Sensors and Materials, Vol. 34, No. 8 (2022) 2985–2999
MYU Tokyo

S & M 3018

Reliability Estimation and Filtering of Heart Rate Measurement Using Inertial Sensor during Exercise

Hiroki Yoshikawa,[1,2*] Masayuki Hayashi,[2]
Akira Uchiyama,[2] and Teruo Higashino[1,2]

[1]Faculty of Engineering, Kyoto Tachibana University,
Oyake Yamada 34, Yamashina, Kyoto 607–8175, Japan
[2]Graduate School of Information Science and Technology, Osaka University,
Yamadaoka 1–5, Suita, Osaka 565–0871, Japan

(Received May 16, 2022; accepted July 11, 2022)

Keywords: Kalman filter, motion artifacts, PPG sensor, heart rate measurement

Heart rate (HR) measurement by a wrist-worn device suffers from noises owing to body movement. Even though many researchers have proposed sophisticated methods for the compensation of noises in measurement, such noises corrupt the sensor data itself, leading to difficulty in compensation. In this paper, we design a method for estimating the reliability of HR measurement. Our design principle is based on the fact that the change in HR is correlated with the magnitude of body movement and the current HR. To model the correlation, we construct a modified Kalman filter that estimates the displacement of the HR from the statistical data of the HR measured by a precise heart rate sensor and the variance of acceleration measured by a wrist-worn device. Then, we define the reliability of HR measurement as the absolute error between the output of a modified Kalman filter and the HR measured by the wrist-worn device. For evaluation, we compare our method with conventional outlier removal and smoothing after compensation using one of the state-of-the-art methods based on deep learning. Our method successfully removes 18.9% of the measurements with low reliability while achieving a mean absolute error of 6.25 bpm, a superior value to the conventional methods, for a single subject. For multiple subjects, our method decreases the mean absolute error by 13.1% on average.

1. Introduction

Wrist-worn devices such as smartwatches and smart wristbands have become popular, providing various smart applications in daily life. Specifically, most wrist-worn devices are capable of monitoring the heart rate (HR) for healthcare and sports. Various learning-based applications that need to collect physiological data including HR have been proposed.(1,2) In such applications, we may be able to accumulate big physiological data using the devices from many participants. In big data analysis, the reliability of the data is essential to develop machine-learning-based applications because noisy data reduces performance.

*Corresponding author: e-mail: yoshikawa-h@tachibana-u.ac.jp
https://doi.org/10.18494/SAM3969

ISSN 0914-4935

To measure the HR by wrist-worn devices, a reflectance-type photoplethysmography (PPG) sensor is widely used. The PPG sensor measures intensity changes in the light reflected from the skin. These PPG signals represent the changes in the arterial blood volume between the systolic and diastolic phases of a cardiac cycle. However, the PPG sensor is vulnerable to movement of the wrist-worn device because the light intensity can change owing to the slight movement of the device. A problem is that when attempting to accumulate big data from many participants in daily life, it is difficult to check that they are wearing the devices correctly. The diversity of the device attachment leads to different levels of reliability of the measured data. Furthermore, even if the device is correctly attached, the measurement accuracy of the HR may be significantly reduced owing to motion artifacts[3,4] and intense exercise.[5] Therefore, the reliability of the collected physiological data is nonuniform, which is a serious problem in big data analysis.

To remove noisy data, many researchers and analysts start with data cleansing[6–8] because noisy data commonly arises in big data analysis. Most existing approaches focus on outlier removal[9] based on statistics. However, they may incorrectly remove correct rare data, which is significant for some applications. To solve the problem, frequency-based data correction methods have been proposed,[4,10,11] where the key idea is identifying the frequency band from the acceleration spectrum generated by the user's motion. Finally, the HR is corrected by removing the noise induced by the motion from the PPG spectrum. As a state-of-the-art method for the correction of HR measurement, a deep learning (DL)-based method has been proposed. Chung *et al.* combined long short-term memory (LSTM) and a convolutional neural network (CNN) for correction using acceleration intensity,[12] where a PPG spectrum and an acceleration spectrum are input to the correction method.

However, we found a few cases where the accuracy is still low even if we apply the DL-based correction method to the PPG signal in our dataset composed of HR and acceleration time series collected from running subjects wearing a tightly attached wrist-worn device. Correction failures occurred when the effect of motion artifacts was large due to factors such as high exercise intensity. In such cases, the HR features in the PPG signal are much smaller than other noises. This leads to a decrease in HR correction accuracy because HR features cannot be extracted from the PPG signal. The accuracy also decreased when there was an overlap between the acceleration frequency and HR frequency. Because motion artifacts usually have larger changes in the PPG signal than the HR in the frequency domain, the overlap causes the loss of HR features in the PPG signal. In addition, when participants measure data by themselves, the accuracy of HR correction also depends on the sensing technology of the device.[13] Therefore, even if the DL-based correction method is applied, the HR measurement is not always correct.

