).

The main conclusion from Fig. 8 is that the unfiltered "Red over Infrared" approach is inferior to the other approaches (namely filtered "Red over Infrared" and DST-based algorithm (filtered or unfiltered)). The filtered "Red over Infrared" approach and the unfiltered DST-based algorithm are comparable. The best performance was obtained by the filtered DST-based algorithm.

Figures 9, 10, and 11 are the histograms of SpO<sub>2</sub> levels (1000 SpO<sub>2</sub> calculations each) calculated using "Red over Infrared" approach and the DST-based algorithm, both with and without processing the synthetic PPG signals with the heart-rate tuned comb filter. We show results for a SNR of –10 dB (Fig. 9), 0 dB (Fig. 10), and 10 dB (Fig. 11). The histograms get closer to the SpO<sub>2</sub> ground truth (SpO<sub>2</sub> of 95%) when the synthetic PPG signals are preprocessed by the heart-rate tuned comb filter for both SpO<sub>2</sub> calculation methods ("Red over Infrared" and the DST). We conclude that (1) the use of a heart-rate tuned comb filter has improved the performance of both methods ("Red over Infrared" and the DST), and (2) the best performance is obtained with the

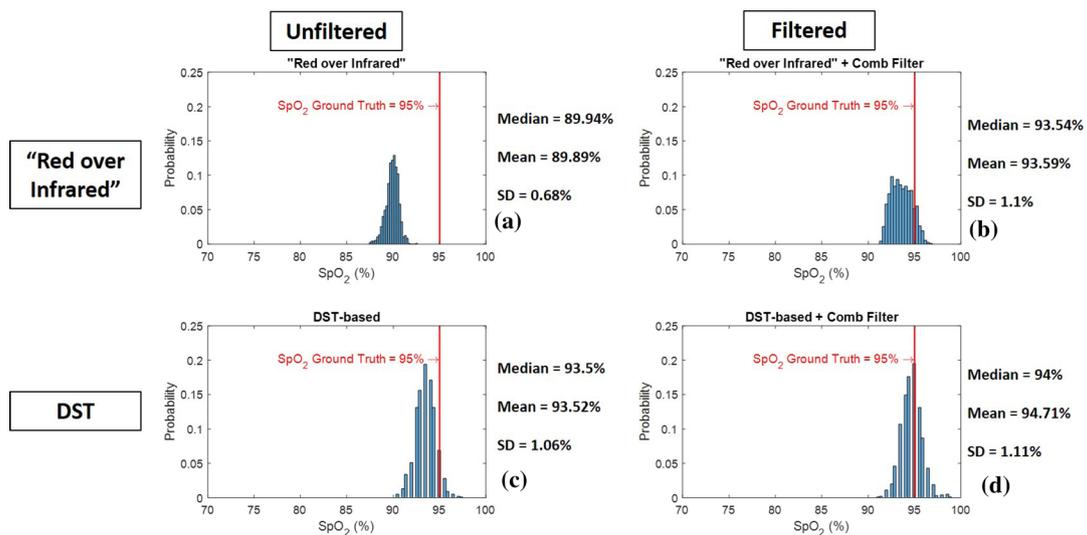
SNR = -10 dB



**Fig. 9** Histogram of 1000 SpO<sub>2</sub> levels calculated from red and infrared PPG signal with SNR = -10 dB using **a** "Red over Infrared" approach—**b** "Red over Infrared" approach preceded by a heart-rate

tuned comb filter—**c** DST-based algorithm—**d** DST-based algorithm preceded by a heart-rate tuned comb filter

SNR = 0 dB



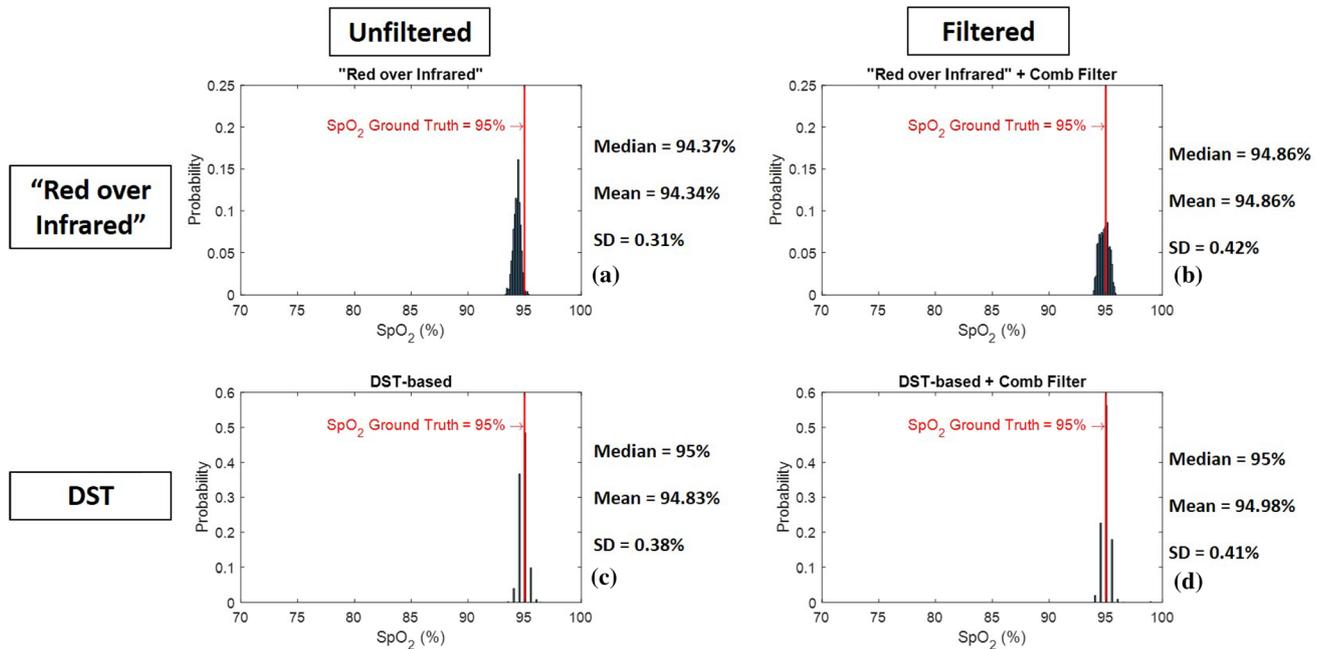
**Fig. 10** Histogram of 1000 SpO<sub>2</sub> levels calculated from red and infrared PPG signal with SNR = 0 dB using **a** "Red over Infrared" approach—**b** "Red over Infrared" approach preceded by a heart-rate

