Article

A jerk-based algorithm ACCEL for the accurate classification of sleep–wake states from arm acceleration

ACCEL

Triaxial acceleration
of arm movement

Jerk:
a derivative of acceleration

Sleep-wake classification
>90% sensitivity, >80% specificity

Koji L. Ode, Shoi Shi, Machiko Katori, ..., Ryo Oguchi, Daisuke Aoki, Hiroki R. Ueda

uedah-tky@umin.ac.jp

Highlights

An algorithm for sleep-wake classification based on arm acceleration is presented

The algorithm only uses a derivative of triaxial arm acceleration (jerk)

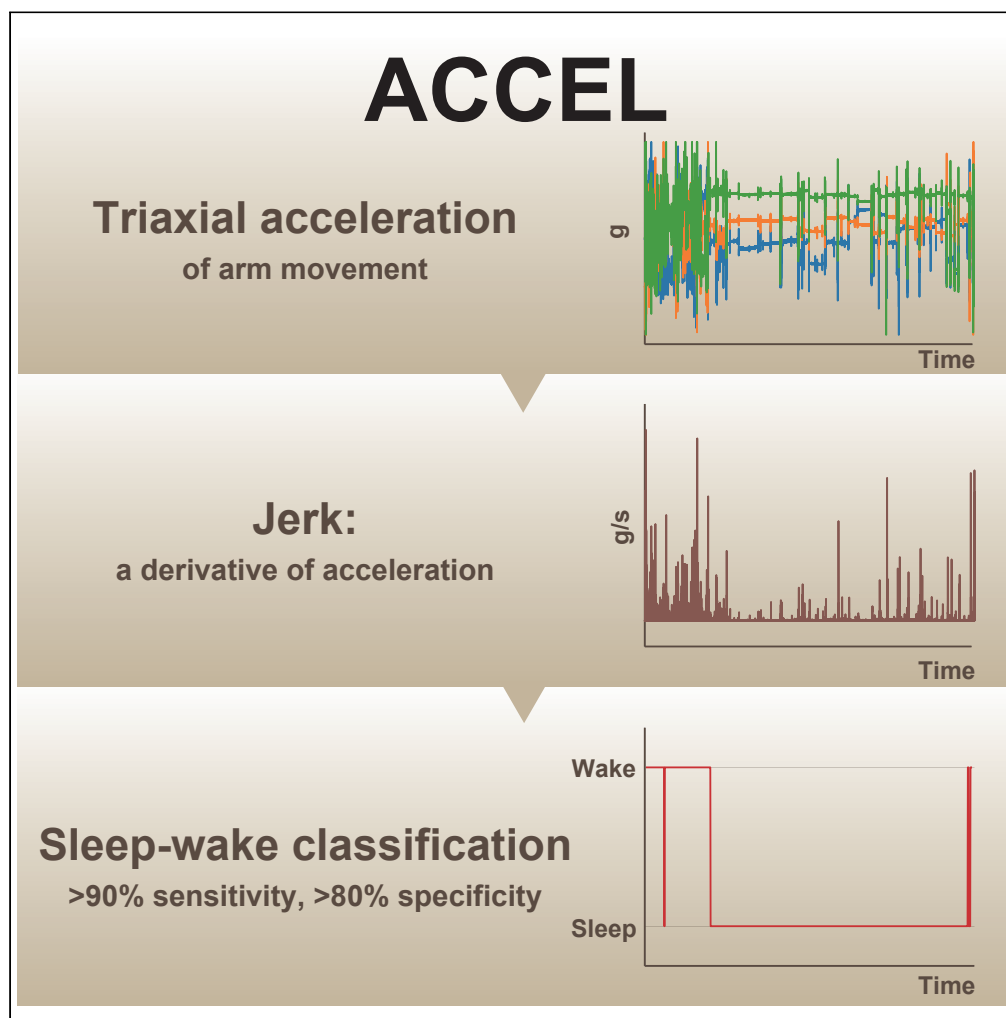The algorithm can accurately detect temporal awake during sleep

Article

A jerk-based algorithm ACCEL
for the accurate classification
of sleep–wake states from arm acceleration

Koji L. Ode,[1,2,6] Shoi Shi,[1,2,6] Machiko Katori,[3,6] Kentaro Mitsui,[3] Shin Takanashi,[4] Ryo Oguchi,[4] Daisuke Aoki,[5] and Hiroki R. Ueda[1,2,3,7,]*

SUMMARY

Arm acceleration data have been used to measure sleep–wake rhythmicity. Although several methods have been developed for the accurate classification of sleep–wake episodes, a method with both high sensitivity and specificity has not been fully established. In this study, we developed an algorithm, named ACceleration-based Classification and Estimation of Long-term sleep–wake cycles (ACCEL) that classifies sleep and wake episodes using only raw accelerometer data, without relying on device-specific functions. The algorithm uses a derivative of triaxial acceleration (jerk), which can reduce individual differences in the variability of acceleration data. Applying a machine learning algorithm to the jerk data achieved sleep–wake classification with a high sensitivity (>90%) and specificity (>80%). A jerk-based analysis also succeeded in recording periodic activities consistent with pulse waves. Therefore, the ACCEL algorithm will be a useful method for large-scale sleep measurement using simple accelerometers in real-world settings.

INTRODUCTION

Sleep disorders and chronic sleep disturbance are associated with various health problems, including lifestyle-related diseases, mental disorders, and more generally, human performance. Assessing sleep status is important not only for patients with specific sleep disorders but also for all people in their daily lives. This is because individuals' sleep–wake behavior is largely affected by social factors such as working/school hours and stress. The problems caused by sleep disorders also affect people in society—for example, it has been estimated that sleep disorders cause a significant amount of economic loss (Bonnet and Arand, 1995; Hillman et al., 2006, 2018; Skaer and Sclar, 2010). Accurate determination of sleep–wake states and measurement of sleep–wake cycles over a long period (e.g., 2 weeks) is required to correctly assess human sleep behavior because sleep rhythms typically differ between weekdays and weekends. A precise and standard method for human sleep measurement is based on polysomnography (PSG) measurements. However, PSG measurements require specialized technicians and are not suitable for continuous and routine sleep measurement for the public.

For this reason, alternative methods have been developed to assess sleep behavior in a nonclinical and noninvasive manner (Perez-Pozuelo et al., 2020). Questionnaires on an individual's sleep routine are one of the easiest alternatives for large-scale data collection. For example, the Munich ChronoType Questionnaire (MCTQ) has been successful in quantitatively estimating an individual's chronotype (Roenneberg et al., 2003, 2015, 2019). The MCTQ also revealed irregular sleep schedules caused by social environments, namely, a social jet lag, in most people in modern society (Roenneberg et al., 2012; Wittmann et al., 2006). Recently, question-based survey methods can be easily distributed around the world using smartphone applications, and such surveys demonstrate the worldwide distribution of different sleep schedules (Walch et al., 2016). Alternatively, activity logs that reflect human behavioral rhythms, even if only indirectly, may be useful for monitoring social sleep schedules. As an example, the analysis of Twitter history was used to estimate regional differences in the presence or absence of social jet lag (Leypunskiy et al., 2018). However, there are limitations with questionnaires and indirect behavioral logs: the questionnaire-based survey can contain biases based on participants' subjective responses, and accurate estimates of sleep schedules are difficult based on indirect behavioral logs. As solutions to these problems, researchers have been

[1]Department of Systems Pharmacology, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

[2]Laboratory for Synthetic Biology, RIKEN Center for Biosystems Dynamics Research, Suita, Osaka 565-0871, Japan

[3]Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

[4]Technology Strategy Office, Sony Mobile Communications Inc., Shinagawa-ku, Tokyo 140-0002, Japan

[5]Product Design Sec. 3, Product Design Department, Product Development Div., Sony Mobile Communications Inc., Shinagawa-ku, Tokyo 140-0002, Japan

[6]These authors contributed equally

[7]Lead contact

*Correspondence:
uedah-tky@umin.ac.jp

https://doi.org/10.1016/j.isci.2021.103727

focusing on the use of wearable devices to measure sleep schedules because these devices can collect behavioral data of subjects more directly (Kim et al., 2020).

