

# The Effect of Automated Preprocessing of RR Interval Tachogram on Discrimination Capability of Heart Rate Variability Parameters

Faezeh Marzbanrad<sup>1</sup>, Herbert Jelinek<sup>2,3</sup>, Ethan Ng<sup>4</sup>, Mikhail Tamayo<sup>4</sup>, Brett Hambly<sup>4</sup>, Craig McLachlan<sup>4</sup>, Slade Matthews<sup>4</sup>, Marimuthu Palaniswami<sup>1</sup>, Ahsan Khandoker<sup>1,2</sup>

<sup>1</sup>The University of Melbourne, Parkville, Australia

<sup>2</sup>Khalifa University, Abu Dhabi, UAE

<sup>3</sup>Charles Sturt University, Albury, Australia

<sup>4</sup>The University of Sydney, Camperdown, Australia

## Abstract

*Heart Rate Variability (HRV) has been extensively investigated for characterizing the autonomic nervous system (ANS) in controlling heart rate. Since ectopic beats, artefacts and noise of the ECG can affect the estimation of HRV features, pre-processing of the RR tachogram can improve the accuracy of HRV analysis and discriminatory power. This paper investigates the effect of different automated preprocessing methods on discriminatory capability of HRV analysis with an example of comparison between different groups of normal and type II diabetic patients with different Angiotensin-Converting Enzyme (ACE) gene polymorphism. Results show that smaller p-values and therefore higher discriminatory capability are found when preprocessing is used, while none of the features can show significant difference if they are estimated from the raw R-R sequence. Secondly, the preprocessing methods do not have the same effect for all HRV features.*

## 1. Introduction

Heart rate variability (HRV) analysis is the basis of various fields of research. Ectopic beats and noise within the ECG can change the estimated value of HRV features and reduce classification accuracy. Pre-processing is aimed to remove noise, artefacts and the beats which are not sourced from sinus node, defined as Non-Sinus Intervals (NSI). Although pre-processing of the RR tachogram can improve the accuracy of HRV analysis and discriminatory power, it is often not performed. Manual techniques for pre-processing are time consuming, difficult to reproduce and diverse automated methods that have been applied do not improve outcomes of HRV analysis.

In this paper, a practical example shows how different automated preprocessing methods can improve

discrimination using HRV features. HRV features are estimated without preprocessing and applying three different automated preprocessing methods. The first method is the exclusion of RR intervals with the beat to beat variation of more than 10%. The exclusion of more than 20% were used in various studies [1-3] and classification of the NSIs based on variation of more than 10%, 20% and 30% were analysed in [3]. The second automated method is to take five consecutive intervals and replace the middle RR interval by the mean of the two preceding and two following RR intervals, if they differ by more than 20%. The third method is adapted from an adaptive preprocessing technique proposed in [4]. In this method the false beats are identified in an adaptive scheme based on a criteria obtained from filtered tachogram.

Different frequency domain and time domain HRV features, tone and entropy, Poincaré indices were analysed for four groups of patients comprising diabetic and control patients, with different Angiotensin-Converting Enzyme (ACE) gene polymorphism corresponding to insertion (I) or deletion (D). The groups may have either D/D or one of the I/D and I/I ACE genotypes.

## 2. Methods

### 2.1. Subjects

Data including ECG recordings, ACE genotypes, diabetes details, other demographic information and biochemistry test results were collected from 231 participants at Charles Sturt University Diabetes Complications Screening Initiative (DiScRi). Written, informed consent approved by the respective local clinical research ethics committees, was obtained prior to be considered in this study.

Since only Type II diabetes (T2D) and normal subjects

were of interest, five cases with type I diabetes or pre-diabetes were excluded. Equal length of 500 samples were analysed for tachogram of each subject, therefore five cases were excluded because of insufficient tachogram length and finally 221 subjects remained for analysis.

## 2.2. Genotyping

The QIAamp DA blood mini kit (from Qiagen) is used to extract genomic DNA from frozen blood samples, according to the manufacturer's instructions. 75 microliter of elution buffer was used instead of 200 microliter as the only modification to the protocol in order to obtain a more concentrated solution of DNA. The extracted DNA was then genotyped using the triple primer method [5] and electrophoresis using a 6% polyacrylamide gel.

## 2.3. Groups

221 subjects were divided four groups according to their ACE genotypes and diabetes status; all subjects were divided into groups of ACE D/D and ACE "I/D + II" types each with subdivisions of T2D and control. The groups and the number of subjects in each group are summarized in table 1.

