# Analysis of Cardiac Dynamics in Patients With Chagas Disease Using PCA

M Vizcardo[†], D Cornejo[†], E Alvarez[‡]

[†]Escuela Profesional de Física, Universidad Nacional de San Agustín de Arequipa
[‡]Escuela de Física, Universidad Central de Venezuela, Venezuela

## Abstract

*Chagas disease American trypanosomiasis is caused by a flagellated parasite: trypanosoma cruzi, transmitted by an insect of the genus Triatoma and also by blood transfusions. In Latin America the number of infected people is approximately 6 million, with a population exposed to the risk of infection of 70 million people. It is our interest to develop a low-cost, non-invasive methodology capable of describing cardiac dynamics within 24 hours and thus finding descriptors of dynamics that allow early detection of cardiac abnormality caused by T. cruzi. We analyzed the first and second principal components of the PCA of the 8 HRV indices of the 24-hour RR records in patients with ECG abnormalities (CH2), patients without ECG abnormalities (CH1) who had positive results Serological findings for Chagas disease and healthy (Control) matched for sex and age. We found significant differences (p-value<0.05) between the three groups Control-CH1, Control-CH2 and CH1-CH2 up to 5 continuous hours in dynamics between the, using the HRV heart rate variability indices and the PCA principal component analysis method, and we also found a lower distance from the mean dynamics in the Control group (0.020267), then the CH1 group (0.027922) and finally in group CH2 (0.034812).*

## 1.    Introduction

In Latin America the number of people infected is approximately 6 million, with a population exposed to the risk of infection of 70 million[1]. These estimates explain why this disease is a serious public health problem in the countries where it is endemic. In the evolution of Chagas disease we can distinguish an initial acute phase of infection and a prolonged intermediate chronic phase, in which the disease is often clinically silent, and the usual diagnostic techniques do not provide a robust criterion to predict whether a seropositive asymptomatic patient will suffer cardiac involvement. During the last twenty years a close association has been observed between the functioning of the autonomic nervous system and mortality due to cardiovascular causes. The experimental evidence that the increase in sympathetic activity or the reduction of parasympathetic disease predict lethal arrhythmias, has generated the intense search for a method that quantifies the influence of the autonomic nervous system on the heart. It is our interest to develop a low-cost non-invasive methodology that allows 24-hour descriptor analysis using HRV and Principal Component Analysis (PCA)

## 2.    Database and Registry

The results of this work were obtained by processing the electrocardiogram (ECG) and obtaining the RR interval, from three different groups of volunteers. The following test: clinical evaluation, serological Machado-Gerreiro test, chest x-rays, echocardiogram, electrocardiogram and ambulatory Holter registration (24 hours), classify volunteers into three groups: 83 healthy people called **Control** group; 102 patients infected with only positive serology (clinical evaluation, chest x-rays, echocardiogram, electrocardiogram and Holter were normal) called **CH1** group; and 107 seropositive patients with incipient cardiac involvement first degree atrioventricular block (BAV), sinus bradycardia (BS) and or right bundle branch block of the bundle of His (BRDHH), that were not being treated with medications, called the **CH2** group. All were outpatients, and informed consent was obtained from all of them. The ECG signals were recorded at 500 Hz with 12 bits of resolution, a set of 288 framer of 5 minutes was obtained.

## 3.    Methods

We have used the database of the Instituto de Medicina Tropical (IMT) of the Universidad Central de Venezuela. For the detection of the QRS complex the program based on the Pan-Tomkip algorithm[2] was used, then the 288 tachograms of the RRs of 5 minutes were generated, they were made a post processed the "ada" filter[3], to eliminate the eptopic beat, and then 8 HRV indices were obtained: intervals NN, standard deviation SDNN, square root of the mean squared differences of successive NN intervals RMSSD, the number of interval differences of successive NN intervals greater than 50 ms pNN50, high frequency

HF, low frequency LF, the ratio LF/ HF and the ultra low frequency VLF[4].

