Heuristic Algorithm for Photoplethysmographic Heart Rate Tracking During Maximal Exercise Test

Sonnia María López-Silva[1] Romano Giannetti[2,*] María Luisa Dotor[1]

Juan Pedro Silveira[1] Dolores Golmayo[3] Francisco Miguel-Tobal[4]

Amaya Bilbao[4] Mercedes Galindo[4] Pilar Martín-Escudero[4]

[1]*Instituto de Microelectrónica de Madrid, CNM-CSIC, C/Isaac Newton 8, Tres Cantos, Madrid 28760, Spain*
[2]*Departamento de Electrónica y Automática, Universidad Pontificia Comillas, C/Alberto Aguilera 23, Madrid 28015, Spain*
[3]*Instituto de Ciencias de Materiales de Madrid, CSIC, C/Sor Juana Inés de la Cruz 3, Madrid 28049, Spain*
[4]*Escuela Profesional de Medicina de la Educación Física y el Deporte, Facultad de Medicina, Universidad Complutense de Madrid, Madrid 28040, Spain*

Received 4 Mar 2011; Accepted 2 Oct 2011; doi: 10.5405/jmbe.898

Abstract

Photoplethysmography (PPG) is a non-invasive optical technique that can be used to quantify the arterial blood pulse rate. Signal corruption by motion artifacts limits the practical accuracy and applicability of instruments for monitoring pulse rate during intense physical exercise. This study develops and validates an algorithm, which is based on linear filtering and frequency-domain and heuristic analyses, for extracting the heart rate from a PPG signal in the presence of severe motion artifacts. The basis of the hearth beat frequency selection is the observed high harmonic content of movement artifact signals with respect to the PPG-derived heartbeat. The algorithm, implemented in an experimental PPG measurement device, is developed by analyzing a set of PPG data recorded from a group of athletes exercising on a treadmill. An extensive set of tests is carried out during maximal exercise tests on a treadmill to validate the proposed algorithm by comparison with a reference electrocardiography measurement system. The Bland-Altman method is used to compare and evaluate PPG signals. The accuracy of the heartbeat measurement is better than ± 6.5 beats per minute (bpm) (≤ 4.2 %) even under maximal exercise conditions.

Keywords: Biomedical signal processing, Photoplethysmography (PPG), Motion artifact, Heart rate

1. Introduction

Photoplethysmography (PPG) offers a simple, useful, and compact way of measuring several clinical parameters, including oxygen saturation, blood pressure, and cardiac output [1,2]. PPG is especially suitable for wearable sensing, which could play an important role in areas ranging from personal health monitoring to sports medicine. Despite the attractive attributes of PPG and the ease of its integration into wearable devices, the PPG signal is known to be fragile and easily corrupted by motion. In most cases, the noise falls within the frequency band of the physiological signal of interest, rendering linear filtering ineffective. Algorithms have been developed to reduce the sensitivity of PPG signals to artifacts commonly encountered in clinical environments and during controlled or moderate motion [1-8]. A non-contact, laser-based remote PPG system has been proposed [9], with positive results obtained in clinical applications. PPG has been also applied to the evaluation of severe exercise during the incremental maximal exercise test (IMET) on a cycle ergometer, where several parameters were monitored, among them oxygen saturation [10,11] and heart rate (HR) [12-15]. Active research efforts are beginning to demonstrate that PPG has utility beyond oxygen saturation and HR determination. For instance, the conditions required for a correct utilization of PPG HR variability (PPGV) have been recently studied [14]; the authors concluded that PPGV cannot be used during slow walking and cycling. Future trends are being heavily influenced by modern digital signal processing [16]. New commercial developments of the PPG-based prototype [17] suggest that a new interest is growing about PPG techniques for HR tracking.

Under severe exercise conditions, motion artifacts present a challenge for PPG analysis. Sensor and packaging designs can help reduce the impact of motion disturbance, but they are

* Corresponding author: Romano Giannetti
Tel: +34-655-092991; Fax: +34-91-5411542
E-mail: romano@dea.icai.upcomillas.es

rarely sufficient for noise removal. Advanced signal processing techniques are often required to remove motion artifacts under vigorous activity. Several techniques have been developed to deal with such artifacts, such as processing context information from additional on-body sensors and light sources [18,19] and adaptive noise cancellation using accelerometers as a noise reference [20]. Very recently, heart rate measurements have been taken using a PPG sensor integrated with an adaptive noise cancellation device during common physical activity, from walking to running at up to 8 km/h [21].

Spectral analysis, such as the traditional fast Fourier transform (FFT), is a simple and inexpensive tool for separating motion artifacts and cardiac physiologic spectra. However, techniques based on spectral analysis are not applicable to spectra that contain frequency bands close to each other. More sophisticated algorithms have shown significant improvement over FFT for measurements obtained during finger bending manoeuvres [2]. The present study develops and validates an algorithm for HR tracking from PPG signals during IMETs on a treadmill. The real-time motion discriminator algorithm, which is based on linear filtering and frequency-domain and heuristic analysis, allows the HR to be extracted from a PPG signal in the presence of severe motion artifacts. The basis of the heart beat frequency selection is the observed high harmonic content of movement artifact signals with respect to the PPG-derived heartbeat. A preliminary version of the algorithm has been published, with off-line analysis of both the HR and oxygen saturation [22]. The present study optimizes the algorithm parameters for HR tracking and uses real-time probes. The results are compared to those obtained using a standard technique. The physiological aspects of training are outside the scope of this work. The custom PPG measurement system is based on a transmittance pulse oximetry system [23].