Table 1. The number of patients in each group.

| Groups      | ACE: D/D |     | ACE: I/D + II |     |
|-------------|----------|-----|---------------|-----|
|             | Control  | T2D | Control       | T2D |
| No of cases | 37       | 16  | 119           | 49  |

## 2.4. HRV analysis

ECGs were recorded by trained staff on Macintosh Chart version 5 with the sampling rate of 400 Hz and a notch filter at 50 Hz (ADInstruments, Sydney) using a lead II configuration [6]. Data were collected between 9-11 am for a period of 10 minutes while the patient was resting in a semisupine position. Patients rested for 5 minutes before measurement. The ECGs were edited using the MLS310 HRV module as part of the Maclab Chart recording software (ADInstruments, Sydney).

Using the recorded ECG, the RR intervals were identified as the time difference between two successive R peaks and instantaneous heart rate (IHR) in beats per min (bpm) was determined as 60/RR intervals in seconds based on the algorithm proposed by Pan and Tomkin [7]. In order to maintain consistency, a fixed length of 500 RR intervals samples was used for each subject.

### Pre-processing

Three different methods were applied to the RR interval sequences, before calculating HRV parameters.

**(Method A)** In the first approach, each interval which was deviated by more than 10% of the preceding interval was marked as false beat. After identification of the false beats, they can be excluded or replaced by interpolation. Removal of the false beats results in loss of time dependency of the tachogram, therefore in the first approach the removed beat is replaced by the mean value of the two adjacent beats.

**(Method B)** In the second method, every five consecutive samples of the RR intervals were taken and the mean of all samples excluding the middle sample was calculated. Then this mean value was compared with the middle sample and if they differed by more than 20% of the mean, the middle sample was replaced by the mean of the two preceding and two following samples.

**(Method C)** The third method is comprised of three procedures: deletion of the obvious errors, adaptive percent filter and adaptive controlling filter. The details of the procedures were found in [4] but can be summarized as follows. In the first step the R-R intervals of less than 200 ms length were removed. In the next step, the sequence is filtered by a binomial-7-filter:

$$t_n = \frac{x_{n-3} + 6x_{n-2} + 15x_{n-1} + 20x_n + 15x_{n+1} + 6x_{n+2} + x_{n+3}}{64} \quad (1)$$

The adaptive mean  $\mu_a(n)$  and adaptive SD  $\sigma_a(n)$  values corresponding to each sample of RR sequence are then obtained from the filtered sequence.

At each stage if an RR sample differs from the previous one and the last valid sample by more than 10% plus a generalized average 3 fold of the averaged adaptive SD, it is marked as false. Since the interpolation method for replacement of the removed false beat may result in misleading beat to beat variability indices, in the third method each false beat is replaced by a random number from  $[\mu_a(n) - \frac{1}{2}\sigma_a(n), \mu_a(n) + \frac{1}{2}\sigma_a(n)]$ .

In the last step, the resulting sequence is again filtered and the mean and SD are calculated for each sample. Then if one sample differs from the corresponding mean value for that sample by more than 3 fold SD plus basic variability (here 20 ms) then it is replaced by the respective sample of the filtered sequence.

### HRV parameters

The standard deviation of the processed RR intervals (SDNN) and the square root of the mean squared difference of the successive heart rate samples (RMSSD) were calculated and compared for different groups. Spectral powers in low frequency (LF) band (0.04–0.15 Hz), high frequency (HF) band (0.15–0.40 Hz) were also calculated. Then power in each frequency band and the ratio of low to high frequency power and total power (TP) were analysed. Further details about these methods were

described in [8].

Poincaré indices of RR intervals were also analysed in this paper. The Poincaré plot is a practical way to visualize the dynamic systems in two dimensions based on the intuitive display of the dynamic properties of a system from a time series. Poincare plot is characterized by two indices; the width (SD1) and the length (SD2) of the plot which represent long and short-term variability of the nonlinear dynamic system [9]. For each subject, SD1, SD2 and their ratio SD1/SD2 of the RR intervals are calculated and used for further analysis.

