Article

Missing RRI Interpolation Algorithm based on Locally Weighted Partial Least Squares for Precise Heart Rate Variability Analysis

Keisuke Kamata [1], Koichi Fujiwara [1,2],* , Takafumi Kinoshita [1] and Manabu Kano [1]

[1] The Department of Systems Science, Kyoto University, Kyoto 615-8085, Japan;
kei.kyoto.be.amb@gmail.com (K.K.); phantom.vib@gmail.com (T.K.);
manabu@human.sys.i.kyoto-u.ac.jp (M.K.)

[2] The Department of Material Process Engineering, Nagoya University, Nagoya 464-8601, Japan

* Correspondence: fujiwara.koichi@material.nagoya-u.ac.jp or fujiwara.koichi@i.kyoto-u.ac.jp

check for updates

Abstract: The R-R interval (RRI) fluctuation in electrocardiogram (ECG) is called heart rate variability (HRV), which reflects activities of the autonomic nervous system (ANS) and has been used for various health monitoring services. Accurate R wave detection is crucial for success in HRV-based health monitoring services; however, ECG artifacts often cause missing R waves and deteriorate the accuracy of HRV analysis. The present work proposes a new missing RRI interpolation technique based on Just-In-Time (JIT) modeling. In the JIT modeling framework, a local regression model is built by weighing samples stored in the database according to the distance from a query and output is estimated only when an estimate is requested. The proposed method builds a local model and estimates missing RRI only when an RRI detection error is detected. Locally weighted partial least squares (LWPLS) is adopted for local model construction. The proposed method is referred to as LWPLS-based RRI interpolation (LWPLS-RI). The performance of the proposed LWPLS-RI was evaluated through its application to RRI data with artificial missing RRIs. We used the MIT-BIH Normal Sinus Rhythm Database for nominal RRI dataset construction. Missing RRIs were artificially introduced and they were interpolated by the proposed LWPLS-RI. In addition, MEAN that replaces the missing RRI by a mean of the past RRI data was compared as a conventional method. The result showed that the proposed LWPLS-RI improved root mean squared error (RMSE) of RRI by about 70% in comparison with MEAN. In addition, the proposed method realized precise HRV analysis. The proposed method will contribute to the realization of precise HRV-based health monitoring services.

Keywords: R wave detection; heart rate variability analysis; Just-In-Time modeling; locally weighted partial least squares

1. Introduction

The RR interval (RRI) fluctuation in an electrocardiogram (ECG) is known as heart rate variability (HRV), which is a physiological activity reflecting the cardiovascular control exerted by the autonomic nervous systems (ANS) [1,2]. Since the ANS is a control system that acts largely unconsciously and regulates functions such as the heart rate, digestion, respiration, perspiration, and body temperature, it is related to various diseases. Many types of HRV features have been proposed for evaluation of ANS activities, [3,4], and application areas of HRV analysis has been expanded due to recent advances in machine learning technologies. A brief introduction of HRV analysis is described in Appendix A.

It is known that changes in sleep condition affect HRV [5,6]. An HRV-based drowsiness detection method using linear discriminant analysis (LDA) was developed by Vicente et al. [7]. Patel et al.

proposed a drowsiness detection algorithm based on a neural network (NN) [8]. A drowsy driving detection algorithm was proposed by integrating HRV analysis and multivariate statistical process control (MSPC), which is a well-known anomaly detection method [9,10].

Apnea affects HRV during sleep because it contributes to the future development of cardiovascular events and HRV reflects the cardiovascular control [11,12]. Sleep apnea can also be screened through monitoring HRV. It has been reported that frequency domain features of HRV are useful for apnea screening [13]. Roche et al. developed a sleep apnea detection method based on HRV features [14].

HRV analysis was also used for epilepsy seizure detection [15], since HRV changes in preictal phases [16,17]. In addition, Fujiwara et al. developed an epileptic seizure prediction algorithm utilizing MSPC, which can predict seizure occurrence prior to onset [18].

Precise long-term RRI measurement devices are needed for realizing these HRV-based health monitoring services. The Holter monitor is used for long-term ECG measurement; however, its use in daily life is difficult since the Holter monitor is expensive and requires operation skills. Yamakawa et al. developed a wearable RRI sensor based on ECG for real-time HRV analysis, which can easily measure RRI [19]. A wearable chemical-electrophysiological sensors for health and fitness monitoring was developed by Imani et al. [20]. In addition, tattoo type sensors for ECG monitoring have been investigated [21–23].

Adequate HRV feature extraction for RRI data is required. In particular, R wave detection from ECG is essential because HRV feature extraction is easily fluctuated by artifact contamination in the raw RRI data. R waves are not always detected stably due to ECG artifacts caused by body movement or ECG electrode contact failure. Figure 1 illustrates an example of an ECG trace with artifacts, in which the colored band denotes contaminated artifacts.

Some HRV features easily fluctuate when a missing RRI occurs even if it is just one. Figure 2a shows an example of RRI data with one missing and Figure 2b–d are RMSSD, NN50, and HF that were calculated from the RRI data of Figure 2a. A three-minute time window was used for HRV calculation. The vertical line denotes a missing RRI occurrence point. This example shows that some of the extracted HRV features were significantly altered when the R detection error occurred and that their influence lasted for three minutes in this case, which is the window size of HRV calculation.
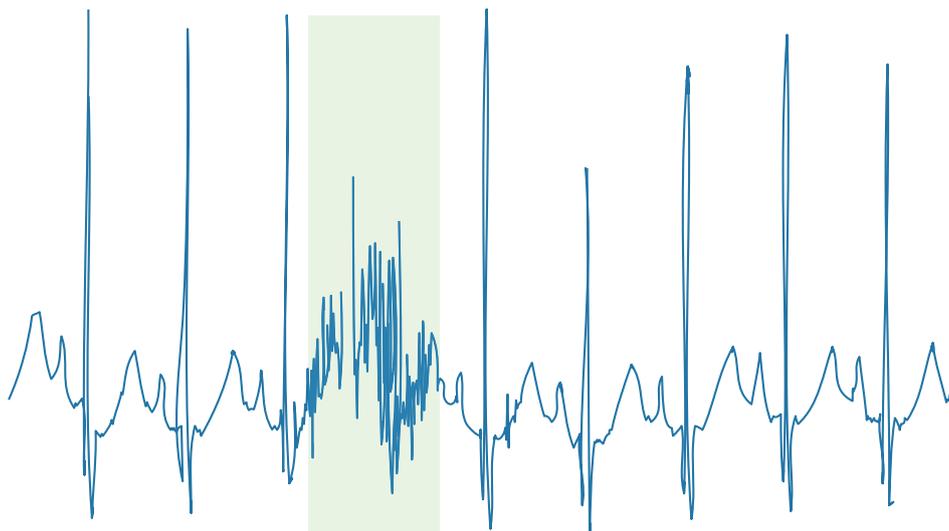

Figure 1. An example of an ECG trace with artifacts (colored band). An R wave cannot be detected due to artifacts.
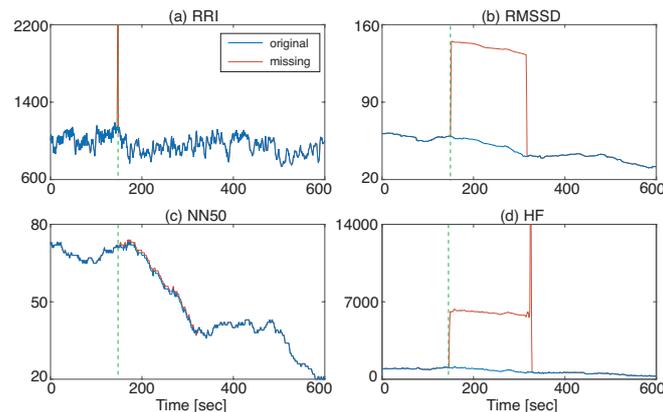
Figure 2. Missing RRI data and its effect on HRV features: (**a**) RRI, (**b**) RMSSD, (**c**) NN50, and (**d**) HF. The HRV features greatly fluctuate when the R detection error occurs.