In this study, we aim to improve the accuracy of the whole dataset after data collection by estimating the reliability of the measured HR. An overview of the proposed method is illustrated in Fig. 1. We focus on the fact that the HR in the following state can be estimated from the current state, i.e., the exercise intensity and HR value. We use the variance of acceleration as an index of the exercise intensity because our preliminary experiment showed that we could distinguish between an exercise state and a nonexercise state where the HR transition differs. On the basis of this observation, we build an HR transition model through data collection by a precise HR sensor, i.e., an ECG-based HR sensor, during exercise. Our reliability estimation

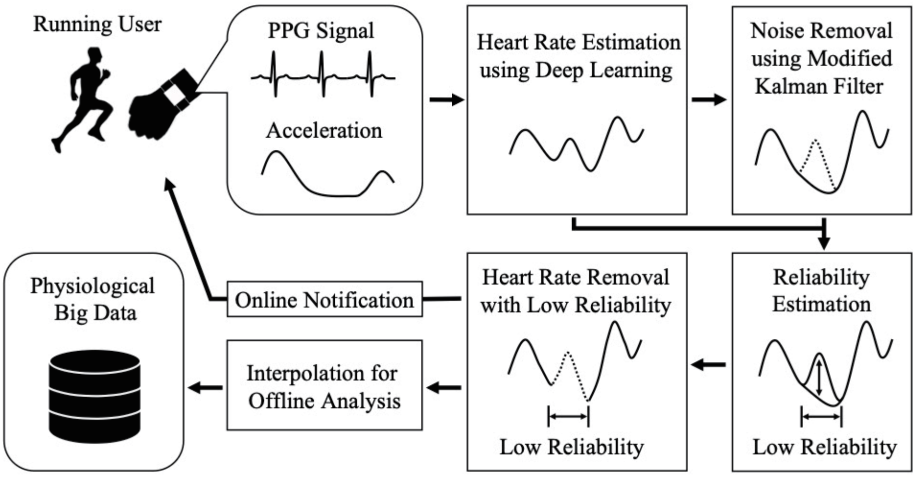

Fig. 1. Proposed method.

method is based on how the corrected HR differs from the transition model. The estimated reliability is useful for accumulating reliable physiological big data or for notifying subjects who incorrectly wear their wrist-worn device. This paper is an extended version of our previous paper, which presented the design of the reliability estimation method with an initial evaluation for one subject.(14) In this paper, we report a further evaluation with 19 subjects to examine the performance of the proposed method for different subjects.

For evaluation, we compare our method with conventional outlier removal(15) and smoothing after compensation using a state-of-the-art method based on DL. Our method successfully removed 18.9% of the measurements with low reliability while achieving a mean absolute error of 6.25 bpm, a superior value to the conventional methods, for a single subject. For multiple subjects, our method decreased the mean absolute error from 6.94 to 6.04 on average.

2. Related Work

As a noninvasive technique, wrist-worn devices utilize PPG, which leverages optical measurement to detect volumetric changes in blood circulation under the skin.(16) To investigate the accuracy of these devices relative to comparable medical-grade technology, many studies have compared the measurement accuracy with an ECG-based HR sensor.(5,17) Weiler *et al.* compared average HR measurements of two measurement technologies, i.e., PPG and ECG, after an interval-style cardio-based workout.(17) They found that when the HR reached around 155–160 bpm, PPG HR readings become less than ECG HR readings with a difference of 5 bpm. Climstein *et al.* assessed the test–retest reliability of a smartwatch and compared its measurement with an ECG-based sensor when measuring HR during treadmill exercise with various intensities.(5) The result also showed that the HR measured by the smartwatch is less that of the ECG-based sensor when the HR reached around 150–180 bpm. Their results showed that PPG-based HR sensors embedded in wrist-worn devices are less reliable than ECG-based HR sensors when the HR is relatively high. Therefore, the reliability of the measurement of wrist-worn devices is important for data collection with small errors.

To mitigate the effect of noise in the ECG signal, many researchers have focused on noise filtering. Eguchi *et al.* proposed an interbeat interval (RRI) outlier removal method for assessing the reliability of ECG signals collected by people in daily life.(18) Salai *et al.* proposed methods to remove outliers of RRI by finding abnormal values.(19) There are also methods to extract the HR from noisy ECG signals. Nakano *et al.* proposed a robust method to detect the instantaneous HR from noisy ECG signals utilizing a short-time autocorrelation technique.(20) Berwal *et al.* also proposed an efficient method for motion artifact removal from ambulatory ECG patterns to calculate RRI.(21)

On the other hand, PPG signal-based HR sensors may include noise generated by motion, especially for a wrist-worn device. For robust PPG sensing, Lee *et al.* implemented a multichannel PPG sensor for accurate estimation during intensive exercise.(22) Also, a PPG measuring system based on multichannel sensors with multiple wavelengths was developed, which employed a motion artifact reduction algorithm.(23) However, the specialized sensor increases the cost of manufacturing and the size of the device. Also, the large sensor may be a burden for the user. Many studies to tackle this problem with a software-based solution have been reported. Zhang *et al.* shared an open dataset containing simultaneously measured acceleration and PPG signals during exercise.(24) This dataset facilitated research on HR estimation from PPG signals. Khan *et al.* proposed rule-based filtering for the PPG signal from running subjects.(25) Fukushima *et al.* proposed an algorithm to estimate the HR using a wrist-worn PPG sensor for running subjects.(26) Also, Salehizadeh *et al.* proposed HR estimation from a PPG signal.(27) Their methods were based on frequency analysis, in which motion artifacts from the power spectrum of a PPG signal are rejected.

