Physical Conditioning Status Stratification Based on Heart Rate Variability: Principal Component Analysis of Power Spectrum Density Function

Olivassé Nasario-Junior[1], Paulo R Benchimol-Barbosa[1,2], Jurandir Nadal[1]

[1]Programa de Engenharia Biomédica, COPPE, Universidade Federal do Rio de Janeiro, RJ, Brasil
[2]Hospital Universitário Pedro Ernesto, Universidade do Estado do Rio de Janeiro, RJ, Brasil

Abstract

The purpose was to investigate the application of principal components analysis (PCA) on power spectral density function (PSD) computed by heart rate variability (HRV) signals aiming at distinguishing physical conditioning status. Ten healthy sedentary volunteers (controls) and ten matched professional long distance runners (athlete) were enrolled. The subjects had their maximal metabolic equivalents (MET) estimated and 20 min resting ECG recorded. In RR intervals (iRR) series, a 5 min segment with the least variance was selected. iRR sequence was interpolated at 1 Hz (cubic spline) to estimate the PSD by fast Fourier transform. PCA was applied to PSD, and the first three principal components were retained (90% of the total variability). A logistic regression model based on Mahalanobis distance determined the optimal separation threshold. The area under the ROC curve (AUC) of the proposed method was compared to standard LF and HF parameters (α=0.05). The proposed method allows to correctly separate all subjects (AUC=1.0), while the HRV parameters in the frequency domain LF and HF provided AUC equal to 0.8 and 0.9 (p>0.05), respectively. The PCA applied to PSD of iRR series properly classified the physical conditioning of controls and athletes with superior performance to standard HRV.

1. Introduction

Regular aerobic exercise provides beneficial changes on the cardiovascular system, characterized by mechanical, autonomic and electrophysiological remodelling [1]. Autonomic remodelling is evidenced by both resting heart rate (HR) reduction and cardiac vagal modulation increase [2]. HR variability (HRV) is a non-invasive diagnostic tool employed to assess autonomic nervous system (ANS) activity. HRV is modulated by both sympathetic and vagal efferent and afferent neural pathways over sinus and AV nodes [3].

The physiological adaptations resulting from physical training are reflected on changes in cardiac autonomic modulation, usually assessed by HRV at rest [2]. One of the methods widely used for the study of the HRV in short duration signals is the spectral analysis, which takes into consideration the very-low-frequency (VLF), low-frequency (LF) and high-frequency (HF) bands obtained from the power spectral density function (PSD) [3].

On the other hand, the principal component analysis (PCA) – a statistical technique to reduce the number of variables with the minimum of information loss, employed to study data variability and provides a privileged data set view – is a tool that allows the classification of characteristics based on the decomposition of the signal into orthogonal vectors [4]. The method admits concomitant appreciation of the morphologic characteristics of the PSD, such as the local variations of the curve as a whole, regardless of the frequency bands definition.

In this context, the purpose of the study was to investigate the application of PCA to identify the variation patterns of the PSD of normal consecutive RR intervals (iRR) aiming at distinguishing physical conditioning status.

2. Materials and methods

2.1. Study population

ECG signals were extracted from an existing high resolution ECG database [1]. The study protocol was approved by Ethics Committee and informed consent was obtained. Ten male elite runners (8.9 ± 3.2 years of training [mean $\pm$ SD]; six to eight training sessions/week; 90 to 120 min/session; 90 to 110 km/week) were enrolled (Athlete group). A group of 10 male healthy sedentary volunteers were included as control (Control group). Inclusion criteria, group separation, physical assessment procedures and maximal oxygen consumption (VO_{2MAX}) estimation protocol have been published elsewhere [1].

VO_{2MAX} was divided by the constant 3.5 ml kg^{-1} min^{-1} to be converted into metabolic equivalents (METs). Control and Athlete groups were separated according to

estimated VO_{2MAX}, by arbitrarily defining as less than 11.5 METs for sedentary controls and more than 16.0 METs for athletes.

Both groups were adjusted by age, and matched by anthropometric data (Table 1). The aim of anthropometric data intergroup matching was to reduce potential effect of thoracic geometry on surface ECG signals.

Table 1: Anthropometric and demographic characteristics (mean ± SD) of the subjects who participated in the study.