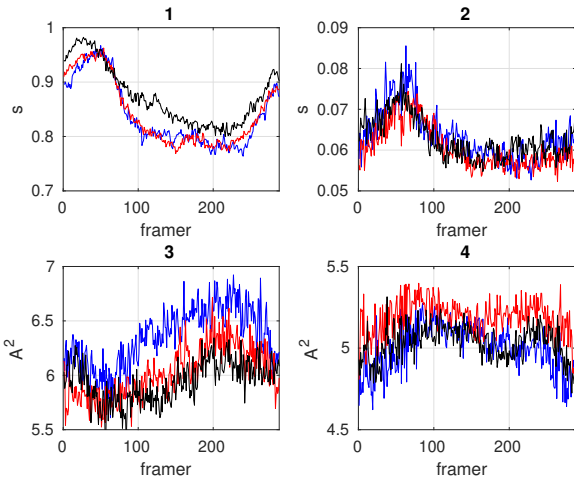In figure 1.1 we have the circadian profile of the aver-



Figure 1. Temporary and Frequency circadian profiles of the average values of: 1) RR intervals, 2) standard deviation, 3) low frequency LF, 4) very low frequency VLF. Groups: Control (blue), CH1 (red) and CH2 (black)

age values of the RR intervals of the three groups where in some frame (hours) there would be a difference between the three groups, but figures 1.2, 1.3 and 1.4 show circadian profiles of HRV where there is a lot of noise even though in 1.3 there are some differences especially with the control group, this may be due to the fact that the ECG data has noise. Faced with this problem, we propose a methodology to eliminate noise and extract the most relevant, uncorrelated information, reduce the dimension and use the 8 HRV indices as a features. For the analysis of the 8 HRV indices we use the PCA Principal Component Analysis method[5] that is to say, reduce the dimension and have an uncorrelated representation. We will mention some characteristics:

• Intuitively, the technique serves to determine the number of underlying explanatory factors after a set of data explaining the variability of said data.
• The PCA seeks the projection according to which the data is best represented in terms of least squares.
• PCA involves the calculation of the decomposition in eigenvalues of the covariance matrix, normally after focusing the data on the average of each attribute.
• The calculation of the covariance matrix is based on the internal or scalar product.
• Limitations: only linearly separable problems
The standard linear PCA algorithm:
• Given a set of observations $x_i \in R^N \quad i = 1, ...m$ which

are centered, $\Sigma x_i = 0$ , The PCA finds the main axes through the diagonalization of the covariance matrix: $C = \frac{1}{m} \Sigma x_j x_j^T$
• $C$ it is a definite positive matrix, the result of the diagonalization leads to eigenvalues non-negative To perform the calculation, the equation of eigenvalues is solved: $\lambda \nu = C\nu$ for non-negative eigenvalues and non-zero eigenvectors $\lambda \geq 0$ . Replacing : $\lambda \nu = C\nu = \frac{1}{m} \Sigma < x_j \cdot \nu > x_j$
• Where you can see that all the solutions $\nu$ con $\lambda \neq 0$ they are in the sub-space generated by $x_1...x_m$, therefore for these solutions the equation of eigenvalues is equivalent to $\lambda < x_i \cdot C\nu > \forall i = 1, ..., m$:
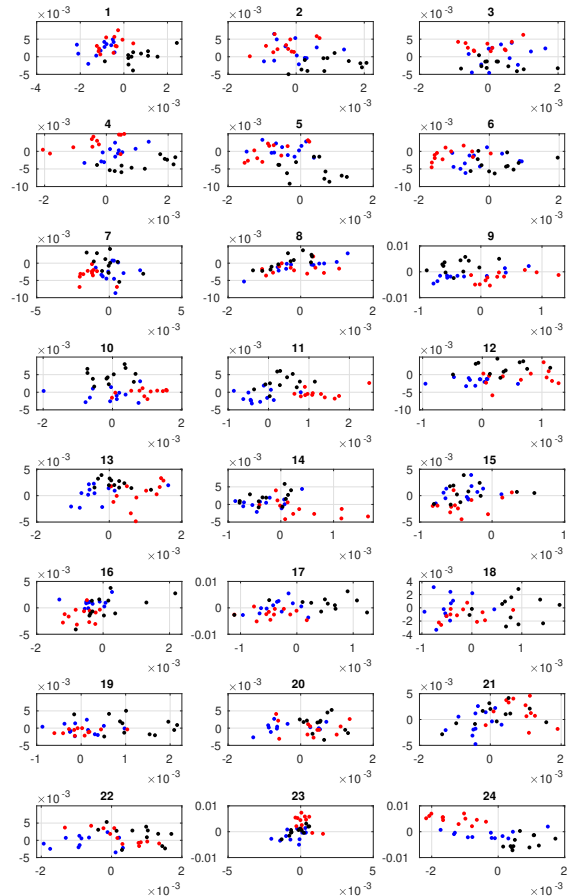