A DL-based method has recently been proposed as a state-of-the-art method to estimate the HR from a PPG signal.(12) The DL-based method reduces the effect of body movement by focusing on the characteristic of PPG sensors using LSTM and CNN layers. However, as mentioned in Sect. 1, noise cancelation based on frequency analysis may inherently fail to cancel noise in some particular cases. The Kalman filter is applied to the estimated HR by removing the motion artifacts from the PPG signal.(28,29) In these studies, there were many trials in which the HR could be measured accurately with the PPG signal. When outliers hardly appear in the estimated HR or the error is relatively small, the HR can be satisfactorily corrected with a conventional simple Kalman filter. However, the filtering-based correction is inherently weak in the case of a sudden change in the time series. In addition, the Kalman filter cannot consider the absolute value of the previous state to correct the current state. For the correction of the HR, the current value can be used to estimate the HR in the next time step. For example, the HR may greatly increase at the beginning of exercise, i.e., a state that the HR is low during exercise.

As mentioned above, frequency-based correction may not work well in particular scenarios. Also, filtering-based correction poorly corrects sudden changes. For machine learning, such corrections may reduce the performance of the model because they embed unnatural trends in the training data. To tackle this problem, we propose a reliability estimation method for removing unreliable data affected by noise. To estimate the reliability, Naeini *et al.* proposed a real-time PPG quality assessment approach using CNNs.(30) They estimated the PPG signal reliability as a binary classification, i.e., reliable or unreliable. In this paper, we estimate the reliability of the HR measurement as a numerical value. Therefore, users can specify their own thresholds for filtering depending on their requirements.

3. Reliability Estimation Method

We propose a reliability estimation method based on the Kalman filter to consider the change in the HR from the previous HR. We define the reliability as the difference between the measured HR and the filtered HR. In this study, we implemented an existing DL-based method(12) to estimate the HR under noise owing to motion artifacts instead of using the measured HR. The output of the DL-based method is hereinafter referred to as the estimated HR. The Kalman filter is often used for correcting observation errors. However, it cannot consider a complicated state transition of the HR. This is because the main parameters of the Kalman filter are the variances in the state equation and measurement equation. In addition, because the filter tends to be influenced by observations, it would be difficult to utilize it for the reliability estimation of observation values (i.e., HR). For example, if the HR is already high, it tends to increase slowly even during high-intensity exercise, which the Kalman filter cannot handle.

Therefore, we design a modified Kalman filter to consider the relationship between the exercise intensity and the current HR as follows:

$$h_{corr}^{i} = h_{est}^{i} + K_{gain}\left(h_{base} - h_{corr}^{i-1}\right), \tag{1}$$

$$K_{gain} = \frac{pv_{est}^{i}}{v_{est}^{i} + \sigma_v}, \tag{2}$$

$$v_{est}^{i} = \left(1 - K_{gain}^{i-1}\right)v_{est}^{i-1} + \sigma_w, \tag{3}$$

where h_{corr}^{i} is the corrected HR at time step i, h_{est}^{i} is the HR estimated from h_{corr}^{i-1} and the current exercise intensity, and h_{base} is the last reliable measurement before time step i. The exercise intensity is determined from the acceleration. The intuition behind our design is that the change in HR in one time step is bounded given the current HR and exercise intensity.

In this paper, we define transition states as six classes to represent the current HR and exercise intensity, and we design the modified Kalman filter on the basis of the conventional Kalman filter. We estimate the HR displacement from h_{corr}^{i-1} to h_{est}^{i} on the basis of the classes. The variables K_{gain}^{i} and v_{est}^{i} are the Kalman gain used by our modified Kalman filter and the corrected variance of the estimation error of the HR, respectively.

In our method, the initial value of v_{est}^{i} is σ_v. The parameters σ_v and σ_w are the variances of the noise of the measurement equation and state equation, respectively. We empirically set $\sigma_v = 2.63$ and $\sigma_w = 498.8$ through our preliminary experiment. The variable p is the likelihood of the transition of the HR from the current time step to the next time step, whose definition is based on the HR change distribution in one time step in the preliminary experiment. To calculate p, we define 13 levels (classes) of the HR transition for the range between +4.38 and −4.38, i.e., the range of each class is 0.674. This range is determined on the basis of our preliminary experiment where the HR transition out of the range was not measured in 2 s, which is the time window used for calculating the HR in our work. The likelihood of the transition is calculated as the ratio of the number of samples in the class to the number in the whole dataset.

We update the last reliable measurement of the HR h_{base} as follows. First, to assess whether the measurement at time step i is reliable, we use a predefined range $[t^i_{lower}, t^i_{upper}]$. The range depends on the transition states. If h^i_{obs} is in the range, h_{base} is replaced with h^i_{obs}. If h^i_{obs} is out of the range, we increase the range of the next step $(i + 1)$ to $[2t^i_{lower}, 2t^i_{upper}]$ and set the likelihood parameter to $p = 0.1$ because the observation is not reliable.

The exercise intensity, i.e., rest or exercise, is estimated using a threshold of the variance of the acceleration magnitude measured in 8 s at 125 Hz frequency. We define the threshold as 0.12 G^2 ($1G \approx 9.8$ m/s^2). The parameters used in the transition states are summarized in Table 1. We define the states from s1 to s6 on the basis of the exercise intensity and current HR h^i_{corr}. The thresholds between s1 and s2, s2 and s3, s4 and s5, and s5 and s6 are 110, 160, 110, and 160 bpm, respectively. We empirically determine the parameters in transition states through a preliminary experiment. Displacement means the difference in the HR between the current state and the next state in Table 1. The parameters are different depending on the person's habit of exercising. Therefore, we evaluate our method in two scenarios. The first scenario is a single subject without an exercise habit. The second scenario is multiple subjects with exercise habits.