Among wearable devices, wristwatch devices have been increasingly regarded as promising tools for behavioral logging. Wristwatch devices, such as smartwatches, have been developed and become more popular. Although research devices for acquiring actograms (e.g., Actiwatch) have been mainly used to acquire personal activity logs, recent development and consumer devices, such as the Apple Watch, have allowed many people to measure personal activity logs in some way over a long period. Consumer devices have been increasingly used in medical research. For example, activity logging with Axivity is involved in the protocol of the UK Biobank project aiming to collect the genetic and phenotypic dataset from half a million UK participants (Bycroft et al., 2018; Sudlow et al., 2015). Simultaneously, the development of algorithms for accurately determining sleep–wake based on wearable devices has been advancing. Wristwatch devices can be equipped with several sensors, such as heart-rate monitors, of which triaxial accelerometers are commonly equipped and used to monitor the wearer's activity status. Accordingly, the classification of sleep–wake states based on acceleration data of the arm can be considered to have a wide range of applications. In addition to the pioneering developments using threshold-based algorithms (Cole et al., 1992), recent applications of deep learning/machine learning (ML) have achieved sleep–wake classification with higher accuracy (Walch et al., 2019). Furthermore, it has been demonstrated that ultradian rhythmicity during the sleep phase can be extracted from the arm movement recordings (Winnebeck et al., 2018), and thus, it is expected that arm acceleration can be used not only to classify sleep and wakefulness but also to evaluate the structure/quality of sleep episodes.

However, there are still several problems to be solved in terms of both devices and algorithms. First, several popular devices such as Apple Watch have a complex structure with multiple sensors, and their internal structure and hardware settings are not completely clear. Thus, it is difficult to keep track of what factors are responsible for differences in measurement accuracy between different consumer devices. The hardware issues also limit the use of proprietary sleep–wake classification methods for medical research. For example, Actiwatch and Fitbit have proprietary sleep–wake algorithms associated with their respective devices (de Zambotti et al., 2016; Kosmadopoulos et al., 2014), but these algorithms are expected to be optimized for the corresponding hardware, and it is unclear how generally applicable they are to different devices. Even algorithms whose details are publicly available, such as algorithms that set specific thresholds to determine sleep–wake are expected to be significantly affected by sensor sensitivity. Furthermore, there is still room for improvement in the accuracy of sleep–wake classification; for example, several recent proprietary or open algorithms showed high accuracy in determining sleep–wake state compared with PSG, but the specificity of classification was relatively low (de Zambotti et al., 2016; Kosmadopoulos et al., 2014; Walch et al., 2019). Low specificity can be a problem in the precise detection of detailed sleep structure metrics such as wake time after sleep onset (WASO).

Overall, standardization is an ongoing challenge in the field of routine sleep measurement with wearable devices. In this study, a wristwatch device is developed that simply records the raw data of a triaxial accelerometer to rule out any uncertain properties of proprietary devices and algorithms. No other functions were actively installed. We have recorded arm movement and PSG simultaneously. By applying ML to features extracted from the power spectrum (PS) of the jerk of triaxial acceleration, we developed an algorithm that accurately classifies sleep and awake stages with high sensitivity and specificity. The ACCEL algorithm can also accurately annotate the total sleep time (TST) and WASO as well as the time window where the device is not equipped. In addition, we found that the accelerometer detects rhythmic jerk signals peaking around 1 Hz, which well match with the frequency of individuals' pulse waves. The detection of a pulse-like signal may be useful for evaluating detailed sleep structures and their abnormality. In summary, we present a jerk-based algorithm that can accurately classify sleep–wake state based on the triaxial accelerometer without the need for any specific features installed in proprietary devices.

RESULTS

Setting up a simple wristwatch accelerometer with open configuration

We used a custom-made wristwatch accelerometer that simply records raw data of a three-dimensional (3D) digital accelerometer LSM6DSM (STmicroelectronics, Swiss Confederation) with a full-scale acceleration range of ± 2 g at a 16-bit resolution to ensure that obtained triaxial acceleration of arm movement is unaffected by device-specific configurations. An analog anti-aliasing low-pass filter was turned off to obtain
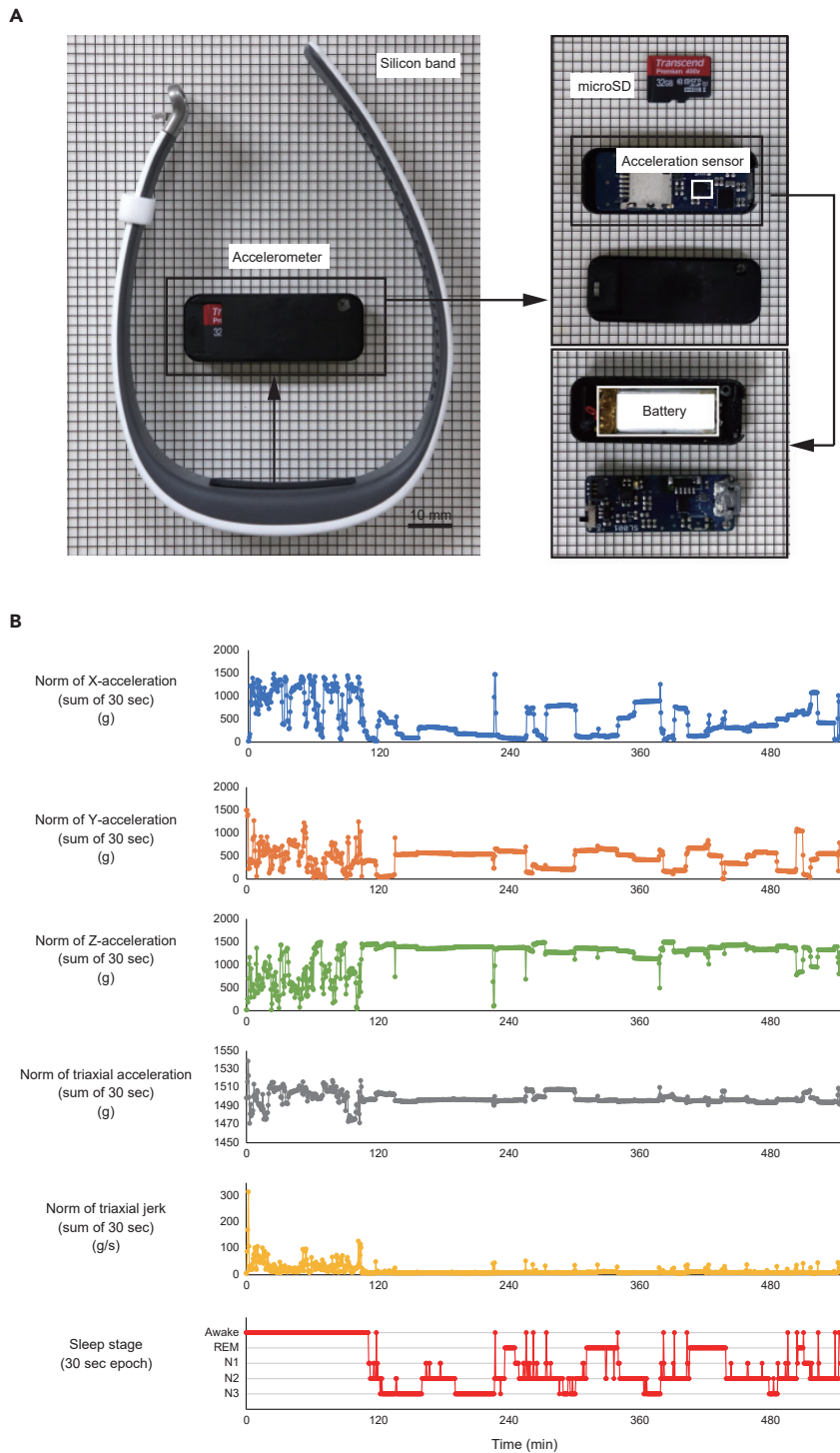
a sensor value that was as unprocessed as possible. The other digital low-pass filter cannot be disabled because of the sensor configuration. Triaxial acceleration values are recorded on the micro SD card every 2.5 min (Figures 1A and S1). Participants were subjected to PSG and accelerometer recordings at the same time for an entire night. According to the questionnaire summarized in Table S1 and Figure S2, the variability in the sleep schedules of all participants can be assumed to be within the range of normal Japanese population variability (Kitamura et al., 2014). The PSG device is wireless, and all participants were asked to sleep or stay awake overnight, so the obtained dataset includes recordings from both sleep and awake states (Figure 1B). The ground-truth annotation of sleep stages was recovered from PSG data with manual curation by experts.

