Heuristic Algorithm for Photoplethysmography Heart Rate Tracking during Athletes Maximal Exercise Test

Sonnia María López-Silva¹  Romano Giannetti²*  María Luisa Dotor¹  Juan Pedro Silveira¹
Dolores Golmayo³  Francisco Miguel-Tobal¹  Amaya Bilbao⁴  Mercedes Galindo⁴  Pilar Martín-Escudero⁴

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.

* Corresponding author: Prof. Romano Giannetti  
Vice-Rector for International Relations  
Universidad Pontificia Comillas  
C/ Alberto Aguilera 23, 28015 Madrid (Spain)  
Tel +34-915422800; Fax +34-915411542  
E-mail: romano@dea.icai.upcomillas.es

RUNNING TITLE: PPG Heart Rate Tracking in Athletes
Abstract

Photoplethysmography (PPG) is a non-invasive optical technique that enables to quantify the arterial blood pulse rate. Signal corruption by motion artifacts is a primary limitation in the practical accuracy and applicability of instruments for monitoring pulse rate during intense physical exercise. Here we develop and validate an algorithm, based on linear filtering, frequency domain and heuristic analysis, capable to extract and display the heart rate from PPG signal in the presence of severe motion artifacts. The basis of the frequency selection is the observed high harmonic content of movement artifact signals with respect to the PPG-derived heartbeat. The algorithm has been implemented in an experimental PPG measurement device and has been developed by analyzing a set of PPG data recorded from a first group of athletes exercise in treadmill. After that, an extensive set of tests has been carried out on real time during maximal exercise tests in treadmill in order to validate the instrument by comparison with a reference electrocardiography measurement system. We have used the Bland and Altman method to compare and evaluate the PPG signal. The accuracy of the heartbeat measurement is better than ± 6.5 bpm (≤ 4.2 %) even under maximal exercise conditions.

Keywords: Biomedical signal processing, Photoplethysmography (PPG), Motion artifact, Detection algorithm, Heart rate
1. Introduction

Photoplethysmography (PPG) is a technique offering a simple, useful and compact way to measure several clinical parameters such as oxygen saturation, blood pressure and cardiac output [1,2]. PPG is especially suitable for wearable sensing, that could play an important role in many 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 modifiable by motion. In most cases, the noise falls within the same frequency band as the physiological signal of interest, rendering linear filtering ineffective. Research carried on motion artifacts in measured PPG signals has reduced the sensitivity to the artifacts common in the clinical environments, as well as in situations of controlled or moderate motion [1-8]. A non-contact, laser emitter remote PPG system, has been proposed [9] and positively evaluated for clinical applications. PPG has been also applied to the evaluation of more severe exercise during incremental maximal exercise test (IMET) on cycle-ergometer, where several parameters are monitored, among them oxygen saturation [10,11] and heart rate [12-15]. Active research efforts are beginning to demonstrate a utility beyond oxygen saturation and heart rate (HR) determination; for instance, the conditions for a correct utilization of PPG HR variability (PPGV) have been recently studied [14] and the authors concluded that PPGV cannot be used during slow walk and cycling exercise. Future trends are being heavily influenced by modern digital signal processing [16]. New commercial developments from the PPG-based prototype [17] support a new interest in the possibilities that PPG technique might offer for HR tracking.

Under severe exercise conditions, motion artifacts present one of the most challenging problems for PPG analysis. Design of sensor and packaging can help to reduce the impact of motion disturbances, but it is rarely sufficient for noise removal. Advanced signal processing techniques are often required to discriminate motion artifacts under vigorous activity. Several techniques have been investigated to deal with those artifacts, such as processing context information by additional on-body sensors and light sources [18,19] or adaptive noise cancellation using accelerometers as a noise reference [20]. Very recently, heart rate measurements using a PPG sensor integrated with an adaptive noise cancellation device has been demonstrated at common physical activity, from walking to running up to 8 Km/h [21].
Spectral analysis, as traditional fast Fourier transform (FFT), is a simple and inexpensive useful tool for separating motion artifact and cardiac physiologic spectra. However, techniques based on spectral analysis will not be 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 bend manoeuvres [2]. Advanced signal processing techniques can help to overcome this problem. The aim of this work is to present and validate an innovative algorithm able for tracking heart rate from PPG signal during athletes IMET on treadmill. The real-time motion discriminator algorithm, based on linear filtering, frequency domain and heuristic analysis, allows us to extract the heart rate value from PPG signal in the presence of severe motion artifacts. The basis of the frequency selection is the observed high harmonic content of movement artifact signals with respect to the PPG-derived heartbeat. A preliminary version has been already presented, with off-line analysis both of heart rate and oxygen saturation [22]; this paper provides an optimization of the algorithm parameters for the heart rate tracking, as well as real time probes and comparison of the results to the values obtained by a standard technique. The physiological aspects of training are out of the scope of this work. The custom PPG measurement system is based on a transmittance pulse oximetry system [23].

