Measuring Heart Rate During Physical Exercise by Subspace Decomposition and Kalman Smoothing

Alessandra Galli, Claudio Narduzzi, and Giada Giorgi

***Abstract*— Monitoring of health parameters during physical exercise is an important aspect of both sports and rehabilitation medicine. Photoplethysmography (PPG) is routinely employed for low-cost heart rate (HR) measurement; however, monitoring during physical exercise is made difficult by the presence of motion artifacts. In this paper, we present an approach that combines denoising by subspace decomposition and Fourier-based HR measurement, and finally, smoothing and tracking by a Kalman filter. Using publicly available real-life PPG traces, we demonstrate accuracy and performance by an extensive set of experimental results, comparing them with similar algorithms proposed in the literature.**

***Index Terms*— Kalman filtering, motion artifact (MA) removal, photoplethysmography (PPG), signal processing, subspace decomposition.**

I. Introduction

MONITORING of health parameters during physical exercise is an important aspect of both sports and rehabilitation medicine. Photoplethysmography (PPG) is routinely employed for heart rate (HR) measurement in low-cost medical devices, since PPG signal periodicity corresponds to cardiac rhythm [1].

PPG is an optical technique that detects the changes in blood volume occurring in tissues at the microvascular level. A pulse oximeter measures changes in the intensity of light reflected from a subject skin when it is illuminated by a light-emitting diode (LED). The resulting waveform presents a pseudoperiodic pulse component attributed to changes in blood volume occurring with each heartbeat. Further components, related to respiration, nervous activity, and thermoregulation, are superimposed on it [2].

HR estimation from PPG signals is a popular research topic, and several algorithms have been presented in the literature. The use of PPG in HR monitoring during physical exercise is notoriously difficult [3], since acquisition of accurate measurement information is made harder by the presence of significant motion artifacts (MAs). During physical exercise, MA characteristics show greater variety, also depending on the kind of physical activity the subject is performing. PPG signals can be considerably degraded as a consequence, to the point that an estimation algorithm might focus on a wrong signal component [3]–[5].

In a wearable data acquisition system, PPG sensors can be complemented by an accelerometric sensor, whose outputs help identify the relevant features of the subject motion. This approach is applied, for instance, in [6], where two-channel PPG signals are acquired from a person wrist band together with the outputs of a tri-axial accelerometer. The auxiliary sensor enables the application of signal processing techniques for denoising and removal of the offending waveform components, but the task is far from simple. Multistage signal processing approaches are often employed (see [6]–[9] and the references therein). The possibilities of lightweight implementations based on machine learning approaches have also been investigated [10], [11], although reported accuracy performances currently appear to fall short of those achieved by other signal processing methods. In general, one of the main difficulties is finding a right balance between algorithm sophistication and computational cost.

An HR estimation algorithm can be seen as a sequence of three signal processing stages, namely, denoising, HR estimation, and tracking. Denoising includes MA removal, which is usually achieved either by adaptive filtering or by subspace-based signal decomposition. In the former case, the accelerometer signals are sent to a time-varying filter with recursively adjusted coefficients. Its output is an estimate of MA-related signal components that is subtracted from the acquired PPG signal. The algorithm is comparatively lightweight, and has been shown to provide acceptably good performance in practice [12]. In subspace-based approaches, principal component analysis (PCA) is usually applied. The rationale is that accelerometer signals are largely uncorrelated with PPG signals, and therefore a suitable choice of principal components allows to separate them. Practical implementation of this criterion, however, is not straightforward. As already noted, spurious signal components may be larger than useful ones, which makes the selection of principal components difficult. Furthermore, during physical exercise, HR and motion are not totally independent, and therefore PPG-related and motion-related signal components are not perfectly separable. Although performances reported in the literature for subspace-based methods are very good [13], their implementation tends to be complex.

Our approach to the HR monitoring problem is characterized by the following features.

Manuscript received July 1, 2017; revised September 4, 2017; accepted October 6, 2017. Date of publication November 24, 2017; date of current version April 5, 2018. The Associate Editor coordinating the review process was Dr. Anirban Mukherjee. *(Corresponding author: Claudio Narduzzi.)*

The authors are with the Department of Information Engineering, University of Padua, I-35131 Padua, Italy (e-mail: narduzzi@dei.unipd.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2017.2770818

1) PPG signal denoising is based on a novel criterion of maximal incorrelation between useful and spurious components, which is both robust and simple.
2) We exploit the redundancy offered by the availability of signals from two PPG sensors to minimize the occurrence of missing measurements.
3) We introduce a simple one-variable dynamic model of HR variation that is employed for prediction and smoothing of discrete Fourier transform (DFT)-based HR estimates by a Kalman filter (KF). This approach allows us to smooth, validate, and track the signal at the same time, producing a much-refined HR estimate. Since only one state variable is involved, KF equations are simple and computational requirements very limited.
4) The proposed algorithm does not require any initialization phase. The HR estimate converges to the correct value after a short transient, during which the subject is already allowed to exercise.

This paper, which extends the results of [14], is organized as follows. In Section II, we discuss relevant signal features, introduce mathematical basics, and present the novel criterion for subspace-based denoising. In Section III, we deal with frequency analysis and introduce a simple model of HR dynamics. This motivates our use of KF as a smoother to refine HR measurements. In Section IV, we define measurement validation criteria that contribute to the robustness of our estimation approach.

Finally, in Section V, we present an extensive set of results, using the signal traces from the PPG trace library provided for the 2015 IEEE Signal Processing Cup (SP Cup) competition [15], which has become a widely accepted benchmark for this kind of algorithm. The MATLAB code of the algorithm described in this paper is freely available in [16].

II. Maximal Incorrelation PPG Denoising

The trace library employed for trials provides a comprehensive set of test cases. Overall, each test lasts about 300 s, producing between 35 000 and 40 000 samples per trace. The database includes 12 traces from male subjects aged from 18 to 35, recorded while running, and 10 traces from further eight subjects, seven male and one female, with ages ranging from 19 to 58 years, recorded during rehabilitation exercises and boxing. All males are healthy, and the female subject is affected by a cardiovascular disease [15]. The test setup described in [6] is composed of two pulse oximeters, using a green LED at 515 nm, and tri-axial accelerometer, carried with a wrist band. Ground-truth data are provided by an associated electrocardiographic (ECG) trace acquired by electrodes placed on a chest band. Each acquisition includes the two PPG records and the outputs of the three accelerometer channels. All signals are sampled at a frequency of $F_s = 125$ Hz.

