WAKE Detection During Sleep using Random Forest for Sleep Apnea Syndrome Patients

Iko Nakari, Yusuke Tajima, Ryo Takano, Akari Tobaru and **Keiki Takadama**
The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo, Japan
{iko0528, y_tajima, takano, tobaru_akari}@cas.lab.uec.ac.jp, keiki@inf.uec.ac.jp

Abstract

This paper proposed the new WAKE detection method for sleep apnea syndrome: SAS patients. In many non-contact method for sleep stage estimation, it is difficult to detect WAKE for SAS patients because it detects WAKE by only one threshold and their Heart Rate Variability: HRV, Body Movements: BM and Respiratory Variability: RV are different from healthy subjects. Furthermore, SAS patients have more sudden WAKE than healthy subjects. In order to detect WAKE for SAS patients, we employed a mattress type pressure sensor which obtains the bio-vibrations, and Random Forest: RF as the detection of WAKE because it is possible to interpret the rules it produces. In detail, the RF learns six features, that labeled with WAKE or Non-WAKE(REM, NREM1 to 4). These features are calculated from sensor value. To verify the effectiveness, the subject experiment was conducted on 9 different SAS subjects. The results revealed that: (1) the top accuracy of the WAKE detection method is 96.0%; (2) extracted rule from the RF is one of the rules that WAKE with weak BM; (3) SAS subjects tends to generate more rules for WAKE detection than healthy subjects. From those results, the contribution of this research is suggesting the way to detect WAKE, and find physiological characteristics that might be useful for SAS discrimination.

Introduction

Recently, the population of people who have been suffered from the sleep disorders has been increased year by year. There is the report that the prevalence of sleeping problems was 56% in the USA, 31% in Western Europe and 23% in Japan (Leger et al. 2008). In order to take appropriate measures for the problems, it is essential to evaluate the quality of sleeping. In the medical field, Polysomnography: PSG is the major method to evaluate the quality of sleeping. Based on the data of PSG, the sleep state is classified into six levels of depth by the international standard Rechtschaffen and Kales: R&K method (Rechtschaffen et al. 1968). However, it is difficult to measure continuously the quality of sleeping because of the following reasons: (1) patients have to be attached lots of electrodes for PSG, which gives a large physical and mental burden; (2) it needs large time for estimation by several professionals.

To tackle these problems, some researcher develop the mattress sensor based sleep stage estimation methods as non-contact method. Watanabe developed mattress sensor and focused on the correlation between heart rate variability: HRV and sleep stage (Watanabe et al. 2001). They reported that their method can extract the macro change of HRV, and filtered HRV is similar to the waveform in the sleep stage. However, the method needs whole data during sleep to estimate sleep stage, and it is difficult to estimate the sleep stage in real time. Based on the Watanabe method, Harada proposed Real-time Sleep Stage Estimation: RSSE that estimates the sleep stage in real time (Harada et al. 2016). To estimate the sleep stage in real time, they construct the trigonometric function regression model from the partially obtained heart rate and use estimated intermediate frequency component of the prospective heart rate. In the RSSE, the sleep stage is estimated by automatically analyzing the sensor value from the mat sensor. As a result, simpler sleep stage estimation was realized in healthy people. Focus on the WAKE judgment, RSSE makes WAKE judgement with one condition concerning Body Movement: BM. Concretely, it will be judged as the WAKE if the standard deviation of BM in the most recent minute is higher than the average value of BM from the time of sleeping to the present. However, RSSE has a tendency for more misjudgment in sleep apnea syndrome: SAS patients because their sleep is shallow and tend to have more BM.

To tackle this issue, this paper propose a novel method to improve the Accuracy and Precision of WAKE Detection based on several features obtained sensor value by non-contact device. We employ Random Forest (Breiman et al. 2001) which is one of machine learning methods as a classifier because it has interpretation and detection by various rules. Furthermore, we focus on that WAKE of SAS patients is different from healthy subjects, and judge SAS or Non-SAS by extract what RF learned.

The rest of paper is organized as follows. First, sleep apnea syndrome is introduced in Section 2. Section 3 describes related work related to the sleep stage estimation. Section 4 describes how to detect WAKE based on sensor value from mat sensor and how to use RF. Section 5 describes the experiments conducted on the subjects, presents the obtained results. Section 6 describes discussion for the results. Finally, the conclusions of this paper are presents in the final section.