tuned comb filter—**c** DST-based algorithm—**d** DST-based algorithm preceded by a heart-rate tuned comb filter

DST-based algorithm preceded by a heart-rate tuned comb filter. Next best is the "Red over Infrared" approach with comb filtering and unfiltered DST-based algorithm (SpO<sub>2</sub>

levels estimated with these two techniques are comparable); the worst performance was the unfiltered "Red over Infrared" approach.

SNR = 10 dB



**Fig. 11** Histogram of 1000 SpO<sub>2</sub> levels calculated from red and infrared PPG signal with SNR = 10 dB using **a** “Red over Infrared” approach—**b** “Red over Infrared” approach preceded by a heart-rate

tuned comb filter—**c** DST-based algorithm—**d** DST-based algorithm preceded by a heart-rate tuned comb filter

## 4.2 SpO<sub>2</sub> level calculations on experimental data

### 4.2.1 Data analysis using the standard model calibration curve equation (2)

We tested the “Red over Infrared” approach and the DST-based algorithm on data collected from human subjects, with and without comb filtering. Data were collected from fourteen (14) human subjects for the regime described in Sect. 3.2 and Fig. 7. Every 2 s, the preceding 10-s long data segment was processed. We show the results of one of the fourteen (1 of 14) subjects on Fig. 12 (we covered the time period from 5 to 35 min of the exercise profile, corresponding to stages 1 to 8 in Fig. 7). A detailed description of the traces in Fig. 12 is provided in Table 2.<sup>8</sup> In all cases, the standard against which the various methods were assessed was the readings of the Nonin 8000R sensor (light blue trace in Fig. 12b, c, d).

The gaps in SpO<sub>2</sub> curves on Fig. 12 (for example: In subplots b, c, and d from 19.5 to 20 min for the “Red over Infrared” approach and the DST-based algorithm, and from around time 20.5 min for Nonin 8000R sensor) indicate a

failure in calculating the SpO<sub>2</sub> level caused by physical loss of the PPG signal. We skipped these gaps in our analysis.

Examination of Fig. 12 suggests that the addition of a comb filter improved performance for both oximeters based on R/IR and oximeters based on DST algorithm (Fig. 12 second and third trace). Also, performance of oximeters employing R/IR + comb filter was very close to performance of oximeters employing DST + comb filter (Fig. 12 fourth trace).

Table 3 provides the overall mean and standard deviation of the SpO<sub>2</sub> error for all fourteen (14) subjects we studied in all eight (8) exercise profile stages (stages 1 to 8 in Fig. 7). The SpO<sub>2</sub> error is defined as absolute difference between SpO<sub>2</sub> levels measured by Nonin 8000R sensor sensor mounted on the subject’s temple (the standard) and SpO<sub>2</sub> levels calculated by the two algorithms. We performed six (6) two-samples t significance tests with level of significance  $\alpha=0.01$ . The tested hypotheses and their corresponding p-values are shown in Tables 4 and 5, respectively.

We concluded from Tables 4 and 5 the following:

- The mean absolute error using R/IR was larger than the mean absolute error using R/IR + Comb in all eight (8) stages;
- The mean absolute error using R/IR was larger than the mean absolute error using DST in all eight (8) stages;

<sup>8</sup> Data for all 14 subjects is available at [<https://github.com/moshekam/PPG-Exercise-Experimental-Data.git>].