The organization of this paper is as follows. Section 2 describes the materials and methods. Section 3 is devoted to the analysis of the elements for the algorithm design through the analysis of PPG recorded data obtained from a first group of athletes during treadmill IMET. Section 4 is devoted to the development of the algorithm. Section 5 is devoted to present the results of validation of the algorithm both on recorded data of a group T1 of athletes and on real time heart rate tracking during IMET on treadmill of a second group T2 of athletes; we have assessed the agreement between heart rate measured by the new and the reference methods, electrocardiography (ECG), by means of Bland and Altman analysis [24,25] obtaining PPG data that lie within ± 6.5 bpm (beats per minute) with respect to the ECG data. In section 6, we present the discussion of the work.

2. Materials and Methods
Athletes of both sex genders with a training schedule exceeding an average of 7 - 8 hours per week participated in IMET performed on treadmill during separate sessions. Data presented and analyzed in this work correspond only to the heart rate measurements; others like oxygen consumption, carbon dioxide production and blood pressure are out of the scope of the present work. The studies were conducted at the laboratories 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

A first group T1 was composed by 10 white males, endurance runners tested on the treadmill ergometer. A second group T2 includes 20 athletes tested on the treadmill: 9 white females (one runner, 8 football players), 9 white male runners and 2 black female basketball players. The development of the algorithm and optimization of its parameters was carried out after data recorded on the first group T1. The algorithm performance has been evaluated on real time during IMET of the independent group of athletes T2.

### 2.2. Experimental protocol

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

Next, a complete maximal exercise test was performed, followed by active recovery. The protocol [26] for the treadmill ergometer (HP Cosmos QUASAR 4.0) test established that after one minute at rest (in order to record basal values) and after warming up, the athlete began to walk at 6 Km/h and 1 % slope during 2 min. The athlete then started the effort phase running at 8 Km/h and 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 constant, while speed was increased every 2 min by 2 Km/h, until they were unable to continue, then the athlete held the protective bars and jumped off the treadmill.
Active recovery was performed during 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 walk on the treadmill and in many cases they took hold of protection bars, which could alter the position of the finger sensor. This will be discussed later. The ECG system computed the heart rate every 10 s, averaging the last eight heart beats. At different stages of the test, once the athlete was running at fixed speed and slope, the full-step rate (SR, in Steps/min) was obtained by counting manually the number of steps in 10 s interval, and the one-foot step was derived from that. When running, steps and waving of the arms are synchronous; hereafter, we will note SR both for the full step and the full arm-waving, which are inseparable issues.

2.3. PPG measurement system

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

This experimental system is connected to a Pentium-class laptop computer that saves files containing the PPG 100 Sa/s data in volts (V). The next stages of the signal processing are carried out digitally. The post-processing of stored signals from the group T1 of athletes is made off-line; once the algorithm has been developed, it has been implemented as a virtual instrument using Lab Windows CVI™, and the validation is made on-line during IMET of the second group T2 of athletes.

2.4. Statistical analysis
Comparisons of HR between the two measuring 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 indication that the two methods evolve in parallel. To compare the values of HR obtained by our PPG technique with the standard ECG, we have used the Bland-Altman (B-A) method. Hereafter, we will use the notation PR and ER for the PPG and the ECG heart rate values, respectively. The differences between the methods (PR-ER, in bpm) are plotted against the average value given by the two methods (PR+ER)/2, in bpm. The results are presented as the differences themselves (PR-ER) and mean difference or bias (M) between the two methods. The standard deviation (Sd) value is used to calculate the limits of agreement (LA) computed as M ± 1.96·Sd, providing an interval within which 95% of differences between 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 limits of agreement. We defined agreement as a difference of (PR-ER) within ±10 bpm [3,12].

