The effect of missing RR-interval data on heart rate variability analysis in the frequency domain

To cite this article: Ko Keun Kim et al 2009 Physiol. Meas. 30 1039

View the article online for updates and enhancements.

Related content
- Effect of missing RR-interval data on heart rate variability analysis in the time domain
  Ko Keun Kim, Yong Gyu Lim, Jung Soo Kim et al.
- Non-constrained heart rate variability analysis using a ballistocardiogram
  Jae Hyuk Shin, Su Hwan Hwang, Min Hye Chang et al.
- HRV during sleep in healthy term newborns in the early postnatal period
  O M Doyle, I Korotchikova, G Lightbody et al.

Recent citations
- Effect of Missing Inter-Beat Interval Data on Heart Rate Variability Analysis Using Wrist-Worn Wearables
  Hyun Jae Baek and JaeWook Shin
- Ectopic beats detection and correction methods: A review
  Dib Nabil and F. Bereksi Reguig
- A careful look at ECG sampling frequency and R-peak interpolation on short-term measures of heart rate variability
  Robert J Ellis et al.
The effect of missing RR-interval data on heart rate variability analysis in the frequency domain

Ko Keun Kim¹, Jung Soo Kim¹, Yong Gyu Lim² and Kwang Suk Park³,⁴

¹ Interdisciplinary Program in Medical and Biological Engineering, Seoul National University, Korea
² Department of Oriental Biomedical Engineering, Sangji University, Korea
³ Department of Biomedical Engineering, College of Medicine, Seoul National University, Korea
E-mail: pks@bmsil.snu.ac.kr

Received 22 May 2009, accepted for publication 6 August 2009
Published 28 August 2009
Online at stacks.iop.org/PM/30/1039

Abstract
In this study, optimal methods for re-sampling and spectral estimation in frequency-domain heart rate variability (HRV) analysis were investigated through a simulation using artificial RR-interval data. Nearest-neighbour, linear, cubic spline and piecewise cubic Hermite interpolation methods were considered for re-sampling and representative non-parametric, parametric, and uneven approaches were used for spectral estimation. Based on this result, the effects of missing RR-interval data on frequency-domain HRV analysis were observed through the simulation of missing data using real RR-interval tachograms. For this simulation, data including the simulated artefact section (0–100 s) were used; these data were selected randomly from the real RR data obtained from the MIT-BIH normal sinus rhythm RR-interval database. In all, 7182 tachograms of 5 min durations were used for this analysis. The analysis for certain missing data durations is performed by 100 Monte Carlo runs. TF, VLF, LF and HF were estimated as the frequency-domain parameters in each run, and the normalized errors between the data with and without the missing data duration for these parameters were calculated. Rules obtained from the results of these simulations were evaluated with real missing RR-interval data derived from a capacitive-coupled ECG during sleep.

Keywords: heart rate variability (HRV), missing data, frequency domain, ubiquitous healthcare

1. Introduction

In order to observe heart rate variability (HRV), analyses are generally performed in the time and frequency domains (Task Force of the ESC and the NASPE 1996). As the time
Table 1. Definitions of short-term frequency domain HRV parameters used in this study (Task Force of the ESC and the NASPE 1996).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF (ms²)</td>
<td>Variance in NN intervals over the temporal segment (≤0.4 Hz)</td>
</tr>
<tr>
<td>VLF (ms²)</td>
<td>Power in very-low-frequency range (≤0.04 Hz)</td>
</tr>
<tr>
<td>LF (ms²)</td>
<td>Power in low-frequency range (0.04–0.15 Hz)</td>
</tr>
<tr>
<td>HF (ms²)</td>
<td>Power in high-frequency range (0.15–0.4 Hz)</td>
</tr>
<tr>
<td>LF/HF ratio</td>
<td>Ratio LF (ms²)/HF (ms²)</td>
</tr>
</tbody>
</table>

domain parameters, MeanNN, SDNN, RMSSD, SDSD and pNN50 are widely used, and these parameters are directly acquired by numerical calculation from raw RR-interval data. Since calculations are performed directly in the beat domain, re-sampling is not necessary to derive the time-domain parameters. However, HRV spectral parameters, TF, VLF, LF and HF (explained in table 1), 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. In this process, several interpolation methods and techniques for spectrum estimation that have been introduced in a wide range of scientific fields are used. Since a combination of the above-mentioned methods is used, the result values can include errors.