**Table 2** Detailed description on all traces shown in Fig. 8

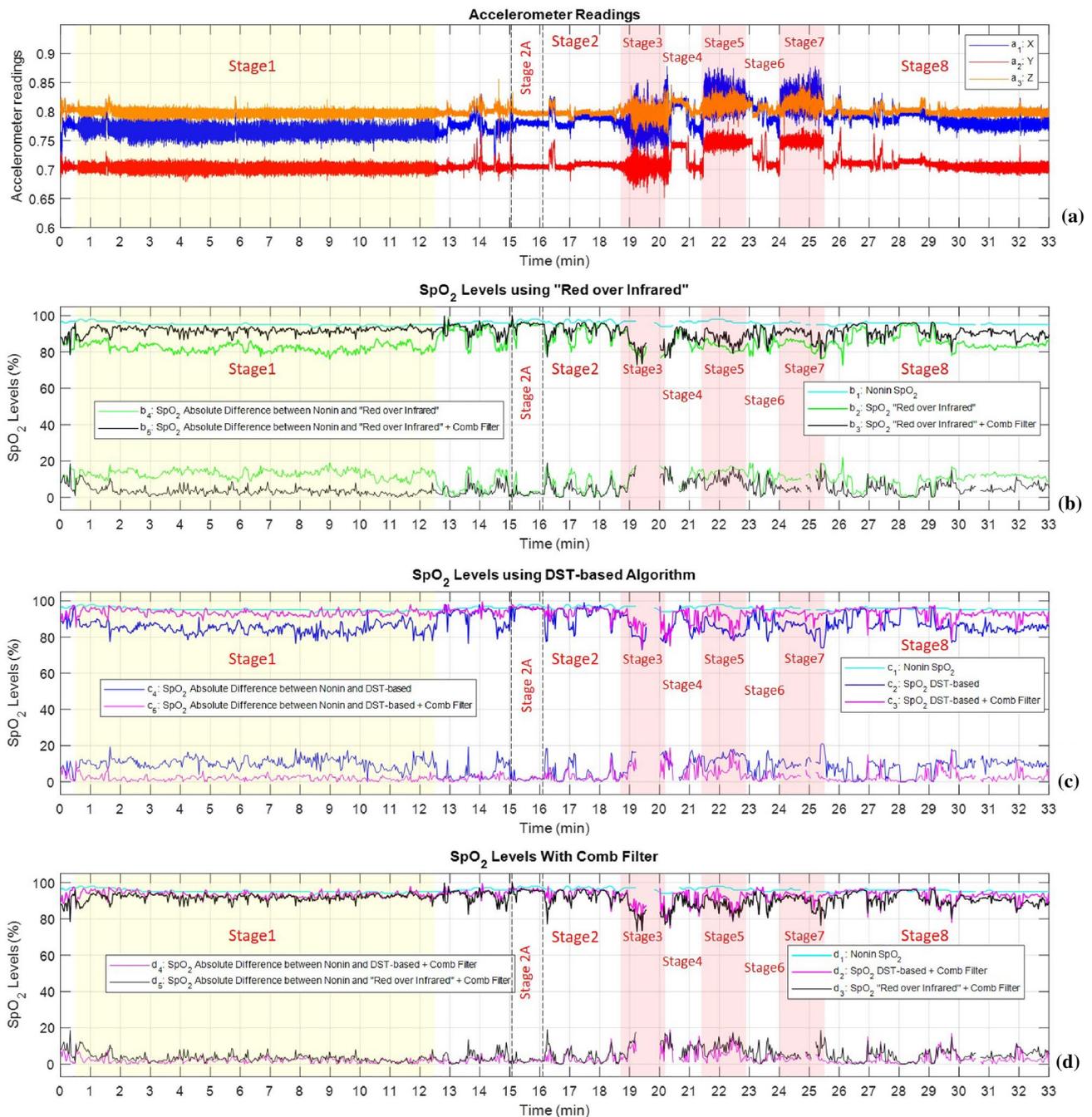
|            | Trace label | Trace color              | Trace name  | Description   |
|------------|-------------|--------------------------|---|---|
| Figure 12a | $a_1$       | Blue                     | X   | x-axis accelerometer  |
|            | $a_2$       | Red                      | Y   | y-axis accelerometer  |
|            | $a_3$       | Orange                   | Z   | z-axis accelerometer  |
| Figure 12b | $b_1$       | Light blue               | Nonin SpO <sub>2</sub>  | SpO <sub>2</sub> levels calculated by Nonin 8000R sensor  |
|            | $b_2$       | Green (top curve)        | SpO <sub>2</sub> “Red over Infrared”  | SpO <sub>2</sub> levels calculated using “Red over Infrared” approach before comb filter  |
|            | $b_3$       | Black (top curve)        | SpO <sub>2</sub> “Red over Infrared” + Comb Filter                              | SpO <sub>2</sub> levels calculated using “Red over Infrared” approach after comb filter   |
|            | $b_4$       | Green (bottom curve)     | SpO <sub>2</sub> Difference between Nonin and “Red over Infrared”               | SpO <sub>2</sub> Difference between levels calculated by Nonin 8000R sensor and “Red over Infrared” approach before comb filter |
|            | $b_5$       | Black (bottom curve)     | SpO <sub>2</sub> Difference between Nonin and “Red over Infrared” + Comb Filter | SpO <sub>2</sub> Difference between levels calculated by Nonin 8000R sensor and “Red over Infrared” approach after comb filter  |
| Figure 12c | $c_1$       | Light blue               | Nonin SpO <sub>2</sub>  | SpO <sub>2</sub> levels calculated by Nonin 8000R sensor  |
|            | $c_2$       | Dark blue (top curve)    | SpO <sub>2</sub> DST-based  | SpO <sub>2</sub> levels calculated using DST-based algorithm before comb filter   |
|            | $c_3$       | Magenta (top curve)      | SpO <sub>2</sub> DST-based + Comb Filter  | SpO <sub>2</sub> levels calculated using DST-based algorithm after comb filter  |
|            | $c_4$       | Dark blue (bottom curve) | SpO <sub>2</sub> Difference between Nonin and DST-based                         | SpO <sub>2</sub> Difference between levels calculated by Nonin 8000R sensor and DST-based algorithm before comb filter          |
|            | $c_5$       | Magenta (bottom curve)   | SpO <sub>2</sub> Difference between Nonin and DST-based + Comb Filter           | SpO <sub>2</sub> Difference between levels calculated by Nonin 8000R sensor and DST-based algorithm after comb filter           |
| Figure 12d | $d_1$       | Light blue               | Nonin SpO <sub>2</sub>  | SpO <sub>2</sub> levels calculated by Nonin 8000R sensor  |
|            | $d_2$       | Magenta (top curve)      | SpO <sub>2</sub> DST-based + Comb Filter  | SpO <sub>2</sub> levels calculated using DST-based algorithm after comb filter.   |
|            | $d_3$       | Black (top curve)        | SpO <sub>2</sub> “Red over Infrared” + Comb Filter                              | SpO <sub>2</sub> levels calculated using “Red over Infrared” approach after comb filter   |
|            | $d_4$       | Magenta (bottom curve)   | SpO <sub>2</sub> Difference between Nonin and DST-based + Comb Filter           | SpO <sub>2</sub> Difference between levels calculated by Nonin 8000R sensor and DST-based algorithm after comb filter           |
|            | $d_5$       | Black (bottom curve)     | SpO <sub>2</sub> Difference between Nonin and “Red over Infrared” + Comb Filter | SpO <sub>2</sub> Difference between levels calculated by Nonin 8000R sensor and “Red over Infrared” approach after comb filter  |

