RESEARCH ARTICLE

PERFORMANCE OF NONLINEAR HEART RATE VARIABILITY PARAMETERS FOR ECG ARTIFACTS

Butta Singh

Department of Electronics and Communication Engineering, Guru Nanak Dev University, Regional Campus, Jalandhar, Punjab, India

ARTICLE INFO

Article History:
Received 14th, April, 2015
Received in revised form 23th, April, 2015
Accepted 13th, May, 2015
Published online 28th, May, 2015

ABSTRACT

In the measurement of bio signals associated with the heart rate, artifacts in the electrocardiogram (ECG) recordings deteriorate the data, yielding ECG artifacts; missing (incomplete) RR interval tachogram. The linear parameters of heart rate variability (HRV) are very sensitive to these missing RR intervals. In this study, the feat of nonlinear measures of HRV is investigated for missing RR interval data, using simulated missing data in real RR interval tachograms. For the simulation, randomly selected data (0–100 RR intervals) were removed from real RR data obtained from the MIT-BIH normal sinus rhythm database. All, 703 tachograms of 1000 RR interval data length were used for this analysis in Approximate entropy (ApEn), sample entropy (SampEn), Poincaré plot indices (SD1 and SD2) and Detrended fluctuation analysis (DFA) were calculated as the nonlinear parameters, and the relative errors between the original and the incomplete tachograms for these parameters were computed. The results of the simulation revealed that nonlinear parameters are more suitable measures than linear parameters of HRV in presence of missing RR interval data.

Key words:
HRV, Approximate entropy, Sample entropy, Detrended fluctuation analysis, Poincaré plot

Copyright © Butta Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The heart rate variability (HRV) is an extended tool to analyze the mechanisms controlling the cardiovascular system. It may be analyzed in either the time, frequency and by using nonlinear approach. The time-domain approach is simple and widely used for clinical applications. Time-domain HRV parameters are easy to calculate and can be computed directly by numerical approach based on tachograms of the RR-interval (Task Force of the ESC and the NASPE 1996; Singh and Singh 2012; Berntson and Bigger et al., 1997). Since calculations are performed directly in the beat domain, re-sampling is not necessary to derive the time-domain parameters. However, HRV spectral parameters, total frequency (TF), very low frequency (VLF), low frequency (LF) and high frequency (HF), are obtained by the sum of the power in the relevant frequency range in the spectrum; this is estimated from data that are regularly re-sampled in the time domain (Singh and Vinod et al., 2004; Singh and Bharti 2015). Biological systems such as the cardiovascular system are comprised of multiple subsystems that exhibit both highly nonlinear deterministic, as well as, stochastic characteristics, and subject to hierarchical regulations (Singh and Singh 2012; Hoyer, Schmidt, Bauer, Zwiener, Kohler, Luthke and Eiselt 1997; Singh and Singh 2011; Veicsteinas and Castiglioni 2006; Singh and Singh 2013). As a result, time series generated by biological systems are often highly nonlinear, non-stationary, random and complex. Therefore, standard linear measures of HRV are not able to detect subtle, but important changes in the heart rate time series. Since the linear parameters of HRV do not provide adequate information on the complexity that lies inside beat-to-beat variability, the application of nonlinear techniques is appropriate. Poincaré plot (Brennan, Palaniswami and Kamen 2001; Kamen, Krum and Tonkin 1996), approximate entropy (ApEn) (Pincus 1991; Singh et al., 2012; Pincus and Goldberger 1994), sample entropy (SampEn) (Lake, Richman, Griffin and Moorman 2002) and detrended fluctuation analysis (DFA) (Peng, Havlin, Hausdorff, Mietus, Stanley and Goldberger 1994; Goldberger, Amaral, Hausdorff, Ivanov, Peng and 2001) are the recently developed nonlinear techniques to quantify the nonlinearity of time series data like heart rate intervals.

In the measurement of biosignals associated with the heart rate, artifacts in the electrocardiogram (ECG) recordings deteriorate the RR interval tachogram, yielding incomplete data. For
example, ECGs obtained under surgical conditions contain electrical artifacts due to the effects of the electro surgical unit (ESU) and motion artifacts during exercise influence ECGs. These artifacts complicate the detection of feature points and yield incomplete RR-interval tachograms (Kim, Lim, Kim and Park 2007: Kim, Lim, Kim and Park 2009). Kim et al. (Kim, Lim, Kim and Park 2007: Kim, Lim, Kim and Park 2009) evaluated the effect of missing RR intervals on linear (time and frequency) domain HRV parameters. They found that missing RR interval affect significantly both time and frequency domain HRV parameters.