Development of sleep–wake classification algorithm

When analyzing sleep from acceleration data, it is common to convert the acceleration time-series data into features and input them into algorithms such as ML algorithms. Activity counts, calculated per epoch as the sum of a maximum value of z axis signals, are traditional features and have been used as input to many sleep–wake classification algorithms (de Zambotti et al., 2016; Kosmadopoulos et al., 2014; Walch et al., 2019). Although existing algorithms have classified sleep and awake episodes with high accuracy (e.g., 90.9%) (de Zambotti et al., 2016), the specificity of sleep classification (i.e., the ability to detect wake) has been relatively low (e.g., 54.1%) (Walch et al., 2019). It means that these algorithms tend to fail in detecting wake during night, including short-term awake, which is one of the major features observed in most insomnia patients (American Psychiatric Association, 2013). These algorithms use only one-dimensional features (i.e., activity count) per epoch for the classification and may be occasionally unable to distinguish between sleep and awake with low activity. Thus, in this study, we tested four features with 60 dimensions: raw norm, raw PS, jerk norm, and jerk PS (Figure 2A, see also Method details). Raw norm and raw PS are based on triaxial acceleration. Jerk norm and jerk PS are based on acceleration deviation. Acceleration data from 32 trials were converted into each feature and used for input to XGBoost, a gradient boosted tree algorithm, to classify sleep and wake for each 30-s epoch. Classification performance was evaluated by comparing the predictions with sleep–wake classification results obtained by PSG. As a result, the algorithm using jerk PS showed the best performance among the four features (Figure 2B). Using all four features did not markedly increase the performance with a slight decrease in the specificity (Figure 2B), suggesting that most of the information used for the sleep-wake classification is covered by jerk PS. We then evaluated whether the XGBoost is superior to other classification algorithms such as linear regression (LR) and multilayer perceptron (MLP). Sleep–wake classification performance was calculated by using the same jerk PS dataset. XGBoost showed highest specificity compared to linear regression, multilayer perceptron (LR: $52.45 \pm 12.12\%$, MLP: $71.63 \pm 12.29\%$, XGBoost: $72.36 \pm 13.27\%$) with comparable accuracy (LR: $82.15 \pm 8.11\%$, MLP: $86.61 \pm 6.10\%$, XGBoost: $86.50 \pm 5.42\%$), F measure (LR: $62.82 \pm 10.44\%$, MLP: $74.38 \pm 11.62\%$, XGBoost: $74.17 \pm 11.48\%$), and sensitivity (LR: $95.54 \pm 1.94\%$, MLP: $92.96 \pm 5.74\%$, XGBoost: $92.28 \pm 4.68\%$). Thus, jerk PS and XGBoost were selected as the feature and classification algorithm for the proposed method. The feature is extracted by two steps shown in Figure 2C. Step 1 is converting triaxial acceleration to jerk, and step 2 is converting jerk to PS (0–2 Hz).

To further improve the algorithm, we considered time-series data that included not only the current epoch but also preceding and subsequent epochs (Figure 2D). The PS of neighboring k ($k = 0, 1, 2, 3, 4, 5$) epochs was concatenated (termed as "large feature") and used as the input to XGBoost. The algorithm's performance improved with larger features and reached saturation after $k = 4$ (Figure 2D). Thus, we chose to use $k = 4$, meaning that our algorithm uses jerk PS feature of nine epochs to classify the sleep status of the target epoch (i.e., one target epoch, four preceding epochs, and four subsequent epochs). The evaluation of XGBoost weightings applied to each neighboring epoch revealed that epochs subsequent to the target epoch, rather than the preceding epochs, have higher contribution to the sleep-wake classification (Figure 2E).
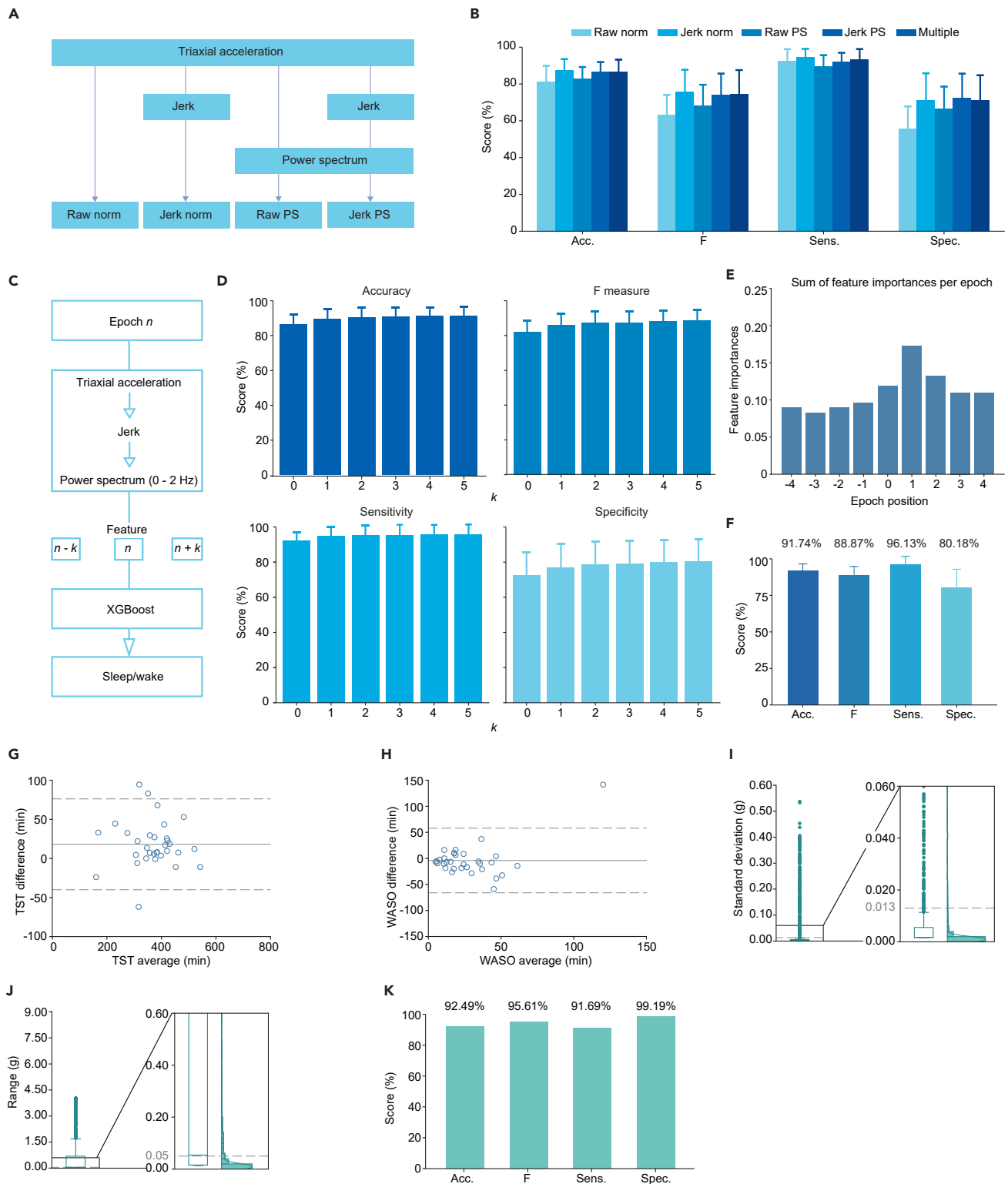
The hyperparameter of XGBoost was further optimized by using Bayesian global optimization with Gaussian process. The six hyperparameters of XGBoost were evaluated and trained by leave-one-out cross-validation (LOOCV) to optimize the summation of accuracy and F measure. LOOCV was conducted by dividing 32 data into training dataset (31 data) and validation dataset (the remaining one data). The performance of XGBoost was calculated by repeating the training and validation 32 times with slecting different one data for the validation such that the evaluation would not be biased toward a specific dataset. Based on the LOOCV, the hyperparameters were updated by Bayesian optimization. We repeated the

Figure 1. Design of simplified wristwatch accelerometer
(A) Device overview. The accelerometer was designed to be worn on the inside of a commercially available silicone wristband. The accelerometer includes a triaxial 3D accelerometer, electric circuit board, battery, and micro SD card to store the acceleration data.
(B) Example plot of simultaneously recorded triaxial acceleration and sleep stages classified by PSG analysis. See also Figures S1 and S2 and Table S1.

Figure 2. Development of jerk-based sleep-wake classification algorithm, ACCEL

(A) Overview of feature extraction.