3. Elements for the Algorithm Design

Fig. 1 presents the FFT spectra corresponding to PPG signals recorded over 10 s time intervals at rest (time interval finishing at 25.96 s, grey) and when running at 12 Km/h (time interval finishing at 435.56 s dark), during the test performed in treadmill by a male athlete A of group T1. Sampling of ECG- and PPG-based measurements are not synchronized, for this reason the match of ECG and PPG readings has been achieved by pairing the nearest values in time.

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

Fig. 2 shows the time evolution of peaks P1, P2 and P3 (frequency vs. time) during IMET of athlete A, from rest to maximal effort; time evolution of ER and full-step rate (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 these lines evolves concomitantly, among them a third line evolves steeply with time; the strongest peak is not always associated with the same line but jumps from a line to another. The steeper line evolves close to ER values, for this reason we have identified these frequencies as the PPG-heart rate measurements PR. The heartbeat harmonics appears sometimes at a noise level at rest and at the beginning of the IMET and goes out of the plotted range of frequency as the heartbeat increases. Regarding the lower and upper lines, it can be observed that the upper one remains always at twice the frequency of the lower one; furthermore, this upper line evolves close to the full-step rate SR, for this reason we have identified these frequencies as the full-step rate SR, the harmonic of the one-step rate, 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 (fundamental and its harmonic, related to the one-foot and the full step rates, respectively) that evolve during exercise at a rate different to the heartbeat.

Results obtained from different athletes show that evolution of the peaks follows always the ER and SR evolutions; in some cases, the heartbeat evolves much more steeply than 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 PR signal and artifacts can overlap.

The arterial pulse wave might have higher frequencies but their power content is low or they can be filtered. At low physical activity, the PR value can be obtained by simply looking for the highest peak in the spectrum. The respiratory rate has been detected several times during exercise, at frequencies around 40-50 min⁻¹ but it has low power content relative to the heartbeat and movement artifact signals.

The algorithm design is based on these observations. Firstly, the algorithm needs a harmonic discriminator, taking into account that movement artifact contribution has higher harmonic content than the heart rate signal. The detection algorithm we have developed is mainly based on analysis of the relative strength of
harmonic content of the PPG signal. Secondly, it is important to have a heuristic algorithm to track the heartbeat, taking into account the past history of the heartbeat itself when PR signal and artifact signal overlap.

### 4. Development of the Algorithm

We have developed the algorithm by analyzing the PPG data recorded from the first group T1 of athletes and comparing them to the ECG ones.

The first step is a linear filtering of the 100 Sa/s PPG signal $E(t)$ through band-pass Bessel filtering (cut-off frequencies $f_1$ and $f_2$, order $n = 6$), in order to get the pulsating component ($E_{ac}(t)$) and to suppress high frequency noise and ripple. Cut-off frequencies are 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 is performed by applying a fast Fourier transform (FFT) to a 10 second, rectangular-shaped, sliding window of the aforementioned data signal, with an overlap of about 75%, delivering a 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.

The spectra stream goes onto a peak search algorithm that selects the peaks 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. We have explored more complex windowing functions and even more complex approaches [27], but the relatively wide window and the experimental tests shows that leakage, albeit present, did not invalidate the effectiveness of the algorithm.

Fig. 3 shows schematically 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 $K_C$ have been tuned in the range 50 - 400 by analyzing and comparing the PPG and ECG data recorded from the group T1 of athletes; $K_C = 100$ was adopted as the default value; it means that the cut-off of minor peaks occurs
when amplitudes are approximately from 7 to 20 times lower than the highest peak. When implemented on the PPG system and in order to simplify the computational cost of the algorithm, the peak search selects only the three main peaks.

Next, the algorithm chooses a “best candidate” $P(t_s, f, p)$ as the heartbeat PR, with a reliability factor $C(t_s)$ which express, in the range (0-1), the degree of confidence that $P(t_s)$ was the heart rate value, not a movement artifact. To achieve this, the list of three peaks is fed into three classification engines: energy, harmonic and historic classification. The task of the classification engine is to associate to each peak the reliability factor $C(t_s)$ defined as above, using different criteria. In all the following discussion and formula, reliability factors saturates to 0 or 1, respectively, when the computed value exceeds these limits. Fig 4b and 4c shows the flow diagram of the heuristic algorithm.