Sleep Apnea Syndrome

Medical Definition

Sleep Apnea Syndrome: SAS is a sleep disorder characterized by breathing stops during sleep. Breathing stops for more than 10 seconds is said to be apnea. It is diagnosed as SAS by a professional physician using data from specialized instruments. The severity of symptoms is as follows: if the apnea is happened

- 5 to 14 times per hour is mild;
- 15 to 29 times per hour is moderate;
- more than 30 times per hour is severe.

Examination Method

The examination methods of SAS are divided into two stages, they are simple examination which can be done at home and hospitalized examination.

First, if there is a possibility of SAS by interview from a doctor, do the simple examination at home. In the simple examination, attaching specialized instruments to the wrist, fingers and nose for collect respiration, snoring, SpO_2 (blood oxygenation level) and heart rate data before going to bed. The doctor analyzes the data to determine whether it is SAS.

Second, if doctor diagnose that it is SAS by the result of simple examination, get hospitalized to do a highly accurate examination by PSG for observe condition of sleeping quality and respiration.

Characteristics and Influence on human body

SAS patients cause breathing stops during sleep and blood oxygenation level drops, their sleep stage become WAKE and sleep again. They repeat this many times. For this reason, as shown in Figure1, comparing the overnight sleep stages of (a) SAS patient and (b) healthy subject, SAS patient has more frequent of WAKE than healthy subject, as can be seen from the red circle. The vertical axis shows the sleep stage, and the horizontal axis shows time.

Frequent WAKE during sleep has a bad influence on the quality of sleeping and invites drowsiness during the day. If drowsiness is invited during driving, the risk of traffic accidents increases. In fact, an accident happened that the Shinkansen stopped suddenly in Japan. The Shinkansen driver was suffering from SAS (Washizaki et al. 2010). In addition, it is thought that SAS causes not only accidents, but also the lifestyle disease such as hypertension, heart failure, diabetes and so on from the load on the heart due to stoppage of respiration. Actually there are many people who are suffering from SAS in lifestyle disease patients.

In recent Japan, it is said that the number of patients who have SAS requiring treatment exceeds 3 million, and it is regarded as one of the modern disease. However, the number of patients receiving medical treatment is only about 400,000 people because people hesitate to go to see a doctor and can not realize yourself apnea during sleep.

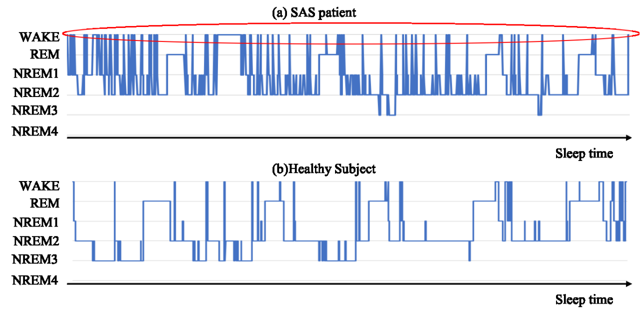

Figure 1: Comparing sleep Stages: (a) SAS patient; (b) healthy subjects

Related Works

Rechtschaffen & Kales Method

The Rechtschaffen & Kales Method: R&K method is the international standard method to classify sleep stage into six levels. The sleep stage is an objective indicator of sleeping depth and is classified into six stages of WAKE, REM, NREM-1 to 4 in order from shallow sleep to deep sleep. The method defines the sleeping state by the biological changes obtained the data from PSG, and the data consists of three pieces of information, electroencephalogram: EEG, electrooculogram: EOG, electromyogram: EMG. Because of high accuracy rate of sleeping stage, this method has been widely used in medical front. However, subjects needs to be attached lots of electrodes on their head and body and it is stressful to subjects. For this reason, it is difficult to measure continuously the sleep stage for healthcare.

Non-contact Method for Sleep Stage Estimation

Some resercher develop the matress sensor for sleep stage estimation as non-contact method. Watanabe focused on the circadian rhythm, which is an indicator of human daily life rhythm, correlates with depth of sleeping through HRV (Watanabe et al. 2001). They extracted intermediate frequency components of HRV with heart rate obtained from the matress sensor. It shows that this correlates with the sleep stage, and they estimate sleep stage from the intermediate frequency components of HRV. In the detection of WAKE/REM, distribution of body movements is used as well as heart rate information.