(B) The performance of the sleep–wake classification algorithms with different features or by using all the four features (Multiple) shown as average ± SD. Acc: Accuracy. (F) F measure. Sens: Sensitivity. Spec: Specificity.

Figure 2. *Continued*

(C) Overview of the sleep–wake classification algorithm.

(D) Effect of surrounding epochs on the performance of the algorithm. The accuracy (top left), F measure (top right), sensitivity (bottom left), and specificity (bottom right) of sleep-wake classification were calculated by using jerk PS with different number of neighboring epochs, k, and shown as average $\pm$ SD.

(E) The contribution of each neighboring epoch at a different position relative to the target n^{th} epoch on the sleep-wake classification of ACCEL. Feature importances were calculated by taking the summation of trained XGBoost weights assigned to the 60-dimensional features in each epoch position.

(F) The performance of the ACCEL algorithm shown as average $\pm$ SD. Acc: Accuracy. (F) F measure. Sens: Sensitivity. Spec: Specificity.

(G and H) Bland–Altman plots of TST, and WASO measured by PSG recording and ACCEL. The solid gray lines represent the mean, and the dashed lines represent the mean $\pm$1.95 standard deviation.

(I and J) The histogram and distribution of standard deviation and range of each accelerometer, calculated by shifting the 60-min block by 15 min. For each epoch, three values of standard deviation and range can be calculated corresponding to x, y, and z axis. The figures plot the values of the second largest axis. When either of the second-largest standard deviation or range was less than or equal to the threshold shown as the gray dashed lines, the block was considered as a non-wear period.

(K) The performance of the non-wear detection algorithm. Acc: Accuracy. (F) F measure. Sens: Sensitivity. Spec: Specificity. See also Table S2.

optimization process 2,000 times. With the optimization process, our algorithm achieved high sensitivity (96.13 $\pm$ 5.61%) and specificity (80.18 $\pm$ 12.71%) (Figure 2F). The algorithm with high specificity, ACCEL, demonstrates that it can detect short-term awake during sleep with high reliability, which was difficult to achieve in previous studies using arm/finger-movement-based algorithms (de Zambotti et al., 2016, 2019; Kosmadopoulos et al., 2014; Markwald et al., 2016; van Hees et al., 2015; Walch et al., 2019) (Table S2). Although we do not exclude the possibility that the data difference (e.g., different devices for the recording) among the listed studies affect the performance of each algorithm, it should be noted that window size for taking the averaged arm/finger-movement does not affect the performance of ACCEL algorithm because jerk PS does not take the average of raw data.

The difference between the sleep time obtained by ACCEL and the ground-truth sleep time obtained by PSG was 18.06 $\pm$ 29.64 min (Figure 2G). Similarly, the difference for WASO was 4.02 $\pm$ 31.62 min (Figure 2H). Cohen's kappa value was calculated to be 77.91%, indicating that the ACCEL-based and PSG-based sleep–wake classifications agree well. Thus, the proposed algorithm has high performance in classifying sleep–wake and accurately detects short-term awakeness.

To investigate the advantage of calculating jerk in sleep–wake classification, we focused on the differences between each measurement. For each measurement, the raw norm and jerk norm values were averaged for sleep and wake epochs. The individual differences in jerk norm in sleep epochs were lower than those of raw norm (p <0.001, Student's t test) (Figure S3A). This result indicates that individual differences in sleep epochs are mitigated by converting raw data into jerk data.
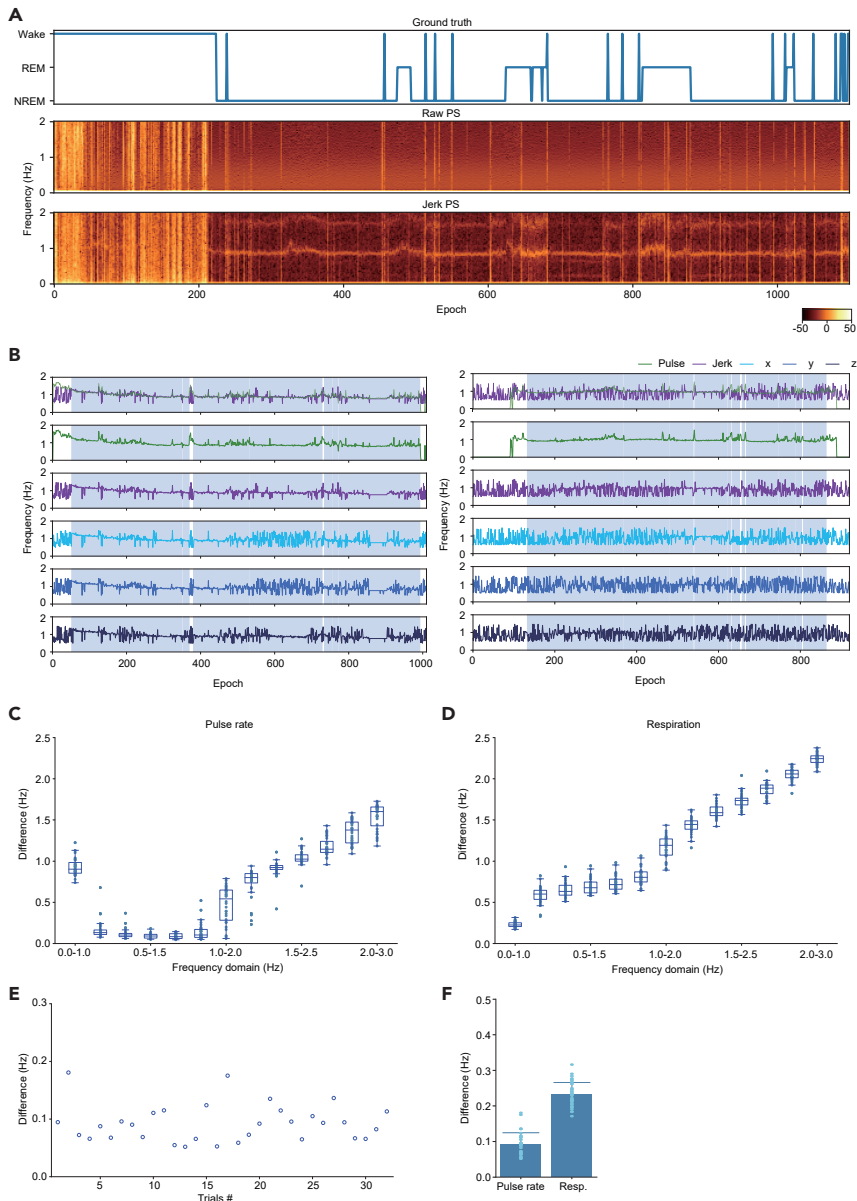
Application of non-wear detection algorithm

Using the thresholds of two features, namely, standard deviation and acceleration range, a previous study predicted non-wear epochs, when a subject does not wear the wristwatch (van Hees et al., 2013). To evaluate whether van Hees's algorithm applies to our device, we collected acceleration data from five subjects (60.03 days data in total) and recorded the timestamps when the subjects were not wearing the device. An analysis of the distribution of the two features during the non-wear period showed that almost all data were below the thresholds, indicating that the thresholds proposed in a previous study (van Hees et al., 2013) apply to our device (Figures 2I and 2J). The non-wear detection algorithm was evaluated using four scores, namely, accuracy, F measure, sensitivity, and specificity, for each individual. In this case, sensitivity and specificity show the ability to detect non-wear and wear periods, respectively. The non-wear detection algorithm achieved high accuracy (92.49%) and high specificity (99.19%) (Figure 2K). By combining the non-wear detection and sleep–wake classification algorithms, we can recover the sleep–wake behavior from the continuous accelerometer measurement for more than 1 week in the presence of occasional non-wear periods (Figure S3B).

