Actigraphic Detection of REM Sleep Based on Respiratory Rate Estimation

Ken Kawamoto and Hiroyuki Kuriyama
Central Research Lab, Hitachi Ltd., Tokyo, Japan.
Email: {ken.kawamoto.qg, hiroyuki.kuriyama.sr}@hitachi.com

Seiki Tajima
Hyogo Children’s Sleep and Development Medical Research Center, Hyogo Rehabilitation Centre Central Hospital, Kobe, Japan.
Email: s_tajima@hwc.or.jp

Abstract—The use of wrist actigraphy in sleep research has for long been limited to the classification of sleep/wake; little progress has been made in the evaluation of the sleep stages. We propose and evaluate two novel algorithms: a method for respiratory rate estimation based on spectral analysis of actigraphic data, and a method for estimating REM sleep based on the detected respiratory rates. Using simultaneous recordings of polysomnography and actigraphy data acquired from 34 subjects, we found that our proposed method successfully estimated respiratory rate with low mean absolute error (0.52 counts/min), and REM sleep with high positive predictive value (64.5%), but low sensitivity (11.0%). While the low sensitivity hinders the immediate clinical use of our algorithms, our findings are important in indicating for the first time that actigraphs have the potential to detect REM sleep.

Index Terms—actigraphy, respiratory rate, REM sleep

I. INTRODUCTION

A. Background

In recent years, greater understanding of the importance of sleep has given rise to various instruments for its measurement. Polysomnography, which allows the evaluation of the sleep stages (Wake, REM, Non-REM) for identification of various forms of sleep disorders [1]-[4], remains the gold standard for sleep quality evaluation. However, its limitations including expense, the expertise required for measurement, and the physical and psychological stress imposed on the examinee [5], narrows its use to short term measurements to a small selection of patients. It is generally unrealistic to use polysomnography for tracking long-term trends or for executing screenings on a large population, both of which we believe are crucial for furthering our understanding of sleep.

In light of this, an increasing number of sleep research have opted to using actigraphy as a simpler alternative for quantifying sleep [6]-[9]. Actigraphy usually involves the examinee wearing wristband shaped accelerometers on the wrist to measure body movements throughout the night. It has the advantage of being low cost, easily administrable, and non-invasive to the examinee, making it appropriate for long-term use. In fact, its simplicity has even allowed it to be mass-marketed as consumer healthcare products.

However, the majority of such past research target the design [10] and validation [11]-[13] of algorithms for automatically classifying sleep and wake epoches. As far as the authors are aware, there are no reports of successfully using actigraphy for analyzing the sleep stages.

In this article, we describe two new findings; firstly, that actigraphic data contain artifacts of respiratory movement that can be used to give an estimation of respiratory rate, and secondly that the inspection of surge features occurring in respiratory rate allows discovery of REM sleep. We show that the combination of these findings allow REM sleep detection solely from actigraphic data.

We first describe the design of the algorithms based on 24 sets of simultaneous polysomnography and actigraphy recordings. We then describe the evaluation of the algorithms using 10 new sets of recordings.

B. Past Research

In polysomnography, respiration is measured by nasal airflow sensors and chest wall sensors. Such sensors that attach to the face or abdomen restrict normal sleep [14]. Watanabe et al. [15] attempted to solve this problem by using mattress that measures heart rate and breath count.

We are not the first to notice that actigraphic data contain traces of respiratory movement. Sadeh et al. [16] noted that “breathing artifacts” resulting from placing the wrist on the chest can appear in actigraphic data, but dismissed these artifacts as being causes of less accurate sleep/wake scoring.

A non-invasive attempt at estimating REM sleep has been made [17], in which the author proposes an algorithm to detect REM sleep from the finger by measuring the peripheral arterial tone signal. The authors achieve a positive predictive value of 62.5%, the highest for
non-invasive REM classification that we have been able to find.

Our research is novel in two respects. We are the first to demonstrate that respiratory rate estimation is possible from actigraphic data. Also, we are the first to propose and validate an algorithm that classifies REM sleep solely from actigraphic data.