However, since Watanabe method needs whole data during sleep, it is difficult to estimate the sleep stage in real time.

Real-time Sleep Stage Estimation

Based on Watanabe method, Harada proposed Real-time Sleep Stage Estimation: RSSE (Harada et al. 2016). They estimate the sleep stage in real-time from partially obtained heart with mat sensor. Assuming that the intermediate frequency of heart rate is based on a normal distribution, they normalize the frequency, and estimate the sleep stage by discretizing it. To get the intermediate frequency of heart rate, they construct the trigonometric function regression model using only partially obtained heart rate.

In WAKE detection, they focus on the large body movements: BM during sleep. The standard deviation of BM: BM_{stv} in the most recent minute and the average value of BM: BM_{ave} from the time of sleeping to the present are calculated, and when the standard deviation is larger than the average value, the recent one minute is judged as the WAKE.

$$\frac{BM_{stv}}{BM_{ave}} > 1.0 \tag{1}$$

However, there is a tendency to decrease the accuracy in WAKE detection. This is because RSSE will judge as WAKE if there is body movements, and if the formula(1) is satisfied, the recent one minute will be judged as WAKE.

Proposed Method

To tackle with RSSE's WAKE detection problem, we calculate six kinds of feature (explained later) from bio-vibration data of humans sleep, and use the prepared learned RF model for estimating sleep stage (WAKE/Non-WAKE). To get bio-vibration data, we employed mat type sensor that is being marketed. Figure 2 shows the whole flow of the WAKE detection.

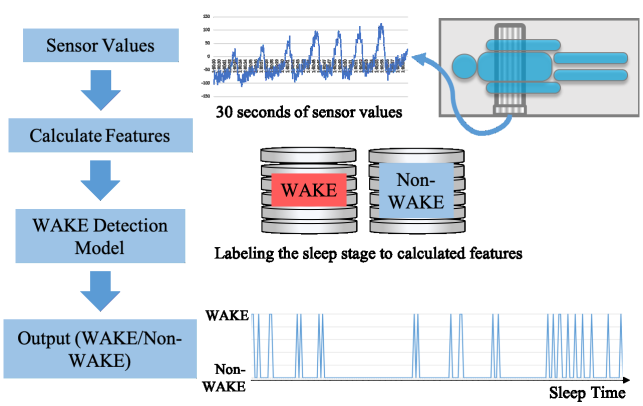

Figure 2: WAKE detection flow

Selecting Non-Contact Device

The device for our method should be low cost and not disturb the patient's sleep. To satisfy these demands, we use the TANITA Sleep Scan; SL-511(Noh et al. 2009)(Fig.3). We put the sensor under the bed mat to get bio-vibration data on the sleeping. The sensor should cover only the area of the chest, and outputs the sensing pressure values 16 times per second.

Processing Sensor Values

The input for the pressure sensor is only vibrations from the patient's body. Figure 4 shows the sensor values of 30 seconds. The wave of vibration includes heart beats, respirations, body movements and noise. To get the characteristics

[1]http://www.tanita.co.jp/product/g/_TSL511WF2/

Figure 3: TANITA Sleep Scan SL-511[1]

of body movements, the proposed method calculates the average of every 1 second sensor values and six kinds of attribute (Table 1) which the average of 30 seconds from the time of predicting the sleep stage. The features are 30 seconds sensor values of standard deviation (SD), difference between maximum value (DIFF), sum (SUM), sum of squares (Square), average of variation (Level Change: LC) and Root-Mean-Square (RMS).