Although missing R waves may easily be found when the ECG data are analyzed offline, such offline analysis cannot be used for online applications like drowsy driving detection and epileptic seizure prediction.

When the $(j+1)$th R wave is not detected, r_j is as

$$r_j = \tilde{r}_j + \tilde{r}_{j+1} \tag{1}$$

where $\tilde{r}_j$ is the true RRI measurement when both the jth and $(j+1)$th R waves are detected correctly.

The simplest way of missing RRI treatment is to remove or ignore r_j [24]. Although Clifford and Tarassenko recommended using the Lomb-Scargle (LS) periodogram after removing r_j for frequency domain feature extraction [25], such treatment cannot be used online because it causes time gaps between real-time. An ectopic RRI modification method from ECG data was proposed by using prior knowledge about arrhythmia [26]. However, its use for missing RRI interpolation is difficult since the effect of R wave detection errors on changes in RRI data is different from that of arrhythmia. Thus, missing RRI should be interpolated appropriately in real-time for precise HRV analysis. Highly adequate interpolation is required particularly in drowsy driving detection and epileptic seizure prediction because their errors caused by missing RRI may lead to severe injuries and accidents.

The electrode contact failure or sensor failure may cause long-term ECG measurement failure. In such cases, HRV analysis and its use for health monitoring should be stopped because we cannot use any information for HRV analysis. Thus, this study focuses on the modification of an isolated R wave detection error.

The present work proposes a new missing RRI interpolation method based on just-in-time (JIT) modeling. JIT modeling or lazy learning is a statistical modeling method that builds a local regression model and that estimates an output using the constructed local model only when an estimate is requested. Although some JIT modeling methods have been proposed [27–30], the present work adopts locally-weighted partial least squares (LWPLS), which utilizes partial least squares (PLS) for local model construction [31,32]. The proposed method, referred to as LWPLS-based RRI interpolation (LWPLS-RI), interpolates missing RRIs by using a local regression model only when an R wave detection error is detected. R wave detection errors are detected by using a threshold of the measured RRI.

This paper is organized as follows. Section 2 introduces LWPLS and proposes new missing RRI interpolation method. Section 3 validates the performance of the proposed missing RRI interpolation method through a case study of RRI data with artificial missing RRI. Finally, the conclusion and future work are described in Section 4. Although a preliminary version of this work has been reported in [33], the data analyzed in [33] was small, and the performance of the proposed method was not compared with other methodologies.

2. Materials and Methods

The present work proposes a new JIT-based algorithm for interpolating missing RRI, which is referred to as LWPLS-based RRI interpolation (LWPLS-RI). This section begins with partial least squares (PLS) and locally-weighted PLS (LWPLS) used in the proposed algorithm.

2.1. PLS

PLS is a widely used linear regression method that can build an accurate model with a small number of latent variables. Given an input data matrix $X \in \Re^{N \times M}$ whose nth row is the nth input sample $x_n \in \Re^M$ and an output data vector $y \in \Re^N$ whose nth element is the nth output sample $y_n \in \Re$. X and y are mean-centered and appropriately scaled. In PLS, the input X and the output y are broken down as follows:

$$X = TP^T + E \tag{2}$$
$$y = Tb + f \tag{3}$$

where $T \in \Re^{N \times K}$ is the latent variable matrix whose columns are the latent variable $t_k \in \Re^N$ $(k = 1, \cdots, K)$, $P \in \Re^{M \times K}$ is the loading matrix of X whose columns are the loading vectors $p_k \in \Re^M$, and $b = [b_1, \cdots, b_K]^T$ is the regression coefficient vector of y. K denotes the number of adopted latent variables. $E \in \Re^{N \times M}$ and $f \in \Re^N$ are errors.

The nonlinear iterative partial least squares (NIPALS) algorithm can be used to construct a PLS model [34]. Suppose that the first to kth latent variables $t_1, \cdots, t_k$, the loading vectors $p_1, \cdots, p_k$ and the loading $b_1, \cdots, b_k$ are given. The $(k + 1)$th residual input and output can be expressed as follows:

$$X_{k+1} = X_k - t_r p_k^T, \tag{4}$$
$$y_{k+1} = y_k - b_k t_k. \tag{5}$$

t_k is a linear combination of the columns of X_k, that is, $t_k = X_k w_k$ where $w_k \in \Re^M$ is the kth weighting vector. It is defined so that the covariance between y_k and t_k is maximized under $||w_k|| = 1$. Using the Lagrange multipliers method, the function to maximize can be defined as

$$G_k = y_k^T t_k - \mu(||w_k|| - 1) = y_k^T X_k w_k - \mu(||w_k|| - 1), \tag{6}$$

where μ is the Lagrange multiplier. By solving $\partial G_k / \partial w = 0$, w_k is derived as

$$w_k = \frac{X_k^T y_k}{||X_k^T y_k||}. \tag{7}$$

The kth loading vector p_k and the kth loading b_k are as follows:

$$p_k = \frac{X_k^T t_k}{t_k^T t_k}, \quad b_k = \frac{y_k^T t_k}{t_k^T b_k}. \tag{8}$$

Finally, the above procedure is repeated until the number of adopted latent variables K is achieved; K can be determined by cross validation.

Instead of using the Lagrange multipliers method, the derivation of the weighting vectors w_k in the NIPALS algorithm can be formulated as an eigenvalue problem. w_k is the eigenvector corresponding to the maximum eigenvalue of the following eigenvalue problem:

$$X_{k-1}^T y_{k-1}^T y_{k-1} X_{k-1} w_k = \lambda w_k, \tag{9}$$

where λ is an eigenvalue.

Algorithm 1 describes the eigenvalue-based NIPALS algorithm for PLS modeling.

Algorithm 1 NIPALS.

1: Set K.
2: $\boldsymbol{X}_0 = \boldsymbol{X} - \bar{\boldsymbol{X}}$.
3: $\boldsymbol{y}_0 = \boldsymbol{y} - \bar{\boldsymbol{y}}$.
4: **for** $k = 1$ to K **do**
5:     Derive the eigenvector $\boldsymbol{w}_k$ which corresponds to the maximum eigenvalue of the following
    eigenvalue problem: $\boldsymbol{X}_{k-1}^T \boldsymbol{y}_{k-1}^T \boldsymbol{y}_{k-1} \boldsymbol{X}_{k-1} \boldsymbol{w}_k = \lambda \boldsymbol{w}_k$.
6:     $\boldsymbol{t}_k = X_{k-1} \boldsymbol{w}_k$.
7:     $\boldsymbol{p}_k = \boldsymbol{X}_{k-1}^T \boldsymbol{t}_k / \boldsymbol{t}_k^T \boldsymbol{t}_k$.
8:     $b_k = \boldsymbol{y}_{k-1}^T \boldsymbol{t}_k / \boldsymbol{t}_k^T \boldsymbol{t}_k$.
9:     **if** $k = K$ **then**
10:       Output $\boldsymbol{P} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_K]$ and $\boldsymbol{b} = [b_1, \cdots b_K]^T$.
11:     **end if**
12: **end for**

2.2. Locally-Weighted Partial Least Squares

In general, a global linear model cannot function well when a system has strong nonlinearity or changes in characteristics with time. The use of nonlinear modeling methods, such as support vector machine (SVM) or artificial neural network (ANN), is the first choice; however, nonlinear modeling methods are not always applicable because a considerable amount of data is required for nonlinear modeling. Also, it is difficult to even for nonlinear models to cope with changes in system characteristics with time.

Another method for dealing with these problems is JIT modeling, which has the following features:

- Store new samples into a database when available.
- Construct a local model by the samples located in the neighboring region around a query and estimate an output only when estimation is required.
- Discard the constructed local model after its use for output estimation.

In JIT modeling, samples for local modeling should be selected appropriately.