The power engine (Fig. 3 (b)) compares each $p_i$ value to the highest value $p_{\text{max}}$ of $F(t_s)$, and assigns to each $(f_i, p_i)$ pair a first reliability factor $C_{pi}$, which expresses how much the peak has a high power content. $C_{pi}$ is computed in the range (0-1) as the amplitude of the peak relative to the 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 frequencies multiple of the peak, with a relative width controlled by a parameter $K_H$, that modulates the tolerance accepted in the frequency comparison. $C_{hi}$ is computed in the range (0-1) based on how much the spectrum energy enters the mask, and its value is low for high harmonic content and high for low harmonic content. When implemented on the PPG system and in order to reduce the computational burden, the algorithm computes how near a frequency is to double another frequency of the list, limiting the analysis to the second harmonic. The parameter $K_H$ is one of the main parameter of the algorithm. $K_H$ has been tuned in the range 5 - 50, and was optimized for a value of 12. For $K_H < 5$, the mask is too wide and all the peaks appear with a 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 under the power engine point of view. On the other hand, at rest, the main peak corresponds to the
heart rate PR, but very small power harmonics can be present; it is the value of $K_C$ of the peak search algorithm which must contribute to eliminate this harmonics of the peak list 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 the frequency $f'_i(t_{s-1})$ of the peak $P(t_{s-1})$ selected as 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 the reliability factor $C_{mi}$. This factor $C_{mi}$ is zero for all the peaks but for that with the frequency closest to $f'_i(t_{s-1})$ selected as PR at previous spectrogram; $C_{mi}$ is computed in the range 0-1, so that a high value means high proximity to the previous PR peak. 

Fig 4 shows, in more detail, the flowchart of the algorithm. Fig 4a shows the general data processing for selecting the three main peaks of a spectrogram $F(t_s)$. Fig. 4b depicts the flowchart of the “power” and “harmonic” engines; Fig. 4c shows the flowchart of the “memory” engine and the final selection step. 

In the end, 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, that is high $C_{pi}$ (amplitude of the peak), high $C_{hi}$ (low harmonic content) and high $C_{mi}$ (high proximity to the previous PR peak). The chosen peak is assigned to the PR($t_s$) value, and the overall confidence index $C(t_s)$ is set to the highest reliability index.

As an exception, when no a peak achieves a reliability factor greater than 0.5 (the highest peak could have been marked as non valid by the harmonic content analysis, and the other peaks are not sufficiently good by any of the classification techniques) the algorithm simply “jumps” over, giving the PR value selected for the last spectrogram, and it restarts with the next one. This step is permitted just once in a row to avoid the possibility of sticking to 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 PR continues to be chosen by means of the memory algorithm, the reliability factor is always decreasing, and this eventually forces the heuristic algorithm to select the PR peak by a different criterion, so that the memory one could not
overtake the others for a long time. This is very important in the case PR peak is 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 does not find a solution and the harmonic engine will search for a new peak.

The parameters $K_C$ and $K_H$ have been optimized by analyzing the recorded data of group T1 of athletes. The highest computational cost of the algorithm is found in the generation of the FFT spectra, for which we used the standard FFT subroutines offered by standard software. As mentioned before, in order to reduce the computational cost, some simplifications have been used: a) at the search peak stage, the number of selected peaks is culled to three, those with highest power density; b) at 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 and allows it to discriminate the movement artifact. When the PR peak is very close to a motion peak, the memory engine helps to follow correct peak, but forcing the harmonic engine to act quickly. When one peak is selected, its frequency is then assumed to be the pulse rate PR, and its value is displayed on the screen.

4.1. Signal to Noise Ratio

An estimation of the signal-to-noise ratio (SNR) of the selected PR data has been done. The selected PR peak has been used as the signal power density and the sum of discarded peaks as the noise power density. The procedure has been repeated for all the PR peaks selected along the exercise and for the whole group T1 of athletes. 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 but just discarded peaks; moreover, the noisiest points have been avoided, those where the algorithm has jumped over, so that the resulting figures are probably an overestimation of the real value. Nevertheless, it constitutes a valuable estimation of the SNR of the selected PR data. We have calculated values of SNR, both averaging over the whole IMET and over the last 20% of IMET time. From the analysis of group T1 recorded data, the averaged values for “full test” and “last 20 %” are -1.6 dB and -2.5 dB, respectively. From that, we have
concluded that the algorithm is able to identify the correct cardiac heartbeat PR even with a SNR as low as -2.5 dB.

