Principal Component Analysis of Heart Rate Variability Data in assessing Cardiac Autonomic Neuropathy

Mika P. Tarvainen[1], David J. Cornforth[2], and Herbert F. Jelinek[3]

Abstract— **Heart rate variability (HRV) is recognized to carry early diagnostic value regarding cardiac autonomic neuropathy (CAN). A number of different HRV analysis algorithms have been proposed for the assessment of CAN, each of them providing partly differing information about HRV time series. Instead of confining to a limited set of HRV features, a multi-dimensional approach incorporating a multitude of HRV parameters could be an optimal way of assessing the changes in HRV related to CAN. In this paper, principal component analysis (PCA) is used for analysing multi-dimensional HRV data of 11 patients with definite CAN and 71 subjects without CAN. Using the two most significant principal components, patients with CAN were separated from subjects without CAN with 87% accuracy.**

I. INTRODUCTION

Heart rate variability (HRV) is commonly used in assessing the functioning of cardiac autonomic regulation. The autonomic nervous system (ANS) regulates heart rate (HR) through sympathetic and parasympathetic branches. Sympathetic activity increases HR and decreases HRV, whereas parasympathetic activity decreases HR and increases HRV. The low frequency (LF, ranging from 0.04-0.15 Hz) component of HRV is influenced by both sympathetic and parasympathetic nervous activities, whereas the high frequency (HF, 0.15-0.4 Hz) component originates solely from parasympathetic nervous activity [1].

HRV is reduced in diabetes mellitus (DM) patients, suggesting dysfunction of cardiac autonomic regulation. Early assessment of cardiac autonomic neuropathy (CAN) and intervention are important for risk stratification and early treatment in preventing sudden cardiac death in diabetic patients. While HRV is recognized to carry early diagnostic value regarding CAN, reduction of HRV has been observed also in patients without clinical evidence of CAN [1], [2]. For the assessment of CAN using HRV analysis, standard time and frequency-domain methods as well as different nonlinear methods have been proposed [3], [4], [5], [6].

It is important to understand that different HRV analysis algorithms can provide different information about the HRV time series. How a pathology effects characteristics of the measured time series may differ between types of pathologies but also between individuals, and thus, use of a limited set of HRV analysis algorithms is not an optimal approach for assessing these changes. It is better to use a multi-dimensional approach incorporating a number of different HRV features when assessing the changes in ANS with respect to changes associated with pathology such as CAN.

In this paper, principal component analysis (PCA) is applied for multi-dimensional HRV data of patients with definite CAN and subjects without CAN.

II. METHODS

A. Subjects and HRV data

The study population consisted of 11 patients with definite CAN (CAN group) and 71 subjects without CAN (Normal group), recruited at the health screening clinic at Charles Sturt University, Australia [7]. CAN was identified using the Ewing battery of tests [8]. According to this test, definite CAN is indicated by any two of three heart rate parameters being abnormal (Valsalva, deep breathing or lying to standing) plus one of the two blood pressure measures (handgrip or lying to standing). Standard exclusion criteria were applied to exclude those with severe heart disease, kidney disease or polypharmacy. All subjects were comparable for age, gender and heart rate.

A 20-minute resting electrocardiogram (ECG) was recorded from all subjects. The ECG was recorded using lead II configuration at 400 Hz sampling rate (Maclab ADInstruments, Australia). For analysis, a 15-minute segment was taken from the middle of the record of 20-minute to avoid movement artifacts at the beginning and end of the recording. The RR interval time series were extracted from the ECG using the QRS detection algorithm of Kubios HRV software [9]. Frequencies below the LF band (below 0.04 Hz) were filtered out from the RR time series by using smoothness priors detrending [10]. Furthermore, the RR interval time series were interpolated using 4 Hz cubic spline interpolation to have evenly sampled data for spectral analysis.

The study was approved by the Charles Sturt University Human Ethics Committee and written informed consent was obtained from all participants.