LWPLS is an expansion of PLS based on the framework of JIT modeling for dealing with nonlinearity and system characteristics change. In LWPLS, a local PLS model is built by weighted samples stored in a database according to the similarity between the query and the weighted samples only when an estimate is requested. The constructed local model represents a nonlinear relationship between the input and the output around the query because a nonlinear relationship can be approximated as a linear relationship in a small region. After being used for estimation, the used local model is purged [31,32].

Let $\boldsymbol{X}$ and $\boldsymbol{y}$ have already been stored in a database. When an estimate is requested for a query $\boldsymbol{x}_q$, the similarity ω_n between $\boldsymbol{x}_q$ and the nth sample $\boldsymbol{x}_n (n = 1, \cdots, N)$ is calculated, and a local PLS model is built by the weighted samples with a similarity matrix $\Omega \in \Re^{N \times N}$ defined as:

$$\Omega = \text{diag}[\omega_1, \omega_2, \cdots, \omega_N]. \tag{10}$$

The similarity between the query $\boldsymbol{x}_q$ and a sample $\boldsymbol{x}_n$ ω_n in this work is defined as:

$$\omega_n = e^{(-d_n \varphi / \sigma_d)}, \tag{11}$$

$$d_n = \sqrt{(\boldsymbol{x}_n - \boldsymbol{x}_q)^T (\boldsymbol{x}_n - \boldsymbol{x}_q)} \tag{12}$$

where σ_d denotes the standard deviation of d_n $(n = 1, 2, \cdots, N)$ and φ is a localization parameter; the similarity decreases steeply when φ is small and gradually when φ is large. When the similarity matrix Ω is an identity matrix, LWPLS becomes the original PLS.

Algorithm 2 describes a procedure of LWPLS based on the NIPALS algorithm. Steps 4–7 derive the latent variable t, the loading vector p, and the regression coefficient vector b iteratively. In step 6, w_k is calculated as the eigenvector of $X_k^T \Omega y_k y_k^T \Omega X_k$, which corresponds to the maximum eigenvalue. The final estimate is output when $k = K$. A localization parameter φ and the number of latent variables K are tuning parameters, which are determined by trial and error or cross-validation.

Algorithm 2 LWPLS.

1: Set K and φ.
2: Calculate the similarity matrix Ω
3: **for** $r = 1$ to K **do**
4:     Calculate X_k, y_k, and $x_{q,k}$;

$$X_k = X - \mathbf{1}_N \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_M \end{bmatrix} \tag{13}$$

$$y_k = y - \mathbf{1}_N \bar{y} \tag{14}$$

$$x_{q,k} = x_q - \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_M \end{bmatrix}^T \tag{15}$$

$$\bar{x}_m = \frac{\sum_{n=1}^{N} \omega_n x_{nm}}{\sum_{n=1}^{N} \omega_n} \tag{16}$$

$$\bar{y} = \frac{\sum_{n=1}^{N} \omega_n y_n}{\sum_{n=1}^{N} \omega_n}. \tag{17}$$

5:     Set $\hat{y}_q = \bar{y}$.
6:     Derive the kth latent variables of X and y, and the rth latent variable of x_q;

$$t_k = X_k w_k, \ t_{q,k} = x_{q,k}^T w_k. \tag{18}$$

7:     Derive the kth loading vectors of X and the rth regression coefficient y;

$$p_k = X_k^T \Omega t_k / t_k^T \Omega t_k, \ d_k = y_k^T \Omega t_k / t_k^T \Omega t_k. \tag{19}$$

8:     Update $\hat{y}_q = \hat{y}_q + t_{q,k} d_k$.
9:     **if** $k = K$ **then**
10:        Output $\hat{y}_q$ as an estimate.
11:    **else**
12:        Calculate X_{k+1}, y_{k+1}, and $x_{q,k+1}$;

$$X_{k+1} = X_k - t_k p_k^T \tag{20}$$

$$y_{k+1} = y_k - t_k d_k \tag{21}$$

$$x_{q,k+1} = x_{q,k} - t_{q,k} p_k. \tag{22}$$

13:    **end if**
14: **end for**

2.3. Missing RRI Interpolation

When the $(j + 1)$th R wave is not detected and the measured rth R wave is expressed as Equation (1), only the jth RRI estimate $\hat{r}_j$ is required because the next RRI estimate $\hat{r}_{j+1}$ can be calculated from r_j and $\hat{r}_j$: $\hat{r}_{j+1} = r_j - \hat{r}_j$. At this time, successive missing occurrences are not considered.

Appropriate input variables should be determined for interpolating missing RRI by LWPLS. Multiple past RRI measurements $r_{j-1}, \cdots, r_{j-L+1}$ and the current measurement r_j are used as input variables;

$$\boldsymbol{x}_j = [r_j/2, r_{j-1}, \cdots, r_{j-L}] \tag{23}$$

where L is the number of past measurements. Although the current measurement may be useful for interpolation, the measured r_j is about double of $\tilde{r}_j$. Thus, its average $r_j/2$ is used as input.

The procedure of the proposed LWPLS-RI is described in Algorithm 3. Before missing RRI interpolation starts, the initial RRI buffer has to be stored for more than the buffer size W. In step 4, $\bar{r}$ is the threshold for finding an R wave detection error. When the RRI measurement r_j exceeds $\bar{r}$, it is determined that an R wave detection error has occurred. The threshold $\bar{r}$ is a predetermined parameter to be tuned beforehand; however, the default value of $\bar{r}$ can be set to 1,500 msec. In step 5, the newly measured RRI r_j is queued into the RRI buffer in a first-in-first-out (FIFO) manner when an R wave is detected correctly. On the other hand, $\hat{r}_j$ is estimated by LWPLS when an R wave detection error occurs in steps 8–10. Finally, r_j is replaced by the estimated $\hat{r}_j$ and $\hat{r}_{j+1}$ which are queued into the RRI buffer in a FIFO manner in steps 11 and 12. The proposed LWPLS-RI has four tuning parameters: the localization parameter φ, the numbers of latent variables K and past RRI measurements used for an input L, and the buffer size W.

Algorithm 3 LWPLS-RI.

1: Set ϕ, R, $\bar{r}$, and l.
2: **while do**
3:　　Measure the jth RRI r_j.
4:　　**if** $r_j \leq \bar{r}$ **then**
5:　　　　Enqueue r_j into the RRI buffer in the FIFO manner.
6:　　　　Wait until the next RRI r_{j+1} is measured.
7:　　**else**
8:　　　　Construct an input $\boldsymbol{x}_j$ as Equation (23) from the RRI buffer.
9:　　　　Estimate the jth RRI $\hat{r}_j$ from $\boldsymbol{x}_j$ by using LWPLS.
10:　　　　Calculate the $j+1$ RRI estimate; $\hat{r}_{j+1} = r_j - \hat{r}_k$.
11:　　　　Replace r_j by $\hat{r}_j$ and $\hat{r}_{j+1}$.
12:　　　　Enqueue $\hat{r}_j$ and $\hat{r}_{j+1}$ into the RRI buffer in the FIFO manner.
13:　　　　Wait until the next RRI r_{j+2} is measured.
14:　　**end if**
15: **end while**

In the proposed algorithm, the RRI data collected from any persons can be used for the initial RRI buffer. Even when the RRI data collected from persons other than users are stored in the initial RRI buffer, the stored RRI data are replaced by the RRI data measured from users themselves through the FIFO manner in steps 11 and 12 in Algorithm 3.

3. Results and Discussion

This section evaluates and discusses the interpolation performance of the proposed LWPLS-RI through its application to RRI data in which missing RRIs were introduced artificially. Long-term ECG measurement failure and an ectopic RRI caused by arrhythmia were not considered here.