Extraction of pulse-like signal from the acceleration data

During the course of jerk PS analysis, we noticed a characteristic rhythmic signal at a frequency of approximately 1 Hz (Figure 3A). Such a ~1 Hz rhythmic signal becomes visible by calculating the jerk PS, and was not found in raw PS. This ~1 Hz signal is clearly observed during sleep when overall arm movement is small. We assumed that this rhythmicity may be related to pulse rate, whose frequency is also approximately 1 Hz (i.e., 60 beats per minute). Our PSG recording includes the pulse rate recorded by a finger sensor for pulse
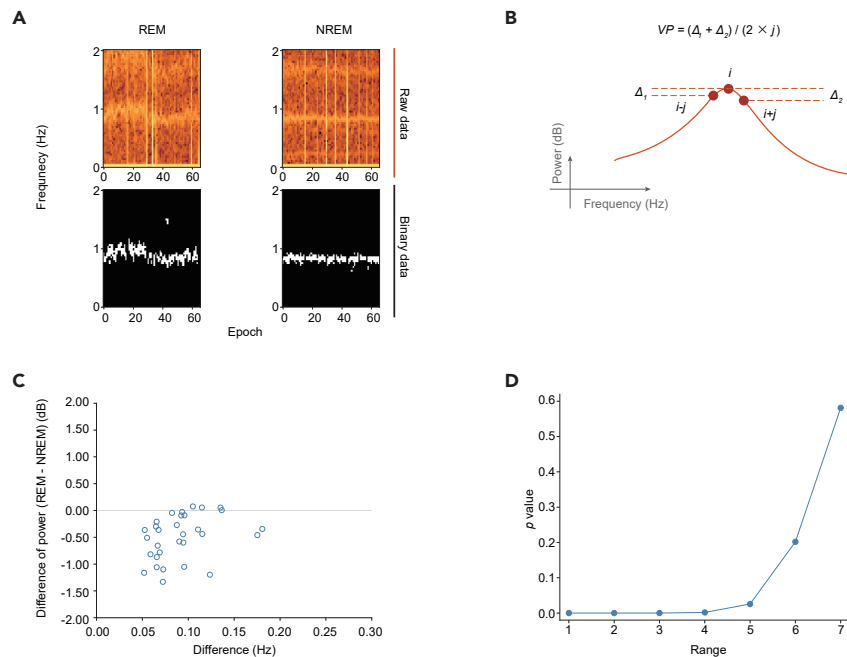
Figure 3. Validation of jerk-based pulse detection

(A) Example time-series plot of ground truth, the power spectrum of acceleration signal, and the power spectrum of jerk.

(B) Example time-series plots of pulse rate and frequency with the maximum power in the 0.5–1.5 Hz of jerk, jerk of x-, y-, or z axis. Blue-shaded area indicates epochs used for the comparison between jerk and pulse rate, where the epochs are classified as sleep and the averaged pulse rate is nonzero value. The plot on the left shows an example trial where the arm movement was generally stable, whereas the plot on the right shows a trial where the jerk showed a large value in most of the sleep epochs.

(C) Boxplot of the difference between pulse rate and jerk frequency with the maximum power in the indicated frequency bands.

(D) Boxplot of the difference between respiration rate and jerk frequency with the maximum power in the indicated frequency bands.

(E) Dot plot of the difference between pulse rate and pulse-like signals based on jerk.

(F) Bar plot of the difference between pulse-like signal and other signals shown as average ± SD. Resp: Respiration signal. See also Figure S3.

Figure 4. Variety of pulse-like signal between REM and NREM sleep epochs

(A) Heatmap of jerk signal during REM and NREM sleep.

(B) Method for calculating the variance of pulse-like signal (*VP*).

(C) Scatterplot of the difference between VP_N and VP_R. It was significant against zero (p <0.01, Student's t test). The horizontal axis shows the average difference between pulse rate and pulse-like signals.

(D) Plot of p value corresponding to the value of *j* for calculating *VP*. The range of the horizontal axis corresponds to the *j* of panel (B).

oximeters. When we plotted the time-series of the frequency at which the jerk PS is maximum within 0.5–1.5 Hz and pulse rate, the similarity between the two signals was observed during the stable epoch, where the overall arm acceleration is relatively small and mostly classified as sleep (Figure 3B). We quantitatively compared the frequency bands of the jerk signal to pulse rate and respiration signal obtained from an airflow sensor attached to the nose to confirm whether the rhythmic jerk signal corresponds to pulse or other periodic movements around 1 Hz. The signal of jerk matched well with the signal of pulse in the range of 0.5–1.5 Hz, corresponding to the pulse rate of the resting state (Figure 3C). In contrast, the epoch-by-epoch difference was significantly larger in the comparison between the jerk signal and the respiration signal at any of the tested frequency domains (Figures 3D and S3C). Finally, we calculated the difference between jerk signals in the 0.5–1.5 Hz range and pulse signals for each trial (Figures 3E and 3F) and found that the difference between the two signals is significantly smaller than the difference between jerk signals and respiration signals. Overall, these evaluations indicate that the rhythmic jerk signal can be attributed to the pulse signal and indicate that a simple accelerometer device can acquire pulse information at least when a subject is in the resting state. In the later sections, we will call the ∼1 Hz rhythmic signal of jerk PS as a "pulse-like signal."

Pulse data in NREM and REM sleep

Pulse signals have been reported to show different characteristics in REM and NREM sleep, where the width of the pulse wave (peak to peak) is regular in NREM sleep but becomes irregular in REM sleep owing to the changes in the autonomic nervous system (Dehkordi et al., 2013; Garde et al., 2014; Khandoker et al., 2011). Thus, in this study, we investigated whether this feature is also preserved in a pulse-like signal acquired by the accelerometer. Extracted frequency bands around the pulse-like signal during NREM and REM periods shown in Figure 4A demonstrate that the pulse-like signal is observed in a wider frequency range during REM sleep. This may be attributed to the irregular nature of pulse during REM sleep. To quantitatively compare the variance of the pulse-like signal between NREM and REM epochs, we measured the variance of the pulse-like signal by calculating the difference between the points before and after the peak (Figure 4B). In addition, we

calculated $VP_N - VP_R$, where VP_N and VP_R represent the average of the variance of the pulse-like signal during NREM and REM epochs, respectively, and showed that the value was significantly negative (i.e., the variance of the pulse-like signal is larger during REM epochs) (Figure 4C). We also confirmed that the significant difference between NREM and REM epochs is independent of the number of points used in calculating the variance of the pulse-like signal (Figure 4D), excluding the possibility that the difference is because of the different number of NREM and REM epochs observed during the recording sessions, and thus used for calculating the variance. Therefore, we concluded that the pulse-like signal obtained from the acceleration signal can capture the different characteristics in REM and NREM as the pulse signal.

DISCUSSION

A jerk-based algorithm provides robust sleep–wake classification with a simple accelerometer device

Accurate and simple measurement of the sleep–wake cycle is a promising direction for human sleep research, especially for large-scale human sleep studies. With the recent development of ML algorithms, various alternative sleep monitoring methods have been proposed. Building on the correlation between sleep stages and fluctuations of heart rate variability (HRV), electrocardiogram-based monitoring has been performed (Aktaruzzaman et al., 2015; Xiao et al., 2013). Respiration-based sleep monitoring (Long et al., 2014; Sun et al., 2020) is one of the major methods because breathing rhythms are more stable during sleep than during wake (Krieger, 1985). Another direction is to use the acceleration of body movements. For example, smartphone-based (Toon et al., 2016) and wristband-type sleep monitoring (de Zambotti et al., 2016; Kosmadopoulos et al., 2014; Walch et al., 2019) have been proposed. Among them, wristband-type accelerometers are particularly suitable for large-scale sleep studies because of their long-term measurement capability, including daytimes, and ease of use. Actiware (Mini-Mitter Philips Respironics, Inc., Sunriver, OR, USA) is a software used to classify wake and sleep from Actiwatch (Mini-Mitter Philips Respironics, Inc., Sunriver, OR, USA), which is a major wristband-type accelerometer. Version 3.4 of the software employs a threshold-based algorithm, where each 30 s and the surrounding 2 min of data are used for estimation (Kushida et al., 2001). This software has several threshold choices, and the trade-off between specificity (the ability to detect wake) and accuracy (supported by sensitivity, i.e., the ability to detect sleep) has been demonstrated with 38 PSG data (Kosmadopoulos et al., 2014). The highest threshold among the tested thresholds showed the highest accuracy (88.0%) and the lowest specificity (26.9%), whereas the lowest threshold showed the lowest accuracy (61.5%) and the highest specificity (61.5%). Fitbit and Apple Watch are the famous wristband implementations for sleep monitoring. The sleep monitoring of FitbitChargeHR™ (FitBit Inc., San Francisco, CA, USA) uses the heart rate measured by PurePulse™ LED lights in addition to acceleration. Validation with 30 PSG data showed low specificity (42.4 ± 15.9%) but high accuracy (90.9 ± 4.7%) (de Zambotti et al., 2016). The combination of high accuracy and low specificity has also been demonstrated with Apple Watch, for which accuracy was ∼90% but specificity was ∼60% (Walch et al., 2019). The low specificity means that the algorithms fail to capture wake during night including short-term awake, leading to the inaccurate measurement of difficulty maintaining sleep, which is the major sleep pattern feature of patients suffering from insomnia (American Psychiatric Association, 2013).