B. HRV analysis

Several different time-domain, frequency-domain and nonlinear HRV parameters were computed following the guidelines given in [1]. The parameters selected for this study are briefly described below.

Most of the time-domain parameters are computed straight from the RR time series. These include the mean RR interval, standard deviation of normal-to-normal RR intervals

[1]M.P. Tarvainen is with the Department of Applied Physics, University of Eastern Finland and Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, 70211 Kuopio, Finland (Phone: +358 40 355 2369, Email: mika.tarvainen@uef.fi).

[2]D. Cornforth is with the Applied Informatics Research group, University of Newcastle, Australia.

[3] H.F. Jelinek is with the Centre for Research in Complex Systems and School of Community Health, Charles Sturt University, Albury, Australia.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

(SDNN), root mean square of successive RR interval differences (RMSSD) and percentage of successive RR intervals that differ by more than 50 ms (pNN50). In addition, two parameters computed from the RR interval histogram were considered. The HRV triangular index (HRVtri) is the integral of the RR interval histogram (i.e. the number of RR intervals within the time series) divided by the height of the histogram (i.e. number of RR intervals at modal bin). The triangular interpolation of RR interval histogram (TINN) is the baseline baseline width of a triangle fitted to the histogram.

Frequency-domain parameters, such as low frequency (LF), high frequency (HF) and total spectral power, were extracted from power spectral density estimates of the RR interval time series. The LF (ranging from 0.04–0.15 Hz) and HF (ranging from 0.15–0.5 Hz) component powers were computed both in absolute units (ms^2) and in normalized units (n.u.). Normalized powers were obtained by dividing the absolute powers with total spectral power and multiplying by 100 to give values as a percentage. In addition, the LF/HF power ratio was also computed.

Several nonlinear methods were also applied on the RR time series data. The Poincaré plot is a simple scatter plot and provides indexes for short term variability (SD1) and long term variability (SD2) which are both nonlinearly connected to time-domain parameters [11]. Sample entropy (SampEn) [12] and correlation dimension (D2) [15] are both measures of signal complexity. Detrended fluctuation analysis (DFA) measures correlations of short term (α_1, within range 4-16 beats) and long term (α_2, within range 16-64 beats) fluctuations within the RR time series [13]. Recurrence plot analysis (RPA) is an approach to measure recurring structures within the time series, where the number of recurrences is quantified by parameters such as recurrence rate (REC), determinism (DET) and mean diagonal line length (Lmean) [14]. A multi-scale Renyi entropy $H(\alpha)$ was computed according to

$$H_\alpha(\mathbf{x}) = \frac{1}{1-\alpha}\log_2\left(\sum_{i=1}^{N} p(x_i)^\alpha\right) \tag{1}$$

for orders $\alpha = -5 \ldots 5$. Above, p is the probability density function and $p(x_i)$ is the probability for $\mathbf{x} = x_i$. The probabilities were estimated using the methods outlined in [16].

C. Principal component analysis

Principal component analysis (PCA) is a multivariate statistical procedure where the random observations are transformed into a smaller set of uncorrelated variables called principal components (PCs) [17]. In other words, the original data is presented as a weighted sum of orthogonal basis vectors, where the basis vectors are the eigenvectors of data covariance or correlation matrix and the weights are the PCs. Typical applications of PCA include data reduction, feature extraction and visualization of multidimensional data.

In this paper, PCA is used for analyzing correlations between 32 HRV parameters and for visualizing the multidimensional HRV data. Let us denote the HRV data (M parameters for N subjects) in matrix form as

$$Z = \begin{pmatrix} z_1^{(1)} & z_2^{(1)} & \cdots & z_M^{(1)} \\ z_1^{(2)} & z_2^{(2)} & \cdots & z_M^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(N)} & z_2^{(N)} & \cdots & z_M^{(N)} \end{pmatrix} \tag{2}$$

where $z_i^{(j)}$ is the value of the ith HRV parameter for the jth subject. Before PCA, HRV parameters were normalized such that the median of each parameter over the N subjects was 0 and width of the 50% confidence interval equal to 1.