2. Materials and methods

Athletes of both genders with a training schedule exceeding an average of 7 hours per week participated in IMETs performed on a treadmill during individual sessions. The data presented and analyzed in this work correspond only to the HR measurements; other parameters such as oxygen consumption, carbon dioxide production, and blood pressure are outside the scope of the present work. The tests were conducted at the Laboratory of Effort Physiology of the Professional School of Physical Education of Complutense University, Madrid, Spain, after approval by the local Research Ethical Committee. All the athletes gave their written informed consent.

2.1. Subjects

One group of athletes (T1) comprised 10 white male endurance runners, who were tested on a treadmill ergometer. A second group of athletes (T2) comprised 20 athletes (9 white females (one runner, 8 soccer players), 9 white male runners, and 2 black female basketball players), who were tested on the treadmill. The development of the algorithm and the optimization of its parameters were carried out after the data was recorded for the T1 group. The algorithm performance was evaluated in real-time during IMETs of the T2 group.

2.2. Experimental protocol

The HR was monitored via electrocardiography (ECG, Burdick, Inc, Model Quest Exercise Stress System) and used as a reference. The PPG sensor was placed on one finger with special care taken to avoid excessive compression of the tissues. Arm movement was not restricted.

A complete maximal exercise test was performed, followed by active recovery. The protocol [26] for the treadmill ergometer (HP Cosmos QUASAR 4.0) test was as follows. After one minute at rest (in order to record basal values) and after warming up, the athlete began to walk at 6 km/h on a 1 % slope for 2 min. The athlete then started the effort phase, running at 8 km/h on a 1 % slope. During the effort phase, the speed was increased every 2 min by 2 km/h. The maximum speed achieved varied among individuals. When the speed of 14 km/h (for female athletes) or 16 km/h (for male athletes) was reached, the slope was increased to 3 %. Afterwards, the slope was maintained at a constant level, while the speed was increased every 2 min by 2 km/h until the athlete was unable to continue. Then, the athlete held onto the protective bars and jumped off the treadmill. Active recovery was performed for 2 min at 8 km/h with a slope of 0 %. At the beginning of the exercise, when walking at 6 km/h, most of the athletes were adapting to walking on the treadmill and in many cases they held onto the protection bars, which could have altered the position of the finger sensor. This issue will be discussed later. The ECG system computed the HR every 10 s, averaging the last eight heart beats. At different stages of the test, once the athlete was running at a fixed speed and slope, the step rate (SR, in steps/min) was obtained by counting manually the number of steps in a 10-s interval. The one-foot step was derived from the SR. During running, the movements of the legs and arms are synchronous; hereafter, we will call SR both the step and the arm-waving rates, which are inseparable issues.

2.3. PPG measurement system

The custom PPG measurement system comprises a transmittance sensor, sensor electronics, a data acquisition board (DAQ), and a laptop computer. The emitter of the sensor is a laser diode with a peak wavelength in the near-infrared region (850 nm), which matches that of usual PPG wavelengths. Three p-i-n silicon photodiodes (BPW34) connected in parallel and aligned to increase the detection area were used as a photodetector. The emitter driver, further amplification and filtering stages, and timing and sample-and-hold (S&H) circuits were connected to the fingertip probe by cables. The photodetector signal was filtered by an anti-aliasing analog low-pass filter at 300 Hz, then sampled with a high-speed S&H circuit and fed into the analog inputs of a 12-bit DAQ (DAQ1200, National Instruments) to be digitized at 1000 samples/s (Sa/s). Finally, a ten-sample moving average followed by a ten-to-one decimation was performed, which resulted in 100-Sa/s PPG data.

The experimental system was connected to a Pentium-class laptop computer, which stored files containing the PPG 100-Sa/s data (in volts). Subsequent stages of the signal processing were carried out digitally. The post-processing of the stored signals from the T1 group was made off-line. Once the algorithm had been developed, it was implemented as a virtual instrument using Lab Windows CVI™. The validation was conducted on-line during the IMETs of the T2 group.

2.4. Statistical analysis

Comparisons of HR between the two measurement methods (PPG and ECG) were performed using the Student *t*-test for paired data. Linear Pearson's correlation analysis was used to correlate quantitative variables, as an indicator of the two methods evolving in parallel. The Bland-Altman (B - A) method was used to compare the values of HR obtained by the proposed PPG technique and the standard ECG. Hereafter, PR and ER are used for the PPG- and ECG-derived HR values, respectively. The differences between the values (PR - ER, in beats per minute, bpm) are plotted against the average value of the two methods ((PR + ER)/2, in bpm). The results are presented as the actual differences (PR - ER) and mean difference or bias (M) between the two methods. The standard deviation (std) value is used to calculate the limits of agreement (LAs), computed as M ± 1.96 std, providing an interval within which 95 % of the differences between the measurements by the two methods are expected to lie. The standard errors and confidence intervals (CIs) were determined for the mean bias and for the upper and lower LAs. Agreement is defined as a difference (PR - ER) of within ± 10 bpm [3,12].

3. Algorithm design

Figure 1 shows FFT spectra corresponding to PPG signals recorded over 10-s time intervals at rest (time interval ending at 25.96 s, grey) and when running at 12 km/h (time interval ending at 435.56 s, dark), during the test performed on a treadmill by male athlete A of the T1 group. The sampling processes of ECG- and PPG-based measurements were not synchronized; the ECG and PPG readings were matched by pairing the nearest values in time.