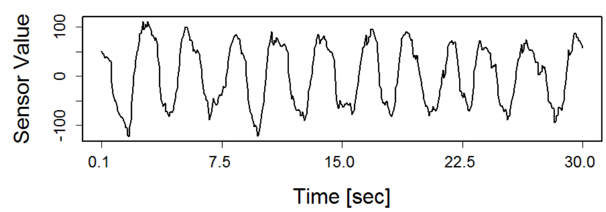

Figure 4: Sensor values of 30 sec

Table 1: Six features from 30 seconds sensor values

feature	Formula
SD	$\sigma(x)$
Range	$max(x) - min(x)$
SUM	$\sum_{n=1}^{N}(x_n)$
Square	$\sum_{n=1}^{N}(x_n^2)$
LC	$\frac{1}{N-1}\sum_{n=1}^{N-1}(x_{n+1} - x_n)$
RMS	$\sqrt{\frac{1}{N}\sum(x^2)}$

Classifier for WAKE Detection

To classify the WAKE/Non-WAKE based on these six features, we select Random Forest: RF model (Breiman et al. 2001) as classifier because of its interpretation and detection by various rules. Concretely, six features are labeled with sleep stage (WAKE/Non-WAKE) measured by PSG, and input it to RF.

RF is one of machine learning algorithms, and it is an ensemble learning algorithm integrating decision trees that are weak learners. The model repeats random sampling from learning data, randomly construct decision trees with different conditional branches, and classify them by majority rule of those results. In this research, Gini impurity is the splitting condition, it becomes low when all the samples contained in each node of the decision tree are the same. RF processing is as follows:

1. Generate bootstrapped sample: S_j from training data set: S.
2. One-third of the original data is called Out-Of-Bug: OOB, and it is used for constructing decision tree. Each node processing is as follows:
   (a) Extract m_{try} features randomly with not allowing duplicate value.
   (b) Choose the feature that minimizes Gini impurity, and divide nodes.
3. Repeat 1. to 2. N_{tree} times.

Where N_{tree} is the number of decision trees to be generated. In the classification problem, it is recommended to use the square root of the total number of features for the variable m_{try}, which used to divide the nodes of decision trees.

In order to extract the WAKE detection rule, we analyzed trees which were used to judge characteristic WAKE (awake with small BM etc.) from generated model for SAS discrimination.

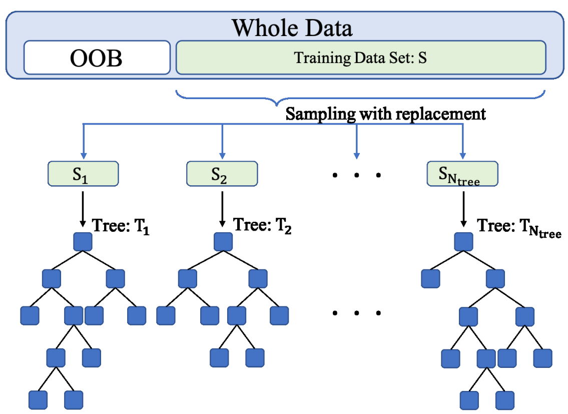

Figure 5: Learning process of Random Forest

Experiments

To investigate the effectiveness of the WAKE detection based on features related to bio-vibration data, we conducted the human subject experiment on nine of SAS subjects. In addition, to compare WAKE characteristics of SAS subjects and healthy subjects, we conducted the human subjects experiment on nine of healthy subjects. Table 2 and Table 3 show the details of nine SAS subjects and nine healthy subjects. The row of "Num of epoch" reprezents the number of epochs of one night sleep, and epoch is 30 seconds. The row of "WAKE" reprezents the number of WAKE epochs of one night sleep. In Table 2, there is no information of age because these are data of patients so that personal information are not disclosed.

Table 2: Details of SAS subjects.

SAS subject ID	Severity	Num of epoch	WAKE
A	moderate	912	51
B	moderate	977	146
C	moderate	865	174
D	moderate	953	66
E	mild	1031	140
F	mild	825	66
G	moderate	934	191
H	mild	954	48
I	mild	952	60

Table 3: Details of healthy subjects

Healthy subject ID	Age	Num of epoch	WAKE
a	20	848	46
b	20	584	53
c	30	704	103
d	40	607	75
e	40	595	44
f	40	420	34
g	40	651	53
h	50	860	35
i	60	720	98

Setup

Each subject wore an electro-encephalograph for PSG and put the mat type sensor (TANITA Sleep Scan) under the bed mat to get bio-vibration data. The data measured by PSG is used to estimate sleep stage by the R&K method, whereas the data measured by mat type sensor is used to estimate sleep stage by the proposed method. In the R&K method, medical specialists determine the sleep stage every 30 seconds of sleeping. RSSE, which is compared with the proposed method, needs the information of body movement, so we used standard deviation of every second bio-vibration data obtained from mat type sensor instead. Since the proposed method and R&K method are determine the sleep stage every 30 seconds of sleeping, we changed RSSE to estimate every 30 seconds from one minute. Concretely, when the number of WAKE, that judged every 30 seconds from the starting time, exceeds 15, it is judged that the section of 30 seconds is WAKE.