	Controls	Athletes
Age (years)	29.0 ± 5.4	24.4 ± 7.2
BMI (Kg/m²)	23.8 ± 3.8	20.7 ± 1.9
BSA (m²)	1.8 ± 0.2	1.8 ± 0.2
APTD (cm)	21.3 ± 1.9	21.1 ± 1.2
LLTD (cm)	28.1 ± 3.2	28.0 ± 1.2
METs	8.7 ± 1.9	19.6 ± 1.3 *

BMI = body mass index; BSA = body surface area; APTD = anteroposterior thoracic diameter; LLTD = laterolateral thoracic diameter; METs = metabolic equivalents. * $p = 0.001$.

2.2. Signal acquisition, processing and wave detection

ECG was acquired shortly after anamnesis, application of a questionnaire and physical examination. Before a 15 min continuous signal acquisition, subjects remained in the supine position for 5 min for stabilization of autonomic modulation after the change from the orthostatic position [5].

High-resolution ECG signals were acquired using modified bipolar *Frank XYZ* orthogonal leads and digital data were processed with custom-made pattern recognition software and detailed description of the equipment was previously described [6]. The R wave detection was carried out with the signal low-pass filtered at 15 Hz (Butterworth, 4th order). For the analysis of the *RR*-interval length, artefacts and ectopic beats were excluded by correlation, precocity and visual inspection, and confirmed by one expert. The detection of the RR intervals was performed using a custom-made program written in Matlab [7].

2.3. HRV spectral analysis

After visual inspection of the ECG signal for exclusion of ectopic beats and artefacts, a 5 min segment of consecutive normal RR intervals (NN intervals) was selected based on least variance criteria [8], aiming at identifying a stationary segment of the time series for spectral analysis. Trend component removal of the time series was carried out according to 'a priori' smoothing method [9], and interpolation using cubic splines at a frequency of 1 Hz was, thus, applied to extract equally spaced samples. From the selected interval, the estimated power spectral density function (PSD) was calculated by fast Fourier transform (FFT) approach [3]. For such, a 128 samples Hanning window was placed over the selected *iRR* series. These pre-processing procedures reduced eventual biases caused by low-frequency trends and assured adequate spectral estimates at all frequency bands. At the end, all the PSDs were normalized, thus equalling their values of maximum amplitude.

For the standard HRV spectral analysis of the HRV, the low (LF - 0.04 to 0.15 Hz) and high frequency (HF - 0.15 to 0.4 Hz) bands of the normalized PSD function were considered, which represent sympathetic and vagal influences (former band) and the vagal modulation on the sinus nodule (later band) [3].

2.4. Principal components analysis

After obtaining the PSD of each signal, the data were organized in a matrix (20 x n), in which the number of lines corresponds to each individual (10 controls and 10 athletes), and 'n', to the number of PSD values ([128/2] + 1 = 65 variables).

The covariance matrix obtained from the E matrix was used for the PCA [4], extracting the vectors and respective values through the method of singular values decomposition of Matlab v. 2007 (The Mathworks, U.S.A.). The first vector, which has the greatest value, corresponds to the first Principal Component (PC) and so forth. The approach used to choose the ideal number of PCs obtained in the analysis was based on the percentage of the accumulated total variance (≥ 90%).

The PC coefficients quantified the contribution of the components for the representation of each original signal (PSD), after the removal of the mean values of each subject [4]. These coefficients were used in the classification, which was accomplished by calculating the standard distance of the data of each volunteer in relation to the central point of the investigated sample. This parameter corresponds to the square root of the Mahalanobis distance, which measures the difference (standard distance) between each observation (from control and athlete groups) and the ensemble average in the space of PC coefficients normalized by their respective variances [4]. These values allowed the stratification of the groups by means of a threshold, which better separates the groups, calculated by logistic regression, thus allowing the determination of sensitivity, specificity, and total accuracy.

2.5. Statistical analysis

The comparison of the four methods (i.e., LF and HF HRV standard parameters; PC analyses in HRV signals and VO_{2MAX} values), was carried out by the receiver operating characteristic curves (*ROC*) comparison (*c-*

statistic). The *alpha* error level was set at 0.05.

3. Results

The pairwise comparisons of the ROC areas of each method are presented in Figure 1. The c-statistics were reported as follows: PC = 1.0; HF = 0.9 and LF = 0.8. No significant differences were observed for all comparisons (p = NS).
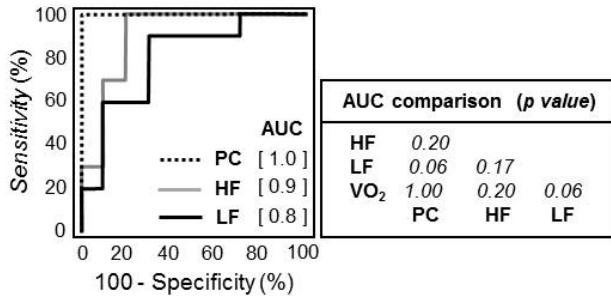

Figure 1: ROC curve of each approach (PC, HF, LF and VO_{2MAX}) using proposed configuration. Table inset depicts p-value in pairwise comparisons of c-statistics (p = NS). VO_{2MAX} and PC matched perfectly. PC = principal component; HF = high frequency; LF = low frequency and VO_2 = VO_{2MAX}.