FFT spectra are plotted as the power density (V^2/Hz) versus frequency in the range of 40-240 bpm, corresponding to 0.66 to 4 Hz. For each time interval, the spectra show three main peaks, which are respectively labeled according to their power as P1 (the strongest peak), P2, and P3 (the weakest peak). For data obtained at rest, the spectrum shows P1 at a frequency coinciding with the recorded ER value for the given time interval, and two tiny peaks corresponding to the harmonics of P1 at frequencies 2 x P1 and 3 x P1. For data obtained during running at 12 km/h, the spectrum exhibits a different peak distribution; the recorded ER value coincides with the P2 frequency. The one-foot-step rate coincides with the P3 frequency, whereas P1 is the harmonic 2 x P3 corresponding to the steps per minute (SR).

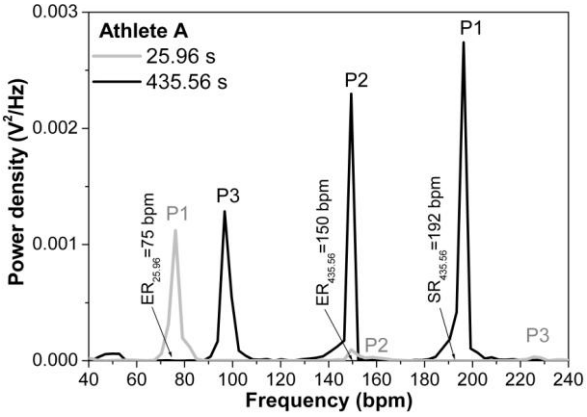


Figure 1. FFT spectra corresponding to PPG signals recorded over 10-s time intervals (grey: at rest, dark: during exercise, running speed: 12 km/h); the three main peaks P1, P2, and P3 are labeled according to decreasing power density; the values of ER (ECG-derived heart-rate) and SR (full-step rate) for the given time intervals are also shown.

Figure 2 shows the time evolution of peaks P1, P2, and P3 (frequency vs. time) during the IMET of athlete A, from rest to maximal effort; time evolution of ER and the SR are also shown, as well as the time interval of walking or running and the stop time. There are three well-defined lines with a frequency that increases with time, two of which evolve concomitantly, with the third increasing steeply with time; the strongest peak is not associated with a given line, but instead jumps from one line to another. The steep line evolves close to the ER values, and thus these frequencies are identified as PR. Heartbeat harmonics appear sometimes as a noise level at rest and at the beginning of the IMET, and move out of the plotted range of frequency as the heartbeat increases. The upper line is always at twice the frequency of that of the lower one. Furthermore, the upper line evolves close to the SR, and thus these frequencies are identified as the SR, a harmonic of the one-step rate, which is identified as the lower line. In brief, the time evolution of the spectra peaks defines a pattern composed of three branches, one ascending central branch corresponding to the PR and two branches related to the periodic movement of feet and arms. The fundamental frequency and its harmonics, related to the one-foot and full step rates, evolve during exercise at a rate different from that of the heartbeat.

Results obtained from different athletes show that the evolution of the peaks always follows the ER and SR evolutions. In some cases, the heartbeat increases much more steeply than the SR and the corresponding line intercepts one of the motion branches at a certain moment of the exercise, and there are time intervals where the frequency domains of the PR signal and artifacts overlap.

At low physical activity, the PR value can be obtained by simply finding the highest peak in the spectrum. The respiratory rate was detected several times during exercise, at frequencies around 40-50 min^{-1}, but it has low power content relative to those of the heartbeat and movement artifact signals.

The algorithm design is based on these observations. The algorithm has a harmonic discriminator that takes into account

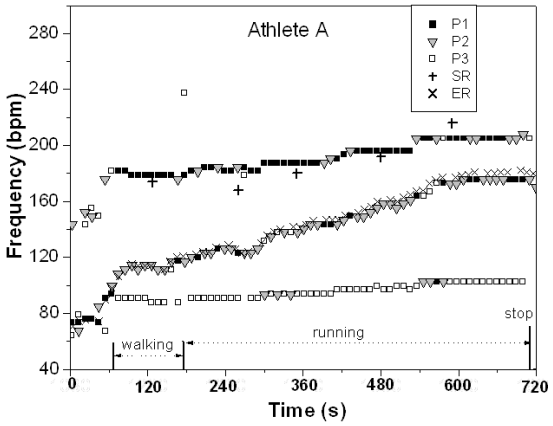


Figure 2. Frequency-time evolution of the main peaks (P1, P2, P3) over the whole test; the evolutions of ER (ECG-derived heart-rate) and SR (full-step rate) are also plotted.

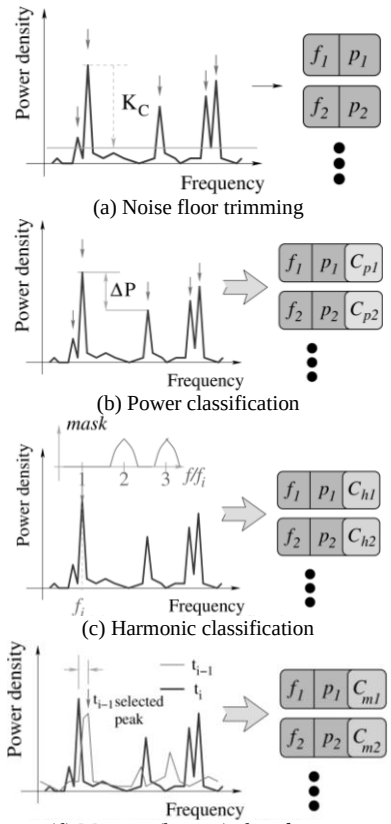
that the movement artifact contribution has higher harmonic content than that of the heart rate signal. The developed detection algorithm is mainly based on an analysis of the relative strength of the harmonic content of the PPG signal. A heuristic algorithm is used to track the heartbeat to take into account the past history of the heartbeat itself when the PR signal and the artifact signal overlap.