We developed the ACCEL algorithm using accelerometer data, which is largely unprocessed. Therefore, the algorithm does not necessarily require expensive and sophisticated devices but can be applied to simple accelerometers that are probably inexpensive to produce. Furthermore, the algorithm has superior performance. We demonstrated that accurate sleep–wake determination with high accuracy (91.71 ± 4.80%) and specificity (80.18 ± 12.71%), especially the sensitive detection of awakeness during sleep. The high specificity of our algorithm may be because of the use of F measure as well as accuracy as the targets for optimization in hyperparameter tuning of ML. The overall performance should be improved by calculating jerk and extracting the frequency components that mitigate the variance between devices and/or measurements. Indeed, jerk PS is also used for the sleep-wake classification in other neural-network based algorithms developed by Eric Canton (https://github.com/ericcanton/sleep_classifiers_accel_ML). Although there are considerable differences between ACCEL and Canton's algorithm (e.g., Canton's code uses the daytime/nighttime information of jerk PS, whereas ACCEL only uses acceleration data), the jerk PS of arm's acceleration may be an important alternative to the EEG-EMG for robust sleep-wake classification. The application of the non-wear detection algorithm proposed by van Hees et al. (van Hees et al., 2013) is also successful. Non-wear detection is important because subjects may not always wear the device reliably in large-scale sleep measurements. These features are

essential to make large-scale sleep–wake measurements in a real-world setting more feasible, and our study suggests that this can be achieved by a simple device that essentially only records the raw value of the triaxial accelerometer.

Triaxial accelerometer can be used to detect pulse signal: potential merit

The possibility of capturing pulse waves, even with a simple accelerometer, would also lead to a more detailed assessment of the sleep state, such as the depth of sleep and NREM–REM cycle. Sleep–wake transition and the transition between sleep stages are accompanied by altered activity in the autonomic nervous system (Mendez et al., 2009; Scholz et al., 1997; Welch and Richardson, 1973). The autonomic nervous system affects the heart rate. Accordingly, differences in HRV are observed between NREM sleep and REM sleep (Mendez et al., 2009), and several algorithms have been developed for classifying sleep stages using HRV (Mendez et al., 2009; Willemen et al., 2014). A simple measurement of HRV can be achieved by many devices, including a belt worn on the chest or a high-performance wearable device such as Apple Watch (Aktaruzzaman et al., 2017; Shcherbina et al., 2017). However, such devices are unsuitable for long-term continuous measurement because it is difficult to install batteries that can measure HRV continuously for more than 2 weeks. In this study, we extracted pulse-like signals from simple acceleration data and found differences in pulse-like signals between NREM and REM sleep. Based on these differences, it will be possible to develop a sleep-stage classification algorithm from acceleration data, which will enable us to obtain long-term time-series data of sleep stages.

Standardized sleep–wake quantification in real life

The precise estimation of the algorithm's general applicability to other devices, especially those that are already commercially available and widely used, needs to be further investigated. Thus, this study will be useful information for standardizing sleep–wake determination algorithms and devices for large-scale sleep measurement.

Standardized and accurate sleep phenotyping is important for each individual in our society to accurately monitor their sleep states and to detect abnormalities in their early stages. In addition, large-scale and accurate analysis of human phenotypes, along with the accumulation of genetic information, will make it possible to conduct reverse genetics and reverse phenotyping approaches to verify human phenotypes based on genotypes (Ozcelik and Onat, 2016; Patke et al., 2017; Saleheen et al., 2017). These efforts may reveal genetic factors that have a significant impact on the sleep phenotype of people in the real world.

Limitation of the study

Currently, the ACCEL algorithm is specialized in classifying sleep and wake epochs, and has not succeeded in distinguishing sleep stages with high sensitivity and specificity. Not only the pulse-like signal found in the jerk PS, but also other features that could be abstracted from arm movement data such as "Locomotor Inactivity During Sleep" will be important for the further in-depth classification of sleep stages (Winnebeck et al., 2018).

Although ACCEL algorithm provides accurate sleep-wake classification, the classification of sleep-wake alone does not provide information that can lead to personalized medicine, such as whether a subject's sleep state is abnormal, the predicted amount of sleep debt of the subject, or when is the appropriate time to take medication including on-line advice to adjust the subject's misaligned chronotype. Such application requires the combination of accurate sleep-wake measurement and quantitative models to predict the internal sleep-wake dynamics (Kim et al., 2020). Several models have been proposed in recent years to represent not only the day-night sleep cycle but also the NREM-REM sleep structure (Booth and Diniz Behn, 2014). Recent studies also have shown that the combination of measured data and mathematical models can predict temporary sleep-wake transitions, such as daytime sleepiness (Hong et al., 2021). It is an interesting future prospect to combine such mathematical models to evaluate the personal sleep state based on the long-term trends of sleep characteristics that are difficult to capture only by self-reported sleep diary, such as short and temporal awakeness during sleep.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Wristwatch accelerometer
  - Subject detail
  - Questioners
  - PSG and arm acceleration recording
  - Sleep staging based on PSG
  - Sleep–wake classification algorithm
  - Non-wear detection
  - Pulse analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103727.

AUTHOR CONTRIBUTIONS

K.L.O., S.S., M.K., and H.R.U. designed the research. S.T., R.O., and D.A. developed the device. K.L.O., S.S., and M.K. performed the data acquisition. S.S., M.K., K.M. analyzed the data. K.L.O, S.S., M.K., and H.R.U. wrote the paper.

DECLARATION OF INTERESTS

H.R.U. and S.T. conducted a collaborative research project between Sony Mobile Inc. and the University of Tokyo. S.T., R.O. and D.A. are employees of Sony Mobile Communications Inc. The company provided support in the form of salary for S.T., R.O. and D.A., and research grant to H.R.U used for this study. However, the company did not have additional roles in the study design and scientific interpretation and decision regarding this study. H.R.U. is the founder and Chief Technology Officer of ACCELStars Inc. HRU is a member of the iScience's editorial advisory board.

The algorithm described in this study has been patent pending as PCT/JP2020/026345 and 2019-125950 by the University of Tokyo. H.R.U., K.L.O., S.S., M.K. and K.M. are coinventors.

REFERENCES

American Psychiatric Association (2013). Diagnostic and Statistical Manual of Mental Disorders : DSM-5 (American Psychiatric Association).

Aktaruzzaman, M., Migliorini, M., Tenhunen, M., Himanen, S.L., Bianchi, A.M., and Sassi, R. (2015). The addition of entropy-based regularity parameters improves sleep stage classification based on heart rate variability. Med. Biol. Eng. Comput. 53, 415–425.

Aktaruzzaman, M., Rivolta, M.W., Karmacharya, R., Scarabottolo, N., Pugnetti, L., Garegnani, M., Bovi, G., Scalera, G., Ferrarin, M., and Sassi, R. (2017). Performance comparison between wrist and chest actigraphy in combination with heart rate variability for sleep classification. Comput. Biol. Med. 89, 212–221.

Bonnet, M.H., and Arand, D.L. (1995). We are chronically sleep deprived. Sleep 18, 908–911.

Booth, V., and Diniz Behn, C.G. (2014). Physiologically-based modeling of sleep-wake regulatory networks. Math. biosci 250, 54–68.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209.

Cole, R.J., Kripke, D.F., Gruen, W., Mullaney, D.J., and Gillin, J.C. (1992). Automatic sleep/wake identification from wrist activity. Sleep 15, 461–469.

de Zambotti, M., Baker, F.C., Willoughby, A.R., Godino, J.G., Wing, D., Patrick, K., and Colrain, I.M. (2016). Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol. Behav. 158, 143–149.

de Zambotti, M., Rosas, L., Colrain, I.M., and Baker, F.C. (2019). The sleep of the ring: comparison of the OURA sleep tracker against polysomnography. Behav. Sleep Med. 17, 124–136.

Dehkordi, P., Garde, A., Karlen, W., Wensley, D., Ansermino, J.M., and Dumont, G.A. (2013). Pulse rate variability compared with Heart Rate Variability in children with and without sleep disordered breathing. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2013, 6563–6566.