- The mean absolute error using R/IR + Comb was larger than the mean absolute error using DST + Comb in all eight (8) stages;
- The mean absolute error using DST was larger than the mean absolute error using DST + Comb in all eight (8) stages.

Additionally, we concluded from Tables 4 and 5 (tests III-a and III-b) that the mean absolute errors using R/IR + Comb and using DST are comparable during most stages.

#### 4.2.2 Data analysis using alternate calibration curves eqs. (3) and (4)

Appendix 1 replicates Table 5 (p-values) for the alternate calibration curves (3) [Lambert-Beer calibration curve] Table 7 in Appendix and (4) [underestimation calibration curve] Table 8 in Appendix. These tables support the conclusion in Table 4 for the alternate calibration curves as well (see Table 4 footnotes a and b). The relative performance of the SpO<sub>2</sub> calculation techniques and impact of the comb filter are qualitatively the same, regardless of the specific calibration curve (be it (2), (3), or (4))



**Fig. 12** SpO<sub>2</sub> level calculations for an exercising subject on stages 1 to 8 (see Sect. 3.2 and Fig. 7). We used the “Red over Infrared” approach and DST-based algorithm with and without preprocessing of the PPG signals with the heart-rate tuned comb filter. SpO<sub>2</sub> levels

calculated using the abovementioned algorithms is compared to the SpO<sub>2</sub> levels calculated by Nonin 8000R sensor. The legends are fully explained in Table 2

### 4.3 Computational complexity

#### 4.3.1 Red over infrared approach

For red and infrared signals of length  $L$  ( $L$  sampled values) the number of operations required for a typical “Red over

Infrared” calculation is  $2L$  multiplications,  $2(L-1)$  additions, three divisions and two square root operations. For each  $L$  (in our case in the thousands), the required computation time is proportional to  $L$ .

**Table 3** Overall Mean and standard deviation (SD) of the SpO<sub>2</sub> error calculated for “Red over Infrared” approach and DST-based algorithm with and without comb filtering for all 14 exercise subjects

|         | R/IR  |      | R/IR + comb |      | DST   |      | DST + comb |      |
|---------|-------|------|-------------|------|-------|------|------------|------|
|         | Mean  | SD   | Mean        | SD   | Mean  | SD   | Mean       | SD   |
| Stage 1 | 14.03 | 1.54 | 9.71        | 1.72 | 11.34 | 1.52 | 7.55       | 1.64 |
| Stage 2 | 7.05  | 1.37 | 4.74        | 1.15 | 4.76  | 1.1  | 3.34       | 0.93 |
| Stage 3 | 10.49 | 0.87 | 7.59        | 1.04 | 8.82  | 1.02 | 5.64       | 1.04 |
| Stage 4 | 9.82  | 1.56 | 6.18        | 1.29 | 5.85  | 1.23 | 4.07       | 1.13 |
| Stage 5 | 11.6  | 1.11 | 7.49        | 1.09 | 8.32  | 1.03 | 5.01       | 0.99 |
| Stage 6 | 9.01  | 1.42 | 4.86        | 1.11 | 5.52  | 1.19 | 3.09       | 0.9  |
| Stage 7 | 11.2  | 0.96 | 7.72        | 1.49 | 7.91  | 1.02 | 4.46       | 1.03 |
| Stage 8 | 8.14  | 1.49 | 5.03        | 1.02 | 6.1   | 1.58 | 3.45       | 0.99 |