We analyze each signal employing an observation window of length $T_W = 8$ s, resulting in a segment of $N = 1000$ samples. Consecutive windows have a 75% overlap (i.e., 6 s), so that the algorithm provides a new HR estimate every $T_R = 2$ s, corresponding to a reporting rate of 0.5 Hz.

For each data segment $s_1(nT_s)$ or $s_2(nT_s)$ of length N, obtained from either of the two PPG sensors, and for each triaxial accelerometer output, respectively, $a_x(nT_s)$, $a_y(nT_s)$, and $a_z(nT_s)$, we form the corresponding trajectory matrices, respectively: $\mathbf{D}_{s_1}$, $\mathbf{D}_{s_2}$, and $\mathbf{D}_{a_x}$, $\mathbf{D}_{a_y}$, and $\mathbf{D}_{a_z}$.

For a PPG sensor output $s(nT_s)$, vectors $\underline{\mathbf{s}}(j)$ are defined as: $\underline{\mathbf{s}}(j) = [s(jT_s)s((j+1)T_s)\dots s((j+L-1)T_s)]^{\mathrm{T}}$, with $L < N$, where the superscript "T" denotes transposition. The corresponding vectors $\underline{\mathbf{d}}_s(j)$ are obtained by subtraction of the mean $\mu_s(j) = (1/L)\sum_{i=1}^{L} s[(j+i)T_s]$ from each element of $\underline{\mathbf{s}}(j)$.

The trajectory matrix $\mathbf{D}_s$ is then

$$\mathbf{D}_s = [\underline{\mathbf{d}}_s(1)\underline{\mathbf{d}}_s(2)\dots\underline{\mathbf{d}}_s(J)] \tag{1}$$

with $1 \le j \le J = N - L + 1$, and it can be noted that the matrix product: $(1/(L-1))\ \mathbf{D}_s^{\mathrm{T}}\mathbf{D}_s$ is an estimate of the autocovariance of $s(nT_s)$. Likewise, $\mathbf{D}_a$ is formed with accelerometer output samples arranged into vectors $\underline{\mathbf{a}}(j)$.

Singular value decomposition (SVD) yields for any of the trajectory matrices

$$\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} \tag{2}$$

where $\mathbf{U} = [\underline{\mathbf{u}}(1)\underline{\mathbf{u}}(2)\dots\underline{\mathbf{u}}(J)]$ contains the left singular vectors of $\mathbf{D}$, matrix $\mathbf{V}$ is formed by the right singular vectors, and Σ is the diagonal matrix of singular values, in decreasing order.

The subspace approach applied to $s(nT_s)$ aims at finding an index subset $J_{\mathrm{PPG}} \subset \{1,\dots J\}$ such that all useful information from the PPG sensor output is preserved in the matrix

$$\mathbf{D}_{\mathrm{PPG}} = \sum_{j \in J_{\mathrm{PPG}}} \underline{\mathbf{u}}_s(j) \cdot \Sigma_s(j,j) \cdot \underline{\mathbf{v}}_s(j)^{\mathrm{T}} \tag{3}$$

while interference effects originating from the MA component are minimized. Component selection in PCA is usually based on thresholding of eigenvalues and/or grouping of PPG signal eigenvectors [6].

The proposed criterion for principal signal component selection is based instead on the degree of incorrelation with accelerometer signals. Specifically, we consider the left eigenvector matrices $\mathbf{U}_s$ and $\mathbf{U}_a$, whose columns are direction vectors for the principal components of $s(nT_s)$ and $a(nT_s)$, respectively. In fact, PPG-related components of $s(nT_s)$ are those whose projection along the acceleration principal components is minimal. Accordingly, for each accelerometer axis, we associate to the ith principal component of $s(nT_s)$ the index

$$C_{s,a}(i) = \max_{1 \le j \le J} \underline{\mathbf{u}}_s^{\mathrm{T}}(i)\underline{\mathbf{u}}_a(j). \tag{4}$$

Then, we form J_{PPG} by selecting the principal components of $s(nT_s)$ for which the sum of indices (4), computed over the three accelerometer axes, is below a certain threshold τ_{pr}

$$J_{\mathrm{PPG}} = \{i : C_{s,a_x}(i) + C_{s,a_y}(i) + C_{s,a_z}(i) < \tau_{pr}\}. \tag{5}$$

This criterion is a refined version of the one presented in [14]. In our experiments, we set $\tau_{pr} = 0.6$.

The process is carried out on both PPG channels and the selected index subsets J_{PPG_1} and J_{PPG_2} allowing to reconstruct denoised versions of the two PPG sensor outputs. An example of denoised signal reconstruction is presented in Fig. 1,

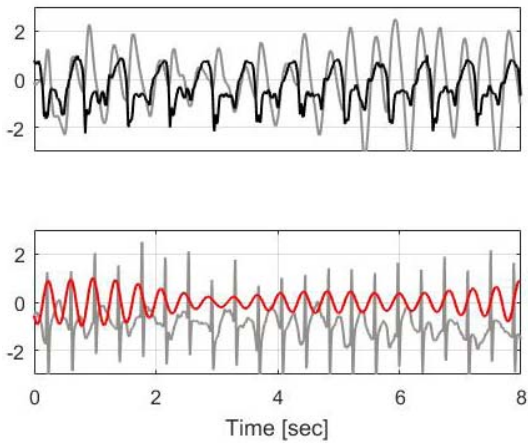

Fig. 1. Top: comparison between the PPG signal (black curve) and accelerometer (gray curve). Both signals are periodic with almost the same pseudoperiod. Bottom: ECG signal (gray curve) and reconstructed PPG signal (red curve).