Figure 2. First principal components vs second principal components of the 24 hours. Groups: Control (blue), CH1 (red) and CH2 (black)
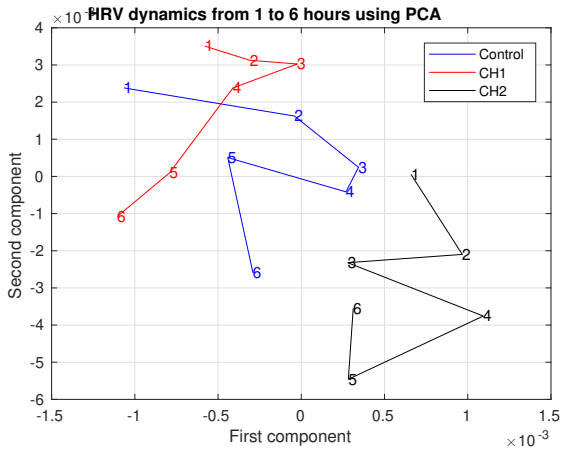
Figure 3. First principal components vs second principal components of the 24 hours, from 01 al 06 hours, groups: Control (blue), CH1 (red) and CH2 (black)



Figure 5. First principal components vs second principal components of the 24 hours, from 13 to 18 hours, groups: Control (blue), CH1 (red) and CH2 (black)



Figure 4. First principal components vs second principal components of the 24 hours, from 07 to 12 hours, groups: Control (blue), CH1 (red) and CH2 (black)
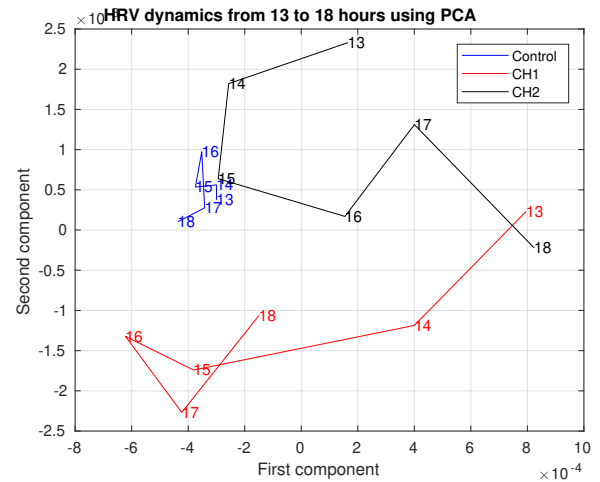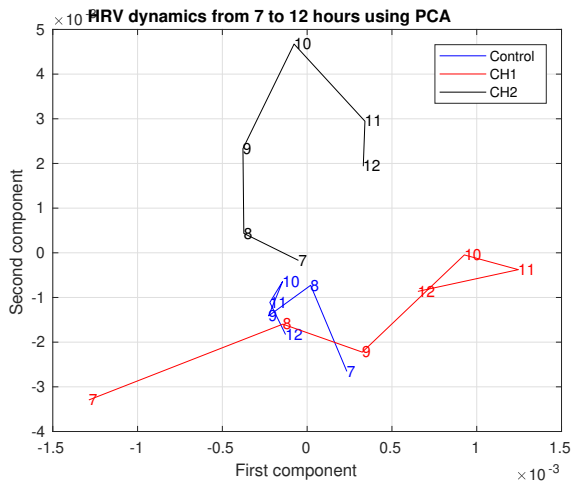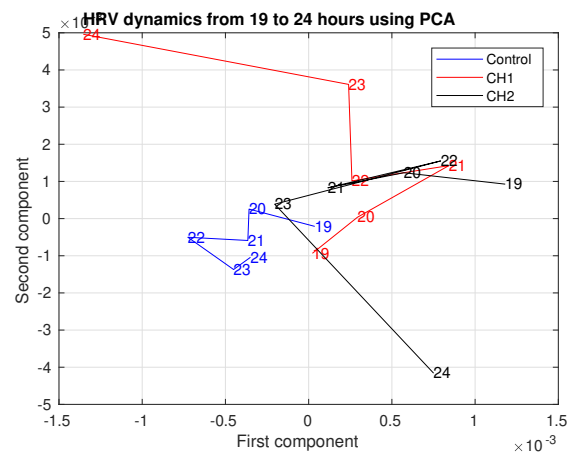