The artefacts can contaminate the bio-signals related to the heart rate during the measurement of these signals, and the RR-interval data can be missed after all. The unconstrained measurement of cardiac-related signals, such as electrocardiograms (ECGs), ballistocardiograms (BCGs) and photoplethysmograms (PPGs), introduced recently by several researchers can derive some motion artefacts. These examples are mentioned in a previous article regarding the missing data effect for HRV analysis in the time domain (Kim et al 2007). In this situation, the HRV parameters will be able to include some errors by the spectral estimation and interpolation methods, and the effect of the missing data needs to be verified.

In this study, the best combination of methods for estimating accurate values of the HRV spectral parameters TF, VLF, LF, HF and the LF/HF ratio will be determined through Monte Carlo simulations using artificial RR-interval models that are not missing data but complete data. In these simulations, several interpolation methods and spectral estimation techniques will be considered. In addition, for investigating the effect of the missing data, the long-term RR-interval tachograms of healthy subjects will be used for the simulation of missing data. The simulation results will be then applied to the actual missing RR-interval data obtained from unconstrained ECG measurements carried out using a capacitive-coupled ECG (CC-ECG) system (Lim et al 2007).

2. Optimal spectral analysis for HRV parameters in the frequency domain

2.1. Artificial RR-interval data

In this study, artificial RR-interval data applying McSharry’s model (McSharry et al 2003) are used for the simulations. By adding the VLF component, we can express the power spectrum of the model as follows:

\[
S(f) = \frac{P_{\text{VLF}}}{\sqrt{2\pi c_{\text{VLF}}^2}} \exp\left(\frac{(f - f_{\text{VLF}})^2}{2c_{\text{VLF}}^2}\right) + \frac{P_{\text{LF}}}{\sqrt{2\pi c_{\text{LF}}^2}} \exp\left(\frac{(f - f_{\text{LF}})^2}{2c_{\text{LF}}^2}\right) + \frac{P_{\text{HF}}}{\sqrt{2\pi c_{\text{HF}}^2}} \exp\left(\frac{(f - f_{\text{HF}})^2}{2c_{\text{HF}}^2}\right)
\]
The effect of missing RR-interval data on HRV analysis in the frequency domain

Figure 1. Example of the artificial RR-interval model: (a) randomly generated power of each HRV element, (b) power spectrum $S(f)$ with VLF, LF HF components, (c) RR-interval data $RR(t)$ having sampling frequency $f_s$ for ECG (---) and $RR(t_n)$ having heartbeat time index $t_n$ (●), and (d) power spectral density of $RR(t_n)$ using the Lomb periodogram (---) and of 4 Hz re-sampling data by cubic spline interpolation of $RR(t_n)$ using the FFT periodogram (---).

where $P_{VLF}$, $P_{LF}$ and $P_{HF}$ represent the power generated randomly; $f_{VLF}$, $f_{LF}$ and $f_{HF}$ are the centre frequency, and $c_{VLF}$, $c_{LF}$ and $c_{HF}$ are the standard deviation of each component, respectively. As this power spectrum $S(f)$ has random phases, the RR-interval time series can be obtained by using the inverse Fourier transform. In addition, the offset values (DC components) are selected randomly for this time series data, which are generally set to a value in range of 0.7–1.3 s.

Figure 1 shows a random selection of these model parameters with $f_{VLF} = 0.02$, $f_{LF} = 0.1$, $f_{HF} = 0.25$, $c_{VLF} = 0.005$, $c_{LF} = 0.01$ and $c_{HF} = 0.01$. In each simulation, the parameters of power $P_{VLF}$, $P_{LF}$ and $P_{HF}$ are changed randomly. From this model, RR-interval data are generated by inverse Fourier transform with sampling rate $f_s$ (=128 Hz in this study), and the irregularly sampled RR-interval data $RR(t_n)$ are calculated for each beat.