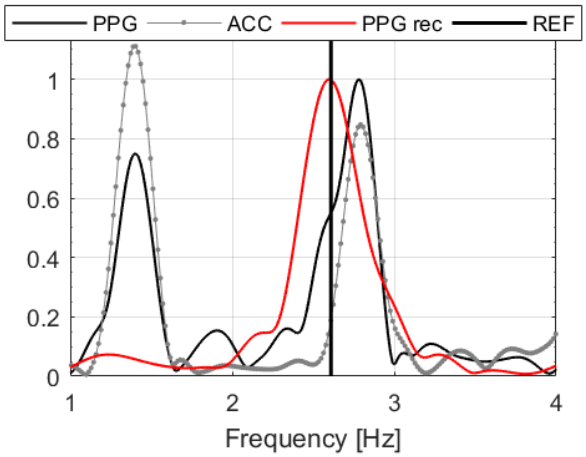

Fig. 2. Comparison among frequency spectra of raw PPG (black curve), accelerometer (gray curve), and reconstructed PPG (red curve). The black vertical line is the ECG reference value. Main peaks of the raw PPG spectrum correspond to MAs, and the main spectral peak of the denoised PPG corresponds to the ECG value.

where the "ground-truth" ECG recording is shown as a reference in Fig. 1(bottom).

It is important to emphasize differences between our selection criterion and the approach used in [6]. Although both algorithms use SVD, components affected by MAs are identified in [5] by frequency-domain analysis. In our approach, we employ instead time-domain signal subspace analysis.

III. Heart Rate Estimation and Tracking

A. Frequency Analysis

The initial HR measurement is obtained by frequency analysis of the denoised PPG signal $x(nT_s)$. We aim at the ability of measuring to a resolution of one beat per minute (± 1 bpm) that translates into the need to detect a minimum frequency shift $\Delta_f < (1/60)$ Hz. This is 7.5 times better than the frequency grid step obtained by computing a DFT with the 8-s observation window. On the other hand, extending this length to the required 60 s would be impractical. For this reason, data are zero-padded to extend DFT vector length to 8192 samples. Zero padding allows to obtain a finer grid step $F \cong 15$ mHz with a moderate increase in computation. Of course, the capability to resolve closely spaced frequency components remains lower bounded by the reciprocal of the sample sequence length T_W. Since the spectrum of PPG signals is mostly made up of harmonic components, this should not cause problems, their separation being usually good enough to resolve them.

The frequency spectra of the traces of Fig. 1 are presented in Fig. 2 and evidence that denoising reduced the MA component to the point that HR frequency can be estimated from the location of the largest spectral peak in the reconstructed signal spectrum

$$m_{peak} = \arg\max_m |X(mF)| \tag{6}$$

where $X(mF)$ is the zero-padded DFT of $x(nT_s)$. The frequency estimate for the kth observation window is $z(kT_R) = m_{peak}(kT_R)F$. Initially, the range of acceptable values of $z(kT_R)$ is set in accordance with physiological parameters.

Depending on the type and intensity of physical exercise, an MA component may get close enough to the HR frequency that the denoising approach described in Section II actually removes both the artifact and the useful PPG component. This could leave the second harmonic (i.e., the PPG component at twice the HR frequency) as the most likely candidate output. To avoid any gross inaccuracy, the frequency estimate obtained from the analysis of denoised data is compared with the estimate from unprocessed data. The occurrence of this specific condition can thus be detected and, accordingly, the correct result obtained as the DFT-based HR estimate divided by two.

HR estimates obtained by frequency analysis are reasonably accurate in the majority of cases; however, a limited but significant proportion deviates from the reference value by a rather large amount. This is evidenced by the histogram of Fig. 3 that refers to the signal acquisitions composing the training set for [15]. The graph summarizes a total of over 3000 HR estimates, obtained from two distinct PPG traces for each of the 12 acquisitions. Although most estimates are within ± 10 bpm of the reference, in a few cases deviation is 10 times larger, evidencing the need for a follow-up processing stage aimed at measurement, validation, and smoothing.

B. Dynamic HR Model and Kalman Filtering

HR measurements provided by DFT-based frequency analysis are still regarded as "raw measurements." The algorithm stage that follows is based on a simple signal model of HR variation and deals with validation, filtering, and smoothing, ultimately providing a considerably more accurate HR measurement.

Considerations based on physiology, combined with the need to employ simple equations in the interest of robustness, led us to the representation of the dynamic behavior of human HR by a discrete-time model based on the discrete-time random walk process

$$\theta(kT_R) = \theta[(k-1)T_R] + w(kT_R) \tag{7}$$

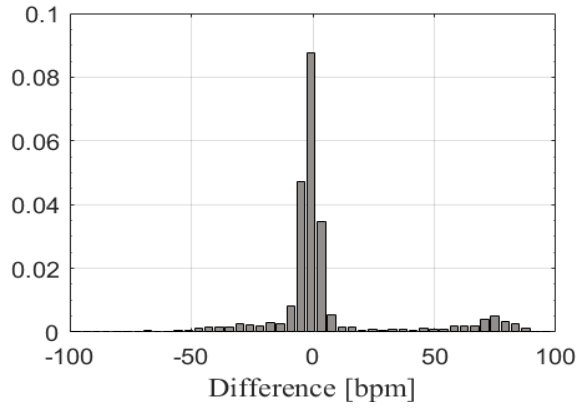

Fig. 3. Deviations of DFT-based HR estimates from reference values. The histogram evidences the presence of a nonnegligible number of outliers.

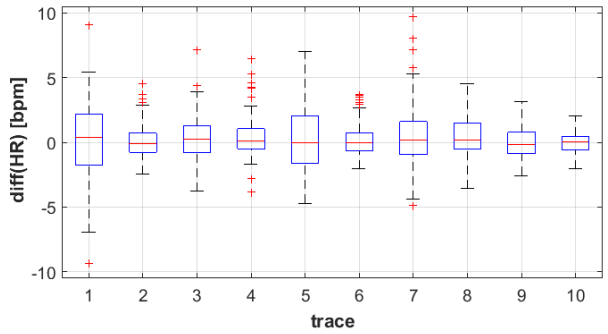

Fig. 4. Tukey boxplot summarizing statistical properties of the process noise for different time series.

where $\theta(kT_R)$ is the current HR value and $w(kT_R)$ is a zero-mean white noise random process with finite variance σ_w^2. It is recalled that T_R is the time interval between two consecutive overlapped trace segments.

The refined HR estimate is obtained by a Kalman filtering approach, in which (7) represents the state equation and the measurement equation is simply

$$z(kT_R) = \theta(kT_R) + v(kT_R). \tag{8}$$

Measurement noise and uncertainty of raw HR estimates are represented by the zero-mean white noise random process $v(k)$, whose finite variance is σ_v^2.