3.1. Simulation Procedure

This case study used the MIT-BIH Normal Sinus Rhythm Database for objective data construction [35]. ECG data measured from subjects 1–18 were clipped from the database and ECG data containing strong artifacts were eliminated. The R waves in the clipped ECG data were detected by using the peak detection algorithm, and each RRI was calculated. Each of eighteen pieces of RRI

data, $r^{[s]}$, was divided into three datasets for parameter optimization of LWPLS $r_o^{[s]}$, initial RRI buffer in Algorithm 3, $r_b^{[s]}$, and validation $r_v^{[s]}$, where $s(s = 1, \cdots, 18)$ denotes the subject index. The numbers of samples in $r_o^{[s]}$, $r_b^{[s]}$, and $r_v^{[s]}$ were 6000, 500, and 5000, respectively. There was no R wave detection error in any of these datasets.

Missing RRIs were artificially introduced to the parameter optimization dataset $r_o^{[s]}$ and the validation datasets $r_v^{[s]}$ in a random manner as $r'_j = r_j + r_{j+1}$ and eliminated r_{j+1}. There were no successive missing RRIs, and the missing rates were $\alpha = 0.3\%$, 0.5%, and 1%. Since the ordinal heart rate of a healthy adult is about 60–80 bpm, 1% missing means that R wave detection error occurs about every 1.5 min. Figure 3 illustrates an example of RRI data with $\alpha = 0.5$ which was generated from the RRI data of subject 16 $r_v^{[16]}$.
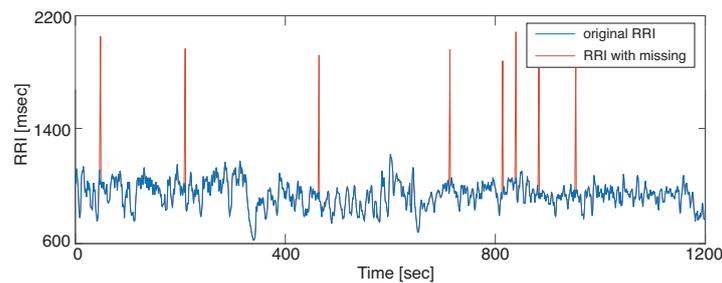

Figure 3. Example of missing RRI with $\alpha = 0.5\%$: some RRIs were intentionally eliminated randomly.

In this case study, the HRV features described in Appendix A were extracted. A rectangular sliding window whose size is three minutes was used. The time domain features were extracted directly from the raw RRI data. For frequency domain feature extraction, the RRI data were interpolated by the third-order spline and resampled at 4 Hz. An AR model of order 40 was used to calculate frequency domain features. These HRV feature extraction settings were determined based on [18]. Figure 4 shows NN50, RMSSD, HF, and LF/HF extracted from the data in Figure 3. HRV features except for NN50 change greatly due to missing RRIs and such a situation may deteriorate the performance of HRV-based monitoring services.
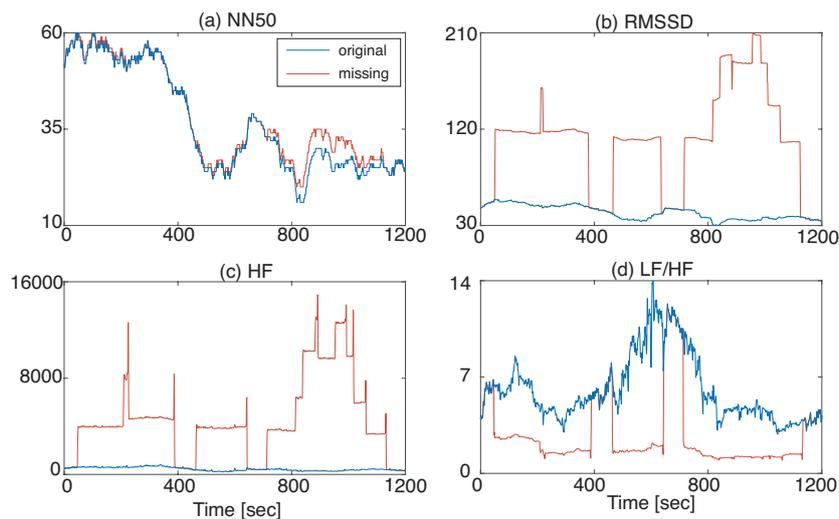

Figure 4. HRV features extracted from the missing RRI: (**a**) NN50, (**b**) RMSSD, (**c**) HF, and (**d**) LF/HF. There were large errors in HRV features due to frequent missing RRI

Before applying the proposed LWPLS-RI to the validation datasets $r_v^{[s]}$, appropriate parameters in LWPLS were determined using the parameter optimization datasets $r_o^{[s]}$. 1% of R wave detection errors artificially occurred in all $r_o^{[s]}$ in a random manner. The localization parameter φ and the number of

latent variables *K* in LWPLS were determined so that the root mean squared error (RMSE) between the true and interpolated RRIs was minimized. Here, the RRI buffer size *W* and the number of past RRIs used for input *L* were fixed to 500 and 3. The determined parameters were $\varphi = 1.3$ and $K = 3$, respectively.

The proposed method with the determined parameters was applied to the validation datasets $r_v^{[s]}$ ($s = 1, \cdots, 18$). Although step 4 in Algorithm 3 judges missing RRI occurrences, missing RRI positions were known in this case study for interpolation performance evaluation. The following three interpolation methods were tested for comparison:

- MEAN: Replace the missing RRI r_j by a mean of x_j.
- Equal Division (ED): Replace r_j by the value of $r_j/(q+1)$, where q is the number of successive R wave detection errors.
- PLS-RI: Replace r_j by an output of a PLS model whose input is x_j.

The ectopic RRI remove method [24] was not tested in this case study because it can not be used in online applications. After interpolation by these methods, RMSE between the true and interpolated RRIs and RMSE between HRV features derived from the true and interpolated RRIs were calculated.

In this case study, missing RRI generation in the validation datasets $r_v^{[k]}$ and missing RRI interpolation were repeated 30 times for precise performance evaluation.

3.2. Interpolation Results

Figure 5 shows application results of four missing RRI interpolation methods to RRI data with $\alpha = 0.5\%$ in Figure 3 and HRV extraction from the interpolated RRI data. These RMSEs of the four methods were calculated through a simulation repeated 30 times using all of the eighteen pieces of validation data. RMSEs of PLS-RI and the proposed LWPLS-RI were lower than those of MEAN and ED, and LWPLS-RI was the best. The proposed LWPLS-RI improved RMSE about 70% in comparison with MEAN when $\alpha = 0.5\%$. In addition, the interpolation performance hardly changed even if the error rate increased.
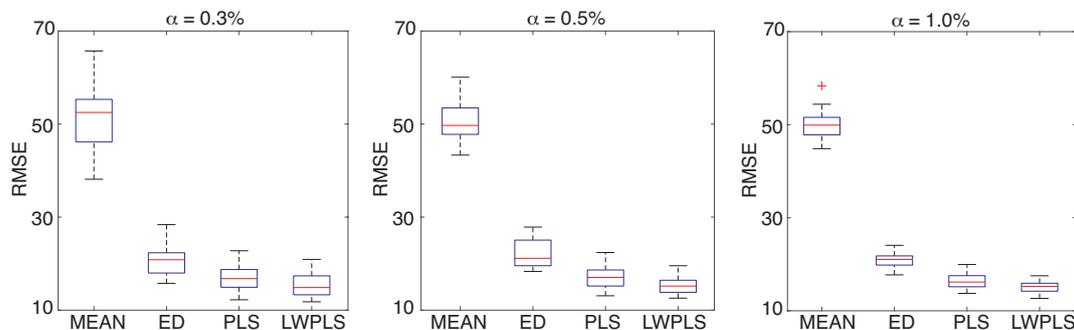

Figure 5. RMSE of RRI interpolated by MEAN, ED, PLS, and LWPLS when $\alpha = 0.3\%$ (**left**), 0.5% (**center**), and 1% (**right**) when a simulation was repeated 30 times using all of the eighteen pieces of validation data. The proposed LWPLS-RI achieved the best interpolation performance.