2.2. Spectral methods for estimating HRV parameters

In order to obtain the power spectrum, regularly sampled data are required. Therefore, the irregularly sampled RR-interval data $RR(t_n)$ should be interpolated and re-sampled to obtain regularly sampled RR-interval data $RR(t')$. The interpolation can be processed generally by the nearest neighbour (NNR), linear, cubic spline (Spline) or piecewise cubic Hermite (PCH) interpolation methods. In this study, each of these interpolation methods is applied to obtain the re-sampled RR-interval data $RR(t')$, and these processes can be performed simply by a MATLAB function (interp1). During these processes, the re-sampling frequency is selected as 4 Hz, which is adequate to satisfy the Nyquist theory, as the maximum frequency considered in the process estimating HRV spectral parameters is 0.4 Hz.
Power spectral density (PSD) is estimated from a spectral analysis usually by both non-parametric and parametric methods. In the non-parametric method, the PSD can be calculated directly from the signal data itself. One such common method is the fast Fourier transform (FFT) periodogram, based on the Fourier transform. In the modified periodogram method, the non-rectangular window is applied to the signal for the PSD. Welch’s averaged modified periodogram method can be used for the PSD estimation (Welch 1963). This method is practiced to relieve the power loss at the edges of the window when the data length is sufficient to obtain the required frequency resolution.

The parametric PSD method uses the autoregressive (AR) model estimated from a signal, which is assumed to be the output of a linear system driven by white noise. The frequency response of the transfer function of the model is used for calculating the PSD. To estimate the AR model, the Yule–Walker and Burg methods are generally used. The Yule–Walker method utilizes the autocorrelation function of a signal, which is complex because of the calculation of the inverse matrix. The Burg algorithm based on minimizing the forward and backward prediction errors, while satisfying the Levinson–Durbin recursion, is computationally efficient and more accurate than the Yule–Walker method (Marple 1987, Percival and Walden 1993). The order of the AR model may influence the estimation accuracy. In a previous study on the parametric model order for HRV spectral analysis, it is reported that the optimum order is around \( p = 16 \) for the 4 Hz re-sampled RR-interval data (Boardman et al 2002). Pichon et al (2006) reported that the FFT and AR analyses are not interchangeable through the experimental comparisons of HRV spectral parameters obtained from RR-interval data measured in several positions. This implies that one of the spectral analyses may derive incorrect PSD values.

The Lomb–Scargle (LS) periodogram for PSD estimation was proposed as a more appropriate method for an unevenly sampled signal, such as RR-interval data (Lomb 1976, Scargle 1982, Laguna et al 1998). While the FFT periodogram overestimates the LF/HF ratio in HRV spectral analysis because of the re-sampling process, which adds to the LF component and reduces the HF content, the LS periodogram estimates the LF/HF ratio more accurately (Clifford and Tarassenko 2005).

In this study, the non-parametric methods of FFT periodogram, the modified FFT periodogram with Hamming or Hanning window and Welch’s method; and the parametric methods of Yule–Walker and Burg are used for carrying out simulations. In these estimations, re-sampled RR-interval data \( RR(t') \) are used because of the required interpolation and re-sampling processes. The LS periodogram is applied on irregularly sampled RR-interval data \( RR(t_n) \) without the re-sampling process.