**Table 4** Tested hypotheses and results

|             | Is the SpO <sub>2</sub> mean absolute error of ...                         |                | SpO <sub>2</sub> mean absolute error of ...?                            |   |  |
|-------------|--|----------------|---|---|--|
| Test I:     | “Red over Infrared” approach (R/IR) <i>without</i> comb filter ( $\mu_1$ ) | Greater than   | “Red over Infrared” approach (R/IR) <i>with</i> comb filter ( $\mu_2$ ) | $H_0: \mu_1 - \mu_2 = 0$<br>$H_a: \mu_1 - \mu_2 > 0$    | Yes. (test I, Tables 5, 7, and 8 in Appendix)                  |
| Test II:    | “Red over Infrared” approach (R/IR) <i>without</i> comb filter ( $\mu_1$ ) | Greater than   | “DST-based” algorithm (DST) <i>without</i> comb filter ( $\mu_3$ )      | $H_0: \mu_1 - \mu_3 = 0$<br>$H_a: \mu_1 - \mu_3 > 0$    | Yes. (test II, Tables 5, 7, and 8 in Appendix)                 |
| Test III-a: | “Red over Infrared” approach (R/IR) <i>with</i> comb filter ( $\mu_2$ )    | Different than | “DST-based” algorithm (DST) <i>without</i> comb filter ( $\mu_3$ )      | $H_0: \mu_2 - \mu_3 = 0$<br>$H_a: \mu_2 - \mu_3 \neq 0$ | No <sup>a</sup> . (test III-a, Tables 5, 7, and 8 in Appendix) |
| Test III-b: | “DST-based” algorithm (DST) <i>without</i> comb filter ( $\mu_3$ )         | Greater than   | “Red over Infrared” approach (R/IR) <i>with</i> comb filter ( $\mu_2$ ) | $H_0: \mu_3 - \mu_2 = 0$<br>$H_a: \mu_3 - \mu_2 > 0$    | No <sup>b</sup> . (test III-b, Tables 5, 7, and 8 in Appendix) |
| Test IV:    | “Red over Infrared” approach (R/IR) <i>with</i> comb filter ( $\mu_2$ )    | Greater than   | “DST-based” algorithm (DST) <i>with</i> comb filter ( $\mu_4$ )         | $H_0: \mu_2 - \mu_4 = 0$<br>$H_a: \mu_2 - \mu_4 > 0$    | Yes. (test IV, Tables 5, 7, and 8 in Appendix)                 |
| Test V:     | “DST-based” algorithm (DST) <i>without</i> comb filter ( $\mu_3$ )         | Greater than   | “DST-based” algorithm (DST) <i>with</i> comb filter ( $\mu_4$ )         | $H_0: \mu_3 - \mu_4 = 0$<br>$H_a: \mu_3 - \mu_4 > 0$    | Yes. (test V, Tables 5, 7, and 8 in Appendix)                  |

<sup>a</sup>For calibration curve (2) the answers are:

For stage 1—No, stage 2—No, stage 3—Yes, stage 4—No, stage 5—No, stage 6—No, stage 7—No, stage 8—No  
Overall answer: No

For calibration curve (3) the answer is No for all stages

For calibration curve (4) the answer is No for all stages

<sup>b</sup>For calibration curve (2) the answers are:

For stage 1—Yes, stage 2—No, stage 3—Yes, stage 4—No, stage 5—No, stage 6—No, stage 7—No, stage 8—No  
Overall answer: No

For calibration curve (3) the answers are:

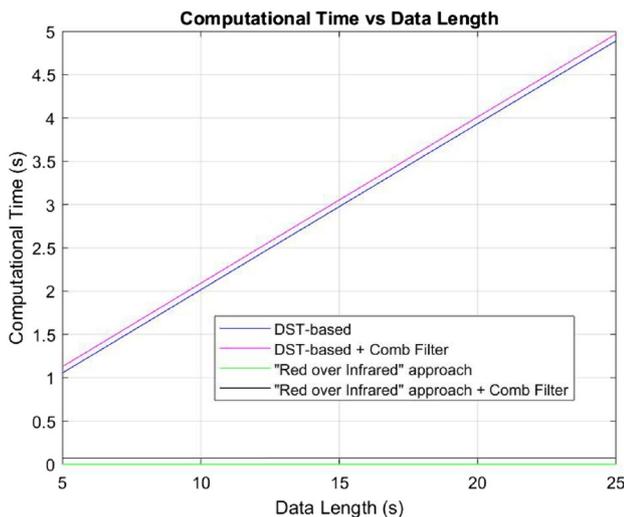
For stage 1—Yes, stage 2—No, stage 3—No, stage 4—No, stage 5—No, stage 6—No, stage 7—No, stage 8—No  
Overall answer: No

For calibration curve (4) the answers are:

For stage 1—Yes, stage 2—No, stage 3—No, stage 4—No, stage 5—No, stage 6—No, stage 7—No, stage 8—No  
Overall answer: No