HRV features were derived from the RRI data interpolated by four methods. Figures 6–8 show RMSEs of eight HRV features calculated by the interpolated RRIs through a simulation repeated 30 times. The proposed LWPLS-RI achieved the highest performance.
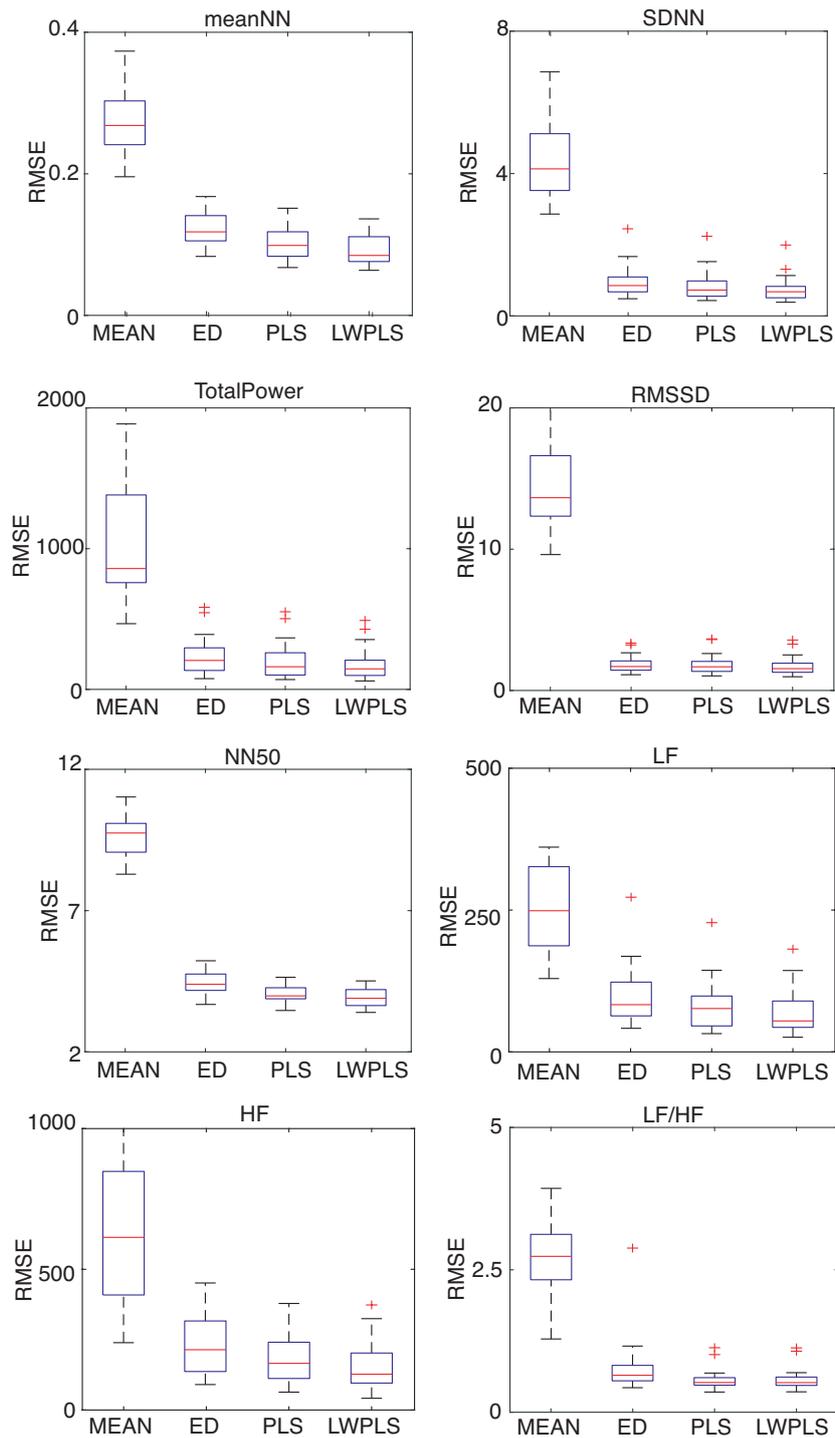
Figure 6. RMSE of HRV features derived from RRIs interpolated by MEAN, ED, PLS, and LWPLS when $\alpha = 0.3\%$. These eight HRV features were explained in the Appendix A. RMSEs of the four methods were calculated through a simulation repeated 30 times using all of the eighteen pieces. The proposed LWPLS-RI achieved the best HRV analysis performance.
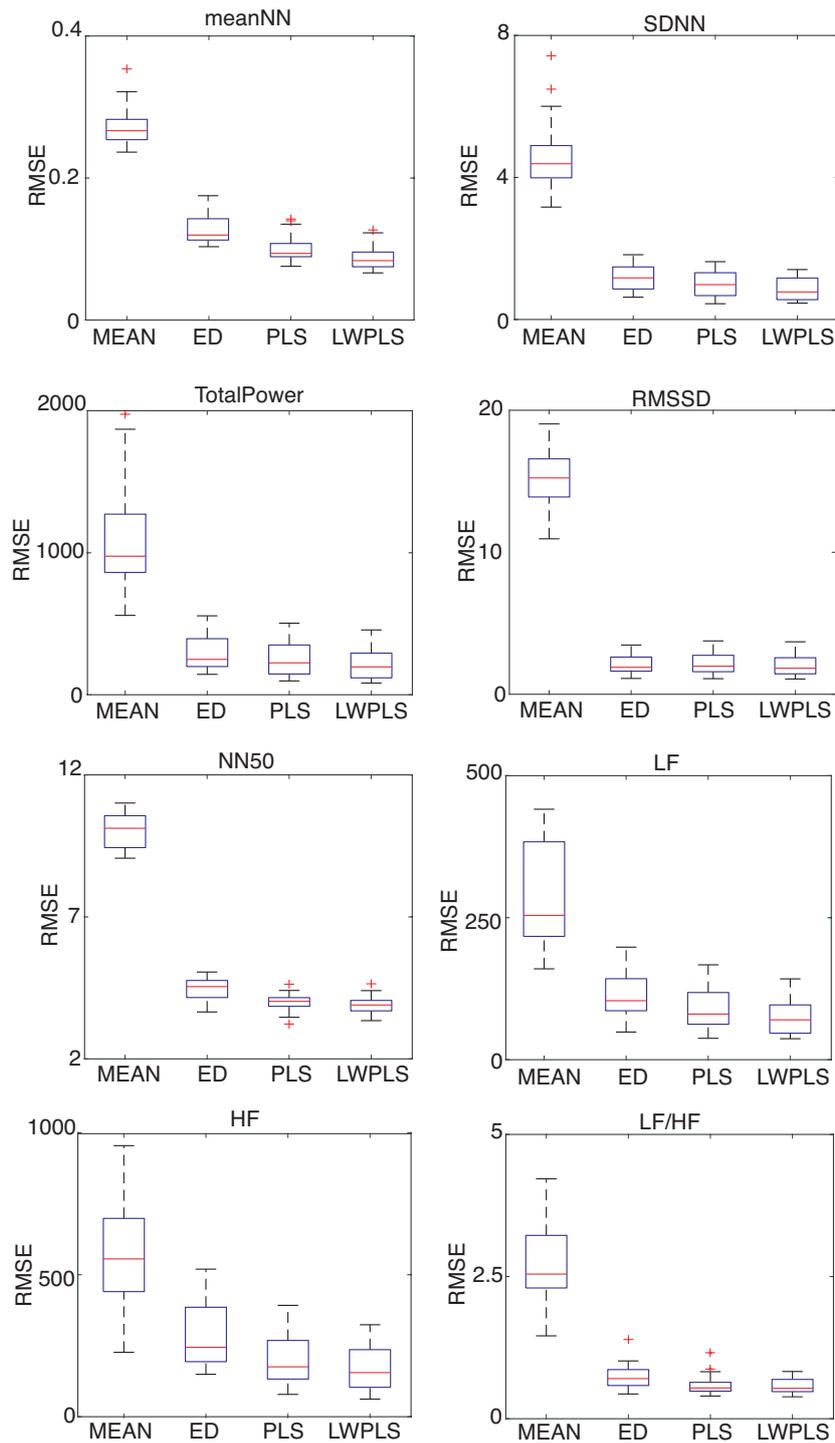
Figure 7. RMSE of HRV features derived from RRIs interpolated by MEAN, ED, PLS, and LWPLS when $\alpha = 0.5\%$. These eight HRV features were explained in the Appendix A. RMSEs of the four methods were calculated through a simulation repeated 30 times using all of the eighteen pieces. The proposed LWPLS-RI achieved the best HRV analysis performance.
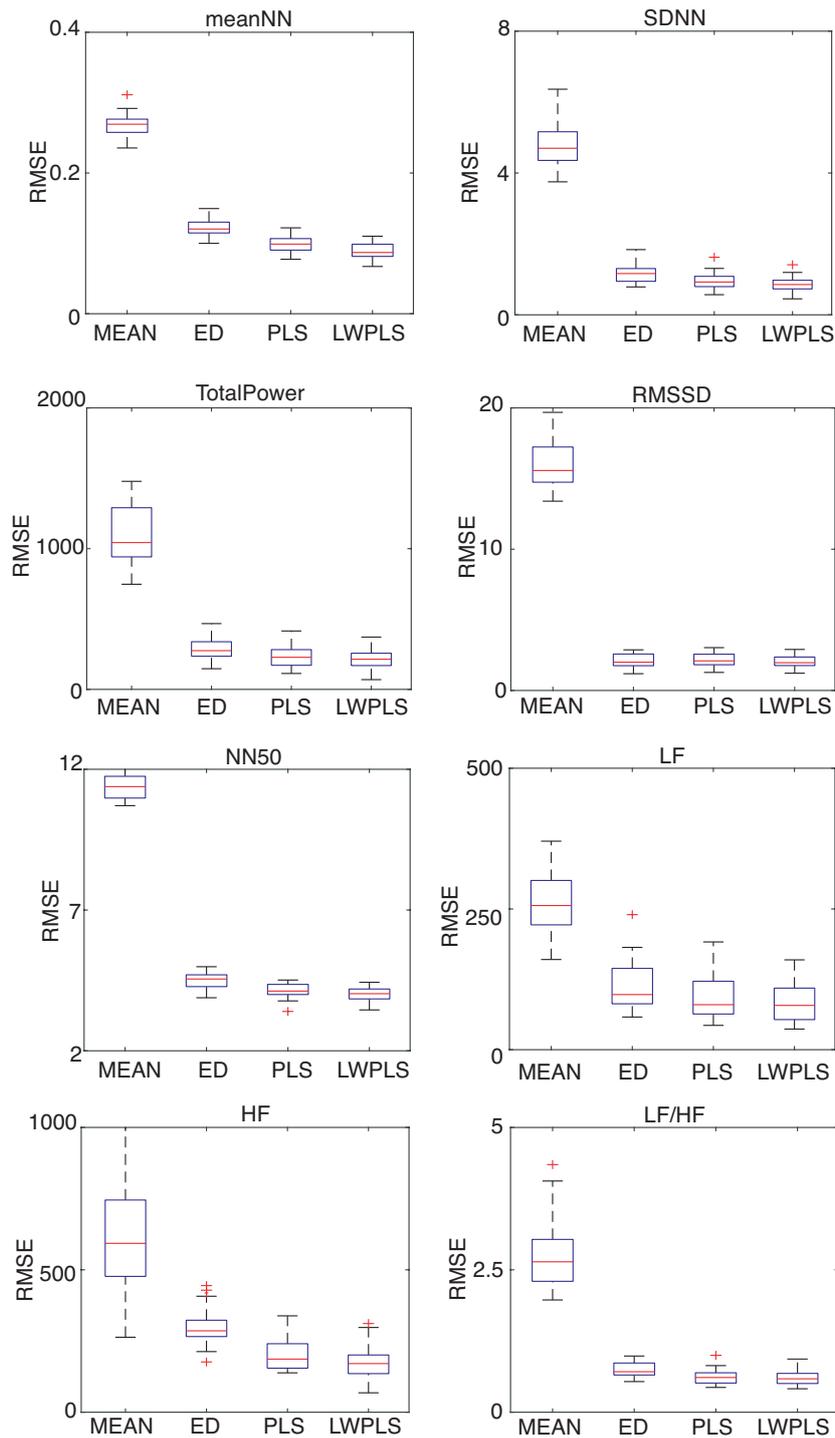
Figure 8. RMSE of HRV features derived from RRIs interpolated by MEAN, ED, PLS, and LWPLS when $\alpha = 1.0\%$. These eight HRV features were explained in the Appendix A. RMSEs of the four methods were calculated through a simulation repeated 30 times using all of the eighteen pieces. The proposed LWPLS-RI achieved the best HRV analysis performance.