4. Development of the algorithm

The proposed algorithm was developed by analyzing the PPG data recorded from the T1 group and comparing them to those obtained using ECG.

The first step is the linear filtering of 100-Sa/s PPG signal $E(t)$ via band-pass Bessel filtering (cut-off frequencies f_1 and f_2, order $n = 6$) to get the pulsating component ($E_{ac}(t)$) and to suppress high-frequency noise and ripple. Cut-off frequencies were chosen in the ranges 0.1~0.3 Hz for f_1 and 5~30 Hz for f_2. After filtering, a frequency-domain analysis was performed by applying the FFT to a 10-s, rectangular-shaped, sliding window of the aforementioned data signal, with an overlap of about 75 %, delivering spectrogram data $F(t_s)$, expressed in terms of power density (p_i) versus frequency (f_i). In this step, the electrical PPG signal is transformed into a sequence of FFT spectra, in which each spectrum is computed every 2.56 s and contains data from a time interval of 10.24 s, which gives a frequency resolution of 0.58 bpm. The instant in time t_s to which the spectrum is assigned corresponds to the end of the interval. A peak search algorithm is applied to the stream of spectra to select peaks that are above the noise floor. In this step, the Gibbs phenomenon produces lateral lobes that, in theory, could create phantom peaks due to side-lobe leakage; this is taken into account in the peak search algorithm, which never selects peaks from adjacent FFT bins. The authors explored more complex windowing functions and even more complex approaches [27], but the relatively wide window and the experimental tests show that leakage, although present, did not invalidate the effectiveness of the algorithm.

Figure 3 schematically shows the heuristic algorithm. The peak search algorithm eliminates all the peaks that are lower than the noise floor of $F(t_s)$, (Fig. 3(a)), and selects those peaks that could be the most relevant. p_{Max} is the highest power value of $F(t_s)$ and K_C is the cut-off parameter of the algorithm. Values for K_C were tuned in the range of 50-400 by analyzing and comparing the PPG and ECG data recorded from the T1 group; $K_C = 100$ was adopted as the default value, meaning that the cut-off of minor peaks occurs when amplitudes are approximately 7 to 20 times lower than the highest peak. To simplify the computational cost of the algorithm, the peak search selects only the three main peaks when implemented on the PPG system.



(a) Noise floor trimming
(b) Power classification
(c) Harmonic classification
(d) Memory (history) classification

Figure 3. Description of the heuristic algorithm. After the band-pass linear filtering, (a) a set of significant peaks is computed, represented by a list of frequency-peak values (power density PD). The selection is made by listing all the peaks that are higher than a noise floor, computed by dividing the highest peak by parameter K_c. The values are then fed into the three mechanisms that assign a reliability value to each pair. (b) Peaks are classified according to their relative amplitude. (c) Their harmonic relationship is computed (see text). (d) Historic record is taken into account, assigning higher reliability values to peaks that are near the previously selected value.

Next, the algorithm chooses the best candidate $P(t_s, f, p)$ as the heartbeat PR, with a reliability factor $C(t_s)$ which expresses, in the range (0-1), the degree of confidence that $P(t_s)$ is the HR value, not a movement artifact. To achieve this, the three peaks are fed into three classification engines: energy, harmonic, and historic classification. The task of the classification engines is to assign each peak the reliability factor $C(t_s)$ defined above using various criteria. The reliability factor saturates to 0 or 1, respectively, when the computed value exceeds these limits.

Figs. 4(b) and (c) show a flow diagram of the heuristic algorithm.

The power engine (Fig. 3(b)) compares each p_i value to the highest value (p_{Max}) of $F(t_s)$, and assigns to each (f_i, p_i) pair reliability factor C_{pi}, which expresses the power content of the peak with respect all others peaks. C_{pi} is computed in the range (0-1) as the amplitude of the peak relative to the next highest peak.

The harmonic engine (Fig. 3(c)) computes the probability that a peak is part of a signal with high harmonic content, assigning to each of the peaks a second reliability factor C_{hi}. The engine compares pairs of frequencies. For each peak, a sliding mask is superposed on the spectrum. The mask is formed by parabolic segments centered on the frequency multiples of the peak, with a relative width controlled by parameter K_H, which modulates the tolerance accepted in the frequency comparison. C_{hi} is computed in the range (0-1) based on how much of the spectrum energy enters the mask; its value is low for high harmonic content and high for low harmonic content. To reduce the computational burden, the algorithm computes how near a frequency is to being double of another frequency on the list, limiting the analysis to the second harmonic, when implemented on the PPG system.

The parameter K_H is one of the main parameters of the algorithm. K_H was tuned in the range of 5-50, and was optimized for a value of 12. For $K_H < 5$, the mask is too wide and all the peaks have harmonic content. For $K_H > 50$, the mask is too narrow and the algorithm never finds any harmonics. When a peak is found to have a strong second-harmonic counterpart in the spectrum, it is automatically invalidated for the selection process, even if it has a very high confidence value assigned by the power engine. At rest, the main peak corresponds to the PR, but very small power harmonics can be present; the value of K_C in the peak search algorithm eliminates the harmonics in the peaks entering the heuristic algorithm.

Finally, the memory or historic engine (Fig. 3(d)) compares frequency $f_i(t_s)$ of peaks P_i (t_s) of $F(t_s)$, with frequency $f_{i'}$ (t_{s-1}) of peak $P(t_{s-1})$ selected as the PR in the previous spectrogram $F(t_{s-1})$. A third reliability factor C_{mi} is computed by taking into account the proximity of the frequency f_i (t_s) to $f_{i'}(t_{s-1})$ as well as the former value of reliability factor C_{mi}. C_{mi} is zero for all the peaks except that whose frequency is closest to the $f_{i'}(t_{s-1})$ selected as the PR in the previous spectrogram; C_{mi} is computed in the range of 0-1, so a high value means high proximity to the previous PR peak.