Garde, A., Dehkordi, P., Karlen, W., Wensley, D., Ansermino, J.M., and Dumont, G.A. (2014). Development of a screening tool for sleep disordered breathing in children using the phone Oximeter. PloS one 9, e112959.

Hillman, D., Mitchell, S., Streatfeild, J., Burns, C., Bruck, D., and Pezzullo, L. (2018). The economic cost of inadequate sleep. Sleep 41. https://doi.org/10.1093/sleep/zsy083.

Hillman, D.R., Murphy, A.S., and Pezzullo, L. (2006). The economic cost of sleep disorders. Sleep 29, 299–305.

Hong, J., Choi, S.J., Park, S.H., Hong, H., Booth, V., Joo, E.Y., and Kim, J.K. (2021). Personalized sleep-wake patterns aligned with circadian rhythm relieve daytime sleepiness. iScience 24, 103129.

Khandoker, A.H., Karmakar, C.K., and Palaniswami, M. (2011). Comparison of pulse rate variability with heart rate variability during obstructive sleep apnea. Med. Eng. Phys. 33, 204–209.

Kim, D.W., Zavala, E., and Kim, J.K. (2020). Wearable technology and systems modeling for personalized chronotherapy. Curr. Opin. Syst. Biol. 21, 9–15.

Kitamura, S., Hida, A., Aritake, S., Higuchi, S., Enomoto, M., Kato, M., Vetter, C., Roenneberg, T., and Mishima, K. (2014). Validity of the Japanese version of the munich chronotype questionnaire. Chronobiology Int. 31, 845–850.

Kosmadopoulos, A., Sargent, C., Darwent, D., Zhou, X., and Roach, G.D. (2014). Alternatives to polysomnography (PSG): a validation of wrist actigraphy and a partial-PSG system. Behav. Res. Methods 46, 1032–1041.

Krieger, J. (1985). Breathing during sleep in normal subjects. Clin. chest Med. 6, 577–594.

Kushida, C.A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O., and Dement, W.C. (2001). Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med. 2, 389–396.

Leypunskiy, E., Kiciman, E., Shah, M., Walch, O.J., Rzhetsky, A., Dinner, A.R., and Rust, M.J. (2018). Geographically resolved rhythms in twitter use reveal social pressures on daily activity patterns. Curr. Biol. 28, 3763–3775 e3765.

Long, X., Fonseca, P., Foussier, J., Haakma, R., and Aarts, R.M. (2014). Sleep and wake classification with actigraphy and respiratory effort using dynamic warping. IEEE J. Biomed. Health Inform. 18, 1272–1284.

Markwald, R.R., Bessman, S.C., Reini, S.A., and Drummond, S.P. (2016). Performance of a portable sleep monitoring device in individuals with high versus low sleep efficiency. J. Clin. Sleep Med. 12, 95–103.

Mendez, M.O., Matteucci, M., Cerutti, S., Aletti, F., and Bianchi, A.M. (2009). Sleep staging classification based on HRV: time-variant analysis. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng 2009, 9–12.

Ozcelik, T., and Onat, O.E. (2016). Genomic landscape of the greater Middle East. Nat. Genet. 48, 978–979.

Patke, A., Murphy, P.J., Onat, O.E., Krieger, A.C., Ozcelik, T., Campbell, S.S., and Young, M.W. (2017). Mutation of the human circadian clock gene CRY1 in familial delayed sleep phase disorder. Cell 169, 203–215 e213.

Perez-Pozuelo, I., Zhai, B., Palotti, J., Mall, R., Aupetit, M., Garcia-Gomez, J.M., Taheri, S., Guan, Y., and Fernandez-Luque, L. (2020). The future of sleep health: a data-driven revolution in sleep science and medicine. NPJ digital Med. 3, 42.

Roenneberg, T., Allebrandt, K.V., Merrow, M., and Vetter, C. (2012). Social jetlag and obesity. Curr. Biol. 22, 939–943.

Roenneberg, T., Keller, L.K., Fischer, D., Matera, J.L., Vetter, C., and Winnebeck, E.C. (2015). Human activity and rest in situ. Methods Enzymol. 552, 257–283.

Roenneberg, T., Pilz, L.K., Zerbini, G., and Winnebeck, E.C. (2019). Chronotype and social jetlag: a (self-) critical review. Biology 8, 54.

Roenneberg, T., Wirz-Justice, A., and Merrow, M. (2003). Life between clocks: daily temporal patterns of human chronotypes. J. Biol. rhythms 18, 80–90.

Saleheen, D., Natarajan, P., Armean, I.M., Zhao, W., Rasheed, A., Khetarpal, S.A., Won, H.H., Karczewski, K.J., O'Donnell-Luria, A.H., Samocha, K.E., et al. (2017). Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. Nature 544, 235–239.

Scholz, U.J., Bianchi, A.M., Cerutti, S., and Kubicki, S. (1997). Vegetative background of sleep: spectral analysis of the heart rate variability. Physiol. Behav. 62, 1037–1043.

Shcherbina, A., Mattsson, C.M., Waggott, D., Salisbury, H., Christle, J.W., Hastie, T., Wheeler, M.T., and Ashley, E.A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J. personalized Med. 7, 3.

Skaer, T.L., and Sclar, D.A. (2010). Economic implications of sleep disorders. PharmacoEconomics 28, 1015–1023.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e1001779.

Sun, H., Ganglberger, W., Panneerselvam, E., Leone, M.J., Quadri, S.A., Goparaju, B., Tesh, R.A., Akeju, O., Thomas, R.J., and Westover, M.B. (2020). Sleep staging from electrocardiography and respiration with deep learning. Sleep 43, zsz306.

Toon, E., Davey, M.J., Hollis, S.L., Nixon, G.M., Horne, R.S., and Biggs, S.N. (2016). Comparison of commercial wrist-based and smartphone accelerometers, actigraphy, and PSG in a clinical cohort of children and adolescents. J. Clin. Sleep Med. 12, 343–350.

van Hees, V.T., Gorzelniak, L., Dean Leon, E.C., Eder, M., Pias, M., Taherian, S., Ekelund, U., Renstrom, F., Franks, P.W., Horsch, A., et al. (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PloS one 8, e61691.

van Hees, V.T., Sabia, S., Anderson, K.N., Denton, S.J., Oliver, J., Catt, M., Abell, J.G., Kivimaki, M., Trenell, M.I., and Singh-Manoux, A. (2015). A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PloS one 10, e0142533.

Walch, O., Huang, Y., Forger, D., and Goldstein, C. (2019). Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. Sleep *42*, zsz180.

Walch, O.J., Cochran, A., and Forger, D.B. (2016). A global quantification of "normal" sleep schedules using smartphone data. Sci. Adv. *2*, e1501705.

Welch, A.J., and Richardson, P.C. (1973). Computer sleep stage classification using heart rate data. Electroencephalogr Clin. Neurophysiol. *34*, 145–152.

Willemen, T., Van Deun, D., Verhaert, V., Vandekerckhove, M., Exadaktylos, V., Verbraecken, J., Van Huffel, S., Haex, B., and Sloten, J.V. (2014). An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification. IEEE J. Biomed. Health Inform *18*, 661–669.

Winnebeck, E.C., Fischer, D., Leise, T., and Roenneberg, T. (2018). Dynamics and ultradian structure of human sleep in real life. Curr. Biol. *28*, 49–59 e45.

Wittmann, M., Dinich, J., Merrow, M., and Roenneberg, T. (2006). Social jetlag: misalignment of biological and social time. Chronobiology Int. *23*, 497–509.

Xiao, M., Yan, H., Song, J., Yang, Y., and Yang, X. (2013). Sleep stages classification based on heart rate variability and random forest. Biomed. Signal Process. Control *8*, 624–633.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python 3.8.3	Python Software Foundation	https://www.python.org/; RRID:SCR_008394
NumPy 1.18.5	NumPy	https://numpy.org/; BRID: SCR_008633
pandas 1.0.5	pandas	https://pandas.pydata.org/; BRID: SCR_018214
Scikit-Learn 0.23.1	Scikit-Learn	https://scikit-learn.org/stable/index.html; BRID: SCR_002577
XGBoost 1.2.0	xgboost	https://xgboost.readthedocs.io/en/stable/; BRID: SCR_021361
Bayesian-optimization 1.1.0	Bayesian Optimization	https://github.com/fmfn/BayesianOptimization
Other		
SOMNOscreen plus	SOMNOmedicsGmbH	https://somnomedics.de/en/solutions/sleep_diagnostics/stationary_sleep_lab_psg/somnoscreen-plus/
LSM6DSM	STmicroelectronics	https://www.st.com/ja/mems-and-sensors/lsm6dsm.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hiroki R. Ueda (uedah-tky@umin.ac.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Any additional information including code and human data except for private information reported in this paper is available for non-commercial use from the lead contact upon reasonable request.