These results demonstrate the usefulness of the proposed LWPLS-RI for HRV analysis from RRI data with missing RRIs.

3.3. Discussion

According to the simulation, the proposed LWPLS-RI achieved the best performance of missing RRI interpolation, of which the mean of RMSEs calculated from all subjects was 15.1 when the error rate was $\alpha = 1\%$. However, there were differences among subjects as shown in Figure 9. RMSE calculated from subject 6 was worse than the other subjects. None of the interpolation methods could improve his/her RMSE. The standard deviations of the raw RRI data of subject 6 and subject 5 whose RMSE was small were 61.8 and 140.4, respectively, which clearly showing that RRI fluctuation of subject 6 was much larger than that of subject 5 although there were no RRI detection errors. This indicates that a missing RRI cannot be interpolated appropriately when the fluctuation range of RRI is huge. LWPLS uses the past information most similar to a sample to be estimated; however, the number of past similar RRIs for interpolation becomes small when RRIs fluctuate largely.
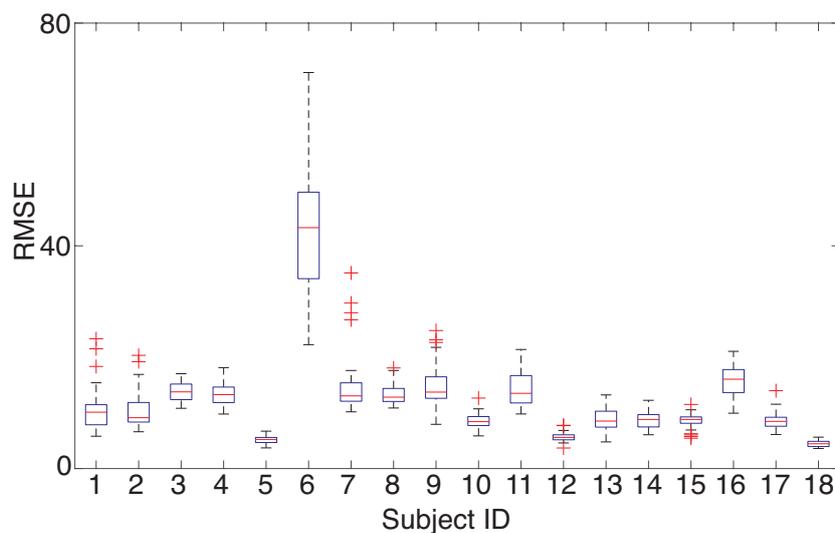

Figure 9. RMSE of RRI interpolation in each subject by LWPLS when $\alpha = 0.5\%$. Although RMSEs of almost all subjects except Subject 6 were less than twenty, that of Subject 6 was more than forty.