2.3. Monte Carlo simulations

The optimal spectral estimation method has been determined through this simulation. Ten thousand artificial RR-interval models are used for the simulation. These are randomly generated from modified McSharry’s models. The theoretical true values of frequency parameters, VLF0, LF0 and HF0 for each model can be calculated as the sum of the power in each frequency range in \( S(f) \). TF0 is the sum of VLF0, LF0 and HF0. If one of these true values is \( X_0 \), the true value set obtained from the 10000 simulations can be expressed as \( [X_0^1, X_0^2, \ldots, X_0^N] \). \( N = 10000 \). The experimental values are obtained from the various spectrums with several interpolation methods (in the re-sampling process). These values can be expressed as follows:
The effect of missing RR-interval data on HRV analysis in the frequency domain

\[ X = \begin{bmatrix} X_1^1 & X_1^2 & \cdots & X_1^n & \cdots & X_1^N \\ X_2^1 & X_2^2 & \cdots & X_2^n & \cdots & X_2^N \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_k^1 & X_k^2 & \cdots & X_k^n & \cdots & X_k^N \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_K^1 & X_K^2 & \cdots & X_K^n & \cdots & X_K^N \end{bmatrix} \]

where \( K \) is the number of spectral methods times the number of interpolation methods. In this study, the value of \( K \) is 21 (= 5 \times 4 + 1).

The optimal method for estimating HRV frequency parameters can be found by evaluation of the root-mean-squared relative error (RMSRE). The mean value of absolute relative errors of certain parameters between theoretical true values and experimental values, for 10,000 artificial RR-interval models, is given as follows:

\[
\text{RMSRE}[X_k](\%) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \text{RE}[X_n^k]^2},
\]

\[
\text{RE}[X_n^k](\%) = \frac{X_n^k - X_n^0}{X_n^0} \times 100.
\]

When the RMSRE value of \( X_k \) is the smallest, the method corresponding to \( k \) will be the optimal spectral analysis.

2.4. Simulation results

Table 2 shows the results of the above-mentioned simulations. TF and VLF can be obtained most accurately by the FFT periodogram as indicated in table 2. The parametric spectrum methods dramatically increase the error of TF and VLF. This implies that parametric methods do not estimate the spectrum of the VLF component properly. The results of the Lomb periodogram for TF and VLF also have large errors, as the power of the DC component that has a frequency of zero is not calculated in the estimation process of the Lomb periodogram (Lomb 1976). The errors of LF and HF obtained by Lomb and FFT periodograms with Spline interpolation are smaller than those of LF and HF obtained by other spectral methods. On the basis of these results, Lomb and FFT periodograms with Spline interpolation can be concluded to be the references spectral methods of choice to calculate the HRV frequency-domain parameters of data without missing RR intervals.

3. Simulations of missing data

3.1. Real RR-interval data

Long-term RR tachograms were obtained from the MIT-BIH normal sinus rhythm RR-interval database for this study (http://www.physionet.org/physiobank/database/nsr2db/). The beat annotation files included in this database were used; the original ECGs were recorded for more than 24 h from subjects having no significant arrhythmias. The subjects were 30 men (aged 28.5–76) and 24 women (aged 58–73). Further, the ECGs were digitized at the rate of 128 samples s\(^{-1}\). In all, 7182 RR-interval data sets including only normal beats and having 5 min durations were collected for the short-term HRV analysis (Task Force of the ESC and the NASPE 1996).
### Table 2. RMSRE results of simulations carried out to determine the optimal spectral methods to obtain HRV frequency parameters. The grey part is not thought to be meaningful since the values are greater than 99%.