Figure 4 shows a detailed flowchart of the algorithm. Fig. 4(a) shows the general data processing for selecting the three main peaks of spectrogram $F(t_s)$. Fig. 4(b) shows flowcharts of the power and harmonic engines. Fig. 4(c) shows a flowchart of the memory engine and the final selection step.

For each spectrogram $F(t_s)$, the heuristic algorithm compares all the peaks and selects one of them according to the assigned reliability factors. The selection is made on the basis of the maximum reliability factor value, i.e., a high C_{pi} (amplitude of the peak), a high C_{hi} (low harmonic content), and a high C_{mi} (high proximity to the previous PR peak). The chosen peak is assigned to the PR(t_s) value, and the overall

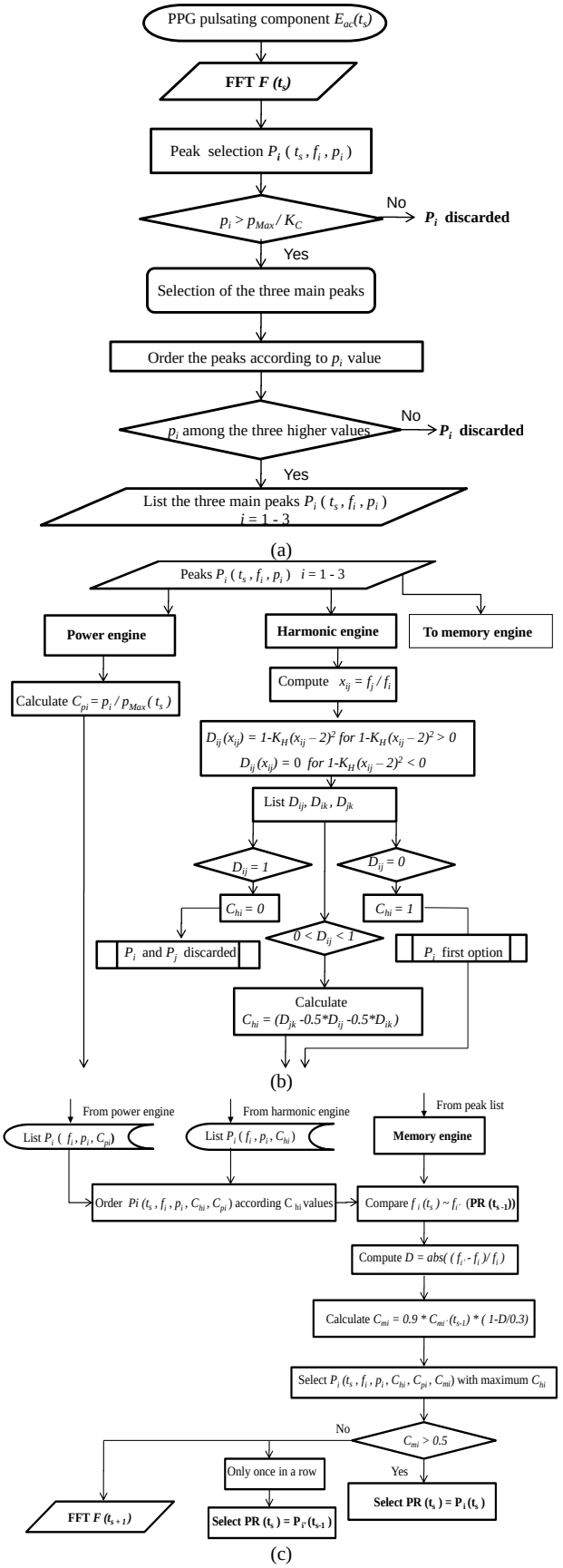


Figure 4. Flowchart of the proposed algorithm. (a) General data processing, (b) power and harmonic engines, and (c) memory engine and the final selection step.

confidence index $C(t_s)$ is set to the highest reliability index.

If no peak obtains a reliability factor greater than 0.5 (the highest peak could have been marked as invalid by the harmonic content analysis, whilst the other peaks could have been determined to be insufficient by any of the classification techniques), the algorithm simply jumps, giving the PR value selected for the previous spectrogram, and it restarts with the next one. This step is permitted just once in a row to avoid the possibility of becoming stuck at a wrong value; i.e., at the next spectrogram, if the reliability factor is again lower than 0.5, the selection is made as described above, even with bad reliability factors. This step allows the algorithm to ignore especially bad points without corrupting the memory part of the algorithm.

Notice that the memory engine assigns to the peak selected in each spectrogram a reliability factor smaller than the one assigned in the previous spectrogram. In this way, if the peak chosen as the PR continues to be chosen by the memory algorithm, the reliability factor decreases. Eventually this approach forces the heuristic algorithm to select the PR peak using a different criterion, so that the one in memory does not overtake others for a long time. This is very important in the case of the PR peak being very close to a motion peak, because when the algorithm selects a high-power peak, both the power engine and the memory engine will tend to follow this high peak. If the reliability factor C_{mi} decreases continuously, the memory engine will not find a solution and the harmonic engine will search for a new peak.

The parameters K_C and K_H were optimized by analyzing the recorded data of the T1 group. The highest computational cost of the algorithm was found in the generation of the FFT spectra, for which the standard FFT subroutines offered by standard software were used. As mentioned before, in order to reduce the computational cost, some simplifications were used: a) at the peak search stage, the number of selected peaks was reduced to the three with the highest power density; b) in the harmonic analysis, only the second harmonic is taken into account, given that higher order harmonics are normally outside the measurement range, and c) the harmonic analysis compares only frequencies, independently of their power density.