The data correlation matrix R was then estimated as

$$R = \frac{1}{N} Z^T Z \tag{3}$$

and eigenvectors of R solved from eigendecomposition. When K most significant eigenvectors (corresponding to largest eigenvalues) are taken as basis vectors, i.e. $H = (v_1, v_2, \ldots, v_K) \in \mathbb{R}^{M\times K}$, the first K PCs can be solved using least squares estimation as

$$\hat{\theta}^{\mathrm{PC}} = H^T Z^T \tag{4}$$

where $\hat{\theta}^{\mathrm{PC}} \in \mathbb{R}^{K\times N}$ such that the kth row includes the kth PCs (weights related to kth eigenvector) for the N subjects.

III. RESULTS

Most of the HRV parameters showed a clear difference between the Normal group (subjects without CAN) and CAN group (subjects diagnosed with definite CAN using Ewing battery of tests) as can be seen in Fig. 1. HRV was overall reduced in CAN shown by decreased SDNN, RMSSD, pNN50, HRVtri, TINN, LF and HF power as well as total spectral power, and SD1 and SD2 in the CAN group. No difference between the groups was observed in normalized LF and HF powers or in LF/HF ratio. Correlation dimension D_2 was lower for the CAN group and Renyi entropy values in the CAN group were lower for negative and higher for positive orders.

The HRV data shown in Fig. 1 was then written in matrix form according to equation (2) and the correlation matrix was estimated using (3). Furthermore, the most significant eigenvectors (corresponding to largest eigenvalues) were computed using eigendecomposition. The correlation matrix and the two most significant eigenvectors are illustrated in Fig. 2. Positive correlations are observed between most of the time-domain parameters and absolute power values. In addition, positive correlations are observed within negative scales of Renyi entropy as well as within positive scales.

The two eigenvectors illustrated in Fig. 2 were able to represent ∼75% of the variance in the HRV data, whereas the first six eigenvectors were able to represent as much as ∼ 95% of the variance. In order to evaluate how these eigenvectors fit into individual data, the first six PCs were computed for all subjects using (4) and the values of these