To estimate the reliability of the HR measurement, we calculate the degree of discrepancy, which is the difference between the filtered HR and the estimated HR. The degree of discrepancy is used for the estimation of the reliability of HR, i.e., HR with a large degree of discrepancy has low reliability. On the basis of a threshold of the degree of discrepancy, we can exclude the HR with low reliability to accumulate reliable big data of HR.

Table 1
Parameters used in six transition states.

Exercise intensity	State	Displacement	t^i_{upper}	t^i_{lower}
Rest	s1	+0.04	+3.0	−3.0
	s2	−0.60	+1.0	−2.0
	s3	−0.62	+2.0	−3.0
Exercise	s4	+2.81	+5.0	−1.0
	s5	+1.29	+4.0	−1.0
	s6	+0.27	+4.0	−4.0

4. Evaluation

4.1 Experiment for single subject over multiple trials

4.1.1 Setting

For evaluation, a subject conducted five trials of running activity. Each trial was 5 min long. The chest-worn sensor myBeat WHS-3 measured the HR as a ground truth, and the wrist-worn sensor E4 measured PPG and acceleration. For training and testing, we performed leave-one-trial-out evaluation. To train the DL-based method,[12] we empirically set the parameters to a sliding window size of 8 s and a stride of 2s. Each trial was trained with an appropriate number of epochs in which the loss hardly decreased, and the batch size was 32. Our modified Kalman filter was applied to the outputs of the DL-based method. We also set a cutoff threshold for filtering HR estimation with low reliability. If the reliability of the HR estimation by the DL-based method was lower than the threshold, the samples were filtered. We calculated the absolute error between the ground truth and the estimated HR after the filtering. To observe the effect of filtering using the estimated reliability, we also applied linear interpolation to the remaining estimated HR. Similarly, the conventional Kalman filter was applied to the estimated HR as a baseline. Also, the existing outlier removal method[15] was applied after the conventional Kalman filter. For comparison, the absolute error was also calculated for the remaining HR after applying the conventional Kalman filter and removing outliers. We performed linear interpolation for the baseline and calculated its absolute error. We evaluated the proposed method by comparing the trade-off between the percentage of remaining HR samples after filtering and their absolute error. We summarize the definition of each HR in Table 2.

4.1.2 Existing outlier removal method

We implemented the existing outlier removal method[15] for the baseline. First, by applying the upper and lower limits of RRI, obvious outliers were removed. Next, HR per minute, i.e., 60000 ms/RRIms, was calculated from the RRI. On the basis of the assumption that the HR does not change suddenly, this method removes values of HR that suddenly change from the

Table 2
Summary of the definition of each HR.

HR	Description
Ground truth HR	HR measured by the chest-worn sensor.
Estimated HR	HR estimated by the DL-based existing method for the estimation of HR from PPG signal measured by the wristband sensor.
Baseline HR	HR after applying the conventional Kalman filter and the existing outlier removal method to the estimated HR.
Filtered HR	HR filtered by the proposed method, which is a combined method of the DL-based HR estimation method and our modified Kalman filter.
Remaining HR	HR after our HR removal method with low reliability.

previous value. However, this idea cannot consider the whole trend of the HR change because it compares the current HR with the HR in the previous time step. Therefore, to consider the trends of the changes in the HR in a wider range, the moving average was used to determine whether the HR should be removed. Through this filtering, outliers were removed from the time series of the original HR. We empirically set all parameters for the existing outlier removal method through our preliminary experiment.

4.1.3 Results

We defined the difference between the estimated HR and the filtered HR obtained by the proposed method as the degree of discrepancy. The relationship between the degree of discrepancy and the absolute error of the estimated HR is shown in Fig. 2.

The correlation coefficient is 0.839. Because of this strong positive correlation, the degree of discrepancy is used as an index of reliability as mentioned earlier. If the degree of discrepancy is large, the reliability is low, and vice versa.

Next, the absolute error and the percentage of the remaining HR samples (remaining rate) of each trial are shown in Figs. 3 and 4, respectively. These figures show the averages of all trials.

The remaining rate of the proposed method is higher than that of the baseline, but the mean absolute error of the proposed method is also larger than that of the baseline. This result indicates a trade-off between the remaining rate and the absolute error. The proposed method helps retain HR samples during exercise, simultaneously suppressing the increase of the absolute error to 1.2 bpm on average.

For further evaluation, we performed linear interpolation for the remaining HR of the two approaches. The mean absolute error for each trial is shown in Fig. 5.

The averages of the mean absolute error of the proposed method and the baseline are 6.25 and 8.01 bpm, respectively. The proposed method is superior to the baseline in the mean absolute error for the average of all trials after performing linear interpolation. This means that the proposed method more successfully removes estimated HRs greatly affected by the noise than the baseline.

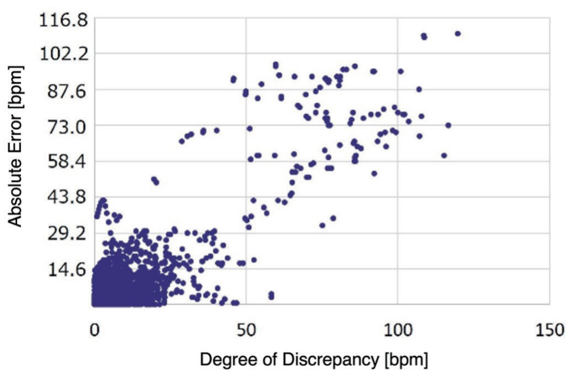

Fig. 2. (Color online) Correlation between degree of discrepancy and absolute error.