Careful KF tuning is critical to performance, and for this reason, the values of KF parameters σ_w^2 and σ_v^2 have been determined from the analysis of experimental data. The process noise variance σ_w^2 accounts for the goodness of fit between actual HR evolution and the model introduced in (7). An evaluation of the contribution of $w(kT_R)$ can be obtained by considering HR reference values and extracting the statistical properties of the difference time series. This analysis was carried out on the set of 10 acquisitions that compose the SP Cup "testing" set, whose difference time series is characterized by greater variability. Results are reported in Fig. 4.

It can be noted that almost all difference values are distributed in the range $|w(kT_R)| < w_{\mathrm{MAX}} = 7$ bpm. Taking this bound, we conservatively assumed a uniform distribution and accordingly considered $\sigma_w = (w_{\mathrm{MAX}}/\sqrt{3}) = 4$ bpm as a suitable value. With this and $T_R = 2$ s, model (7) is responsive enough to represent changes related to the intensity of physical exercise, but rules out unrealistically sudden variations.

TABLE I
KF EQUATIONS FOR REFINED HR ESTIMATION

Equation	Description
$\theta^-(k) = \theta^+(k-1)$	state prediction
$P^-(k) = P^+(k-1) + \sigma_w^2$	state prediction variance
$y(k) = z(k) - \theta^-(k)$	innovation
$S(k) = P^-(k) + \sigma_v^2$	innovation variance
$K(k) = \dfrac{P^-(k)}{P^-(k) + \sigma_v^2}$	Kalman gain
$\theta^+(k) = \theta^-(k) + K(k)y(k)$	updated state estimate
$P^+(k) = K(k)\sigma_v^2$	updated state estimation variance

Similarly, the variance of measurement noise process $v(kT_R)$ can be evaluated by considering the time series corresponding to the difference between the reference HR values and the estimated values provided at the output of the SVD algorithm. Performances of the algorithm are unaffected by slight variations of σ_v, as long as the order of magnitude is correct. Consequently, based on the histogram in Fig. 3, we calculated the standard deviation of values within ± 50 bpm from zero, to discard possible outliers. This yields $\sigma_v \cong 10$ bpm.

The whole set of KF equations are summarized in Table I for conciseness and, from here on, time interval T_R is dropped from equations for simplicity. At the end of the kth algorithm iteration, the KF-smoothed HR measurement is, of course, the *a posteriori* state estimate $\theta^+(k)$.

IV. Validation of DFT-Based Measurements

A KF iteration takes the raw HR measurements $z_1(k)$ and $z_2(k)$, obtained from the two denoised PPG sensor traces, and processes them to provide an accurate estimate of the actual HR value. Iteration steps are as follows:

1) measurement, where an HR estimate is obtained for each PPG channel from a zero-padded DFT computed on the given time window;
2) validation, where suitable checks determine whether new measurements are acceptable; if so, measurement information is passed to the next step;
3) update, where a refined *a posteriori* HR estimate is provided;
4) prediction, where *a priori* estimates are obtained both for the HR value and its variance, in preparation for the next algorithm iteration.

It can be seen that, besides the straightforward KF iteration steps, we introduced a *measurement validation* step, specifically to deal with possible measurement inconsistencies originating from residual MA effects, as well as from the possibility of degradation in the acquired signals.

A preliminary check considers possible variability among measurements $z_1(k)$ and $z_2(k)$ from the two PPG channels. For this reason, their variances, computed over a window including stored previous values over a length of 3 min,

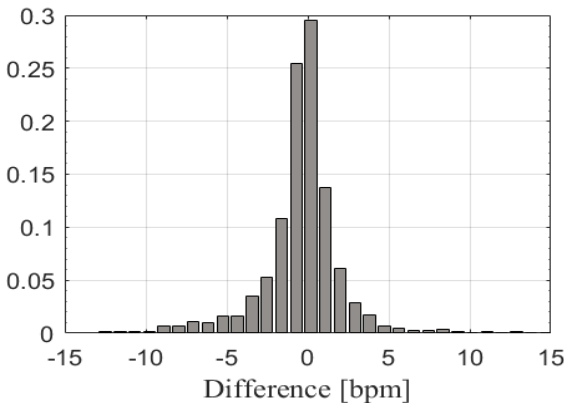

Fig. 5. Deviations of KF-smoothed HR estimates from reference values.

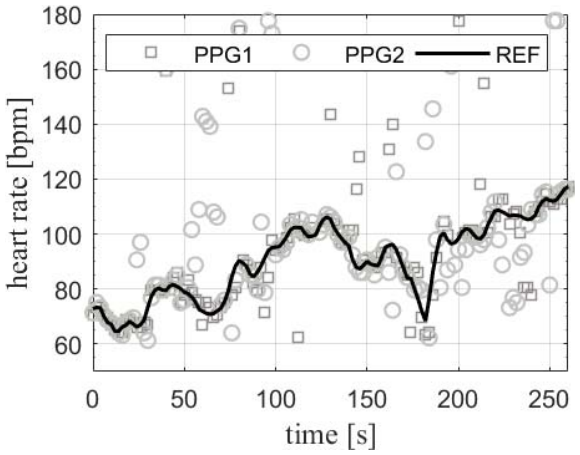

Fig. 6. Raw HR estimates from denoised PPG traces (gray symbols) for a subject performing a rehabilitation exercise.

are compared. Mutual disagreement is considered too great when either of the two variances is greater than twice the other value. Accordingly, the corresponding measurement is flagged as "unavailable" and discarded from the following steps.

The validation check proper is related to KF innovation and leads to the acceptance of a new measurement only if the following condition is satisfied:

$$|z_i(k) - \theta^-(k)| \leq \kappa \cdot \sqrt{S(k)} \quad \text{with: } i = 1,\ 2. \tag{9}$$

The factor κ can be associated with a suitable confidence level. For the results presented in the following, $\kappa = 2$ was applied, which corresponds to a level of confidence of about 95%, assuming innovation to be approximately Gaussian. It can be seen from (9) that the check is applied individually to both PPG measurements (provided both are available), and therefore it might be passed by both $z_1(k)$ and $z_2(k)$, failed by either of the two, or by both. Consequently, the outcome of the validation process must be defined in the three cases.