<table>
<thead>
<tr>
<th>Spectral methods</th>
<th>TF</th>
<th>VLF</th>
<th>LF</th>
<th>HF</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+NNR</td>
<td>0.05</td>
<td>0.04</td>
<td>3.9</td>
<td>20.4</td>
</tr>
<tr>
<td>+Linear</td>
<td>0.05</td>
<td>0.04</td>
<td>7.2</td>
<td>36.3</td>
</tr>
<tr>
<td>+Spline</td>
<td>0.03</td>
<td>0.03</td>
<td>2.5</td>
<td>5.2</td>
</tr>
<tr>
<td>+PCH</td>
<td>0.03</td>
<td>0.03</td>
<td>2.6</td>
<td>19.7</td>
</tr>
<tr>
<td>mFFT (with Hanning)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+NNR</td>
<td>0.12</td>
<td>0.12</td>
<td>28.1</td>
<td>31.2</td>
</tr>
<tr>
<td>+Linear</td>
<td>0.13</td>
<td>0.12</td>
<td>27.9</td>
<td>41.0</td>
</tr>
<tr>
<td>+Spline</td>
<td>0.12</td>
<td>0.12</td>
<td>28.8</td>
<td>28.8</td>
</tr>
<tr>
<td>+PCH</td>
<td>0.12</td>
<td>0.12</td>
<td>28.4</td>
<td>31.8</td>
</tr>
<tr>
<td>Welch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+NNR</td>
<td>0.10</td>
<td>0.11</td>
<td>55.5</td>
<td>21.2</td>
</tr>
<tr>
<td>+Linear</td>
<td>0.11</td>
<td>0.11</td>
<td>52.1</td>
<td>28.8</td>
</tr>
<tr>
<td>+Spline</td>
<td>0.10</td>
<td>0.10</td>
<td>58.8</td>
<td>24.4</td>
</tr>
<tr>
<td>+PCH</td>
<td>0.10</td>
<td>0.10</td>
<td>56.5</td>
<td>21.6</td>
</tr>
<tr>
<td>Yule–Walker</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+NNR</td>
<td>99.8</td>
<td>99.0</td>
<td>2.2×10³</td>
<td>438.2</td>
</tr>
<tr>
<td>+Linear</td>
<td>101.8</td>
<td>101.1</td>
<td>2.2×10³</td>
<td>428.7</td>
</tr>
<tr>
<td>+Spline</td>
<td>101.1</td>
<td>100.4</td>
<td>2.2×10³</td>
<td>458.7</td>
</tr>
<tr>
<td>+PCH</td>
<td>101.0</td>
<td>100.2</td>
<td>2.2×10³</td>
<td>443.5</td>
</tr>
<tr>
<td>Burg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+NNR</td>
<td>3.5×10³</td>
<td>3.5×10³</td>
<td>120.8</td>
<td>31.9</td>
</tr>
<tr>
<td>+Linear</td>
<td>12.3×10³</td>
<td>12.3×10³</td>
<td>46.3</td>
<td>21.0</td>
</tr>
<tr>
<td>+Spline</td>
<td>2.2×10³</td>
<td>2.2×10³</td>
<td>76.3</td>
<td>12.5</td>
</tr>
<tr>
<td>+PCH</td>
<td>9.3×10³</td>
<td>9.3×10³</td>
<td>36.5</td>
<td>15.2</td>
</tr>
<tr>
<td>Lomb</td>
<td>99.9</td>
<td>99.9</td>
<td>2.9</td>
<td>6.3</td>
</tr>
</tbody>
</table>

### 3.2. Spectral methods for HRV parameters

In section 2, FFT periodogram with Spline interpolation and Lomb periodogram were selected as the optimal spectral methods. The reference values of LF and HF were calculated with these methods. In case of VLF and TF, only the FFT method was used since the Lomb periodogram cannot yield these values. For estimating the HRV spectral parameters in RR-interval tachograms having missing data, each representative spectral method of non-parametric methods with four interpolation methods and uneven methods without interpolation process, mentioned in section 2, was applied as follows: the FFT periodogram for a non-parametric method and the Lomb periodogram for an uneven method. In this process, TF and VLF values could not be obtained with a Lomb periodogram. In the same manner as that described in section 2, the missing RR-interval data were re-sampled with several interpolation modes and at a frequency of 4 Hz. When the Lomb method was applied, this re-sampling process was unnecessary.

### 3.3. Monte Carlo simulations

Consecutive RR-interval data were removed by random selection, and the removed duration was increased from 0 to 30 s and 30 to 100 s in increments of 1 s and 5 s, respectively, for each dataset. In all, 316,008 datasets (7182 datasets × 44 missing data) were used in these
The effect of missing RR-interval data on HRV analysis in the frequency domain

Figure 2. Log-normal distributions of original HRV parameters in the frequency domain for all subjects, and the geometric mean value $\mu$ and the multiplicative standard deviation $\sigma$ of each parameter. The 95% confidence interval (CI) is calculated to be $[\mu/\sigma^2, \mu \cdot \sigma^2]$.

simulations. In each case, the HRV spectral parameters were calculated by 100 Monte Carlo runs using a MATLAB program in a manner similar to that used in the previous study on the missing data effect of HRV parameters in the time domain (Kim et al 2007).