Once the IMET starts, the harmonic engine is the principal component of the algorithm, allowing it to remove movement artifacts. When the PR peak is very close to a motion peak, the memory engine helps the algorithm to follow the correct peak by forcing the harmonic engine to act quickly. When one peak is selected, its frequency is then assumed to be the PR, and its value is displayed on the screen.

4.1. Signal-to-noise ratio

The signal-to-noise ratio (SNR) of the selected PR data was estimated. The selected PR peak was used as the signal power density and the sum of discarded peaks was used as the noise power density. The procedure was repeated for all the PR peaks selected in the exercise and for the T1 group. The highest 10 % and the lowest 10 % SNR values were discarded and the remaining data were averaged. The SNR estimated in such a way does not take into account the ground floor noise; it just discards peaks. The noisiest points were avoided, those over which the algorithm jumped, so that the resulting figures are likely an overestimation of the real values. Nevertheless, it is a valuable estimation of the SNR of the selected PR data. The values of SNR were calculated, both as averages over the whole IMET and over the last 20 % of the IMET time. From the analysis of the T1 group data, the average values for the full test and the last 20 % are -1.6 dB and -2.5 dB, respectively. The algorithm can thus identify the correct PR even with an SNR of as low as -2.5 dB.

5. Results

5.1. Validation of the algorithm using recorded data of T2 group

To validate the algorithm, each PR value was compared with the temporally nearest ER, within intervals in which the athletes were running, i.e., from 8 km/h until the maximal effort. The initial periods were not included due to dissimilarities in the athlete actions. At 6 km/h, most of the athletes were adapting to walking on the treadmill and in many cases they held onto the protection bars, which could have altered the position of the finger sensor and therefore distort the measurement.

The regression lines of ER and PR measurements were calculated for a number of athletes (A = 10) and a number of (PR, ER) pairs (N = 477), with a correlation R = 0.994 and $p < 0.0001$. This correlation is an indication that the two parameters evolve in parallel. The histogram of differences between the two methods confirmed the normal distribution. The B-A plot of the difference (PR-ER) against the average value given by the two methods ((PR + ER)/2) is shown in Fig. 5. The bias is M = -1.10 bpm, the standard deviation is

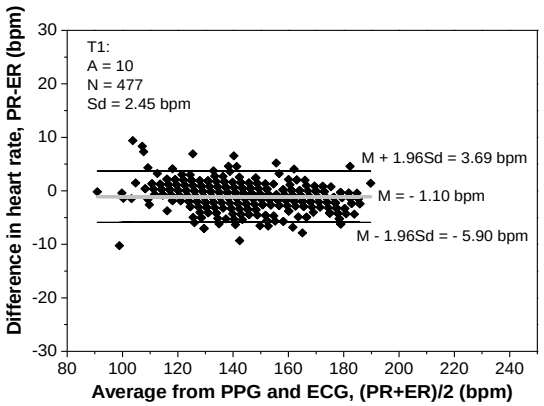


Figure 5. Bland-Altman plot of differences in recorded heart rate measurements versus average values for the T1 group, with bias value (M) and limits of agreement (M ± 1.96 std). Std is the standard deviation, A is the number of athletes, and N is the number of pairs (PR, ER).

std = 2.45 bpm, and the LAs are in the range of -5.90 bpm to + 3.69 bpm within the *a priori* set value of ± 10 bpm. The 95 % confidence intervals are -1.32 to -0.88 bpm for the bias, -6.28 to -5.52 bpm for the lower LA (LLA), and 3.31 to 4.07 bpm for the upper LA (ULA). All these intervals are

reasonably narrow; the width of the LAs is 9.59 bpm and more than 95 % of all (PR-ER) differences fall within the LA. The plot reveals that there is no relationship between the differences and the bias. Plotting the differences as percentages is useful when there is an increase in the variability of the differences as the magnitude of the measurement increases. With percentages, the values obtained are M = -0.7 %, LLA = -4.2 %, and ULA = 2.8 %.

5.2. Real-time HR tracking during IMET on treadmill

The algorithm was implemented in the experimental PPG system and validated using real-time measurements obtained during treadmill ergometer IMETs for the T2 group. The SNR values are -0.1 dB for the full test and -2.1 dB for the last 20 %. Despite the low SNR, the algorithm follows the correct PR value.

The plot of PR versus ER for the T2 group gives a correlation R = 0.992 and $p < 0.0001$; the number of athletes is A = 20 and the number of (PR,ER) pairs is N = 1061. The normal distribution of differences between the two methods was confirmed by their histograms. The (PR,ER) pairs obtained are shown in the B-A plots in Fig. 6, as the differences (PR-ER) versus the mean value (M) of the differences; the value M = -0.70 bpm indicates a low negative bias and the standard deviation is std = 2.92 bpm. The width of the LAs is 11.42 bpm, between -6.41 bpm (LLA) and + 5.01 bpm (ULA), within the *a priori* set value of ± 10 bpm. More than 95 % of all the differences obtained for the athletes tested on the treadmill fall within ± 6.5 bpm, even for HRs as high as about 200 bpm. The 95 % confidence intervals are -0.88 to -0.53 bpm for bias, -6.72 to -6.12 bpm for LLA, and 4.71 to 5.31 bpm for ULA. For the T2 group, the values expressed as percentages of the average between the two techniques are M = -0.4%, LLA = -4.2 %, and ULA = 3.4 %.