In this study, the effects of consecutive missing RR interval data on nonlinear HRV analysis are investigated by simulating missing data in RR interval tachograms recorded from healthy subjects.

Data
In this study, 18 long-term RR tachograms belonging to the MIT-BIH normal sinus rhythm database (http://www.physionet.org/physiobank/database/nsrdb) were used. The RR tachograms were extracted from annotations in the database in which the sampling rate for recording was 128 Hz. These data were recorded over 24 h from subjects who did not exhibit any significant arrhythmias; these subjects included 5 men (age: 26–45 years) and 13 women (age: 20–50 years). In all, 703 RR-interval data sets of data length N=1000 were collected for HRV analysis; they included only normal beats (Task Force of the ESC and the NASPE 1996). In each data set, consecutive RR-interval data were randomly selected for removal, and the data length removed was increased from 0 to 100 RR intervals in an increment of 5. Therefore, the number of data sets used in these simulations was 14 060 (=703 data sets × 20 missing data). Two random selections of RR-intervals were made to analyze the average effects of the missing data. In each case, the nonlinear HRV parameters were calculated; the evaluated nonlinear HRV parameters in this study are the Poincaré plot indices (SD1 and SD2), ApEn, SampEn and DFA. A total of 28.12 × 10^3 calculations were performed using MATLAB for each nonlinear parameter. The HRV measures the spontaneous variability between successive beats, as they are revealed by the presence of an R wave in the ECG surface signal. It has been shown that HRV signal changes can be related to the activity of several physiological control mechanisms of different nature. Their interaction produces changes in the beat rate assuring the control activity reacts efficiently to various incoming stimuli.

Nonlinear Analysis OF HRV
The development of the nonlinear dynamical system analysis has led to the introduction of a large amount of signal analysis techniques aimed at the extraction of nonlinear parameters from experimental time series. The original objective was the evaluation of the generating system characteristics in order to better understand its nature. In many cases however the generation system is unknown and the output signal is the only information we can have about the system itself. This is precisely the case of the human life support systems among which the heart plays a dominant role. It has been shown that HRV signal changes can be related to the activity of several physiological control mechanisms of different nature. Their interaction produces changes in the beat rate assuring the control activity reacts efficiently to various incoming stimuli.

Poincaré plot
Poincaré plot is a visual tool in which each RR interval is plotted as a function of previous RR interval. Analysis of RR intervals with the use of standard deviations, histograms and spectral techniques provide an assessment of overall variability but obscures instantaneous beat-to-beat changes However, Poincaré plot provides summary information as well as detailed beat-to-beat information on the behavior of heart. Beat-to-beat variation can be easily displayed for visual assessment by graphing of each RR interval against the subsequent RR interval. The problem regarding Poincaré plot use has been lack of obvious quantitative measures that characterize the salient features of Poincaré plots. To characterize the shape of the plot mathematically, most researchers have adopted the technique of fitting an ellipse to the plot, as shown in Figure 1. A set of axis oriented with the line of identity is defined (Brennan, Palaniswami and Kamen 2001). The axes of the Poincaré plot are related to the new set of axis by a rotation of θ =π/4 radian as shown in equation (i).

In the reference system of the new axis, the dispersion of the points around the X1-axis is measured by the standard deviation denoted by SD1. This quantity measures the width of the Poincaré cloud and, therefore, indicates the level of short-term HRV (Brennan, Palaniswami and Kamen 2001). The length of the cloud along the line of identity measures the long-term HRV and is measured by SD2, which is the standard deviation around the X2-axis (Kamen, Krum and Tonkin 1996; Brennan, Palaniswami and Kamen 2001). These measures are related to the standard HRV measures by equation (ii).

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} RR_n \\ RR_{n+1} \end{bmatrix}$$ \hspace{1cm} (i)

$$SD_1^2 = \text{Var}(x_1) = \text{Var}(\frac{1}{\sqrt{2}}RR_n - \frac{1}{\sqrt{2}}RR_{n+1}) = \frac{1}{2}\text{Var}(RR_n - RR_{n+1}) = \frac{1}{2}\text{SD}_2^2$$ \hspace{1cm} (ii)
Where $\text{Var}(x)$ denotes the variance of $x$ sequence and $SDSD$ denotes the standard deviation of successive differences of RR interval series. Thus, the $SD_1$ measure of Poincaré width is equivalent to the standard deviation of the successive difference of intervals, except that it is scaled by $\sqrt{7}$. Further we can relate $SD_1$ to the autocovariance function by equation (iii).

$$SD_1^2 = \Phi_{RR}(0) - \Phi_{RR}(1)$$

... (iii)

where $\Phi_{RR}(0)$ and $\Phi_{RR}(1)$ are the autocovariance functions.