Experiment 1: Proposed method vs. RSSE

To prove the effectiveness of the proposed method, we compared between WAKE Detection of the proposed method and RSSE. In the proposed method, we generate nine of different RF models with the following parameters: (1) tree's max depth is 5; (2) the number of tree is 300; (3) the number of variables used to generate the tree is 3. The training data of each RF model is eight of SAS subjects, and the validation data is the other SAS subject. The ratio of WAKE and

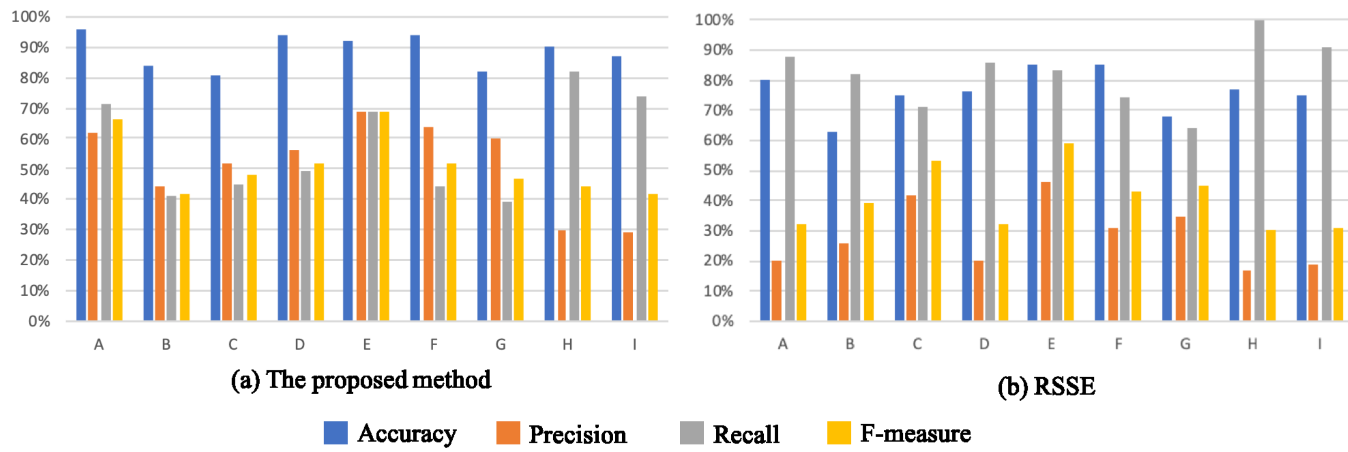

Figure 6: Indices of estimated results

Non-WAKE of learning data is 1:4, because there are five stages of REM, NREM1 to 4 except WAKE.

Experiment 2: Comparing SAS and Healthy subjects

To compare WAKE characteristics of SAS subjects and healthy subjects, we generated three of RF model for each nine SAS subjects and nine healthy subjects, and the total count of models is 54. As the depth of the RF model gets deeper, the kinds of rules generated from RF model increase. To compare the kinds of rules, we set the parameters of the three models are as follows:

1. tree's max depth is 3;
2. tree's max depth is 4;
3. tree's max depth is 5,

and the number of trees and the number of variables used to generate the tree are 100 and 3 respectively. The ratio of WAKE and Non-WAKE of learning data is also 1:4, like Experiment 1. After generating the RF model, extract all WAKE data from same subject, which is learning data, and input to the model to analyze the decision path of each decision tree. Then, extract only the path that the RF model could actually judge as WAKE, and count the kinds of combination of the rules. For example, there are four combinations (top node to bottom node) as follows:

1. SD, AVE, Square;
2. SD, DIFF, LC;
3. SD, AVE, Square;
4. Square, AVE, SD.

In this case, (1) and (3) is same, and (1), (2), (4) are different from each other (also consider the order), so the number of kinds of feature combination is three. However, the decision paths, that can only judge WAKE with less than 10, is excluded.