The figure 2 (a) shows the mean PSD function computed from all subjects signals (mean ± SD) and figure 2 (b) brings the graphical representation of the firsts three PCs (each eigenvector).
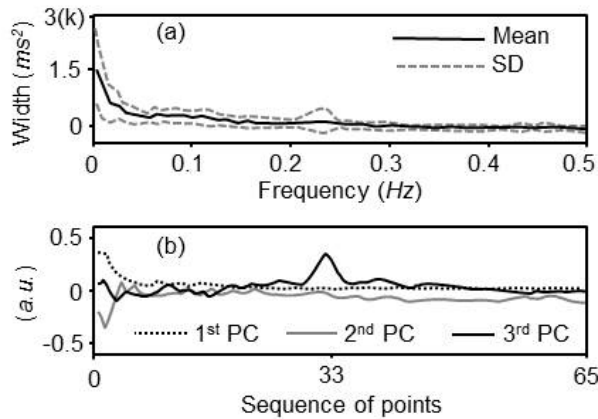

Figure 2: (a) Mean PSD function computed from all signals (mean ± SD); (b) Graphical representation of the first three PCs (each eigenvector). PC = principal component; SD = standard deviation.

Using the optimal configuration, Figure 3 shows the intergroup separation threshold for both VO_{2MAX} and PC approach. The PC threshold method was defined by logistic regression using the standard distance, taking into account the firsts three PCs.
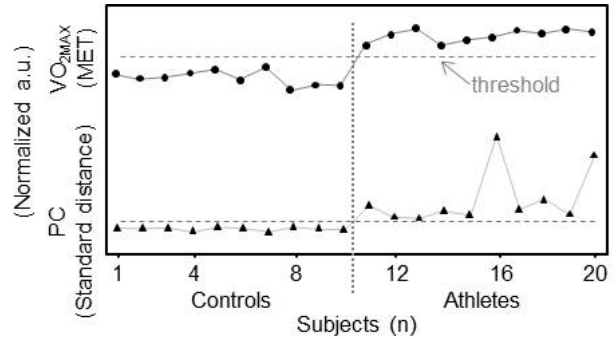

Figure 3: Intergroup separation threshold for both VO_{2MAX} and PC approach. ● = VO_{2MAX} (MET); ▲ = PC (standard distance). (controls on the left and athletes on the right)

4. Discussion

This paper describes a method for assessing PSD pattern related to physical conditioning status based on PC analysis of HRV signals. The advantage over the classical spectral analysis methods alone is that PCA-based classification does not depend on frequency band delimitation and takes into account the morphology of the PSD function.

Some studies that have evaluated the cardiac autonomic modulation throughout different physical conditioning status groups using HRV indexes obtained from the spectral analysis. Those studies highlighted an HRV increase with the advancing of VO_{2MAX} induced by physical fitness [1, 2]. The results of the present study corroborate those findings when identifying a significant increase in the HRV measures in elite professional runners (men) when compared with the healthy sedentary volunteers. Such finding justifies the study of the HRV in athletes and controls for the identification of variation standards in the PSD, by means of the PCA, which may highlight those changes.

PCA based on a QRS complex segmentation for pattern recognition was first proposed by Nasario-Junior *et al.* (2014) to detect pathological intra-QRS potentials [7]. The authors also applied PCA on HRECG signals, which enabled assessing intra-QRS potentials as markers of cardiac electrical remodeling in athletes, suggestive of myocardial hypertrophy [10]. Recently, the same group demonstrated that PC analysis method appropriately stratified the PSD function in two groups of healthy women of different age groups, being an effective method for identification of cardiac autonomic deterioration due to ageing [11]. PC analysis is a multivariate statistical technique employed to reduce the number of variables

without compromising information, and also provides a privileged data set view [4]. In Cardiology, the PC analysis can be used for different purposes, however the best of the author's knowledge, this is the first study to use PC analysis of the PSD function to differentiate healthy sedentary subjects from athletes.