Also $\Phi_{RR}(0) = E[(RR_i - \bar{RR})^2]$ i.e., variance of RR intervals. Similarly,

$$SDSD = \text{RMSSD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_i - RR_{i+1})^2}$$

but $\bar{RR}_{n}^2=0.$ for stationary intervals. Thus

$$SDSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_i - RR_{i+1})^2}$$

With a similar argument, it may be shown that the length of the Poincaré cloud is related to the autocovariance function by equation (iv)

$$SD_2^2 = \Phi_{RR}(0) + \Phi_{RR}(1)$$

... (iv)

By adding (iii) and (iv) together, we get

$$SD_1^2 + SD_2^2 = 2 \times SDRR^2$$

... (v)

where $SDRR$ denotes the standard deviation of RR interval series. Finally

$$SD_2^2 = 2 \times SDRR^2 - \frac{1}{2} SDSD^2$$

... (vi)

Thus equation (vi) represents $SD_2$ in terms of existing indices of HRV. Fitting an ellipse to the Poincaré plot does not generate indices that are independent of the standard time domain HRV indices.

**Approximate Entropy**

Approximate entropy (ApEn) is a statistical index to quantify the complexity of a signal. It has been widely adopted by many researchers especially in the field of heart rate variability. The popularity of approximate entropy stems from its capability to provide quantitative information about the complexity of the experimental data that are short in data length (Pincus 1991; Pincus and Goldberger 1994). ApEn measures the (logarithmic) likelihood that runs of patterns that are close for $m$ observations remain close on next incremental comparison. Greater likelihood of remaining close, i.e., high regularity, produces smaller ApEn values (Pincus 1991; Pincus and Goldberger 1994). While implementing ApEn, calculation requires a priori specification of two unknown parameters: $m$, the embedding dimension and $r$, a threshold, which is in effect a noise filter. Pincus (1991), who developed the ApEn method, suggested that $r$ should be 0.1 to 0.25 times, the standard deviation of the data, and that $m$ be 1 or 2 for data lengths $N$ ranging from 100 to 5,000 data points. Given a signal $u(1), u(2), \ldots u(N)$, where, $N$ is the total number of data points. Fix $m$, a positive integer and $r$, a positive real number. For our study we have choose $r$ equal to 20% of standard deviation and $m = 2$. ApEn algorithm can be summarized as follows.

1. Form $m$ vectors $X(1)$ to $X(N-m+1)$ defined by $X(i) = [u(i), u(i+1), \ldots u(i+(m-1))]$ $1 \leq i \leq N-m+1$
2. Define the distance $d[X(i), X(j)]$ between the vectors $X(i)$ and $X(j)$ as the maximum absolute difference between their respective scalar components:

$$d[X(i), X(j)] = \max_{k=1, 2, \ldots, m} |u(i+k-1) - u(j+k-1)|$$

3. Define for each $i$, for $i=1, 2, \ldots, N-m$:

$$C_i^r(r) = \frac{v^r(i)}{N-m+1}$$

Where $v^r(i) =$ number of $d[X(i), X(j)] \leq r$
4. Take the natural logarithm of each $C_i^r(r)$ and average it over $i$

$$\phi^r = \frac{1}{N-m+1} \sum_{i=1}^{N-m} \ln(C_i^r(r))$$

5. Increase the dimension to $m+1$ and repeat steps 1 to 4.
6. Calculate ApEn value for a finite data length of $N$:

$$\text{ApEn}(m, r, N) = \phi^r - \phi^{r+1}(r)$$

A high degree of regularity means that sequences, which are similar for $m$ points, are likely to be similar for the next $m+1$ points, while this is unlikely to occur for irregular time series. Thus low values of ApEn reflect high regularity.

**Sample entropy**

The Sample entropy (SampEn) is a modification of ApEn. The differences with respect to ApEn are: (i) self-matches are not counted (ii) only the first $N-m$ vectors of length $m$ are considered (Singh et al., 2012; Lake, Richman, Griffin and Moorman 2002). SampEn algorithm can be summarized as follows:

1. Form $m$ vectors $X(1)$ to $X(N-m+1)$ defined by $X(i) = [u(i), u(i+1), \ldots u(i+(m-1))]$ $1 \leq i \leq N-m+1$
2. Define the distance $d[X(i), X(j)]$ between the vectors $X(i)$ and $X(j)$ as the maximum absolute difference between their respective scalar components:

$$d[X(i), X(j)] = \max_{k=1, 2, \ldots, m} |u(i+k-1) - u(j+k-1)|$$

3. Define for each $i$, for $i=1, 2, \ldots, N-m$:

$$B_i^r = \frac{v^r(i)}{N-m+1}$$

Where $v^r(i) =$ number of $d[X(i), X(j)] \leq r$
4. Define for each $i$, for $i=1, 2, \ldots, N-m$:

$$A_i^r = \frac{v^{r+1}(i)}{N-m+1}$$

Where $v^{r+1}(i) =$ number of $d[X(i), X(j)] \leq r$
5. Define:

$$B^r = \frac{\sum_{i=1}^{N-m} B_i^r}{N-m}$$

$$A^r = \frac{\sum_{i=1}^{N-m} A_i^r}{N-m}$$
6. SampEn for a finite data length of \( N \) can be estimated as
\[
\text{SampEn}(m,r,N) = -\ln\left(\frac{A^{m}(r)}{B^{m}(r)}\right)
\]
Similarly to ApEn, we estimated SampEn with \( r \) equal to 20% of standard deviation and \( m = 2 \).

**Detrended fluctuation analysis**

Detrended fluctuation analysis (DFA) is a technique to quantify the fractal scaling properties of RR interval time series. The concept of a fractal is most often associated with irregular geometric objects that display self-similarity. Fractal forms are composed of subunits (and sub-sub-units, etc.) that resembles or show correlation with the structure of the overall object. Similarly, times series extracted from physical or biological systems contain hidden long-range correlation that can provide interesting and useful information on the structure and evolution of the dynamical system. To test whether heartbeat time series exhibit fractal behavior and to determine their correlation properties, we can apply the DFA algorithm. DFA developed by Peng et al. (1994) is a simple and efficient scaling method commonly used for detecting long-range correlations. This technique is a modification of root-mean-square analysis of random walks applied to nonstationary signals (Pena, Echeverria, Garcia, and Gonzalez-Camarena 2009; Rodriguez, Echeverria and Alvarez-Ramirez 2007). The root-mean-square fluctuation of an integrated and detrended time series is measured at different observation windows and plotted against the size of the observation window on a log–log scale. First, the RR interval series (of total length \( k \)) is integrated using the equation:
\[
y(k) = \sum_{i=1}^{k} RR(i) - RR_{av}
\]
Where \( y(k) \) is the \( k^{th} \) value of the integrated series, \( RR(i) \) is the \( i^{th} \) inter beat interval, and the \( RR_{av} \) is the average inter beat interval over the entire series. Then, the integrated time series is divided into windows of equal length, \( n \). In each window of length \( n \), a least-squares line is fitted to the RR interval data (representing the trend in that window). The \( y \) coordinate of the straight line segments are denoted by \( y_n(k) \). Next, we detrended the integrated time series, \( y_n(k) \), in each window. The root-mean-square fluctuation of this integrated and detrended series is calculated using the equation:
\[
F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} \left[ y(k) - y_n(k) \right]^2}
\]
The procedure is repeated for different boxes size or time scales. Finally, the relationship on a double-log graph between fluctuations \( F(n) \) and time scales \( n \) can be approximately evaluated by a linear model \( F(n) \sim n^\alpha \) that provides the scaling exponent \( \alpha \). White Gaussian noise (totally random signal) results in an exponent value of 0.5, and a Brownian noise signal with spectrum rapidly decreasing power in the higher frequencies results in an value of 1.5 (Alvarez-Ramirez, Rodriguez and Echeverria 2005; Huikuri, Makikallio, Peng, Goldberger, Hintze and Moller 2000). The \( \alpha \) can be viewed as an indicator of the ‘roughness’ of the original time series: the larger the value of the \( \alpha \) the smoother the time series. The fractal scaling (\( \alpha \)) for the normal subjects (healthy young) is closer to 1, and this value falls in different ranges for various types of cardiac abnormalities. This slope is very low for very highly varying signals like pre-ventricular contraction (PVC), left bundle branch block (LBBB), atrial fibrillation (AF) and

![Figure 2 Effect of missing RR interval of healthy subjects on HRV parameters Poincaré plot indices (a, b), approximate entropy (c), sample entropy (d) and DFA (e)](image-url)
ventricular fibrillation (VF). But for rhythmically varying signals like sick sinus syndrome (SSS), complete heart block (CHB) and Ischemic/dilated cardiomyopathy this value is slightly higher (comparable to 1) (Acharya, Kannathal, Sing, Ping and Chua 2004).

STATISTICAL METHODS

The nonlinear HRV parameters of complete RR interval series were compared to the HRV parameters of RR interval series with missing data. In addition independent samples t test was used to analyze the percentage differences in nonlinear parameters with missing RR intervals. A value of P<0.05 was considered statistically significant.