Tone and entropy are the other parameters which were analysed. The RR interval time series is defined as follows:

$$RR \equiv (RR_1, RR_2, \dots, RR_N) \quad (2)$$

where, N is the total number of RR intervals. The difference of the consecutive RR intervals is associated with heart rate acceleration or inhibition. For example, a heart rate acceleration occurs when  $RR_{i+1}$  becomes shorter than  $RR_i$ . Namely, heart rate acceleration and inhibition are represented by plus and minus difference of RR intervals, respectively. Normalized variation in RR interval is preferred to examine the variability, in order to diminish the impact of heart rate variation over a long period of time and different subjects. In Tone-Entropy analysis, the percentage of the difference of two successive RR intervals divided by the first one is defined as the percentage index (PI):

$$PI(i) = \frac{RR_i - RR_{i+1}}{RR_i} \times 100 \quad (3)$$

The first order moment (arithmetic average) of the PI is derived as Tone:

$$T = \frac{1}{N-1} \sum_{i=1}^{N-1} PI(i) \quad (4)$$

Tone shows the balance between accelerations ( $PI > 0$ ) and inhibitions ( $PI < 0$ ) of the heart rate and it is closely associated with the sympatho-vagal balance as shown in previous studies [10,11].

Entropy is determined using the probability distribution of PI, based on Shannon's formula [12]:

$$E = -\sum_{i=1}^n p(i) \log_2 p(i) \quad (5)$$

where  $p(i)$  is a probability of PI with values in the range of  $i < PI < i+1$ , ( $i$  is an integer). The entropy evaluates total acceleration/inhibition or total variations of heart period, in the unit of bit.

## 2.5. Statistical analysis

Because the distribution of the variables is non-Gaussian as ascertained by the Chi-square goodness-of-fit test, non-parametric statistical analysis is used. Kruskal-Wallis test is employed to compare HRV parameters across different

groups. The p-value of  $p < 0.05$  is assumed significant.

## 3. Results

An example of RR interval tachogram preprocessed by three different methods is shown in figure 1. As it is inferred from the figure, if there is no abnormal jump or drop in the sequence, the preprocessing methods do not change the sequence. But in the case of an abnormal change, different methods modify the abnormal samples in different ways.

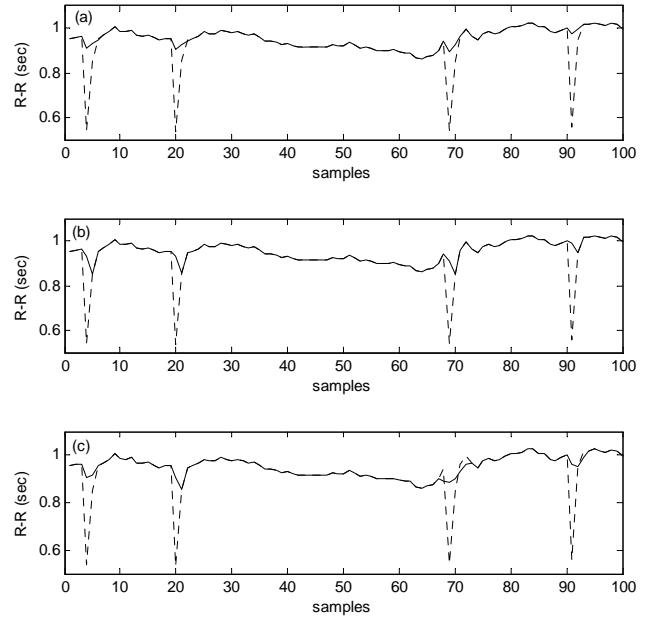


Figure 1 An example of preprocessing of a R-R interval tachogram. Dashed traces show the raw R-R sequence. Solid traces illustrate the processed sequence by (a) method A, (b) method B and (c) method C.

Kruskal-wallis was used to analyse the differences of parameters across groups. P-values are demonstrated in table 2.

## 4. Discussion

As it is evident from the results in table 2, preprocessing has a significant effect on discrimination based on HRV parameters. Without preprocessing, no HRV feature could uncover differences between different groups. For example the total power which has been interpreted as a discriminatory index of cardiac parasympathetic tone [8,13] is not significantly different across the groups if unprocessed RR tachogram is used for analysis. However when any of the three preprocessing methods is applied to the RR interval sequence, total power can demonstrate significant differences.