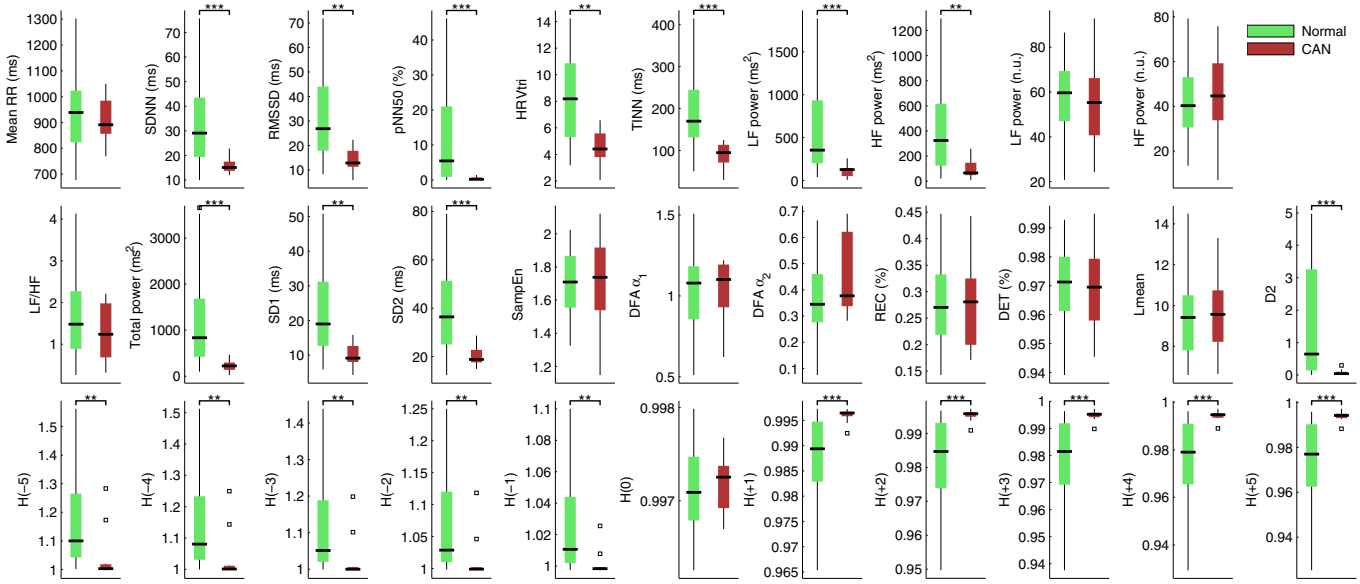

Fig. 1. Box plots of HRV parameters for Normal and CAN groups. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme parameter values excluding outliers (plotted as squares). Significant differences between the groups were tested using the Wilcoxon rank sum test (*$p \leq 0.05$, **$p \leq 0.001$, ***$p \leq 0.0001$).

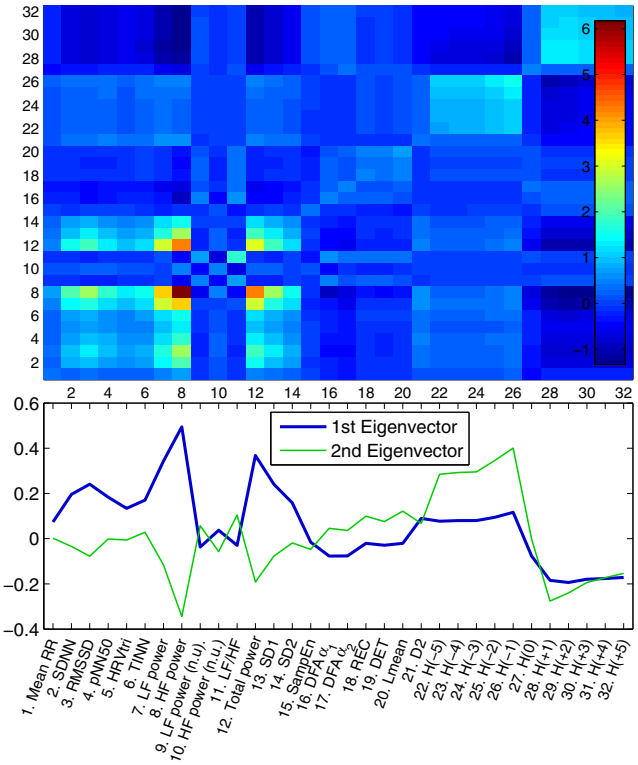

Fig. 2. Correlation matrix of HRV data (top) and the two most significant eigenvectors (bottom).

PCs are shown in Fig. 3. The first two PCs are both smaller for CAN group when compared to Normal group, which indicates that these two PCs can potentially separate the two groups.

Representations of the HRV data using the first two PCs

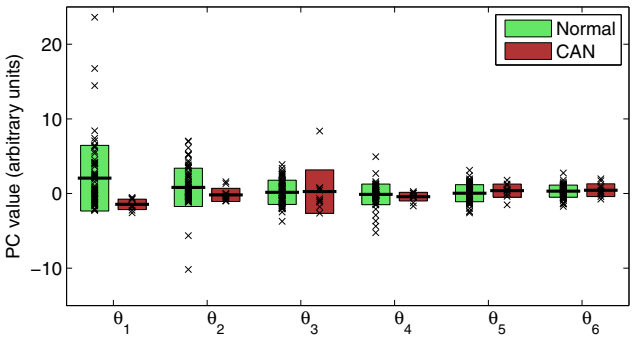

Fig. 3. First six PCs for Normal and CAN groups. The bars indicate Mean±SD intervals and crosses (×) indicate individual values.