Evaluation Criteria for Experiment 1

The correct answer is the sleep stage measured by R&K method and sleep stage was classified as WAKE or not-WAKE because the R&K method is international standard method. We evaluated by the four of indices, Accuracy, Precision, Recall and F-measure, for the WAKE detection, and compared these evaluation indices of the proposed method and that of RSSE. It can be said that the proposed method is effective when the Accuracy and Precision of proposed method are higher than that of RSSE.

Evaluation Criteria for Experiment 2

We focus on the WAKE of healthy subjects can be judged relatively easily than SAS subjects. If the WAKE is difficult to judge, the kinds of feature combination, that generated from RF model, will increase, and therefore it can be said that the subject is SAS, if the number of kinds of feature combination is high.

Results

Result 1: Proposed method vs. RSSE

Figure 6 shows the results which was acquired by the each method based on 4 evaluation indices, Figure 6(a) is the results of the proposed method and Figure 6(b) is that of RSSE. In the proposed method (Figure 6(a)), labels on the horizontal axis shows combinations of training data and validation data. For example, "A" represents that the training data is subject "B" to "I", and the validation data is subject "A". Combination A is the top Accuracy which percentage is 96.0% and Precision of all combinations are higher than 40% except combination "H" and "I". In combination H and I, Recall is larger than Precision. In RSSE (Figure 6(b)), labels on the horizontal axis shows each subject ID. In all subjects, Recall is larger than Precision, and Precision is lower than 40% except subject "C" and "E".

Figure 7 shows the average of all indices. The blue bar shows average of each indices of the proposed method and orange wavy bar shows that of RSSE. Comparing with the

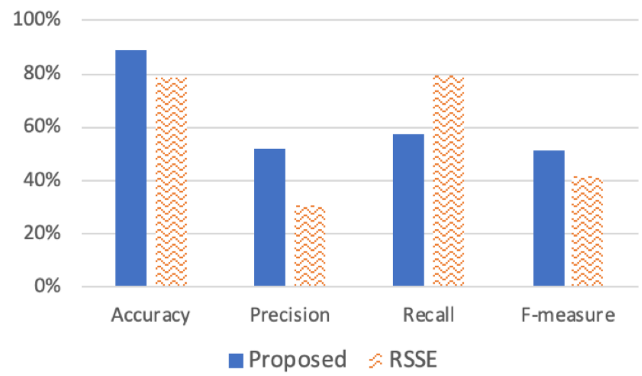

Figure 7: Average of result 1

proposed method and RSSE, Accuracy and Precision of the proposed method are higher than that of RSSE, and Recall of the proposed method is lower than that of RSSE. However, F-measure of the proposed method is higher than that of RSSE because of improving of Precision.

Result 2: Comparing SAS and Healthy subjects

Figure 8 shows the average of kinds of feature combination extracted from each generated model. The vertical axis shows the average of kinds of feature combination, and the horizontal axis shows the max depth of RF model. The blue bar represents the average of all SAS subjects, and orange wavy bar represents that of Healthy subjects. As the graph shows, the number of feature combinations of SAS subjects is higher than that of healthy subjects in all cases. As the max depth of RF model gets deeper, the difference between SAS and healthy subjects manifests al lot.

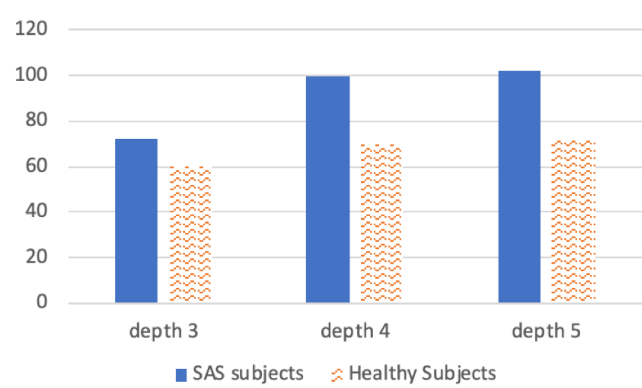

Figure 8: The average of kinds of feature combination extracted from each RF model.