Table 2. Kruskal-Wallis p-value for four group comparison of HRV parameters computed based on the tachograms which are preprocessed by different methods (\* significant  $p < 0.05$ )

| HRV Features | Without Preprocessing | Method A | Method B | Method C |
|--------------|-----------------------|----------|----------|----------|
| SDNN         | 0.0928                | 0.0222*  | 0.0630   | 0.0411*  |
| RMSSD        | 0.1314                | 0.0596   | 0.1253   | 0.0383*  |
| LF           | 0.1442                | 0.0614   | 0.2120   | 0.0663   |
| HF           | 0.4887                | 0.6480   | 0.5266   | 0.5277   |
| LF/HF        | 0.4887                | 0.6480   | 0.5266   | 0.5277   |
| TP           | 0.0543                | 0.0128*  | 0.0378*  | 0.0110*  |
| SD1          | 0.1314                | 0.0596   | 0.1253   | 0.0383*  |
| SD2          | 0.0681                | 0.0201*  | 0.0569   | 0.0381*  |
| SD1/SD2      | 0.1000                | 0.2167   | 0.1001   | 0.1425   |
| tone         | 0.3274                | 0.3623   | 0.5953   | 0.6765   |
| entropy      | 0.2037                | 0.0693   | 0.0220*  | 0.0403*  |

On the other hand different preprocessing approaches result in different discrimination results, for example SDNN and SD2 are significantly different across groups when method A or C is used, however Method B results in higher p-values which do not show significance. Entropy is another feature which does not seem to demonstrate any significant differences when raw RR intervals are used ( $p=0.20$ ), while only method B and C can uncover significant differences.

The Method C could help to reveal variation of more parameters across the groups. RMSSd and SD1 are the features that could be used for discrimination of the groups only when method C was used for preprocessing. However part of our finding is dependent on how the HRV feature is calculated, whether it is sensitive to beat by beat variations or global change over several beats.

## 5. Conclusion

The effect of preprocessing on discrimination capability of HRV analysis was investigated with an example. According to the results, smaller p-values and therefore higher discriminatory capacity were found when preprocessing was used. For example, total power could show significant difference of the groups only when one of the preprocessing approaches was used. However the preprocessing methods do not have the same effect for all HRV features and depends on how the variation is analysed using each feature. Overall it is recommended to use appropriate preprocessing before obtaining each HRV feature to optimize classification.

## References

- [1] Pipilis A, Flather M, Ormerod O, Sleight P. Heart rate variability in acute myocardial infarction and its association with infarct site and clinical course. *The American journal of cardiology* 1991;67: 1137-1139.
- [2] Kleiger R E, Miller J P, Krone R J, Bigger J T. The independence of cycle length variability and exercise testing on predicting mortality of patients surviving acute myocardial infarction. *The American journal of cardiology* 1990; 65(7): 408-411.
- [3] Haaksma J, Dijk W A, Brouwer J, Van den Berg M P, Crijns H J G M, Lie K I. Effects of automatic ectopy exclusion on the analysis of heart rate variability using a percentile exclusion rule. *Computers in Cardiology* 1995; 22:197-200.
- [4] Wessel N, Voss A, Malberg H, Ziehm C, Voss H U, Schirdewan A, Meyerfeldt U, Kurths J. Nonlinear analysis of complex phenomena in cardiological data. *Herzschrittmachertherapie und Elektrophysiologie* 2000; 11(3): 159-173.
- [5] Evans AE, Poirier O, Kee F, Lecerf L, McCrum E, Falconer T, et al. Polymorphisms of the angiotensin converting-enzyme gene in subjects who die from coronary heart disease. *Q J Med* 1994; 87(4): 211-4.
- [6] ADInstruments. Powerlab. 1.0 ed. Australia.
- [7] Pan J, Tompkins WJ. Real time QRS detector algorithm. *IEEE Trans Biomed Eng* 1985; 32:230-323.
- [8] TFESC/NASPE. 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. *European Heart Journal* 1996;17(3):354-81.
- [9] Brennan M, Palaniswami M, Kamen P: Do existing measures of Poincaré' plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans Biomed Eng* 2001; 48:1342-1347.
- [10] Amano M, Oida E, Moritani T. Age-associated alteration of sympatho-vagal balance in a female population assessed through the tone-entropy analysis. *Eur J Appl Physiol* 2005;94:602-610
- [11] Oida E, Moritani T, Yamori Y. Tone-entropy analysis on cardiac recovery after dynamic exercise. *J Appl Physiol* 1997;82: 1794-1801
- [12] Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948; 27:379-423
- [13] Malliani A, Pagani M, Lombardi F, Cerutti S. Cardiovascular neural regulation explored in the frequency domain. *Circulation* 1991;84(2): 482-492.

Address for correspondence.

Dr Ahsan Khandoker  
Dept. of Biomedical Engineering  
Khalifa University of Science, Technology and Research  
(KUSTAR), Abu Dhabi, UAE.  
ahsan.khandoker@kustar.ac.ae