The current RRI measurement $r_j (= \tilde{r}_j + \tilde{r}_{j+1})$ is used for utilizing missing information. To investigate the effect of using the current measurement, RRI interpolation without r_j was tested. The number of past RRIs used for input L was $L = 4$ so that the total number of input variables did not change. As a result, the means of RMSE by MEAN, PLS-RI, and LWPLS-RI were 49.1, 34.2, and 31.9, respectively, which are worse than RRI interpolation with r_j. This shows that the current RRI measurement r_j should be used for interpolation in any method.

In the proposed LWPLS-RI, the RRI buffer size W and the number of past RRIs used for input L were fixed to 500 and 3 in this case study. To investigate the effect of these parameters on the interpolation performance of the proposed LWPLS-RI, different parameter settings were compared. The RRI buffer size W was changed to $W = 100, 300, 500$, and 1000 and other parameters were fixed to the same values as Section 3.1. RMSEs were calculated using validation data with $\alpha = 0.5\%$ in a simulation repeated 30 times. The means of RMSE were 15.9, 15.4, 15.7, and 15.1 when $W = 100, 300, 500$, and 1000, respectively. These numbers show that the interpolation performance was not improved when a large buffer size was selected, and in fact became worse when $W = 100$. Thus, $W = 300$ is sufficient for RRI interpolation.

The number of past RRIs used for input L was changed to $L = 2, 3, 4$, and 5, and other parameters were fixed to the same values as Section 3.1, except the number of latent variables K which was determined using the parameter optimization dataset for every different L. RMSEs were calculated using validation data with $\alpha = 0.5\%$ in a simulation repeated 30 times. The means of RMSE were 12.7,

12.8, 12.9, and 12.8 when $L = 2, 3, 4$, and 5, which shows that the interpolation performance did not change, regardless of which L was selected.

In addition, the proposed LWPLS-RI was applied to the RRI data containing two successive missing RRIs and compared with the single RRI missing case. In this simulation, two successive missing RRIs were randomly generated at five different points in each validation dataset, and they were interpolated sequentially by LWPLS-RI whose parameters were the same as Section 3.1. The following input was used for the interpolation instead of Equation (23):

$$x_{j,1} = [r_j/3, r_{j-1}, \cdots, r_{j-L}].$$ (24)

The interpolated RRI for the first missing $\hat{r}_{j,1}$ was added to the input for the second interpolation as

$$x_{j,2} = [(r_j - \hat{r}_{j,1})/2, \hat{r}_{j,1}, r_{j-1}, \cdots, r_{j-L+1}].$$ (25)

Using these inputs, the RRI data containing two successive missing RRIs were interpolated by LWPLS-RI sequentially, and RMSEs of the first and the second interpolations were calculated. This procedure was repeated 30 times. The mean of RMSE of the first missing interpolation was 24.9, which was worse than that of the single missing interpolation. More information about RRI was lost in the successive missing case than the single missing case, which makes it difficult to interpolate RRI. Thus, it becomes more difficult to interpolate missing RRIs when more than three successive missing RRIs occur. However, the mean of RMSE of the second missing interpolation was 15.2 which is almost the same as the single missing interpolation case. This may have been caused by input variable construction for the second interpolation. Equation (25) uses the interpolated first RRI $\hat{r}_{j,1}$. Information lost through successive missing RRIs was recovered to some extent by the first interpolation and $\hat{r}_{j,1}$ contained such recovered information that may be useful for the second missing interpolation.

The proposed LWPLS-RI can be easily implemented in mobile computers such as a smartphone because the computational load is much lighter than methods that need to process ECG signals directly. This is important from the viewpoint of practical use. An HRV-based epileptic seizure prediction smartphone app has already been developed and tested in hospitals [18]. In addition, an HRV-based drowsy driving detection smartphone app has been commercialized [9]. The performances of these smartphone apps will be improved through appropriate RRI interpolation by the proposed method.

It is concluded that the proposed LWPLS-RI has the potential for realizing highly accurate HRV-based health monitoring services in the future.

4. Conclusions

A missing RRI interpolation method was proposed utilizing the framework of JIT modeling, whereby a local model is constructed, and missing RRI is estimated by the constructed model only when R wave detection errors occur. The result of applying the proposed LWPLS-RI to real RRI data with artificial missing RRIs showed that it interpolated missing RRI more appropriately than other methods. The proposed method is not applicable to long-term ECG measurement failure since it modifies only one RRI. HRV analysis should be stopped during long-term ECG measurement failure.

The limitations of this study include the properties of the objective data collected from the Physionet database, such as the limited number of subjects and the fact that all subjects were healthy and did not have arrhythmia or other cardiovascular diseases.

In the future work, we will develop a unified framework for interpolating and modifying ectopic RRI data caused by arrhythmia as well as R wave detection errors.

Author Contributions: K.K. and K.F. developed the proposed method and wrote the initial draft of the manuscript. T.K. analyzed the data. M.K. contributed to data collection and analysis and assisted in the preparation of the manuscript. All authors approved the final version of the manuscript, and agree to be accountable for all aspects of the work.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Heart Rate Variability Analysis

This appendix introduces HRV analysis. A typical ECG trace consists of some peaks, and the highest peak is called an R wave. The RR interval (RRI) [ms] is defined as an interval between an R wave and the next R wave. The R wave is detected by using a derivative-based peak detection algorithm [36]. Every RRI changes due to ANS activities, and this phenomenon is called HRV. Standard HRV features are classified into time domain features and frequency domain features [1].

The time domain features are obtained from the original RRI data.

- meanNN: Mean of RRI.
- SDNN: Standard deviation of RRI.
- Total Power (TP): Variance of RRI.
- RMSSD: Root means square of the difference of adjacent RRI.
- NN50: Number of pairs of adjacent RRI, whose difference is more than 50 ms.

Frequency domain features are defined based on the power spectrum density (PSD) of the resampled RRI data, which is derived by Fourier analysis or an autoregressive (AR) model.

- LF: Power of the low-frequency band (0.04–0.15Hz) in PSD. LF reflects the activity of both the sympathetic and parasympathetic nervous systems.
- HF: Power of the high-frequency band (0.15–0.4Hz) in PSD. HF reflects the parasympathetic nervous system activity.
- LF/HF: Ratio of LF to HF. LF/HF expresses the balance between the sympathetic nervous system activity and the parasympathetic nervous system activity.

For frequency analysis, the raw RRI data are interpolated by using spline and are resampled at equal intervals, because they are not sampled at equal intervals. Appropriate settings of HRV analysis should be determined depending on applications; however, the HRV analysis guideline recommends that RRI data should be measured for more than two minutes and that the sampling rate should be at least 200 Hz [1].

References

1. Variability, H.R. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation* **1996**, *93*, 1043–1065.
2. Karemaker, J.M. An introduction into autonomic nervous function. *Physiol. Meas.* **2017**, *38*, R89. [CrossRef] [PubMed]
3. Kleiger, R.E.; Miller, J.P.; Bigger, J.T.; Moss, A.J. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am. J. Cardiol.* **1987**, *59*, 256–262. [CrossRef]
4. Malliani, A. Cardiovascular neural regulation explored in the frequency domain. *Circulation* **1991**, *84*, 482–492. [CrossRef] [PubMed]
5. Versace, F.; Mozzato, M.; De Min Tona, G.; Cavallero, C.; Stegagno, L. Heart rate variability during sleep as a function of the sleep cycle. *Biol. Psychol.* **2003**, *63*, 146–162. [CrossRef]
6. Jurysta, F.; van de Borne, P.; Migeotte, P.F.; Dumont, M.; Lanquart, J.P.; Degaute, J.P.; Linkowski, P. A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men. *Clin. Neurophysiol.* **2003**, *114*, 2146–2155. [CrossRef]
7. Vicente, J.; Laguna, P.; Bartra, A.; Bailon, R. Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* **2016**, *54*, 927–937. [CrossRef] [PubMed]
8. Patel, M.; Lal, S.K.L.; Kavanagh, D.; Rossiter, P. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert. Syst. Appl.* **2011**, *38*, 7235–7242. [CrossRef]