1) If both measurements are accepted, the actual measurement input passed to the KF is the one whose innovation is smaller.
2) When one of the measurements fails condition (9), it is flagged as "unavailable" and only the other is passed on for KF processing.
3) When both new measurements fail the check and are flagged as "unavailable," the current state prediction is kept for the next KF iteration, i.e., $\theta^+(k) = \theta^-(k)$ and $P^+(k) = P^-(k)$. However, after new measurements have been "unavailable" for five consecutive iterations, the next *a posteriori* KF state estimate is set to the average between $\theta^-(k)$ and either or both of the validated new measurements. This is essential to avoid deadlock and bring KF estimates progressively back to correct values.

The smoothed HR estimate $\theta^+(k)$ is significantly more accurate than raw HR measurements, as shown by the histogram of Fig. 5 that refers to the same set of measurements as Fig. 3. Referring again to over 3000 occurrences, HR estimates turn out to be practically unbiased (mean value < 0.1 bpm). It should be stressed this is a cumulative assessment, and therefore standard deviation (about 3.7 bpm) is representative of the varied features of PPG traces.

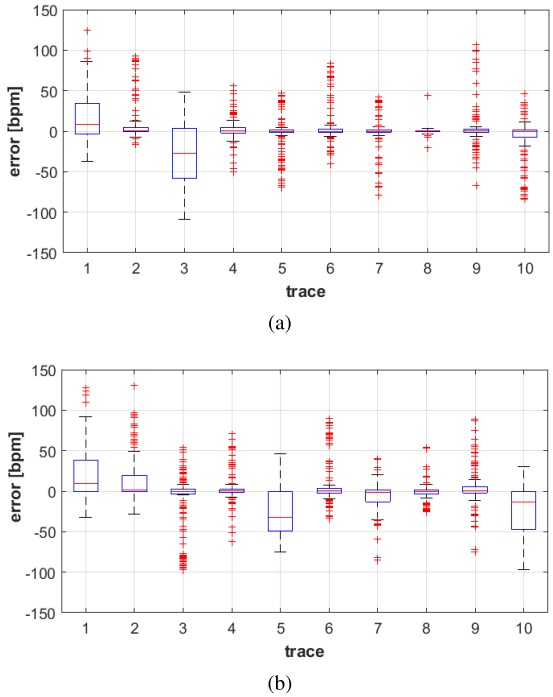

Fig. 7. Tukey box plots summarizing statistical properties of raw HR measurement noise for the 10 acquisitions in the "testing" SP Cup set. (a) PPG1 sensor output. (b) PPG2 sensor output.

For a visual assessment of the proposed algorithm behavior, it can be seen from Fig. 6 that raw HR estimates are occasionally affected by large deviations, as already shown by the spread of the histogram in Fig. 3. However, only in a few cases are estimates from *both* PPG traces far from the reference. To better understand this, Fig. 7 presents Tukey box plots for the outputs of the two sensors in each of the SP Cup "testing" set of trace acquisitions. It can be noted that in most cases, 50% of the raw HR estimates are distributed in a narrow range around the reference value for at least one of the two PPG sensors. This means that at least one sensor can provide an acceptably accurate raw HR estimate. Trace no. 1 is slightly more critical in this regard.

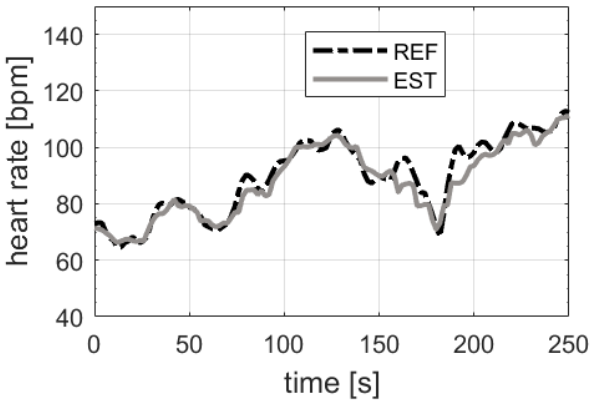

Fig. 8. KF-smoothed and validated HR estimates (continuous gray line) for a subject performing a rehabilitation exercise.

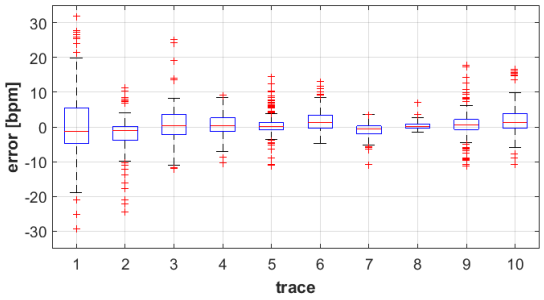

Fig. 9. Tukey box plot summarizing variability of KF-smoothed HR measurements for the ten acquisitions in the "testing" SP cup set.

The effect of the validation criteria introduced above is evidenced in the smoothed HR trace of Fig. 8, for which average absolute deviation from reference values is 2.6 bpm. The post-smoothing Tukey box plot is shown in Fig. 9, and it can be noted that the range of variation has been reduced from ± 120 to ± 30 bpm. It can be observed that, for all traces, box widths tend to coincide with the narrower among the corresponding two boxes of Fig. 7. Even in the case of trace no. 1 in Fig. 7, the selection rule for raw measurements led to an improvement, although less marked.

Uncertainty at approximately 99% confidence level can be assessed by the extension of box plot whiskers, ranging from ± 3 bpm in the best case (trace no. 8) to ± 20 bpm in the worst case (trace no. 1). It should be remarked that, in terms of measurement accuracy, this indication of uncertainty is much more severe than the parameters usually employed to assess algorithms of this kind. Performance metric E_3 is introduced in the next section for the same purpose.

V. Results

We refer in this section to traces from [15], employed in [6] as well as in a number of subsequent works for performance evaluation. Those traces cover a set of subjects of various ages, undertaking physical exercises of different intensity. HR varies in accordance with physical stress, though individual responses are closely related to specific physical conditions of the subject. Reference HR values, obtained from the simultaneously recorded ECG trace and indicated in the following as $B_{\text{ref}}(k)$, were compared with the estimate $B_{\text{meas}}(k) = \theta^+(k)$.

It should be remembered that, for this kind of HR measurement, uncertainty is primarily a consequence of MAs. Thus, even though the set of conditions provided by [15] does not cover all combinations of gender, age, and health conditions, the variety of MAs found in the traces is enough to support the claim that reported performance is representative. As in similar works, we consider two different performance metrics:

1) the average *absolute* deviation from reference values

$$E_1 = \frac{1}{N}\sum_{k=1}^{N} |B_{\text{meas}}(k) - B_{\text{ref}}(k)|$$

that is expressed in bpm;

2) the average *relative* deviation from reference values

$$E_2 = \frac{1}{N}\sum_{k=1}^{N} \frac{|B_{\text{meas}}(k) - B_{\text{ref}}(k)|}{B_{\text{ref}}(k)}$$

that is given in percent.