METHOD DETAILS

Wristwatch accelerometer

A custom-made accelerometer, including device selection, design of system block diagram (Figure S1) electric circuit diagram, and printed circuit board (PCB) pattern was developed by Sony Mobile Inc. The 3D acceleration sensor LSM6DSM (STmicroelectronics, Swiss Confederation), battery, and other devices were assembled on the PCB fabricated by Sony Mobile Inc. and operated through a firmware developed by Sony Mobile Inc. The accelerometer was housed in a plastic case fit into a commercially available rubber band (SWR122, Sony). The weight of the housed device (excluding the rubber band) was ∼6 g. A rechargeable Li-ion battery with a capacity of 3.7 V and 95 mAh, which is sufficient to continuously measure acceleration over 14 days was used by the device. No waterproof function is provided.

Subject detail

Twenty-five participants were recruited for this study, all of whom stated that they had no present and past diagnosis of any type of sleep disorder, mental illness, and neurodegenerative disease. Persons who had been on regular sleep or anti-anxiety medication in the past were excluded during the recruitment process. The participants were asked if they temporally take medicines that may affect sleep (e.g., cold medicine or hay fever medicine), and no participant declared that s/he took medication. The participants were allowed to participate in multiple rounds of sleep recording trials. Trials with errors in the recording of PSG/accelerometer data were excluded from the analysis. Accordingly, a total of 32 sleep recordings from 21 participants were obtained through successful sleep staging, pulse rate, and accelerometer data collection (15 participants participated in a single trial, 3 participants participated in two trials, 1 participated in three trials, and 2 participated in four trials) (Table S1). The protocol has been approved by the Research Ethics

Committee, Graduate School of Medicine, The University of Tokyo (No. 11653-[1]). All participants provided written informed consent for each trial.

Questioners

All participants were asked their sex, age at the date of measurement, height, body weight, regular bed-in and wake-up time of weekdays (working days), and regular bed-in and wake-up time of weekends (free days). The results of the questioners are summarized in Table S1 and Figure S2. Midpoint of sleep in workdays (MSW) and midpoint of sleep in free days (MSF) were caliculated according to the MCTQ method (Kitamura et al., 2014; Roenneberg et al., 2003, 2015).

PSG and arm acceleration recording

Participants were asked to arrive at the sleep recording room at the University of Tokyo at 20:00-22:00. Participants were equipped with a wireless portable PSG system (SOMNOscreen plus, SOMNOmedics GmbH, Germany). PSG data are stored in the CF card storage in the portable system as well as in the laptop computer in the sleep measurement room wirelessly. Participants were also equipped with a custom-made accelerometer that records the triaxial acceleration simultaneously on either wrist. After being outfitted with the PSG system and the accelerometer, the participants were free to sleep or stay awake in the measurement room. They may turn on and off the lights in the room and may go to bed and wake up at any time. After waking up, the participants removed the PSG system and the accelerometer by themselves to complete the measurements. Some participants were required to continue wearing the accelerometer only for 15 days. Because the accelerometer was not waterproof, participants could put it on and take it off several times during the days when it was supposed to be worn.

Sleep staging based on PSG

Sleep staging for every 30-s epoch was conducted with manual analysis of the PSG dataset by trained experts. The staging service is provided by Fukuda Denshi Co. Ltd (Japan).

Sleep–wake classification algorithm

Triaxial acceleration was converted into jerk data (a derivative of acceleration) and then converted into PS (0–2 Hz in 1/30 Hz segments), which was calculated for each 30 s. The 30-s epoch size was determined to be matched with the epoch size of PSG-based sleep staging. The feature extraction process includes jerk signal development and PS conversion) (Figure 2A). With and without the two processes, a total of four types of features were extracted as follows. Raw norm represents the mean values of the L2 norm of triaxial signals within every 0.5 s. Jerk norm represents the mean values of jerk signals within every 0.5 s. Raw PS represents the PS (0–2 Hz) of the L2 norm of triaxial signals. Jerk PS represents the PS (0–2 Hz) of jerk signals. Therefore, all of raw norm, jerk norm, raw PS and jerk PS features are 60 dimensions, allowing us to compare different features under the same input dimension size. The large feature was obtained by adding PS before and after k = 1, 2, 3, 4 or 5 epochs. The final ACCEL algorithm employs k = 4. Linear regression (LR), multilayer perceptron (MLP), and XGBoost were used as the classifiers. Scikit-learn packages (https://scikit-learn.org/stable/) were used for each implementation. All evaluation of algorithm were calculated with 32 data from one-night measurements (Table S1) by LOOCV. The comparison among LR, MLP, and XGBoost was performed by using scikit-learn and the parameters was set as default value of the package. XGBoost has six hyperparameters: *learning_rate*, *gamma*, *colsample_bytree*, *subsample*, *max_depth*, and *min_child_weight*. These parameters were optimized using Bayesian global optimization with gaussian processes (https://github.com/fmfn/BayesianOptimization) (iteration = 2,000). Each parameter could have a value in the range [0, 1], [0, 5], [0.01, 1], [0.01, 1], [1, 30], and [1, 30]. The parameter set of six hyperparameters was evaluated as the summation of accuracy and F measure by LOOCV. In this process, 32 data was divided into training data set (31 data) and validation data set (1 data) and the performance of XGBoost was obtained by repeating the training and validation 32 times with different validation data. The predicted sleep–wake data were compared with PSG-based sleep–wake data (ground truth) by epoch to epoch. F measure is calculated as the (2 × *precision* × *recall*)/(*precision* + *recall*), where wake is calculated as true and sleep as false. Sensitivity and specificity represent the performance of sleep and wake detection, respectively. TST for predicted sleep–wake data was calculated with the TST between sleep onset and offset, where the first sleep epoch of more than 15 min is defined as the sleep onset and the first wake epoch of more than 1 h as sleep offset. WASO was calculated as the length of wake duration between sleep and sleep offset.

Non-wear detection

The standard deviation and value range of each accelerometer axis were calculated by shifting the 60-min block by 15 min, and if the standard deviation was less than or equal to a threshold value (standard deviation threshold equals 13 mg) for at least two axes, or if the value range was less than or equal to a threshold value (value range threshold equals 50 mg) for at least two axes, the block was considered a non-wear period. To validate the non-wear detection, we acquired a dataset consisting of five independent, continuous triaxial accelerometer recordings: the recording duration are 14.56 days, 12.09 days, 8.75 days, 11.02 days, and 13.60 days, respectively. In this measurement, the participants were asked to record the non-wear period, which was used as the ground truth. The non-wear detection algorithm was evaluated using four scores, accuracy, F measure, sensitivity, and specificity, for each participant by comparing their timestamps. In this case, sensitivity and specificity show the ability to detect non-wear and wear periods, respectively.

Pulse analysis

The pulse rate data acquired by the PSG system were used for analysis. The sampling rate is 4 Hz, and each epoch has 120 points. The average of pulse rate data for each epoch was calculated (P_{ave}). Sleep epochs with $P_{ave} > 0$ were used for pulse analysis. If <50% of the total sleep epochs met the criteria, the data from the trial was removed from the pulse analysis. To calculate pulse-like signals, the absolute time derivative of each axis data (x, y, and z) and jerk were converted into a PS. We set up a 1-Hz-wide window and calculated the value when we moved the window from 0 to 3 Hz in 1/6-Hz steps. For each epoch, the frequency with the highest power in the interval of 0.5–1.5 Hz is defined as a pulse-like signal. The breathing data acquired by the PSG system were divided into epochs and converted into a PS. The frequency with the highest power in the range of 0–2 Hz is defined as a respiration signal. When the pressure equals zero, the epoch was excluded from the analysis. The variance of the pulse-like signal (VP) of each epoch was calculated using the following equation: $VP = \{ (a_i - a_{(i-j)}) + (a_i - a_{(i+j)}) \}/(j \times 2)$, where a_i represents the maximum value in the range 0.5–1.5 Hz of the jerk signal and $a_{(i-j)}$ and $a_{(i+j)}$ represent the value of jerk signal j ($j = 1, 2, …, 7$) points before and after a_i.