**Table 5** p values of all six tests

|         | p value (significance level was $\alpha = 0.01$ ) |           |            |            |           |           |
|---------|---|-----------|------------|------------|-----------|-----------|
|         | Test I  | Test II   | Test III-a | Test III-b | Test IV   | Test V    |
| Stage 1 | < 0.00001   | 0.00023   | 0.01975    | 0.00987    | 0.00237   | < 0.0001  |
| Stage 2 | 0.00016   | 0.00015   | 0.01004    | 0.48161    | 0.00181   | 0.00136   |
| Stage 3 | < 0.00001   | 0.00022   | 0.00753    | 0.00377    | 0.00013   | < 0.00001 |
| Stage 4 | < 0.00001   | < 0.00001 | 0.25116    | 0.25025    | 0.00025   | 0.00077   |
| Stage 5 | < 0.00001   | < 0.00001 | 0.05893    | 0.02941    | < 0.0001  | < 0.00001 |
| Stage 6 | < 0.00001   | < 0.00001 | 0.15245    | 0.07654    | 0.00024   | < 0.0001  |
| Stage 7 | < 0.00001   | < 0.00001 | 0.35143    | 0.35009    | < 0.00001 | < 0.00001 |
| Stage 8 | 0.000011  | 0.00192   | 0.02643    | 0.26483    | 0.00056   | < 0.0001  |

**Fig. 13** Comparison of computational time of the “Red over Infrared” approach and the DST-based Algorithm with and without comb filtering

#### 4.3.2 DST algorithm

At every time instance, the DST algorithm generates a family of  $M$  reference signals (in our study we used  $M=101$  reference signal for each optical density ratio of  $\text{SpO}_2$  level, ranging from 50% to 100% with an increment of 0.5%). For each one of the  $M$   $\text{SpO}_2$  levels, the DST algorithm uses an ANC filter and additional computations to create one point on the DST graph of power vs.  $\text{SpO}_2$  level (e.g. Fig. 6b in this paper and Fig. 3 in [2]). For the ANC we have used the QR-decomposition-based least-squares lattice (QRD-LSL) adaptive filter algorithm, on account of its relative computational efficiency. The QRD-LSL requires  $O(R)$  operations per time instance (we use  $L$  instances for block of data), where  $R$  is the number of taps in the adaptive filter [30]. The computational times (in seconds) of “Red over Infrared” approach and the DST-based algorithm (with and without comb filtering) are shown on the ordinate of Fig. 13. The abscissa of Fig. 13 is the red and infrared signals’ length in

second. On Table 6 we show the computational time and the  $\text{SpO}_2$  RMSE (in %) calculated on a 10-s long synthetic PPG signals. We use the “Red over Infrared” approach and the DST-based algorithm, both before and after comb filtering in each case, for different SNR values (–10 dB, 0 dB, and 10 dB).<sup>9</sup> Results presented in Fig. 13 and Table 6 indicate that the computational costs of the DST-based algorithm are considerably larger than the “Red over Infrared” approach. The effects of adding the comb filter are much smaller.

#### 4.4 Implementation

The comb filter could be integrated into commercial systems due to the light hardware and computational requirements. We demonstrated this point by implementing the comb filter on an ARM Cortex-M4 based processor (model MK70FN1M0VMJ12).

### 5 Discussion and conclusion

Preprocessing PPG signals with a heart-rate tuned comb filter improved the performance of the two tested  $\text{SpO}_2$  calculation algorithms (namely, “Red over Infrared” approach and DST-based algorithm). We tested both algorithms on synthetic and experimental data.

- The most accurate technique was the filtered DST-based algorithm. At very low signal to noise (SNR) environments the filtered DST-based algorithm performed somewhat better on synthetic data compared to the other methods (up to 6% improvement in accuracy at minus 10 dB SNR over the unfiltered DST algorithm and the filtered “Red over Infrared” approach). However, this technique was costly in computations.

<sup>9</sup> The results were generated by MatLab R2018a on a personal computer, with an Intel Core™ i5-8500 CPU running at 3.00 GHz, 8GB RAM and Windows 10 operating system.

**Table 6** Computational time and SpO<sub>2</sub> RMSE calculated on a 10-s long PPG signals using “Red over Infrared” approach and DST-based algorithm before and after comb filtering for a SNR of –10dB, 0dB, and 10dB

|                            | SpO <sub>2</sub> RMSE (%) |            |             | Computational time (second) |
|----------------------------|---------------------------|------------|-------------|-----------------------------|
|                            | SNR = –10 dB              | SNR = 0 dB | SNR = 10 dB |                             |
| “Red over infrared”        | 17.4811                   | 5.1564     | 0.7301      | $2.463 \times 10^{-5}$      |
| “Red over infrared” + comb | 9.4388                    | 1.7844     | 0.4425      | $7.248 \times 10^{-2}$      |
| DST-based                  | 10.2173                   | 1.8191     | 0.4135      | 2.017                       |
| DST-based + comb           | 4.6482                    | 1.1431     | 0.4056      | 2.092                       |

- Next best were the filtered “Red over Infrared” approach and the unfiltered DST-based algorithm, which provided similar accuracies. However, the DST-based algorithm was much costlier in computations compared to the filtered “Red over Infrared” approach.
- The least accurate performance was of the unfiltered “Red over Infrared” approach.