5. Results

5.1. Validation of the algorithm on recorded data of group T1 of athletes.

To validate the algorithm, each PR value was compared with the timely 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. The reason is that at 6 Km/h most of the athletes were adapting to walk on the treadmill and in many cases they took hold of protection bars, which could alter the position of the finger sensor and therefore distort the measurement.

We have calculated the regression line of ER and PR measurements, 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 has confirmed the normal distribution. The Bland and Altman 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, Sd = 2.45 bpm is the standard deviation, and the limits of agreement LA are in the range (-5.90 bpm, +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 limit of agreement (LLA), 3.31 to 4.07 bpm for the upper limit of agreement (ULA). All these intervals are reasonably narrow, the width of limits of agreement is 9.59 bpm and more than 95% of all (PR-ER) differences fall within LA. The plot reveals that there is no relationship between the differences and the bias. The option of plotting the differences as percentages is useful when there is an increase in variability of the differences as the magnitude of the measurement increases. When plotting as percentages, the values obtained are M = -0.7%, LLA = -4.2%, ULA = 2.8%.

5.2. Results on real time heart rate tracking during IMET on treadmill.

The algorithm has been implemented in the experimental PPG system, and has been validated on real time measurements during treadmill ergometer IMET on the second group of athletes T2. The SNR values are -0.1
dB for “full test” and -2.1 dB for “last 20%” respectively. Here again, in spite of the low SNR, the algorithm is able to follow the correct PR value, as shown below.

The plot of PR versus ER for the group T2 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. We have confirmed the normal distribution of differences between the two methods by their histograms. The (PR, ER) pairs obtained have been displayed in Bland-Altman plots in Fig. 6, as the differences (PR - ER) versus mean value (M) of the differences; the value M = -0.70 bpm indicates a low negative bias, the standard deviation is Sd = 2.92 bpm. The width of limits of agreement 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 HR 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 group T2, the values expressed as percentage of the average between the two techniques are M = -0.4%, LLA = -4.2% and ULA = 3.4%.

6. Discussion

The algorithm for the automatic detection of the HR from the PPG signals was found to be efficient, despite the small signal-to-noise ratio. Comparison to ECG measurements, the reference standard, shows very high agreement. The low negative bias obtained on recorded data of group T1 of athletes may be caused by slight differences in beat-to-beat averaging times used in the ECG monitor and the PPG system; if the subject's heart rate is changing rapidly, the different averaging times would influence the results and amplify the calculated errors. Moreover, the small bias might be due to possible errors when collecting data from a variety of sources in which time is 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 in activities as IMET. The narrow width of limits of agreement, narrower than ±10 bpm, is an acceptable LA for interchangeability between PR and ER. It should be noted that parameters optimization, carried out on the recorded data, has been performed in a “first step” way. Optimization with an iterative mathematical program should give the best values, but it is not the aim of the present work.
Overall, the system shows a general success to track heart rate during the treadmill probe. The system has also shown good results for IMET on cycle-ergometer [29]. When compared to traditional FFT methods, the presented algorithm shows significant improvements for separating motion artifact and cardiac physiologic spectra. The limits of agreement are comparable to the mean absolute error obtained using an algorithm including time-frequency techniques based on the smoothed Wigner Ville distribution [2,8], where six subjects participated in the study realizing four kinds of motions at time, frequency and intensity fixed.

When comparing our method with results from a PPG sensor integrated with an adaptive noise cancellation device, we obtained LA values between -6.41 bpm and +5.01 bpm (or -4.2 % and 3.4 %), better than values (-21.15% and 20.52 %) obtained for subjects running at 8 Km/h [21].

The algorithm has been developed by analyzing the recorded data of a group of white men, all of them runners; and it has been evaluated on a very dissimilar group, composed of men and women, black and white skins. The algorithm parameters we have used have been the same for female subjects or for black skins and the accuracy of the technique might be higher when adapted for each case. But in the current form the algorithm is suitable to people without discrimination neither gender or skin color.