Discussions

Discussion 1: Proposed method vs. RSSE

Figure 9 shows the part of estimated results of sleep stage, Figure 9(a) is result of subject "E" and Figure 9(b) is result of subject "I". Both of them indicate result of RSSE on the upper side and that of the proposed method on the lower side. The vertical axis shows sleep stage, and the horizontal axis shows time. In all graphs, blue line shows the sleep stage determined by R&K method, gray line shows RSSE's estimated sleep stage, green line shows the proposed method's estimated sleep stage, and orange line shows 60 seconds of standard deviation of BM or 30 seconds of standard deviation of bio-vibration data. In Figure 9(a), left side red circles show Non-WAKE with small BM and RSSE made misjudgement while the proposed method made correct judgement. In the proposed method, the reason why the proposed method could decrese the misjudgement is RF generates multiple rules from six features obtained from bio-vibration and it can evaluate data not only large or small of BM, but also from various directions. Focus on subject "I", the reason why the difference between Precision and Recall is larger than else is a part of sensor values had noises and it affects the features like Figure 9(b)'s red circle. It was also seen subject "H". In order to solve this problem, first, the proposed method dose not remove noises, so that we must remove it. Second, to extract correct BM, we can analyze frequency domain by Fourier transform because if the body moved, the shape of the power spectrum will be disturbed.

Discussion 2: Comparing SAS and Healthy subjects

From the results, simply putting the mat sensor under the bed mat and sleeping, there is a possibility to judge SAS or Non-SAS by analyzing the kinds of feature combination extracted from RF model. However, since this SAS judgment method needs the two stage of sleep stage (WAKE/Non-WAKE), further improvement of the sleep stage estimation accuracy is required.

Figure 10 shows the number of kinds of feature combination extracted from each generated RF model. In SAS subjects, subject "A", "H" and "I" have less kinds of feature combination than the others. It is thoght that these subjects has more normal WAKE than WAKE by apnea, and it affects the RF model. In order to solve this problem, it is necessary to separate WAKE by apnea and normal WAKE, and learn two type of WAKE respectively. In healthy subjects. subject "c" and "i" have more kinds of feature combination than the others. What can be said commonly between subject "c" and "i" is they have many WAKE as Table 3 shows. In Subject "c" there was a long terms of WAKE during sleep, so there is a possibility that subject "c" woke up in that terms, and it affects the RF model. In order to solve this problem, it is necessary to distinguish between WAKE and completely awake. Subject "i" is over 60 years old, and elderly people tend to have more WAKE during sleep. To find differences between SAS subjects and healthy elderly subjects, we should analyze what differences exist in the kinds of feature combinations. In addtion, we can analyze not only body movements but also respiratory rate and heart rate by analyzing frequency domain.

Conclusion and Future Works

In this paper, we proposed the WAKE detection method for SAS patients to improve the estimated Accuracy and Precison than RSSE. To improve the Accuracy and Precision,