Results for the full set of 22 trace acquisitions are reported in Table II, where the first 12 traces formed the SP Cup "training set" and the other 10 traces are the "testing set." A label indicates for each trace the corresponding physical activity. As already noted, intertrace variability is greater in the latter set and this is reflected in the slightly larger standard deviation values of metrics E_1 and E_2.

On the basis of these two metrics, we compared our algorithm with some of the better performing ones in the literature. Values referring to [6] (columns labeled "A" in Table II) are provided mainly as a historical reference, since TROIKA was a performance benchmark at the time of the 2015 SP Cup. Since then, developments allowed other algorithms to outperform it, as shown in columns labeled "B," "C," and "D" in Table II. The algorithm we present in this paper compares on even terms with all three, showing good average agreement with values obtained from the ECG reference traces. Specifically, [13] shows remarkably good results, but the algorithm is more complex than the proposed one. Conversely, the accuracy of [12] appears to be slightly inferior, but the algorithm is claimed to be designed for lower computational complexity. A very recently published work [17] is also included in the comparison. Although based on a different approach, this shows some conceptual similarity to our work in the use of methods (Wiener filtering and phase vocoding) that, although somewhat demanding from a theoretical viewpoint, can be implemented with moderate computational complexity.

To better characterize our algorithm and complete the comparison with [17], we added, as a further indication, the *maximum* absolute deviation

$$E_3 = \max_{1\le k\le N} |B_{\text{meas}}(k) - B_{\text{ref}}(k)| \tag{10}$$

that allows an assessment of algorithm pointwise accuracy by showing the worst case local disagreement between PPG and ECG measurements. Metric E_3 provides a useful performance bound that, to the best of the authors' knowledge, is unreported in the literature for most other algorithms.

TABLE II

PERFORMANCE METRICS FOR PROPOSED HR ESTIMATOR AND COMPARISON WITH OTHER ESTIMATORS ***Key***: A = TROIKA [6]; B = SPECTRAP [13]; C = FALLET *et al.* [12]; D = TEMKO [17]

Trace	*Type*	**E1** [bpm]					**E2** [%]					**E3** [bpm]		**E4** [bpm]	
		A	B	C	D	**Ours**	A	B	C	D	**Ours**	D	**Ours**	D	**Ours**
1	R	2.29	1.18	1.75	1.25	**2.72**	2.18	1.04	1.59	1.15	**2.11**	11.82	**12.95**	1.81	**2.51**
2 *	R	2.19	2.42	1.94	1.41	**3.25**	2.37	2.33	1.99	1.30	**3.02**	18.12	**27.25**	2.33	**5.03**
3	R	2.00	0.86	1.17	0.71	**1.40**	1.50	0.66	1.02	0.59	**1.11**	5.85	**8.72**	0.80	**1.51**
4	R	2.15	1.38	1.67	0.97	**1.21**	2.00	1.31	1.51	0.88	**1.04**	12.20	**6.08**	1.43	**1.21**
5	R	2.01	0.92	0.95	0.75	**0.93**	1.22	0.74	0.75	0.57	**0.70**	5.47	**4.80**	0.86	**0.90**
6	R	2.76	1.37	1.22	0.92	**2.12**	2.51	1.14	1.05	0.75	**1.82**	7.52	**32.36**	1.21	**5.43**
7	R	1.67	1.53	0.91	0.65	**1.40**	1.27	1.36	0.72	0.50	**1.04**	3.20	**8.70**	0.53	**1.61**
8	R	1.93	0.64	1.17	0.97	**1.16**	1.47	0.55	1.04	0.83	**0.97**	7.28	**5.31**	1.05	**1.03**
9	R	1.86	0.60	0.87	0.55	**1.17**	1.28	0.52	0.76	0.48	**0.95**	2.26	**6.75**	0.48	**1.54**
10 *	R	4.70	3.65	2.95	2.06	**4.14**	2.49	2.27	1.93	1.29	**2.79**	19.20	**32.30**	2.31	**6.04**
11	R	1.72	0.92	1.15	1.03	**1.38**	1.29	0.65	0.79	0.68	**0.91**	9.08	**9.27**	1.17	**1.86**
12	R	2.84	1.25	1.00	0.99	**1.29**	2.30	1.02	0.79	0.70	**0.92**	10.24	**6.17**	1.05	**1.10**
Average	1 – 12	**2.34**	**1.50**	**1.40**	**1.02**	**1.85**	**1.82**	**1.12**	**1.16**	**0.81**	**1.45**	**9.35**	**13.39**	**1.25**	**2.48**
Std. Dev.		**0.53**	**0.86**	**0.60**	**0.41**	**1.00**	**0.53**	**0.61**	**0.50**	**0.28**	**0.79**	**5.32**	**10.70**	**0.62**	**1.88**
13	T1	6.63	4.89	12.12	9.59	**7.91**	8.76	6.29	16.13	12.2	**10.3**	32.13	**31.89**	8.95	**7.33**
14	T1	1.94	1.58	4.02	2.57	**3.65**	2.56	1.98	5.28	3.16	**4.73**	22.10	**24.38**	3.92	**4.51**
15	T2	7.82	3.05	5.64	3.01	**3.90**	4.88	2.00	3.52	1.99	**2.52**	21.97	**25.10**	3.74	**4.16**
16	T2	2.46	1.62	3.31	2.73	**2.44**	2.00	1.36	2.81	2.29	**1.97**	22.71	**10.30**	3.64	**2.10**
17	T2	1.73	1.24	3.39	1.57	**2.14**	1.27	0.92	2.51	1.15	**1.57**	13.87	**14.41**	2.37	**2.79**
18	T1	3.33	2.04	3.45	2.10	**2.60**	3.90	2.23	4.11	2.41	**2.86**	22.76	**13.10**	2.96	**2.58**
19	T2	2.69	1.16	1.56	1.61	**1.86**	2.12	0.92	1.21	1.26	**1.44**	15.38	**10.86**	2.28	**1.78**
20	T1	0.51	0.66	0.95	0.75	**0.85**	0.59	0.79	1.11	0.88	**0.99**	5.63	**6.97**	0.72	**0.89**
21	T2	1.35	1.83	2.52	2.25	**3.06**	1.04	1.49	2.10	1.87	**2.54**	16.98	**17.79**	3.15	**3.61**
22	T2	3.41	2.49	5.86	3.44	**3.38**	2.43	1.81	3.99	2.45	**2.32**	23.70	**16.46**	4.66	**3.88**
Average	13 – 22	**3.19**	**2.13**	**4.28**	**2.95**	**3.18**	**2.95**	**2.04**	**4.28**	**2.96**	**3.13**	**19.73**	**17.13**	**3.64**	**3.26**
Std. Dev.		**2.32**	**1.21**	**3.16**	**2.46**	**1.89**	**2.41**	**1.60**	**4.37**	**3.31**	**2.72**	**7.13**	**7.80**	**2.16**	**1.80**
Average	Total	**2.73**	**1.79**	**2.71**	**1.90**	**2.45**	**2.33**	**1.87**	**2.58**	**1.98**	**2.21**	**14.06**	**15.08**	**2.33**	**2.83**
Std. Dev.		**1.34**	**1.02**	**1.76**	**1.45**	**1.59**	**1.73**	**1.06**	**2.26**	**2.35**	**2.06**	**6.14**	**9.46**	**1.32**	**1.85**