The effects of the missing data on the HRV spectral parameters were evaluated by the errors between the original parameter value and the value obtained from the missing data. Since each HRV parameter has a wide range, normalization of these parameters is required for the comparison. Therefore, the normalized error (NE) obtained using the confidence interval (CI) was used in this study. It is generally known that the distributions of HRV parameters exhibit a log-normal property as shown in figure 2 (Ziegler et al 1992). NE is defined in this study as follows:

$$NE = \frac{|X_{origin} - X|}{95\% \text{ confidence interval length of } \{X_{origin}\}} \times 100(\%) ,$$

95% CI of the data having a log-normal distribution can be expressed as $[\mu/\sigma^2, \mu \cdot \sigma^2]$, where $\mu$ is the geometric mean and $\sigma$ is the multiplicative standard deviation of the $\{X_{origin}\}$ data. These estimators can be computed as follows:

$$\mu = \exp \left( \frac{1}{Q} \sum_{i=1}^{Q} \log(X_{origin,i}) \right) , \quad \sigma = \exp \left( \frac{1}{Q-1} \sum_{i=1}^{Q} \left[ \log \left( \frac{X_{origin,i}}{\mu} \right) \right]^2 \right)^{1/2} ,$$

where $Q$ is the total number of $\{X_{origin}\}$ data (Aitchison and Brown 1957, Limpert et al 2001). Therefore, NE can be understood as the error normalized by the variation in the original data in this study.

3.4. Simulation results

In the simulation results shown in figure 3, the maximum NE value in 95% of the total errors caused by considering the missing data durations can be observed. The missing data duration axis is in the log scale. In these figures, the NE results for each HRV parameter were changed by various spectral estimation methods and interpolation methods. The results of using the parametric algorithm as the spectral method and of using other algorithms with NNR and linear interpolation methods had relatively large errors; therefore, these were not plotted. In the first row graphs of figure 3, the NE results using the HRV parameters based on the FFT for the original data without any missing data duration as the reference values are shown, and
Figure 3. Simulation results of normalized errors (NEs) for HRV parameters in the frequency domain. The graphs are subdivided by several spectral estimation and interpolation methods and by the spectral methods estimating referenced values of the original data without the missing data durations. The x-axis is in the log scale.

The Lomb periodogram was used for the reference values in the second row plots. Either the Lomb or FFT periodogram as a spectral method for estimating reference values can be selected through the total error graph for LF and HF, as shown in figure 4. The total error $E_{FFT}$ and $E_{Lomb}$ was obtained as the mean of the NE values for the overall combinations of spectral methods (FFT, mFFT, Welch, Yule–Walker, Burg and Lomb) and interpolation methods (NNR, Linear, Spline and PCH) referenced with the HRV parameters based on FFT and Lomb periodograms, respectively, in certain missing data durations. For the most part of missing data durations, the mean of total errors can be expected to be higher when the Lomb periodogram is used for estimating the reference value. Therefore, the FFT periodogram with Spline interpolation is considered to be a better choice for the estimation of the HRV parameters of the original data without any missing data in the frequency domain, and the NE results using this combination for the reference values are shown in the first row of figure 3. This method was used for estimating reference values in further analysis.

For estimating TF and VLF, the FFT with PCH interpolation is considered to be the best method in the entire missing data duration. In the case of LF, the error is the lowest against the referenced value by FFT when FFT with PCH is used. In the case of HF, the choice is dependent on the missing data duration. When the missing data duration is less than 60 s, the FFT with Spline interpolation is the best method, and the FFT with PCH interpolation is the best in other cases. The rules for optimal spectral methods using missing RR-interval data are summarized in table 3. These rules are applied to the analysis of real data containing missing RR-intervals in the next section.