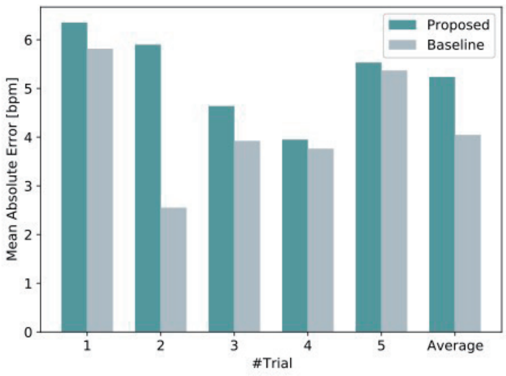

Fig. 3. (Color online) Mean absolute error of remaining HR.

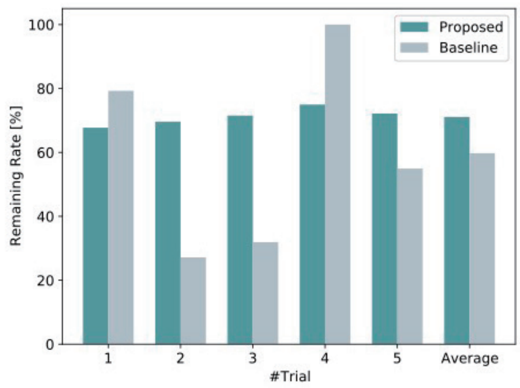

Fig. 4. (Color online) Remaining rate after removal.

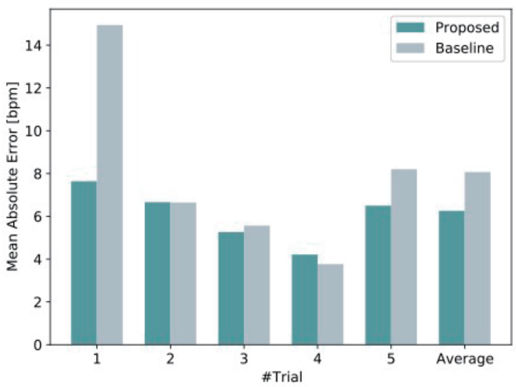

Fig. 5. (Color online) Mean absolute error of remaining HR after linear interpolation.

4.1.4 Discussion

Figure 6 shows a trial in which the conventional outlier removal method incorrectly removed a large part of the data. After the sudden increase in HR marked by the red circle, the whole data was removed as outliers. This is because the conventional outlier removal method is designed to remove instantaneous outliers. In contrast, high-intensity exercise such as running causes continuous outliers. Therefore, the conventional method fails to retain the correctly measured data while running. This is why the remaining rate of the conventional method shown in Fig. 4 is low.

On the other hand, we also found a trial for which the remaining rate is higher than that for the proposed method. We found that this occurs when the estimated HR is close to the ground truth.

As shown in Fig. 7, when the HR estimation does not work correctly, outliers can be corrected to values close to the ground truth using the modified Kalman filter. On the other hand, Fig. 8 shows the time series of the ground truth, the estimated HR, and the HR filtered by the proposed method in a trial where the mean absolute error of the proposed method is larger than that of the

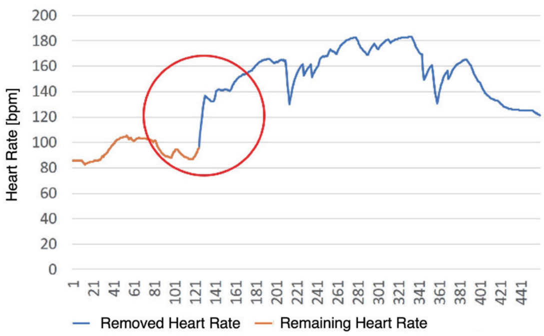

Fig. 6. (Color online) An example of conventional outlier filtering.

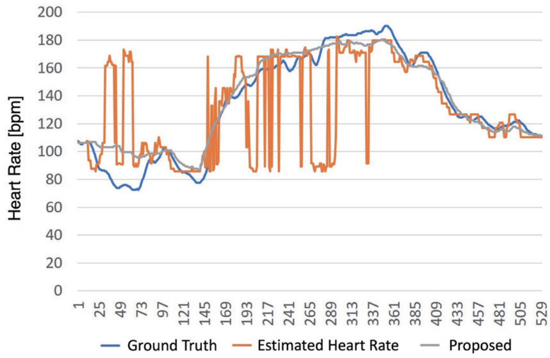

Fig. 7. (Color online) Trial where the proposed method worked successfully.

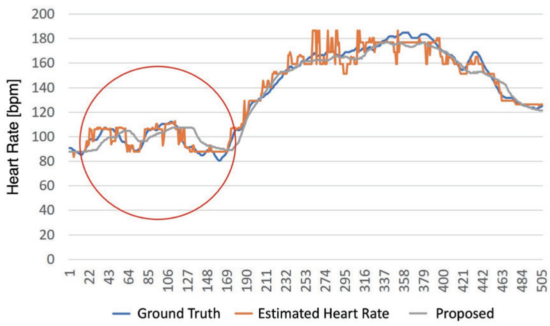

Fig. 8. (Color online) Trial where the proposed method failed.

conventional method. In particular, in the duration indicated by the red circle, the accuracy of the proposed method was low. This is because the ground truth HR suddenly increased even in the rest state. The proposed method attempted to decrease the measured value because the participant was judged as in the rest state. One of the reasons for this was the lack of HR transition data. Such cases can be estimated successfully if a large variety of data and subjects is given.