Current approach, which takes into account the morphology of the whole PSD function, and not only particular bands, may provide additional contribution to the autonomic function study. Moreover, the use of Mahalanobis distance to classify subjects of different physical conditioning status seemed to allow an accurate diagnostic appreciation, given that it was possible to define a threshold value satisfactorily separating health sedentary individuals from athletes.

The proposed method allowed to correctly separate all subjects (AUC=1.0), according to prior classification based on MET. On the other hand, HRV parameters LF and HF provided nonsignificant lower AUC, respectively, 0.8 and 0.9 (Figure 1).

The firsts three PCs retained for analysis comprised 90% of the total ECG signal variance. It may be noted in Figure 2 (b) that the first and second PC accounted for the high variance in the very low frequency region on the spectra (up to 0.1 Hz), which reflect long term HR changes and also is affected by the mean HR. On the other hand, the third PC presents a peak aligned with the region of highest variability in the mean PSD, between 0.2 Hz and 0.3 Hz, representing the variance of the respiratory component of HRV.

It was noteworthy that, starting from predefined threshold, PC analysis was able to correctly separate groups (Figure 3). Particularly, sedentary group showed lower whereas athletes showed higher standard distances, perfectly matching VO_{2MAX} classification.

Although present study did not assess longitudinal effects of aerobic physical conditioning, it is emphasized that the method showed potential applications for assessing physical conditioning status at rest. The advantages of this approach were its non-invasive nature, lower cost, and an intrinsic lower risk for unexpected adverse events during provocative tests.

Limitations of the study include small sample size, the application of the method using two physiologically well-defined groups, and VO_{2MAX} estimated indirectly, according to the 12 min Cooper test. Further studies are needed to confirm the present findings.

This study allow to conclude that principal component analysis applied to PSD function of *iRR* series properly classifies the physical conditioning of healthy sedentary and athletes with superior performance than standard HRV.

Acknowledgements

References

[1] Marocolo M, Nadal J, Benchimol-Barbosa PR. The effect of an aerobic training program on the electrical remodeling of heart high-frequency components of the signal-averaged electrocardiogram is a predictor of the maximal aerobic power. Braz J Med Biol Res 2007;40:199–208.

[2] Melanson EL, Freedson PS. The effect of endurance training on resting HR variability in sedentary adult males. Eur J Appl Physiol 2001;85:442–9.

[3] Task Force. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J 1996;17:354–381.

[4] Jolliffe IT. Principal Component Analysis (2002); Springer, New York. Lopes FL, Pereira FM, Reboredo MM, Castro TM, Vianna JM, Novo JR JM and Silva LP. Redução da variabilidade da frequência cardíaca em indivíduos de meia-idade e o efeito do treinamento de força. Rev Bras Fisioter 2007;11:113–119.

[5] Lipsitz LA, Mietus J, Moody GB, Goldberger AL. Spectral characteristics of HR variability before and during postural tilt. Relations to aging and risk of syncope. Circulation 1990;81:1803–10.

[6] Barbosa PR, Barbosa-Filho J, de Sá CA, Barbosa EC, Nadal J. Reduction of electromyographic noise in the signal-averaged electrocardiogram by spectral decomposition. Biomedical Engineering, IEEE Transactions 2003;50:114–7.

[7] Nasario-Junior O, Benchimol-Barbosa PR, Nadal J. Refining the deceleration capacity index in phase-rectified signal averaging to assess physical conditioning level. Journal of Electrocardiology 2014;47:306–10.

[8] Lopes FL, Pereira FM, Reboredo MM, Castro TM, Vianna JM, Novo Jr JM, Silva LP. Reduction of heart rate variability in middle-aged individuals and the effect of strength training. Braz J Phys Ther 2007;11:113–9.

[9] Tarvainen MP, Ranta-Aho PO, Karjalainen PA. An advanced detrending method with application to HRV analysis. IEEE Trans Biomed Eng 2002;49:172–5.

[10] Nasario-Junior O, Benchimol-Barbosa PR, Trevizani GA, Marocolo M, Nadal J. Effect of aerobic conditioning on ventricular activation: a principal components analysis approach to high-resolution electrocardiogram. Comput Biol Med (2013); 43: 1920–6.

[11] Trevizani GA, Nasario-Junior O, Benchimol-Barbosa PR, Silva LP, Nadal J. Cardiac autonomic changes in middle-aged women: identification based on principal component analysis. Clin Physiol Funct Imaging 2016;36:269–73.

Address for correspondence.
Jurandir Nadal
PO Box 68510; Zip code 21941-972; Rio de Janeiro, RJ, Brasil.
E-mail: jn@peb.ufrj.br