R = subject running on a treadmill

T1 = subject performing arm rehabilitation exercises

T2 = subject performing fast arm movements (boxing)

Finally, E_4 is the standard deviation

$$E_4 = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[B_{\text{meas}}(k) - B_{\text{ref}}(k)]^2} \tag{11}$$

that is computed over each whole trace.

We aimed at a balance of performance factors and built-in robustness that, in a physical exercise environment, is essential to make the algorithm suitable for HR measurement in a wearable unit. In this regard, it is useful to comment on the peculiarities of some of the traces, to explain the reasons of occasionally poorer performance. In particular, Fig. 10 shows plots of trace no. 2 in Table II, where deviations from the HR reference values are localized in two specific places, encircled in Fig. 10. It must be noted that our algorithm converges back to correct values in both cases, after no more than 30 s. Furthermore, in practice, these local events can be detected in the KF measurement validation stage and HR estimates suitably marked. Discarding this short length of time, the value of E_1 for the rest of the trace drops from 3.25 to 0.71 bpm. Likewise, E_1 could drop from 4.14 to 0.96 bpm for trace no. 10 (both traces are marked by an asterisk in Table II). We remark that in this case the average E_1 for the first 12 traces is just 1.37 bpm with a standard deviation of 0.55 bpm, i.e., comparable with [17] and better than any of the other three algorithms.

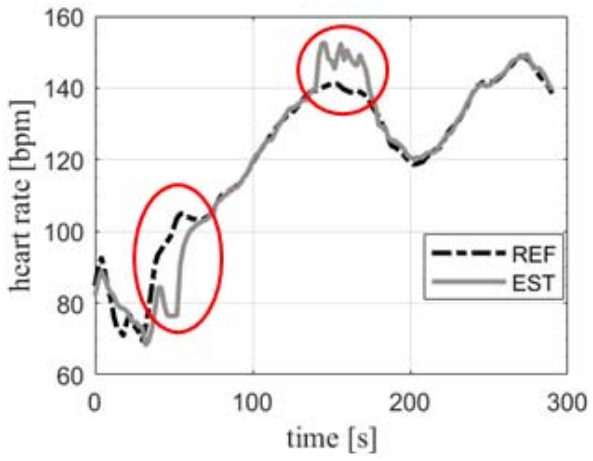

Fig. 10. HR estimate (EST) for a running exercise (trace no. 2 in Table II). Significant deviations from the reference (REF) occur only in limited places.

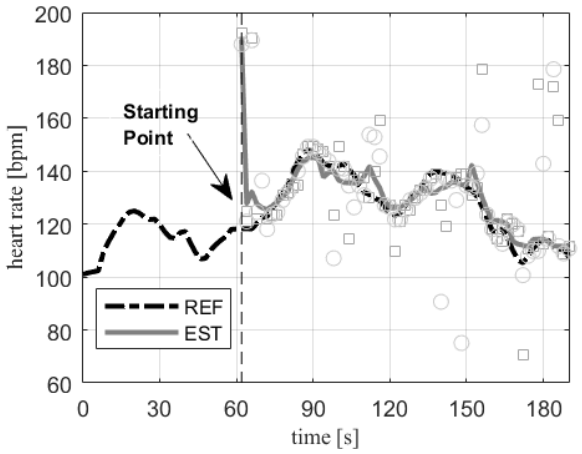

Fig. 11. Estimated HR during boxing exercises. The algorithm starts 64 s after trace recording. KF transient duration: about 10 s.

As emphasized in Section I, another feature that differentiates our approach is that it requires no training phase, unlike many other algorithms discussed in the literature. The training phase is very important, in particular, for algorithms like [6] and [17], that rely on a "search range" defined around the current HR estimate. It is usually carried out under the assumption that the subject does not start exercising for a given time interval, so that the initial PPG-based HR estimate is unaffected by the presence of artifacts. For this reason, the user is required to reduce motion as much as possible for several seconds, until a tracking algorithm is correctly set up, since incorrect user behavior may invalidate training assumptions and severely affect accuracy. Furthermore, initialization needs to be repeated whenever the HR monitoring system is used.

On the contrary, the proposed denoising approach effectively reduces the impact of MAs to the point that no preliminary training phase is needed for correct initialization. The KF estimate quickly converges to the correct HR value after an initial transient, during which the subject is already allowed to exercise. To demonstrate this, we repeatedly ran the algorithm with the same PPG trace, using randomly selected start times so that no training phase could be involved. In a set of 100 randomly chosen start times, HR error was 3.18 bpm on average, and never in excess of 4.52 bpm.

A worst case example is presented in Fig. 11, where the algorithm was started after 64 s, coincident with the arrival of measurement outliers that initially disrupted the PPG-based estimates. Regardless of the particularly adverse initial conditions, the algorithm converged quickly and, after the first iterations, tracked the reference HR values rather closely.

VI. Conclusion

The algorithm presented in this paper employs a zero-padded DFT to obtain raw HR estimates from a denoised PPG trace, and then refines the measurement by a simple KF using just one state variable, tracking the time evolution of the heart rate. The most computationally demanding steps involve SVD and the eigenvector analysis of a number of trajectory matrices that are required for signal denoising.