The overall conclusion is that if an uninterrupted high-quality heart rate measurement is available, then the pre-filtered “Red over Infrared” approach (using a heart-rate tuned comb filter) provides a preferred tradeoff between SpO<sub>2</sub>-level accuracy and computational complexity.

## 6 Limitations

- (1) In this study SpO<sub>2</sub> ground truth was considered the levels calculated using the highly-accurate Nonin 8000R sensor [29], mounted on the subject’s temple. It may have been preferable to draw blood samples from the subjects and measure arterial oxygen saturation directly. However, this invasive procedure was not permitted under the IRB-approved experimental protocol.
- (2) All fourteen (14) tested datasets were collected from healthy and physically fit subjects. Therefore, the effect of irregular heart rhythm on SpO<sub>2</sub> level calculation using our methods was not studied.
- (3) All fourteen (14) datasets available from the experiment were processed. We did not exclude any sample based on its corresponding ECG signal quality or other considerations. In this study, the heart rate derived from measurements by the Nonin sensor was not used in our calculation, nor investigated further.

**Acknowledgements** This work was sponsored in part by the U.S. Office of Naval Research (ONR) under Grant No. N68335-16-C-0085.

## Compliance with ethical standards

**Conflict of interest** Data collection described in Sect. 3.2 involved a product of Athena GTX, of which author S. Mahoney is an employee. All other authors declare that they have no conflict of interest.

**Ethical approval** The experiments described in Sect. 3.2 were approved by Naval Air Warfare Center Aircraft Division IRB, protocol FWR21070114H, original approval date: 12 June 2017. Air Force Research Lab (AFRL) IRB protocols comply with DoD Directive 3216.02, Title 25, CFR 46, and are in compliance with the Declaration of Helsinki Revision 6, 2008.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix 1: p values for the two alternate calibration curves

We re-calculated the p values of Table 5 for the “Lambert-Beer curve” (Eq. (3)), and the “underestimation calibration curve” (Eq. (4)). We show the p-values of the tested hypotheses in Table 7 in Appendix for the “Lambert-Beer curve” and Table 8 in Appendix for the “underestimation calibration curve.” They support the conclusions provided in Table 4.

**Table 7** p values of all six tests for “Lambert–Beer calibration curve” (Eq. (3))

|         | p value (significance level was $\alpha = 0.01$ ) |           |            |            |           |           |
|---------|---|-----------|------------|------------|-----------|-----------|
|         | Test I  | Test II   | Test III-a | Test III-b | Test IV   | Test V    |
| Stage 1 | < 0.00001   | 0.00015   | 0.01147    | 0.00573    | 0.00066   | < 0.00001 |
| Stage 2 | < 0.0001  | < 0.0001  | 0.96102    | 0.48083    | 0.00032   | 0.00032   |
| Stage 3 | < 0.00001   | < 0.0001  | 0.07928    | 0.03958    | < 0.0001  | < 0.00001 |
| Stage 4 | < 0.00001   | < 0.00001 | 0.51485    | 0.25760    | < 0.0001  | < 0.0001  |
| Stage 5 | < 0.00001   | < 0.00001 | 0.10969    | 0.05477    | < 0.00001 | < 0.00001 |
| Stage 6 | < 0.00001   | < 0.00001 | 0.26849    | 0.13412    | < 0.0001  | < 0.00001 |
| Stage 7 | < 0.00001   | < 0.00001 | 0.16402    | 0.08195    | < 0.00001 | < 0.00001 |
| Stage 8 | < 0.00001   | 0.00024   | 0.28919    | 0.14459    | < 0.0001  | < 0.0001  |

**Table 8** p values of all six tests for “underestimation calibration curve” (Eq. (4))

|         | p value (significance level was $\alpha = 0.01$ ) |           |            |            |           |           |
|---------|---|-----------|------------|------------|-----------|-----------|
|         | Test I  | Test II   | Test III-a | Test III-b | Test IV   | Test V    |
| Stage 1 | < 0.00001   | 0.00046   | 0.01024    | 0.00511    | 0.00142   | < 0.00001 |
| Stage 2 | 0.00012   | 0.00010   | 0.98250    | 0.49139    | 0.00041   | 0.00036   |
| Stage 3 | < 0.00001   | < 0.0001  | 0.07269    | 0.03629    | < 0.0001  | < 0.00001 |
| Stage 4 | < 0.00001   | < 0.00001 | 0.49433    | 0.24723    | < 0.0001  | < 0.0001  |
| Stage 5 | < 0.00001   | < 0.00001 | 0.15623    | 0.078109   | < 0.00001 | < 0.00001 |
| Stage 6 | < 0.00001   | < 0.00001 | 0.28380    | 0.14178    | < 0.0001  | < 0.0001  |
| Stage 7 | < 0.00001   | < 0.00001 | 0.18906    | 0.09453    | < 0.00001 | < 0.00001 |
| Stage 8 | < 0.0001  | 0.00038   | 0.24414    | 0.12205    | < 0.0001  | < 0.0001  |