By interpreting the results on the basis of the rules (table 3), we find that the NE values of TF and VLF are very similar to each other for all missing data durations and are below 5% even if the missing data duration is 100 s. The NE of LF is greater than 10% and increases
The effect of missing RR-interval data on HRV analysis in the frequency domain

Figure 4. Difference in mean values of total errors for overall combinations of spectral and interpolation methods when two spectral estimation methods are used for estimating reference values. If the difference is more than zero, the errors caused by using the Lomb method for reference are greater than those caused by using FFT.

Table 3. Rules of using spectral estimation and interpolation methods to estimate missing data durations for HRV spectral parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>No missing data</th>
<th>Missing data duration &lt; 60 s</th>
<th>Missing data duration &gt; 60 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF</td>
<td>FFT + Spline</td>
<td>FFT + PCH</td>
<td>FFT + PCH</td>
</tr>
<tr>
<td>VLF</td>
<td>FFT + Spline</td>
<td>FFT + PCH</td>
<td>FFT + PCH</td>
</tr>
<tr>
<td>LF</td>
<td>FFT + Spline</td>
<td>FFT + PCH</td>
<td>FFT + PCH</td>
</tr>
<tr>
<td>HF</td>
<td>FFT + Spline</td>
<td>FFT + Spline</td>
<td>FFT + PCH</td>
</tr>
</tbody>
</table>

significantly if the missing data duration is longer than 40 s. HF has NE values less than 10% when the missing data duration is less than 30 s.

4. Real missing RR-interval data

In order to investigate the effects of the real missing RR-interval data, the CC-ECG data were used in the same manner as that of the previous study on time-domain HRV parameters (Kim et al. 2007, Lim et al. 2007). RR-interval tachograms were obtained by using an R-peak detection algorithm from both original ECG with Ag–AgCl electrodes and CC-ECG signals with artefact noises. These tachograms were segmented to 5 min RR-interval data with time shifts of 30 s. In each segment, the missing data duration is shown at the top of the right column in figure 5. NEs of each parameter were calculated from the TF, VLF, LF, HF and LF/HF of both original and missing tachograms by the estimation methods on the basis of the above-mentioned rules (table 3), as shown in figure 5. These NEs were normalized by the 95% CI lengths of each parameter distribution estimated from the original RR-interval data.

The RR-interval tachograms of the original data are plotted on the top of the left column in figure 5, and the HRV parameters in the frequency domain of both the original and the missing data are shown in the other left column graphs. NEs for each parameter of the original and the missing data are calculated at the right side of each graph in figure 5. Similar to the simulation results, the TF and VLF have lower error values and the bias of errors of HF is the highest.

The correlations of the patterns of each parameter between the original and the missing data are observed to be high. The correlation coefficients of TF and VLF are the highest.
Figure 5. Frequency-domain HRV parameters of real missing RR-interval data (---) and original RR-interval data (---) obtained using the rules (table 2) from a CC-ECG and a typical ECG, respectively (left column), and the NEs between the original and the missing values (right column).

Table 4. Correlation coefficients between the missing and the original data for TF, VLF, LF, HF, and LF/HF ratio.

<table>
<thead>
<tr>
<th></th>
<th>TF</th>
<th>VLF</th>
<th>LF</th>
<th>HF</th>
<th>LF/HF</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>0.9968</td>
<td>0.9967</td>
<td>0.9805</td>
<td>0.9651</td>
<td>0.9835</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

(0.9968 and 0.9967, respectively), as shown in table 4. The coefficient of HF has the lowest value ($r = 0.9651$), but this is high as usual.

5. Discussion

In the simulation result for the optimal method to obtain the HRV spectral parameters of intact RR-interval data, as given in table 2, the effects of using several spectral estimation
methods using four interpolation modes are indicated. The NNR, linear, and PCH interpolation processes made relatively more errors, especially in HF results. It is thought that these errors in the HF parameters are due to the discontinuity properties of re-sampled data using these interpolation methods.