4.2 Experiment for multiple subjects

4.2.1 Setting

For further evaluation, we collected data from 19 males with a habit of exercising while they ran for 30 min. Because these subjects were in the habit of exercising, we used different parameters from those in Sect. 4.1 as shown in Table 3. Because no high HRs, i.e., larger than 160 bpm, were observed in the experiment, the displacement was not defined.

The subjects wore an E4 wristband and myBeat WHS-3 chest sensor. Also, we implemented the existing DL-based method[12] to estimate the HR under noise owing to motion artifacts. The sliding window size is 8 s and its stride was 2 s. For training and testing, we performed leave-

Table 3
Parameters used in six transition states for athlete subjects.

Exercise intensity	State	Displacement	t^i_{upper}	t^i_{lower}
Rest	s1	−0.29	+3.0	−3.0
	s2	−0.02	+1.0	−2.0
	s3	—	+2.0	−3.0
Exercise	s4	+0.03	+5.0	−1.0
	s5	+0.01	+4.0	−1.0
	s6	—	+4.0	−4.0

one-subject-out evaluation. The batch size and the number of epochs were 512 and 200, respectively. We also set a cutoff threshold for filtering HR estimation with low reliability. If the reliability of the HR estimation by the DL-based method was lower than the threshold, the samples were filtered. We calculated the absolute error between the ground truth and the estimated HR after the filtering. For the filtering, we removed samples whose discrepancy between the estimated HR and the filtered HR was larger than 5.0 bpm.

4.2.2 Results

The performance of each method is shown in Fig. 9. The evaluation metric is the mean absolute error between the output HR of each method and the ground truth. The first method (green bars) is the DL-based method.(12) The second method (red bars) is the use of the modified Kalman filter. The third method (blue bars) is the removal of samples based on the reliability estimated by our method. The outputs of the above three methods are denoted as the estimated HR, filtered HR, and remaining HR, respectively. The average mean absolute errors were 6.94, 6.15, and 6.04 for the estimated HR, the filtered HR, and the remaining HR, respectively. The figure shows an improvement from the DL-based method when we use the modified Kalman filter. In addition, there is a slight improvement in the remaining HR compared with the filtered HR. As observed in the preliminary experiment, the modified Kalman filter can successfully deal with moderate sudden transitions of the HR for multiple subjects.

However, there are some trials where the modified Kalman filter failed to decrease the error, i.e., subjects 2, 8, 9, 13, and 18. The HR time series of subject 18 for each method is shown in Fig. 10. The estimated HR, i.e., the output of the DL-based method, markedly fluctuated in this trial. As a result, the fluctuation caused a large error, and such a trial had few reliable data. Figure 11 shows the remaining rate of the remaining HR. For most of the subjects with a low remaining rate, the modified Kalman filter failed to decrease the error. Therefore, we can remove trials with a low remaining rate as unreliable trials. However, as shown in Fig. 12, subject 13 had a high remaining rate with a large error. Therefore, it cannot be excluded on the basis of the remaining rate only. In such cases, the removal method based on the reliability cannot decrease the error. To decrease the error, we must enhance the performance of the DL-based method. For this purpose, one option is to train the DL-based model with personal data of the estimation target.

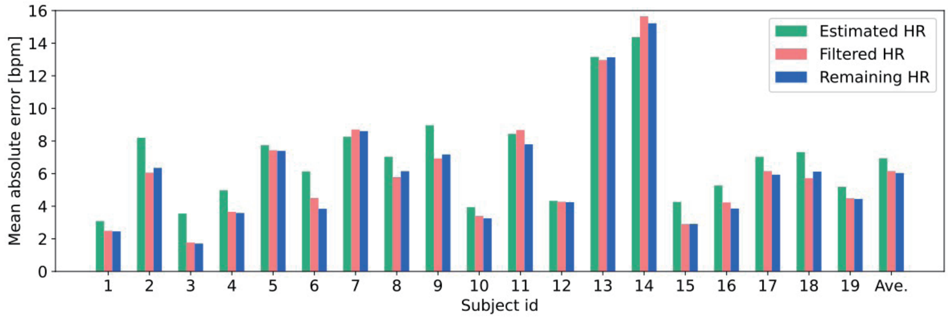

Fig. 9. (Color online) Mean absolute error for each subject using each method.

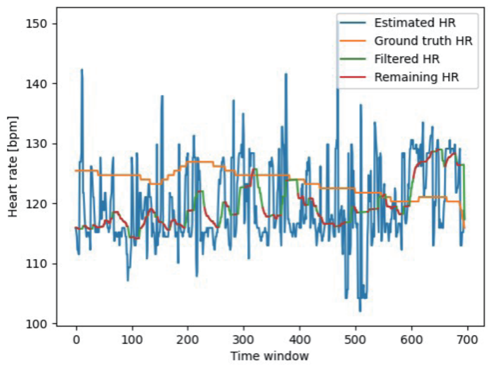

Fig. 10. (Color online) HR time series of subject 18.