Figure 6. First principal components vs second principal components of the 24 hours, from 19 to 24 hours, groups: Control (blue), CH1 (red) and CH2 (black)

## 4. Results

### 4.1. PCA analysis

For the PCA analysis we use the average circadian profiles and the 8 HRV indices as feature of the three groups: Control, CH1 and CH2. The covariance matrix was calculated, then its eigenvalues which were ordered from greater to less with their respective eigenvectors, the first and second eigenvalue will have the greatest variance, that is to rotate and translate two orthogonal axes in the direction of greater variance. These axes are called components principal (PC).

On the first principal component (1st PC) and the second principal component (2nd PC) we project the 12 vectors corresponding to each hour of the three groups in such a way that we will have a representation of the HRV dynamics, this we can see in the figure 2 where we have 24 subfigures that correspond to 24 hours and in turn in each of the 12 points that correspond to every 5 minutes. The Kruskal Wallis test was used to find significant differences (p-value<0.05) between groups: Control-CH1, Control-CH2 and CH1-CH2 in 24 hours. For this, it was carried out between the 1st CP of the three groups in each hour, then it was done in the second CP, Table 1 shows the values of p-value less than <0.05 and with a "NS" where the values they are older. It can be seen that it is enough for

one of the principal components to have p-value<0.05 to be significantly different. We found significant differences between the three groups Control-CH1, Control-CH2 and CH1-CH2 up to 5 continuous hours. We can see that in the projections of the 1CP: Control-CH1 there are 6 continuous hours (9 to 14) where they are significantly different, also in Control-CH2 there are 4 continuous hours (17 to 20) where they are significantly different and in CH1-CH2 there are 15 hours where they are significantly different. Regarding the projections in 2CP: Control-CH2 there are 5 continuous hours (9 to 13) where they are significantly different, also in CH1-CH2 there are 15 continuous hours (1 to 15) where they are significantly different. These zones could be used to stratify risk. We calculate the