(corresponding to two most significant eigenvectors) are shown in Fig. 4. Fig. 4A shows the results when PCA was applied to all HRV parameters and Fig 4B the results when only those HRV parameters showing a significant difference between the groups in Fig. 1 were used for PCA.

IV. DISCUSSION

Principal component analysis of a number of HRV parameters, each providing at least partly differing information about the RR time series, was presented. Using data from 11 patients with definite CAN and 71 subjects without CAN the effects of CAN on different HRV parameters was evaluated. Most of the HRV parameters were significantly different between the groups (CAN vs. Normal). HRV was overall reduced in CAN patients indicated by lowered time-domain HRV parameters and spectral powers. In addition, Renyi entropy and the correlation dimension were clearly able to capture the HRV changes related to CAN.

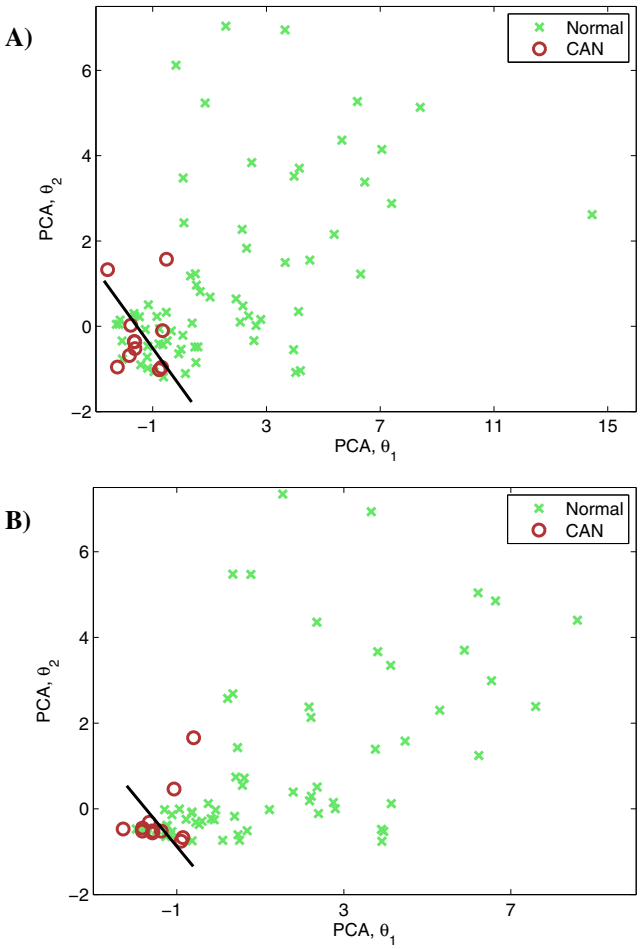

Fig. 4. Representation of HRV data using first two PCs, when A) all HRV parameters are included in the PCA and B) when only those HRV parameters, which showed significant difference between the groups, are included in the PCA.

The correlation matrix of the multi-dimensional HRV data revealed that there are positive correlations between most of the time-domain variables and absolute spectral powers. This was expected because all these variables measure the strength of the variability in one way or another. Weak positive (negative) correlations were also observed between the above-mentioned variability measures and negative (positive) orders of Renyi entropy. The weak correlation indicates that CAN is not represented only as lowered variability, but it also effects RR interval time series characteristics in a way detectable by Renyi entropy.

Due to the fact that many of the HRV parameters were correlated, the multi-dimensional data were possible to present using only the two most significant PCs (which covered ∼75% of the variance of HRV data). By adopting simple linear discrimination on these PCs a good separation between CAN and Normal groups is obtained (see Figs. 4A and 4B). In Fig. 4A (Fig. 4B), 8/11 (7/11) CAN and 11/71 (7/71) Normal subjects are situated below the visually defined discriminant line, this corresponds to 83% (87%) discrimination accuracy.

In conclusion, dimensionality of multi-dimensional HRV data can be reduced using PCA. The first few PCs and the corresponding eigenvectors are able to model most of the information in the original multi-dimensional data, and thus, these PCs can be used to visualize the original data in lower dimensions, e.g. in 2D.