## References

- Webster JG. Design of pulse oximeters. Boca Raton: CRC Press; 1997.
- Goldman JM, Petterson MT, Kopotic RJ, Barker SJ. Masimo signal extraction pulse oximetry. *J Clin Monit Comput*. 2000;16:475–83.
- Lee HW, Lee JW, Jung WG, Lee GK. The periodic moving average filter for removing motion artifacts from PPG signals. *Int J Control Autom Syst*. 2007;5:701–6.
- Grap MJ. Pulse oximetry. *Crit Care Nurs*. 2002;22:69–74.
- Tremper KK. Pulse oximetry. *Chest*. 1989;95:713–5.
- Acharya S, Rajasekar A, Shender BS, Hrebien L, Kam M. Pulse oximeter signal fusion for robust hypoxia detection. *Aerospace Med Hum Perform*. 2015;86:495–6.
- Nihon Kohden. SpO<sub>2</sub> Monitoring Performance Digest; 2008. <https://mx.nihonkohden.com/sites/default/files/u21/SpO2-performance-digest.pdf>
- Stubán N, Masatsugu N. Non-invasive calibration method for pulse oximeters. *Periodica Polytechnica Electr Eng*. 2008;52:91–4.
- Wukitsch MW, Petterson MT, Tobler DR, Pologe JA. Pulse oximetry: analysis of theory, technology, and practice. *J Clin Monit*. 1988;4:290–301.
- Clarke GW. Signal quality analysis in pulse oximetry: modelling and detection of motion artifact (Doctoral dissertation). Carleton University; 2015.
- Widrow B, Glover JR, McCool JM, Kaunitz J, Williams CS, Hearn RH, Zeidler JR, Dong JE, Goodlin RC. Adaptive noise cancelling: principles and applications. *Proc IEEE*. 1975;63:1692–716.
- Stapleton J, Bass S. Adaptive noise cancellation for a class of nonlinear, dynamic reference channels. *IEEE Trans Circ Syst*. 1985;32:143–50.
- Barker SJ. “Motion-resistant” pulse oximetry: a comparison of new and old models. *Anesth Analg*. 2002;95:967–72.
- Lipnick MS, Feiner JR, Au P, Bernstein M, Bickler PE. The accuracy of 6 inexpensive pulse oximeters not cleared by the Food and Drug Administration: the possible global public health implications. *Anesth Analg*. 2016;123:338–45.
- Louie A, Feiner JR, Bickler PE, Rhodes L, Bernstein M, Lucero J. Four types of pulse oximeters accurately detect hypoxia during low perfusion and motion. *Anesthesiology*. 2018;128:520–30.
- Polimeni G, Scarpino A, Barbé K, Lamonaca F, Grimaldi D. Evaluation of the number of PPG harmonics to assess Smartphone effectiveness. In: 2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA), IEEE 2014, p. 1–6.
- Jayadevappa BM, Kiran Kumar GH, Anjaneya LH, Mallikarjun SH. Design and development of electro-optical system for acquisition of PPG signals for the assessment of cardiovascular system. *Int J Res Eng Technol*. 2014;3:520–5.
- Galli A, Narduzzi C, Giorgi G. Measuring heart rate during physical exercise by subspace decomposition and Kalman smoothing. *IEEE Trans Instrum Meas*. 2017;67:1102–10.
- Zhang Q, Zeng X, Hu W, Zhou D. A machine learning-empowered system for long-term motion-tolerant wearable monitoring of blood pressure and heart rate with ear-ECG/PPG. *IEEE Access*. 2017;5:10547–61.
- Pan J, Tompkins WJ. A real-time QRS detection algorithm. *IEEE Trans Biomed Eng*. 1985;32:230–6.
- Sang-Soo O, Praveen A. How to design peripheral oxygen saturation (SpO<sub>2</sub>) and optical heart rate monitoring (OHRM) systems using the AFE4403. Application report SLAA655. Texas Instruments; 2015. <http://www.ti.com/lit/an/slaa655/slaa655.pdf>

22. Kästle S, Noller F, Falk S, Bukta A, Mayer E, Miller D. A new family of sensors for pulse oximetry. *Hewlett Packard J.* 1997;48:39–47.
23. Prael S. Tabulated molar extinction coefficient for hemoglobin in water; 1998. <https://omlc.org/spectra/hemoglobin/summary.html>.
24. Saechia S, Koseeyaporn J, Wardkein P. Human identification system based ECG signal. In: TENCON 2005 IEEE Region 10 conference, IEEE 2005, p. 1–4.
25. Islam MK, Tangim G, Ahammad T, Khondokar MR. Study and analysis of ecg signal using matlab & labview as effective tools. *Int J Comput Electr Eng.* 2012;4:404.
26. Karthik R. ECG simulation using MATLAB; 2003. [www.mathworks.com/matlabcentral/fileexchange/10858](http://www.mathworks.com/matlabcentral/fileexchange/10858).
27. Karvonen J, Vuorimaa T. Heart rate and exercise intensity during sports activities. *Sports Med.* 1988;5:303–11.
28. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. *J Am Coll Cardiol.* 2001;37:153–6.
29. Nonin. Model 8000R Reflectance pulse oximeter sensor; 2016. [https://www.nonin.com/wp-content/uploads/8000R-0658-001-12\\_ENG.pdf](https://www.nonin.com/wp-content/uploads/8000R-0658-001-12_ENG.pdf).
30. Yan YS, Zhang YT. An efficient motion-resistant method for wearable pulse oximeter. *IEEE Trans Inf Technol Biomed.* 2008;12:399–405.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.