The effect of various detrending methods was researched in our preliminary study in which the error values were only slightly different from each other when non-parametric or uneven method was used for the spectral estimation. Although the difference in errors caused by detrending is significant in the results using parametric methods, these methods produce a considerably large error by themselves, and hence, the detrending effects are not considered in this study.

The missing data effects of real RR-interval data are shown in figure 3. In this figure, the results for only Spline and PCH interpolation methods are plotted since the errors for NNR or linear interpolation are relatively large. From these results, the PCH interpolation appears to be a better choice than the Spline method for the entire missing data duration except in the case of HF results. In the case of LF results for the Spline interpolation, especially, the missing data duration section of 5–20 s is observed during which the errors increase significantly. These duration values in the time domain correspond to the LF frequency region in the frequency domain (0.04–0.15 Hz). Thus, it can be considered that the missing data in this region are overestimated in the process of Spline interpolation. In the case of HF, the Spline interpolation produces a low error for small missing data duration of less than 60 s, which is similar to the simulation results, shown in table 2, of the data with no missing data durations. However, if the missing data duration is more than 60 s, the errors caused by PCH interpolation are rather smaller than those caused by Spline. The errors for all HRV parameters increase dramatically if the missing data duration is over a certain point (approximately 20 s), and it can be explained that the data overestimation by Spline interpolation occurs largely due to the lack of continuity, unlike PCH.

From figure 5, it appears that the normalized errors caused by the real missing data of LF, HF and LF/HF ratio are not perfectly correlated with the missing data duration that is different to previous simulation results (figure 3) in which the error values are increased by missing data duration. The correlation coefficients between the missing data duration and the NE values of TF, VLF, LF, HF and LF/HF ratio are 0.7476, 0.7531, 0.5536, 0.6362 and 0.3847, respectively \((p < 0.001 \text{ in all cases})\). This implies that factors other than the missing data duration can affect the HRV parameter estimation and the peculiar RR-interval pattern at a certain point including the missing data caused by motion artefact can be considered an important factor. In fact, the motion is originated generally when the subject is in wake stage in sleep stage. In the wake stage, it is known that heart rate is generally elevated and this high heart rate (low RR-interval value) usually makes the variation smaller. Hence, the errors in this duration can increase because of the unique property of the data even though the missing data duration is short. The HRV parameters between the original and the missing data are shown to be strongly correlated to each other, as shown in the left column of figure 5; the normalized errors of the parameters cannot be ignored. Therefore, the rules for the missing data, given in table 3, are considered useful for the long-term analysis of HRV spectral parameters in unconstrained monitoring of cardiac-related bio-signals.

If only one method can be used for the HRV analysis in the frequency domain with missing RR-interval data, then the FFT periodogram with PCH interpolation is recommended since the error is overall most robust to the missing data duration.

The long-term variation of HF is very similar to the variations of SDSD, RMSSD and pNN50 obtained from the same subjects in a previous study on the effect of missing data on time-domain HRV analysis (Kim et al 2007). This is also in agreement with a report of a Task
Force of the ESC and the NASPE (1996). However, the variation of TF is not similar to that of SDNN, which rather resembles that of LF; this disagrees with the report.

In a future study, the effects of missing data on the estimation of the nonlinear HRV parameters will be investigated. In addition, it is expected that the log-normal properties of HRV parameters observed in this study (figure 2) can be used for characterizing the cardiovascular activities through further studies about the properties in not only healthy subjects but also in patients with certain cardiovascular disease.

Acknowledgments

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Centre) support programme supervised by the IITA (Institute of Information Technology Advancement) (IITA-2009-(C1090-0902-0002)) and a grant from the Advanced Biometric Research Centre (ABRC).

References


Ziegler D et al 1992 Assessment of cardiovascular autonomic function: age-related normal ranges and reproducibility of spectral analysis, vector analysis and standard test of heart rate variation and blood pressure responses Diabetic Med. 9 166–75