| | 1 CP | 2 CP | 1 CP | 2 CP | 1 CP | 2 CP |
|---|---|---|---|---|---|---|
| Hours | C-CH1 | C-CH1 | C-CH2 | C-CH2 | CH1-CH2 | CH1-CH2 |
| 1 | NS | NS | 0.0000 | 0.0284 | 0.0096 | 0.0017 |
| 2 | NS | NS | 0.01298 | 0.0096 | 0.0006 | 0.0000 |
| 3 | NS | NS | NS | NS | NS | 0.00006 |
| 4 | NS | NS | NS | 0.0145 | 0.0015 | 0.0000 |
| 5 | NS | NS | NS | 0.0001 | 0.0009 | 0.0007 |
| 6 | 0.0409 | NS | NS | NS | 0.0001 | 0.0269 |
| 7 | 0.0005 | NS | NS | NS | 0.0075 | 0.0229 |
| 8 | NS | NS | NS | NS | NS | 0.0255 |
| 9 | 0.0241 | NS | NS | 0.0070 | 0.0026 | 0.0009 |
| 10 | 0.0028 | NS | NS | 0.0000 | 0.0010 | 0.0005 |
| 11 | 0.0000 | NS | NS | 0.0001 | 0.0115 | 0.0066 |
| 12 | 0.0005 | NS | 0.0388 | 0.0000 | NS | 0.0193 |
| 13 | 0.0001 | NS | NS | 0.0122 | NS | 0.0205 |
| 14 | 0.0096 | NS | NS | NS | 0.0299 | 0.0034 |
| 15 | NS | 0.0090 | NS | NS | NS | 0.0055 |
| 16 | NS | 0.0048 | NS | NS | 0.0075 | NS |
| 17 | NS | NS | 0.0154 | NS | 0.0037 | 0.0030 |
| 18 | NS | NS | 0.0000 | NS | 0.0070 | NS |
| 19 | NS | NS | 0.0026 | NS | 0.0017 | NS |
| 20 | NS | NS | 0.0085 | NS | NS | NS |
| 21 | 0.0002 | NS | NS | NS | NS | NS |
| 22 | 0.0475 | NS | 0.0004 | NS | NS | NS |
| 23 | NS | 0.0005 | NS | NS | NS | NS |
| 24 | NS | 0.0037 | 0.0205 | NS | 0.0000 | 0.0000 |

Table 1. p-value < 0.05 the Kruskal Wallis test

centroid of each group in each hour and see its evolution in 24 hours that would be the descriptors of the dynamics of each group. We have identified four stages in the dynamic: from 01 to 06 hours (figure 3), from 07 to 12 (figure 4), from 13 to 18 hours (figure 5) and from 18 to 24 hours (figure 6).

Likewise, we calculate the total distance traveled for this we calculate the distance between each hour using the Euclidean distance for each descriptor in the 24 hours, finding that the descriptor of the control group was the lowest, then that of the CH1 group and finally that of the CH2 group, these results agree with the prediction error [6] where the group Control had a lower value, then the CH1 group and the highest was the CH2 group. $d_{Control} = 0.020267$ $d_{CH1} = 0.027922$ $d_{CH2} = 0.034812$

## 5.    Discussion and conclusion

We present a new methodology for the analysis of HRV in patients with chagas disease using the PCA, and introduce the descriptors of the dynamics and a measure is the distance traveled and that agrees with the results of the prediction error[6], where it would be measured regularity, that is, the Control group is more regular and this regularity is lost as the disease progresses

## Acknowledgments

## References

[1] World Health Organization The World Health Report 2017; 2:75,9–16 http://www.who.int/wer
[2] J. Pan and W. Tompkins, A Real-Time QRS Detection Algorithm IEEE Transactions on Biomedical Engineering, Vol. BME-32, NO. 3, March (1985).
[3] N Wessel, A Voss, Renormalised Entropy: A New Method of Non-Linear Dynamics for the Analysis of Heart Rate Variability, Computers in Cardiology 1994
[4] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Circulation, 73, 5, 1044 ( 1996 ).
[5] B. Schlkopf, S. Mika, C. J. Burges, P. Knirsch, K. R. Mller, G. Rtsch, & A. J. Smola, (1999). Input space versus feature space in kernel-based methods. IEEE transactions on neural networks, 10(5), 1000-1017.
[6] Vizcardo M, et.al., Nonlinear characterization of ECGs in patients with Chagas' disease, 2019, Biomedical Physics & Engineering Express, doi = 10.1088/2057-1976/ab03f7, IOP Publishing, vol 5, number 2, pag 025042.
[7] Vizcardo M., Ravelo A., Computing in Cardiology Conference (CinC),2018
[8] S. Guzzetti, D. Iosa, M. Pecis, L. Bonura, M. Prosdocimi and A. Malliani  Impaired heart rate variability in patients with chronic Chagas' disease. Am. Heart J. 1991;121:1727

Address for correspondence:

Miguel Vizcardo Cornejo
Escuela Profesional de Física, Av. Independencia sn, Arequipa, Perú
mvizcardocunsa.edu.pe