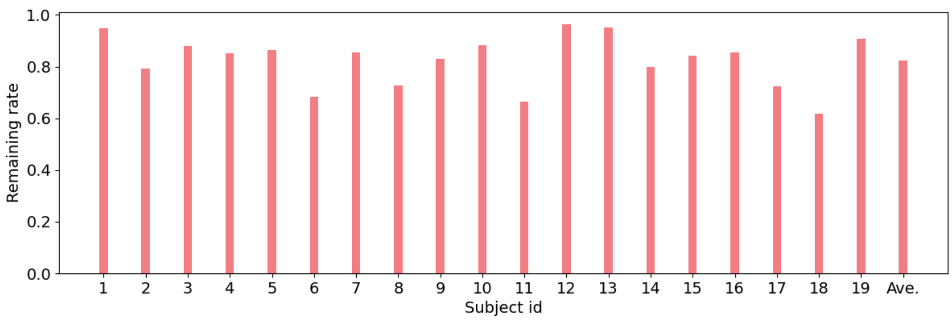

Fig. 11. (Color online) Remaining rate of HR for each subject.

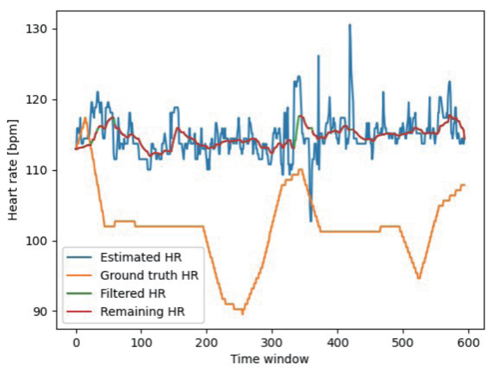

Fig. 12. (Color online) HR time series of subject 13.

5. Conclusions

In this study, we designed a reliability estimation method for HR estimation. We found that the error in the estimation of the HR from a PPG signal includes unrealistic transitions of the HR over time. The proposed method is based on the idea that unrealistic transitions are corrected using the HR filter so that they reflect realistic transitions. For this purpose, we designed a modified Kalman filter using acceleration, and we used the difference between its output and the estimated HR from the PPG signal as an index of reliability.

For evaluation, we compared the proposed method with a conventional outlier removal method using a state-of-the-art method based on DL. After removal, linear interpolation is performed. As a result, our method successfully removed 18.9% of the measurements with low reliability while achieving a mean absolute error of 6.25 bpm when the HR after filtering was interpolated for a single subject. For multiple subjects, our method decreased the mean absolute error by 13.1% on average. Our future work includes improving the estimation of HR displacement. Since the variance of acceleration, which is the basis of our HR displacement estimation, is measured by a wrist-worn device only, we are planning to design an estimation model that considers body movements other than those of the arm.

Acknowledgments

This work was supported by the Initiative for Life Design Innovation (iLDI) Innovation Platform for Society 5.0 (iPLAS 5.0).

References

1 A. Pantelopoulos and N. G. Bourbakis: IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **40** (2010) 1. https://doi.org/10.1109/TSMCC.2009.2032660
2 D. Biswas, L. Everson, M. Liu, M. Panwar, B.-E. Verhoef, S. Patki, C. H. Kim, A. Acharyya, C. V. Hoof, M. Konijnenburg, and N. V. Helleputte: IEEE Trans. Biomed. Circuits Syst. **13** (2019) 282. https://doi.org/10.1109/TBCAS.2019.2892297
3 I. A. Majeed, S. Jos, R. Arora, K. Choi, and S. Bae: Proc. 41st Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (IEEE, 2019) 5576–5580. https://doi.org/10.1109/EMBC.2019.8857131
4 D. Pollreisz and N. TaheriNejad: Mobile Networks Appl. **27** (2019) 728. https://doi.org/10.1007/s11036-019-01323-6
5 M. Climstein, J. L. Alder, A. M. Brooker, E. J. Cartwright, K. Kemp-Smith, V. Simas, and J. Furness: Sports **8** (2020) 117. https://doi.org/10.3390/sports8090117
6 B. Saha and D. Srivastava: Proc. 2014 IEEE 30th Int. Conf. Data Engineering (IEEE, 2014) 1294–1297. https://doi.org/10.1109/ICDE.2014.6816764
7 Z. Guan, T. Ji, X. Qian, Y. Ma, and X. Hong: Proc. IEEE 2017 5th Int. Conf. Applied Computing and Information Technology/4th Int. Conf. Computational Science/Intelligence and Applied Informatics/2nd Int. Conf. Big Data, Cloud Computing, Data Science (IEEE, 2017) 241–247. https://doi.org/10.1109/ACIT-CSII-BCD.2017.49
8 A. Juneja and N. N. Das: Proc. 2019 Int. Conf. Machine Learning, Big Data, Cloud and Parallel Computing (IEEE, 2019) 559–563. https://doi.org/10.1109/COMITCon.2019.8862267
9 C. S. Nunes, S. Bras, and P. Amorim: Eur. J. Anaesthesiology **23** (2006) 37. http://dx.doi.org/10.1097/00003643-200606001-00131
10 S. S. Bashar, M. S. Miah, A. Z. Karim, M. A. Al Mahmud, and Z. Hasan: Proc. 2019 Int. Conf. Electrical, Computer and Communication Engineering (IEEE 2019) 1–5. https://doi.org/10.1109/ECACE.2019.8679356