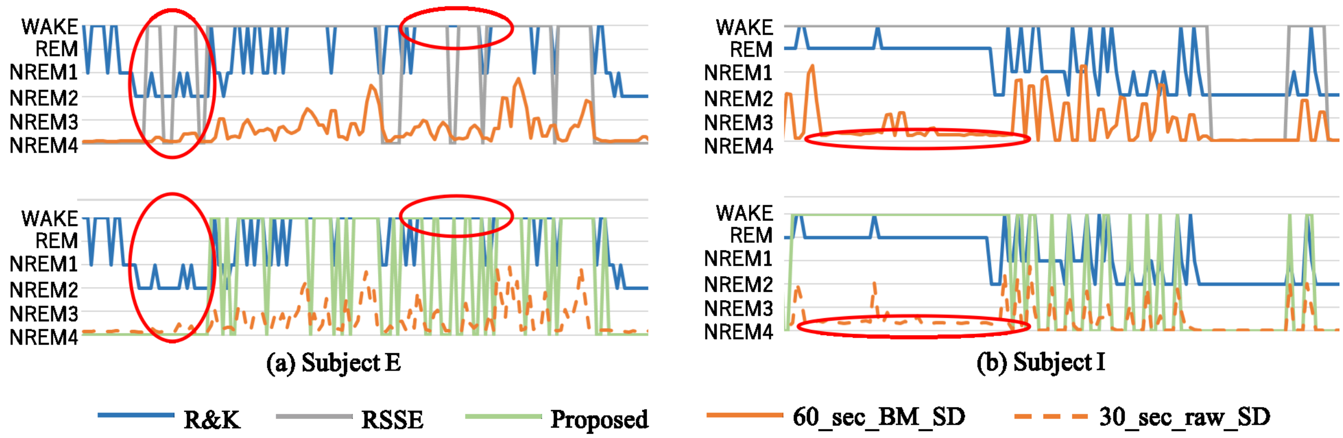

Figure 9: The part of estimated results of sleep stage

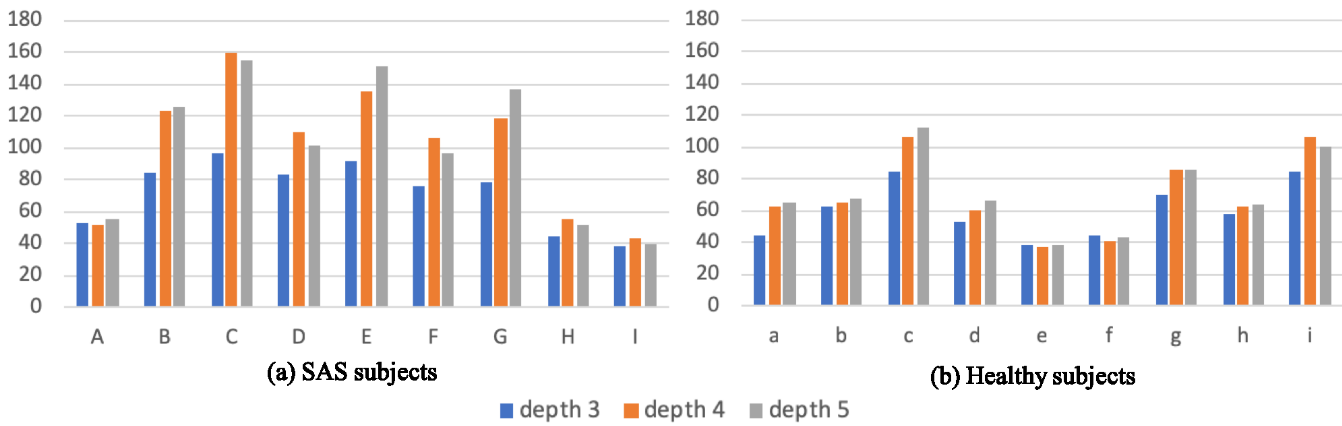

Figure 10: The number of kinds of feature combination extracted from each generated RF model

we calculated six features from bio-vibrations obtained from mat type sensor, and to evaluate from not only single direction but also various directions, we used RF model for classifier. In addtion, to find a possibility to judge SAS or Non-SAS with mat type sensor, we focus on the kinds of feature combination extracted from generated RF model.

To investigate the effectiveness of the proposed method, we conducted the SAS and healthy subjects experiments. We compared the sleep stage determined by R&K method with the sleep stage estimated by the proposed method and RSSE. As a results, the proposed method was effective for improving Accuracy and Precision, and we found the possibility to judge SAS or Non-SAS with mat type sensor.

The future task is following: (1) to improve the Precision and Recall because the method of SAS detection needs correct sleep stage; (2) separate WAKE by apnea and normal WAKE to find more detailed differences between SAS and healthy subjects.

References

Harada T. and Takadama K. 2016. Real-Time Sleep Stage Estimation from Biological Data with Trigonometric Function Regression Model. *AAAI Spring Symposium*

Leger D., Poursain B., Neubauer D. and Uchiyama M. 2008. An international survey of sleeping problems in the general population.

Leo Breiman. 2001. Random Forests. *Machine Learning* 45(1): 5–32.

Rechtschaffen A. and Kales A. 1968. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Washington DC*

Watanabe T. and Watanabe K. 2001. Estimation of the Sleep Stages by the Non-Restrictive Air Mattress Sensor. *The Journal of Transactions of the Society of Instrument and Control Engineers* 37(9): 821–828.

Washizaki M. and Shima S. 2010. The measure of Sleep Apena Syndrome in Railroad. *International Association of Traffic and Safety Sciences* 35(1): 26–42.