RESULTS AND DISCUSSION

The effects of the missing RR interval data on the nonlinear HRV parameters were evaluated based on the relative errors (REs) compared with the parameters calculated from the original, complete RR interval data. When \( \{X_1, X_2, \ldots, X_n\} \) (n = 2 in this study) is obtained for a nonlinear HRV parameter of the data set with a missing duration, and \( X_{\text{original}} \) is the corresponding parameter value of that without any missing data, the relative errors \( \text{RE}_k \) are computed as \( |X_{\text{original}} - X_k|/X_{\text{original}} \times 100 \) (%), where \( k = 1, 2 \). For each HRV parameter and missing duration, 1406 error values were derived and used for the statistical calculations. Figure 2 shows the statistical results for the mean REs in each nonlinear HRV parameter with an increase in the missing data duration. All the nonlinear parameters are found to be robust to missing RR intervals. Mean REs are (<0.5 for SD1, <0.3 for SD2, <0.4 for ApEn, SampEn and <0.3 for DFA) even in the presence of 100 missing RR intervals. Pattern of relative error of ApEn and SampEn are almost similar.

Table 1 Significance level (P value) to reject the alternative hypothesis that the nonlinear HRV parameters of 703 subjects in presence of missing RR intervals is less than mean HRV indices without missing RR intervals.

<table>
<thead>
<tr>
<th>Missing RR</th>
<th>SD1</th>
<th>SD2</th>
<th>ApEn</th>
<th>SampEn</th>
<th>DFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.4644</td>
<td>0.4979</td>
<td>0.594</td>
<td>0.5931</td>
<td>0.4736</td>
</tr>
<tr>
<td>20</td>
<td>0.4661</td>
<td>0.5137</td>
<td>0.5477</td>
<td>0.5149</td>
<td>0.4675</td>
</tr>
<tr>
<td>30</td>
<td>0.4565</td>
<td>0.5221</td>
<td>0.5404</td>
<td>0.4942</td>
<td>0.507</td>
</tr>
<tr>
<td>40</td>
<td>0.4273</td>
<td>0.4971</td>
<td>0.77</td>
<td>0.7463</td>
<td>0.5731</td>
</tr>
<tr>
<td>50</td>
<td>0.4673</td>
<td>0.4928</td>
<td>0.7899</td>
<td>0.7465</td>
<td>0.3841</td>
</tr>
<tr>
<td>60</td>
<td>0.4461</td>
<td>0.5246</td>
<td>0.8834</td>
<td>0.843</td>
<td>0.5966</td>
</tr>
<tr>
<td>70</td>
<td>0.4385</td>
<td>0.4861</td>
<td>0.8669</td>
<td>0.8376</td>
<td>0.5304</td>
</tr>
<tr>
<td>80</td>
<td>0.4566</td>
<td>0.565</td>
<td>0.915</td>
<td>0.8505</td>
<td>0.4848</td>
</tr>
<tr>
<td>90</td>
<td>0.447</td>
<td>0.5228</td>
<td>0.9461</td>
<td>0.8938</td>
<td>0.5286</td>
</tr>
<tr>
<td>100</td>
<td>0.4306</td>
<td>0.5516</td>
<td>0.9084</td>
<td>0.8217</td>
<td>0.5165</td>
</tr>
</tbody>
</table>

CONCLUSION

As compared to time domain (RE: mean <3%; SDSD and SDNN >10%; pNN50 >50%) and frequency domain HRV parameters (RE: VLF 0.04% for FFT; 0.12% for mFFT; 0.11% for Welch; 100.4% for Yule-Walker; 12.3 X 10% for Burg; 99.9% for Lomb, LF 7.2% for FFT; 28.8% for mFFT; 58.8% for Welch; 2.2 X 10% for Yule-Walker; 120.8% for Burg; 2.9% for Lomb, HF 36.3% for FFT; 41.0% for mFFT; 28.8% for Welch; 458.7% for Yule-Walker; 31.9% for Burg; 6.3% for Lomb) (Kim, Lim, Kim and Park 2007; Kim, Lim, Kim and Park 2009), non-linear parameters are found to be very robust to missing RR interval data. Maximum RE for nonlinear parameters remains (<0.5% for SD1; 0.3% for SD2, <0.4 for ApEn and SampEn; <0.3% for DFA) even in the presence of 100 missing RR intervals. Therefore, in case of missing RR interval data, nonlinear measures of HRV are more suitable parameters than time and frequency domain measures.

References


How to cite this article:

******