Much development effort was concentrated on robustness to variability originating in data acquisition during physical exercise, as well as related to intrinsic variation of subjective physiological parameters. In this regard, we proposed an effective principal component selection criterion and applied criteria for accurate KF tuning based on experimental data that significantly improved the filter transient behavior.

Performance-wise, the algorithm is in an intermediate position between complex but accurate offline processing algorithms and lightweight less accurate algorithms, for which implementation on a portable device is among the main objectives. It must be emphasized that, in terms of accuracy, the algorithm runs close to the better ones reported in the literature. On the other hand, its complexity is still manageable even for reasonably low-cost devices (e.g., a smartphone), considering the computing power currently available in such devices.

Computational requirements of our approach are determined by the first denoising stage, where SVD is implemented. Computational cost for each full SVD of a $J \times L$ matrix $\mathbf{U}$ is $O(J^2 \times L + J \times L^2 + L^3)$, using MATLAB function `svd()`, and five of them are needed. In addition, the cost for spectrum calculation is $O(M \log M)$ for each M-sample zero-padded DFT and two, for the raw and denoised PPG, are required for each channel. In comparison, the computational complexity of the KF with just one state variable is negligible since, as can be seen from Table I, each iteration requires five additions, one division, and two multiplications. Even accounting for checks and comparisons for measurement validation, the total cost for KF smoothing, validation, and tracking is negligible anyway.

Summarizing, with $J = 601$, $L = 400$, and $M = 8192$, the computational requirement is about 300×10^6 operations per iteration of the proposed algorithm. Using MATLAB R2015b on a personal computer with Intel core i7 running at 2.4 MHz, 8-GB RAM, and Windows 10 operating system, the computing time was 200 ms per second of recording time, which leaves a healthy margin for slower processors.

Acknowledgment

The authors would like to thank Dr. G. Frigo for his collaboration in the analysis of the SP Cup trace library and useful discussions in the initial stages of this paper.

References

[1] E. Gil, M. Orini, R. Bailón, J. M. Vergara, L. Mainardi, and P. Laguna, "Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions," *Physiol. Meas.*, vol. 31, no. 9, pp. 1271–1290, 2010.

[2] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, 2007.

[3] B. S. Kim and S. K. Yoo, "Motion artifact reduction in photoplethysmography using independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 566–568, Mar. 2006.

[4] E. Gil, P. Laguna, J. P. Martínez, Ó. Barquero-Pérez, A. Garcia-Alberola, and L. Sörnmo, "Heart rate turbulence analysis based on photoplethysmography," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3149–3155, Nov. 2013.

[5] X. He, R. A. Goubran, and X. P. Liu, "Secondary peak detection of PPG signal for continuous cuffless arterial blood pressure measurement," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1431–1439, Jun. 2014.

[6] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.
[7] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 5, pp. 1445–1457, May 2012.
[8] E. Khan, F. A. Hossain, S. Z. Uddin, S. K. Alam, and M. K. Hasan, "A robust heart rate monitoring scheme using photoplethysmographic signals corrupted by intense motion artifacts," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 550–562, Mar. 2016.
[9] Y. Ye, Y. Cheng, W. He, M. Hou, and Z. Zhang, "Combining nonlinear adaptive filtering and signal decomposition for motion artifact removal in wearable photoplethysmography," *IEEE Sensors J.*, vol. 16, no. 19, pp. 7133–7141, Oct. 2016.
[10] E. Grisan, G. Cantisani, G. Tarroni, S. K. Yoon, and M. Rossi, "A supervised learning approach for the robust detection of heart beat in plethysmographic data," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Milan, Italy, Aug. 2015, pp. 5825–5828.
[11] V. Jindal, J. Birjandtalab, M. B. Pouyan, and M. Nourani, "An adaptive deep learning approach for PPG-based identification," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Orlando, FL, USA, Aug. 2016, pp. 6401–6404.
[12] S. Fallet and J.-M. Vesin, "Robust heart rate estimation using wrist-type photoplethysmographic signals during physical exercise: An approach based on adaptive filtering," *Physiol. Meas.*, vol. 38, no. 2, pp. 155–170, 2017.
[13] B. Sun and Z. Zhang, "Photoplethysmography-based heart rate monitoring using asymmetric least squares spectrum subtraction and Bayesian decision theory," *IEEE Sensors J.*, vol. 15, no. 12, pp. 7161–7168, Dec. 2015.
[14] A. Galli, G. Frigo, C. Narduzzi, and G. Giorgi, "Robust estimation and tracking of heart rate by PPG signal analysis," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Turin, Italy, May 2017, pp. 364–369.
[15] *IEEE Signal Process. Cup 2015*. Accessed: 2015. [Online]. Available: http://www.signalprocessingsociety.org/community/sp-cup/
[16] *MATLAB Code*. Accessed: 2017. [Online]. Available: https://github.com/AlessandraGalli/PPG
[17] A. Temko, "Accurate heart rate monitoring during physical exercises using PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, nol. 9, pp. 2016–2024, Sep. 2017, doi: 10.1109/TBME.2017.2676243.

Alessandra Galli received the bachelor's degree in biomedical engineering and the master's degree (*cum laude*) in bioengineering from the University of Padua, Padua, Italy, in 2015 and 2017, respectively, where she is currently pursuing the Ph.D. degree in information and communication science and technologies with the Department of Information Engineering.

Claudio Narduzzi received the Laurea degree in electronics engineering from the University of Padua, Padua, Italy, in 1982.

He is currently a Full Professor in instrumentation and measurement with the Engineering School, University of Padua. He has authored over 150 scientific papers and textbooks for academic courses. His current research interests include digital signal processing applications in measurement and distributed test and measurement systems, spectral analysis, inverse filtering, networked measurement systems, the application of advanced signal processing methods, and compressive sensing.

Giada Giorgi received the Laurea degree in telecommunications engineering and the Ph.D. degree in microelectronics and telecommunications engineering from the University of Padua, Padua, Italy, in 2003 and 2007, respectively.

She is currently an Assistant Professor with the Department of Information Engineering, University of Padua. Her current research interests include the development of signal processing applications and measurement systems for wireless sensor and actuator networks, wireless body sensor networks, and smartphones.