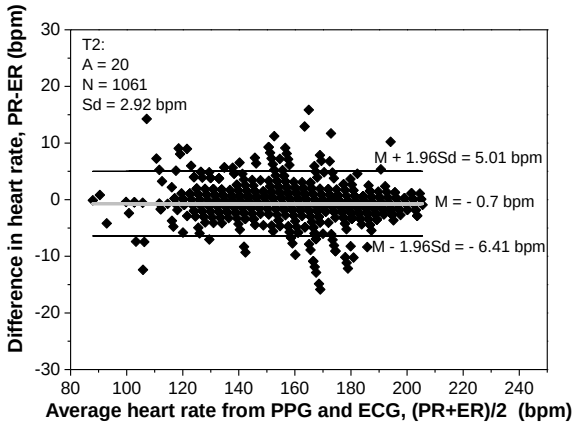


Figure 6. Bland-Altman plot of differences in real-time heart rate measurements versus average values for the T2 group, with bias value (M) and limits of agreement (M ± 1.96 std). Std is the standard deviation, A is the number of athletes, and N is the number of pairs (PR,ER).

6. Discussion

The algorithm for the automatic detection of the HR from PPG signals is efficient, despite the small SNR. A comparison to ECG measurements, the reference standard, shows very high agreement. The low negative bias obtained for the recorded data of the T1 group may be caused by slight differences between beat-to-beat averaging window used in the ECG monitor and the PPG system; if a subject's HR is changing rapidly, the different averaging window can influence the results and amplify the calculation errors. Moreover, the small bias might be due to possible errors during data collection from a variety of sources in which time was recorded by independent and unsynchronized clocks [28] (independent computers for ECG and PPG). The algorithm selects PR values that show high agreement with ER measurements during activities such as the IMET. The narrow width of the LA, narrower than ± 10 bpm, is acceptable for interchangeability between the PR and the ER. It should be noted that parameter optimization, carried out on the recorded data, was performed in a "first step" way. Optimization of the algorithm's parameters with an iterative mathematical program, instead of the approximate selection used in this work, should give better values, but it is not the aim of the present work.

Overall, the system can track the HR during treadmill exercise. The system also shows good results for the IMET on a cycle ergometer [29]. Compared to traditional FFT methods, the proposed algorithm shows significant improvements in the removal of motion artifacts from cardiac physiologic spectra. The accuracy of the measurements are comparable to the mean absolute error obtained using time-frequency techniques based on the smoothed Wigner-Ville distribution [2,8], where six subjects participated in a study realizing four kinds of motion, each one at a time, frequency, and intensity fixed, which are less demanding conditions that the ones presented in this study. Compared to results from a PPG sensor integrated with an adaptive noise cancellation device, the proposed method obtained LA values of between -6.41 bpm and + 5.01 bpm (or -4.2 % and 3.4 %), which are better than the values (-21.15% and 20.52 %) obtained for subjects running at 8 km/h [21], again in less demanding conditions that the IMET.

The algorithm was developed by analyzing the recorded data of a group of white men, all of them runners; it was evaluated on a very dissimilar group, composed of men and women, black and white. The algorithm parameters were the same for all subjects; the accuracy of the technique might be higher if the parameters are tailored for each subject. In its current form, the algorithm is suitable for both men and woman of various races.

The basic premise of the algorithm is that the motion artifacts created in the PPG signal due to running have a high harmonic content with respect to the PPG-derived HR. The use of graded treadmill running (or cycling) in the IMET may bias the motion artifacts in the PPG signal by making them regular and rhythmic in nature. Although this seems to be a limitation of the PPG-based technique, athletes run with a regular and stable movement. A significant advancement of the algorithm utility would be to test its ability to measure HR during road running or running on uneven ground. Another limitation is that the PPG system is a fingertip probe; the algorithm can be implemented in other PPG systems, such as earlobe probes or

ring sensors, provided that a rhythmic movement of the body and/or extremities has high harmonic content. Other error sources, like the movement of the sensor on the finger, do not cause important signals in the FFT analysis when the arms have free movement; otherwise, ECG-derived HR values can also give some errors. When the athlete is sweating or when the fixation of the ECG electrodes is poor, readings may contain errors.

7. Conclusion

This work proposed an algorithm for eliminating signals caused by motion artifacts that limit the use of PPG in quantifying key cardiovascular variables such as the HR during exercise. The proposed algorithm converts PPG signals collected during treadmill exercise into an HR value. The algorithm identifies three frequencies of interest from the PPG signal and uses decision-making logic to identify the two harmonic frequencies that most likely represent motion artifacts, leaving one frequency to represent the PPG-derived HR. The algorithm was tested using a heterogeneous athletic population transitioning from rest through a graded maximal intensity treadmill test. Electrocardiography was used to validate the PPG-derived HRs. The proposed algorithm allows for significant advancement in the area of biosensor development as PPG technology is ideally suited to be incorporated into wearable devices.

Acknowledgments

This work was supported by the Spanish Consejo Superior de Deportes under CSD grants 03/UPB10/06, 01/UPB10/07, and 01/UPB10/08. S. M. López-Silva was supported by the CSIC I3P Program financed by the Fondo Social Europeo. The authors gratefully acknowledge the assistance of Dr. Isabel Gómez Caridad in revising the manuscript.

References

[1] A. A. R. Kamal, J. B. Harness, G. Irving and A. J. Mearns, "Skin photoplethysmography: a review," *Comput. Methods Programs Biomed.*, 28: 257-269, 1989.

[2] J. Allen, "Photoplethysmography and its applications in clinical physiological measurement," *Physiol. Meas.*, 28: R1-R39, 2007.

[3] J. Townshend, B. J. Taylor, B. Galland and S. Williams, "Comparison of new generation motion-resistant pulse oximeters," *J. Paediatr. Child Health*, 42: 359-365, 2006.