References

[1] Task force of the European society of cardiology and the North American society of pacing and electrophysiology. Heart rate variability standards of measurement, physiological interpretation, and clinical use, Circulation, vol. 93, no. 5, pp. 10431065, 1996.

[2] P. Poirier, P. Bogaty, F. Philippon, C. Garneau, C. Fortin, and J. Dumesnil. Preclinical diabetic cardiomyopathy: relation of left ventricular diastolic dysfunction to cardiac autonomic neuropathy in men with uncomplicated well-controlled type 2 diabetes, Metabolism, vol. 52, no. 8, pp. 10561061, 2003.

[3] R. Freeman, J. Saul, M. Roberts, R. Berger, C. Broadbridge, and J. Cohen. Spectral analysis of heart rate in diabetic autonomic neuropathy, Arch Neurology, vol. 48, no. 2, pp. 185190, 1991.

[4] T. Laitinen, I. Vauhkonen, L. Niskanen, J. Hartikainen, E. Lansimies, M. Uusitupa, and M. Laakso. Power spectral analysis of heart rate variability during hyperinsulinemia in nondiabetic offspring of type 2 diabetic patients: evidence for possible early autonomic dysfunction in insulin-resistant subjects, Diabetes, vol. 48, no. 6, pp. 12951299, 1999.

[5] A.H. Khandoker, H.F. Jelinek, and M. Palaniswami. Identifying diabetic patients with cardiac autonomic neuropathy by heart rate complexity analysis, Biomed Eng Online, vol. 8, no. 3, 2009.

[6] D.J. Cornforth, M.P. Tarvainen, and H.F. Jelinek. Using Renyi entropy to detect early cardiac autonomic neuropathy, in Proc 35th Ann Int Conf IEEE EMBS, Osaka, July, 2013.

[7] H.F. Jelinek, C. Wilding, and P. Tinley. An innovative multidisciplinary diabetes complications screening programme in a rural community: A description and preliminary results of the screening. Australian Journal of Primary Health, vol. 12, pp. 14-20, 2006.

[8] D.J. Ewing and B.F. Clarke. Diabetic autonomic neuropathy: present insights and future prospects. Diabetes Care, vol. 9, pp. 648-665, 1986.

[9] M.P. Tarvainen, J.-P. Niskanen, J.A. Lipponen and P.O. Ranta-aho, and P.A. Karjalainen. Kubios HRV - Heart rate variability analysis software, Comp Meth Programs Biomedicine, vol. 113, no. 1, pp. 210-220, 2014.

[10] M.P. Tarvainen, P.O. Ranta-aho, and P.A. Karjalainen. An advanced detrending method with application to HRV analysis, IEEE Trans Biomed Eng, vol. 49, no. 2, pp. 172175, 2002.

[11] M. Brennan, M. Palaniswami, and P. Kamen. Do existing measures of Poincaréplot geometry reflect nonlinear features of heart rate variability, IEEE Trans Biomed Eng, vol. 48, no. 11, pp. 13421347, 2001.

[12] J. Richman and J. Moorman. Physiological time-series analysis using approximate entropy and sample entropy, Am J Physiol, vol. 278, pp. H2039H2049, 2000.

[13] C.-K. Peng, S. Havlin, H. Stanley, and A. Goldberger. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos, vol. 5, pp. 8287, 1995.

[14] N. Marwan, M.C. Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems, Physics Reports, vol. 438, no. 5-6, pp. 237-329, 2007.

[15] P. Grassberger and I. Procaccia. Characterization of strange attractors, Phys Rev Lett, vol. 50, pp. 346349, 1983.

[16] D. Lake. Renyi entropy measures of heart rate Gaussianity, IEEE Trans Biomed Eng, vol. 53, pp. 2127, 2006.

[17] I. Jolliffe. Principal Component Analysis. Berlin, Germany: Springer-Verlag, 1986.