9. Abe, E.; Fujiwara, K.; Hiraoka, T.; Yamakawa, T.; Kano, M. Development of Drowsiness Detection Method by Integrating Heart Rate Variability Analysis and Multivariate Statistical Process Control. *SICE J. Control Meas. Syst. Integr.* **2016**, *9*, 10–17. [CrossRef]

10. Fujiwara, K.; Abe, E.; Kamata, K.; Nakayama, C.; Suzuki, Y.; Yamakawa, T.; Hiraoka, T.; Kano, M.; Sumi, Y.; Masuda, F.; Matsuo, M.; Kadotani, H. Heart Rate Variability-based Driver Drowsiness Detection and its Validation with EEG. *IEEE Trans. Biomed. Eng.* **2018**. [CrossRef] [PubMed]

11. Aeschbacher, S.; Bossard, M.; Schoen, T.; Schmidlin, D.; Muff, C.; Maseli, A.; Leuppi, J.D.; Miedinger, D.; Probst-Hensch, N.M.; Schmidt-Trucksass, A.; Risch, M.; Risch, L.; Conen, D. Heart Rate Variability and Sleep-Related Breathing Disorders in the General Population. *Am. J. Cardiol.* **2016**, *118*, 912–917. [CrossRef] [PubMed]

12. Gong, X.; Huang, L.; Liu, X.; Li, C.; Mao, X.; Liu, W.; Huang, X.; Chu, H.; Wang, Y.; Wu, W.; Lu, J. Correlation Analysis between Polysomnography Diagnostic Indices and Heart Rate Variability Parameters among Patients with Obstructive Sleep Apnea Hypopnea Syndrome. *PLoS One* **2016**, *11*, e0156628. [CrossRef] [PubMed]

13. Alvarez-Estevez, D.; Moret-Bonillo, V. Spectral Heart Rate Variability analysis using the heart timing signal for the screening of the Sleep Apnea-Hypopnea Syndrome. *Comput. Biol. Med.* **2016**, *71*, 14–23. [CrossRef] [PubMed]

14. Roche, F.; Gaspoz, J.M.; Court-Fortune, I.; Minini, P.; Pichot, V.; Duverney, D.; Costes, F.; Lacour, J.R.; Barthelemy, J.C. Screening of Obstructive Sleep Apnea Syndrome by Heart Rate Variability Analysis. *Circulation* **1999**, *100*, 1411–1415. [CrossRef] [PubMed]

15. Jeppesen, J.; Beniczky, S.; Johansen, P.; Sidenius, P.; Fuglsang-Frederiksen, A. Detection of epileptic seizures with a modified heart rate variability algorithm based on Lorenz plot. *Seizure* **2015**, *24*, 1–7. [CrossRef] [PubMed]

16. Behbahani, S.; Dabanloo, N.J.; Nasrabadi, A.M.; Teixeira, C.A.; Dourado, A. Pre-ictal heart rate variability assessment of epileptic seizures by means of linear and non-linear analyses. *Anadolu Kardiyol Derg.* **2013**, *13*, 797–803. [CrossRef] [PubMed]

17. Calandra-Buonaura, G.; Toschi, N.; Provini, F.; Corazza, I.; Bisulli, F.; Barletta, G.; Vandi, S.; Montagna, P.; Guerrisi, M.; Tinuper, P.; Cortelli, P. Physiologic autonomic arousal heralds motor manifestations of seizures in nocturnal frontal lobe epilepsy: Implications for pathophysiology. *Sleep Med.* **2012**, *13*, 252–262. [CrossRef] [PubMed]

18. Fujiwara, K.; Miyajima, M.; Yamakawa, T.; Abe, E.; Suzuki, Y.; Sawada, Y.; Kano, M.; Maehara, T.; Ohta, K.; Sasai-Sakuma, T.; Sasano, T.; Matsuura, M.; Matsushima, E. Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1321–1332. [PubMed]

19. Yamakawa, T.; Kinishita, R.; Fujiwara, K.; Kano, M.; Miyajima, M.; Sakata, T.; Ueda, Y. Accuracy Comparison of Two Microcontroller-embedded R-wave Detection Methods for Heart-rate Variability Analysis. In Proceedings of the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Hong Kong, China, 16–19 December 2015; pp. 1010–1013.

20. Imani, S.; Bandodkar, A.J.; Mohan, A.M.; Kumar, R.; Yu, S.; Wang, J.; Mercier, P.P. A wearable chemical-electrophysiological hybrid biosensing system for real-time health and fitness monitoring. *Nat. Commun.* **2016**, *7*, 11650. [CrossRef] [PubMed]

21. Kim, J.; Sempionatto, J.R.; Imani, S.; Hartel, M.C.; Barfidokht, A.; Tang, G.; Campbell, A.S.; Mercier, P.P.; Wang, J. Simultaneous Monitoring of Sweat and Interstitial Fluid Using a Single Wearable Biosensor Platform. *Adv Sci (Weinh)* **2018**, *5*, 1800880. [CrossRef] [PubMed]

22. Mishr, R.K.; Barfidokht, A.; Karajic, A.; Sempionatto, J.R.; Wang, J.; Wang, J. Wearable potentiometric tattoo biosensor for on-body detection of G-type nerve agents simulants. *Sens. Actuators B Chem.* **2018**, *273*, 966–972. [CrossRef]

23. Choi, J.; Ghaffari, R.; Baker, L.B.; Rogers, J.A. Skin-interfaced systems for sweat collection and analytics. *Sci. Adv.* **2018**, *4*, eaar3921. [CrossRef] [PubMed]

24. Lippman, N.; Stein, K.M.; Lerman, B.B. Comparison of methods for removal of ectopy in measurement of heart rate variability. *Am. J. Physiol.* **1994**, *267*, H411–H418. [CrossRef] [PubMed]

25. Clifford, G.D.; Tarassenko, L. Quantifying errors in spectral estimates of HRV due to beat replacement and resampling. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 630–638. [CrossRef] [PubMed]

26. Mateo, J.; Laguna, P. Analysis of Heart Rate Variability in the Presence of Ectopic Beats Using the Heart Timing Signal. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 334–343. [CrossRef] [PubMed]

27. Atkeson, C.G.; Moore, A.W.; Schaal, S. Locally Weighted Learning. *Artif. Intel. Rev.* **1997**, *11*, 11–73. [CrossRef]

28. Bontempi, G.; Birattari, M.; Bersini, H. Lazy learing for local modelling and control design. *Int. J. Cont.* **1999**, *72*, 643–658. [CrossRef]

29. Cheng, C.; Chiu, M. A new data-based methodology for nonlinear process modeling. *Chem. Eng. Sci.* **2004**, *59*, 2801–2810. [CrossRef]

30. Fujiwara, K.; Kano, M.; Hasebe, S. Soft-Sensor Development using Correlation-Based Just-In-Time modeling. *AIChE J.* **2009**, *55*, 1754–1765. [CrossRef]

31. Kano, M.; Fujiwara, K. Virtual Sensing Technology in Process Industries: Trends and Challenges Revealed by Recent Industrial Applications. *J. Chem. Eng. Jpn.* **2013**, *46*, 1–17. [CrossRef]

32. Kim, S.; Okajima, R.; Kano, M.; Hasebe, S. Development of soft-sensor using locally weighted PLS with adaptive similarity measure. *Chemometr. Intell. Lab. Syst.* **2013**, *124*, 43–49. [CrossRef]

33. Kamata, K.; Fujiwara, K.; Yamakawa, T.; Kano, M. Missing RRI interpolation for HRV analysis using Locally-Weighted Partial Least Squares Regression. In Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Orlando, FL, USA, 17–20 August 2016; pp. 2386–2389

34. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression: a Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]

35. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **2000**, *101*, e215–e220. [CrossRef] [PubMed]

36. Pan, J.; Tompkins, W.J. A Real-Time QRS Detection Algorithm. *IEEE Trans. Biomed. Eng.* **1985**, *32*, 230–236. [CrossRef] [PubMed]