[4] B. S. Kim and S. K. Yoo, "Motion artifact reduction in photoplethysmography using independent component analysis," *IEEE Trans. Biomed. Eng.*, 53: 566-568, 2006.

[5] M. J. Hayes and P. R. Smith, "A new method for pulse oximetry possessing inherent insensitivity to artifact," *IEEE Trans. Biomed. Eng.*, 48: 452-461, 2001.

[6] S. Rhee, B. H Yang and H. H. Asada, "Artifact-resistant power-efficient design of finger-ring plethysmographic sensors," *IEEE Trans. Biomed. Eng.*, 48: 795-804, 2001.

[7] Q. Li, R. G Mark and G. D. Clifford, "Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator," *Biomed. Eng. Online*, 8: 13, 2009.

[8] Y. S. Yan, C. C. Poon and Y. T. Zhang, "Reduction of motion artifact in pulse oximetry by smoothed pseudo Wigner-Ville distribution," *J. Neuroeng. Rehabil.*, 2: 3, 2005.

[9] P. Shi, V. A. Peris, A. Echiadis, J. Zheng, Y. Zhu, P. Y. S. Cheang and S. Hu, "Non-contact reflection photoplethysmography towards effective human physiological monitoring," *J. Med. Biol. Eng.*, 30: 161-167, 2010.

[10] H. Benoit, F. Costes, L. Feasson, J. R. Lacour, F. Roche, C. Denis, A. Geyssant and J. C. Barthélémy, "Accuracy of pulse oximetry during intense exercise under severe hypoxic conditions," *Eur. J. Appl. Physiol.*, 76: 260-263, 1997.

[11] Y. Yamaya, H. J. Bogaard, P. D. Wagner, K. Niizeki and S. R. Hopkins, "Validity of pulse oximetry during maximal exercise in normoxia, hypoxia and hiperoxia," *J. Appl. Physiol.*, 92: 162-168, 2002.

[12] K. Nakajima, T. Tamura and H. Miike, "Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique," *Med. Eng. Phys.*, 18: 365-372, 1996.

[13] Y. Iyriboz, S. Powers, J. Morrow, D. Ayers and G. Landry, "Accuracy of pulse oximeters in estimating heart rate at rest and during exercise," *Br. J. Sports Med.*, 25: 162-164, 1991.

[14] K. Charlot, J. Cornolo, J. V. Brugniaux, J. P. Richalet and A. Pichon, "Interchangeability between heart rate and photo-plethysmography variabilities during sympathetic stimulations," *Physiol. Meas.*, 30: 1357-1369, 2009.

[15] S. H. Liu, K. M. Chang and T. H. Fu, "Heart rate extraction from photoplethysmogram on fuzzy logic discriminator," *Eng. Appl. Artif. Intell.*, 23: 968-977, 2010.

[16] K. H. Shelley, "Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate," *Anesth. Analg.*, 105: S31-S36, 2007.

[17] Centre Suisse d'Electronique et Microtechnique (CSEM), Swiss Center for Electronics and Microtechnology, 2008, Available: http//www.csem.ch

[18] H. H. Asada, P. Shaltis, A. Reisner, S. Rhee and R. C. Hutchinson, "Mobile monitoring with wearable photo- plethysmographic biosensors," *IEEE Eng. Med. Biol. Mag.*, 22: 28-40, 2003.

[19] G. Cennini, J. Arguel, K. Aksit and A. van Leest, "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction," *Opt. Express*, 18: 4867-4875, 2010.

[20] H. Han, M. J. Kim and J. Kim, "Development of real-time motion artifact reduction algorithm for a wearable photo-plethysmography," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 1538-1541, 2007.

[21] M. Z. Poh, N. C. Swenson and R. W. Picard, "Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography," *IEEE Trans. Inf. Technol. Biomed.*, 14: 786-794, 2010.

[22] R. Giannetti, J. P. Silveira, M. L. Dotor, D. Golmayo, P. Martín, A. Bilbao, F. Miguel-Tobal and S. M. López-Silva, "Oxygen saturation measurements in athletes attaining maximal exertion conditions," *Proc. IEEE Instrum. Meas. Technol. Conf. (IMTC)*, 1: 740-744, 2004.

[23] S. M. López-Silva, M. L. Dotor and J. P. Silveira, "NIR transmittance pulse oximetry with laser diodes," *J. Biomed. Opt.*, 8: 525-533, 2003.

[24] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods Med. Res.*, 8: 135-160, 1999.

[25] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Int. J. Nurs. Stud.*, 47: 931-936, 2010.

[26] M. Rabadán and J. C. Segovia, "Pruebas de esfuerzo directas," in: J. C. Segovia, J. C. Legido and F. J. López-Silvarrey (Eds.), *Manual Valoración Funcional: Aspectos Clínicos y Fisiológicos*, Madrid: Elsevier, 267-278, 2008.

[27] C. Pan, "Gibbs phenomenon removal and digital filtering directly through the fast Fourier transform", *IEEE Trans. Signal Process.*, 49: 444-448, 2001.

[28] G. D. Clifford, W. J. Long, G. B. Moody and P. Szolovits, "Robust parameter extraction for decision support using multimodal intensive care data," *Philos. Trans. R. Soc. A*, 367: 411-429, 2009.

[29] S. M. López-Silva, M. L. Dotor, J. P. Silveira, R. Giannetti, F. Miguel-Tobal, A. Bilbao, M. Galindo and P. Martín, "Pulse rate measurement from transmittance photoplethysmography in cycle ergometer test," *Proc. IEEE Instrum. Meas. Technol. Conf. (IMTC)*, 1039-1043, 2010.