11 K. T. Tanweer, S. R. Hasan, and A. M. Kamboh: Proc. 2017 IEEE Int. Symp. Circuits and Systems (IEEE 2017) 1–4. https://doi.org/10.1109/ISCAS.2017.8050418
12 H. Chung, H. Ko, H. Lee, and J. Lee: IEEE Access **8** (2020) 63390. https://doi.org/10.1109/ACCESS.2020.2981956
13 B. W. Nelson and N. B. Allen: JMIR mHealth and uHealth **7** (2019) e10828. https://doi.org/10.2196/10828
14 H. Yoshikawa, M. Hayashi, A. Uchiyama and T. Higashino: Proc. 2021 13th Int. Conf. Mobile Computing and Ubiquitous Network (IPSJ 2021) 1–6. https://doi.org/10.23919/ICMU50196.2021.9638918
15 Sensor products: https://www.uniontool.co.jp/en/product/sensor/ (accessed July 2022).
16 M. Ghamari, D. Castaneda, A. Esparza, C. Soltanpur, and H. Nazeran: Int. J. Biosens. Bioelectron. **4** (2018) 195. https://doi.org/10.15406%2Fijbsbe.2018.04.00125
17 D. T. Weiler, S. O. Villajuan, L. Edkins, S. Cleary, and J. J. Saleem: Proc. Hum. Factors Ergon. Soc. Annu. Meet. **61** (2017) 1292. https://doi.org/10.1177%2F1541931213601804
18 K. Eguchi, R. Aoki, S. Shimauchi, K. Yoshida, and T. Yamada: Adv. Biomed. Eng. **7** (2018) 28. https://doi.org/10.14326/abe.7.28
19 M. Salai, I. Vass´anyi, and I. K´osa: J. Healthcare Eng. **2016** (2016) 5136705. https://doi.org/10.1155/2016/5136705
20 M. Nakano, T. Konishi, S. Izumi, H. Kawaguchi, and M. Yoshimoto: Proc. 2012 Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (IEEE 2012) 6703–6706. https://doi.org/10.1109/EMBC.2012.6347532
21 D. Berwal, C. R. Vandana, S. Dewan, C. V. Jiji, and M. S. Baghini: IEEE Sens. J. **19** (2019) 12432. https://doi.org/10.1109/JSEN.2019.2939391
22 H. Lee, H. Chung, H. Ko, and J. Lee: IEEE Sens. J. **18** (2018) 2983. https://doi.org/10.1109/JSEN.2018.2801385
23 J. Lee, M. Kim, H.-K. Park, and I. Y. Kim: Sensors **20** (2020) 1493. https://doi.org/10.3390/s20051493
24 Z. Zhang, Z. Pi, and B. Liu: IEEE Trans. Biomed. Eng. **62** (2015) 522. https://doi.org/10.1109/TBME.2014.2359372
25 E. Khan, F. A. Hossain, S. Z. Uddin, S. K. Alam, and M. K. Hasan: IEEE Trans. Biomed. Eng. **63** (2015) 550. https://doi.org/10.1109/TBME.2015.2466075
26 H. Fukushima, H. Kawanaka, M. B. Shoaib, and K. Oguri: Proc. 2012 Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (IEEE 2012) 2901–2904. https://doi.org/10.1109/embc.2012.6346570
27 S. M. Salehizadeh, D. Dao, J. Bolkhovsky, C. Cho, Y. Mendelson, and K. H. Chon: Sensors **16** (2016) 10. https://doi.org/10.3390/s16010010
28 A. Galli, C. Narduzzi, and G. Giorgi: IEEE Trans. Instrum. Meas. **67** (2018) 1102. https://doi.org/10.1109/TIM.2017.2770818
29 B. Lee, J. Han, H. J. Baek, J. H. Shin, K. S. Park, and W. J. Yi: Physiol. Meas. **31** (2010) 1585. https://doi.org/10.1088/0967-3334/31/12/003
30 E. K. Naeini, I. Azimi, A. M. Rahmani, P. Liljeberg, and N. Dutt: Procedia Comput. Sci. **151** (2019) 551. https://doi.org/10.1016/j.procs.2019.04.074

About the Authors

Hiroki Yoshikawa received his bachelor's degree in engineering science and master's degree in information science and technology from Osaka University in 2017 and 2019, repectively. Since 2021, he has been an assistant professor at Kyoto Tachibana University. His research interests include physiological sensing and applications. He is a member of IEEE and IPSJ.
(yoshikawa-h@tachibana-u.ac.jp)

Masayuki Hayashi received his B.S. degree from Osaka University in 2019 and M.S. degree in information science and technology from Osaka University in 2021. (m-hayashi@ist.osaka-u.ac.jp)

Akira Uchiyama received his M.E. and Ph.D. degrees in information and computer sciences from Osaka University in 2005 and 2008, respectively. He is an assistant professor at the Graduate School of Information Science and Technology, Osaka University. He was a visiting scholar in the University of Illinois at Urbana-Champaign in 2008 and a research fellow of the Japan Society for the Promotion of Science from 2007 to 2009. His current research interests include mobile sensing and applications in pervasive and ubiquitous computing. He is a member of IEEE, ACM, IEICE, and IPSJ.
(uchiyama@ist.osaka-u.ac.jp)

Teruo Higashino received his B.S., M.S. and Ph.D. degrees in information and computer sciences from Osaka University, Japan in 1979, 1981, and 1984, respectively. He joined the faculty of Osaka University in 1984. From 2002 to 2021, he had been a professor in the Graduate School of Information Science and Technology at Osaka University. He has been a Specially Appointed Professor at Osaka University and a Professor at Kyoto Tachibana University since 2021. His current research interests include design and analysis of distributed systems, communication protocol, and mobile computing. He is a senior member of IEEE, a fellow of IPSJ, and a member of ACM and IEICE of Japan. (higashino-t@tachibana-u.ac.jp)