The basic premise of the algorithm is that the motion artifact created in the PPG signal by human running has a high harmonic content with respect to the PPG derived heart rate. The use of graded treadmill running (or cycling) in IMET may bias the motion artifact in the PPG signal by making it quite regular and rhythmic in nature. Although this seems to be a limitation of the PPG-based technique, athletes running are a regular and stable movement. A significant advancement of the algorithm utility would be to show that it can extract representative heart rates during road running or running on uneven ground. Other apparent limitation is that our PPG system is a fingertip probe; the algorithm can be implemented in other PPG systems, as earlobe probes or ring sensor, provided that a rhythmic movement of the body and/or extremities will have high harmonic contents. It is worth to note that other error sources, like movement of the sensor on the finger, do not cause important signals in the FFT analysis once the athlete arms have free movement; otherwise, ECG-derived HR values can also give some errors, when the athlete is sweating and the fixation of the ECG electrodes could be poor and produce wrong readings.
7. Conclusions

This work addresses the problem that motion artifacts limit the use of photoplethysmography (PPG) to quantify key cardiovascular variables such as heart rate during human exercise. During human movement, the PPG signal is corrupted at frequencies that fall within the range of interest. With appropriate filtering and signal processing, the portion of the PPG signal caused by motion artifact could be removed and the heart rate identified. We have described and evaluated a new algorithm to process PPG signals collected during treadmill exercise into heart rate. 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 artifact leaving one frequency to represent PPG derived heart rate. The algorithm is tested using a heterogeneous athletic population transitioning from rest through a graded maximal intensity treadmill test. Electrocardiography is used to validate the PPG-derived heart rates. This would allow for significant advancement in the area of biosensor development as PPG technology is ideally suited to be incorporated into wearable devices.

Acknowledgment

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 has been supported by CSIC I3P Program financed by the Fondo Social Europeo. The authors gratefully acknowledge the assistance of Dr. Isabel Gómez Caridad in revising the present manuscript.
References


Figure captions

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

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

Fig. 3. Description of the heuristic algorithm. After the band-pass linear filtering, in (a), a set of significant peaks is computed, represented by a list of frequency-peak value (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 a parameter Kc. The list is then fed into the three mechanisms that assign a reliability value to each pair. In (b), peaks are classified in function of their relative amplitude; in (c), their harmonic relationship is computed (see a more thorough explanation in the text); finally, in (d), the historic record is taken into account, assigning higher reliability values to the peaks that are near to the previously selected value.

Fig. 4. Flowchart of the proposed algorithm. In Fig 4a the general data processing is shown. Fig. 4b depicts the “power” and “harmonic” engines, and Fig. 4c the “memory” engine and the final selection step.

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

Fig. 6. Bland and Altman plot of differences in real time heart rate measurements versus average values for group T2, with bias value (M) and limits of agreement (M ± 1.96 Sd). Sd is the standard deviation, A is the number of athletes and N is the number of pairs (PR, ER).
Fig. 1

Athlete A

- 25.96 s
- 435.56 s

Power density ($V^2/Hz$) vs. Frequency (bpm)

- $ER_{25.96} = 75$ bpm
- $ER_{435.56} = 150$ bpm
- $SR_{435.56} = 192$ bpm

Fig. 1
Fig. 2
(a) Noise floor trimming

(k) Harmonic classification

(b) "Power" classification

(d) Memory (history) classification
Fig. 3
Fig. 4a
Fig. 4b
Fig. 4c
Fig. 5

Diff in heart rate, PR-ER (bpm)
Average by PPG and ECG, (PR+ER)/2 (bpm)

$T_1:$
$A = 10$
$N = 477$
$Sd = 2.45 \text{ bpm}$

$M + 1.96 \text{Sd} = 3.69 \text{ bpm}$
$M = -1.10 \text{ bpm}$
$M - 1.96 \text{Sd} = -5.90 \text{ bpm}$
Fig. 6

T2:
A = 20
N = 1061
Sd = 2.92 bpm

M = -0.7 bpm
M - 1.96Sd = -6.41 bpm
M + 1.96Sd = 5.01 bpm

Diff in heart rate, PR-ER (bpm)
Average heart rate by PPG and ECG, (PR+ER)/2 (bpm)