II. METHODS

A. Data Collection

For this study, we made 34 simultaneous polysomnography (PSG) and actigraphy recordings from 34 participants. All were healthy young adults (age 20~40 yrs) screened using the Pittsburgh Sleep Quality Index [18]. The study was performed in two phases. The first phase involving 24 subjects, took place between Oct. 2010 and Mar. 2011. These data (“Training Data”) were used for the design of the algorithm. The second phase involving 10 subjects took place between Feb. and Mar. 2012. These data (“Validation Data”) were used for evaluating the algorithms.

The actigraphic device employed was a “Life Microscope” wristband (Hitachi Ltd.) [19], that measures tri-axial acceleration with a resolution of 11.7mG at 20Hz. All participants were asked to refrain from alcohol for 3 days, and caffeine for 24 hours prior to the night of the PSG recording. The PSG recording took place at the Hyogo Rehabilitation Centre Central Hospital.

In the PSG recording, EEG, EOG, EMG, ECG, body postures, SpO2, and nasal air flow were measured. The participants, after being fitted with the devices, were given 1~2 hours to relax before lights out at 9pm. The measurements were visually scored by technicians according to the R&K criteria [20] for 30 second epochs.

The ground data for respiratory rates were calculated from the nasal airflow measurements as follows. For every 1 minute epoch, we calculated the autocorrelation of the nasal airflow rate for increasing time lag, and chose the first positive peak as the peak period. We arbitrarily defined peaks with correlation below 0.1 to be invalid peaks, and marked that minute as having no valid respiratory rate ground data. We have confirmed the validity of this method by randomly sampling 280 minutes from the test data, and visually counting the number of peaks of airflow in each minute. The calculated rates matched the counted respiratory counts to a mean absolute error of 0.48 counts/minute.

The study was approved by the Hyogo Rehabilitation Center Central Hospital Ethics Committee. Written informed consent was obtained from all participants of this study.

B. Respiratory Rate Estimation

We first discuss the theory behind the proposed method for respiratory rate estimation. Fig. 1 shows the raw acceleration data from 30 minutes of actigraphic recording of sleep (top), as well as its spectrogram (bottom). Red in the spectrogram signifies greater power. Apart from the occasional large movements caused by rollovers, the acceleration data appears still for most of the time. In the spectrogram, some sections of the recording only show white noise [see Fig. 1(a)] originating from sources like random thermal noise. However, other sections display strong peaks in the lower frequency band (around 0.2~0.3Hz) [see Fig. 1(b)]. We hypothesized that this peak is caused by the wristband moving periodically when placed near the chest, leading us to consider estimating respiratory rates by determining its peak frequency.

An issue we face is that such peaks appear only about half of the time during sleep, so simply picking the peak frequency for each timestep will give meaningless results half of the time. We thus evaluate the plausibility of the found frequency peak by calculating a “peak score” which, if above a certain threshold, would signify that the protrusion of this peak is enough for it to have been induced by respiratory movement.

A detailed description of the estimation algorithm is as follows. We put each of the three axes of data through a band pass filter with a passband of 0.1Hz~1Hz, large enough to contain the respiratory rates found in our data. We then calculate the power spectrum of each axis using FFT for every timestep, say 1 minute, with a window of 205 seconds (equivalent to 4096 data points on our 20Hz wristband). We then select the best candidate frequencies; for every time step t, in each axis a, we select the frequency $f^a$ with the largest power $p_{a,t}(f^a)$. We evaluate the protrusion of this frequency by calculating its peak score $s_{a,t}$ using (1). Intuitively, this evaluates how much the power of frequency $f^a$ stands out from the power of all other frequencies.

$$s_{a,t} = \frac{p_{a,t}(f^a) - \text{avg } p_{a,t}(f)}{\text{stddev } p_{a,t}(f)}$$

For each time step t, we select the axis with the highest peak score. If this peak score exceeds a certain threshold (whose selection we describe later on), we accept the found frequency as the respiratory frequency for this time step. Otherwise, the time step is labeled as undetected.
The selection of the peak score threshold is a tradeoff between accuracy and detection ratio; the tighter the threshold, the more accurate the estimation but lower the detection ratio. We chose the threshold as follows. Using the 24 cases of training data, we calculated for varying thresholds the correlation between the estimated data and the ground data (solid line in Fig. 2). We also calculated for each threshold the detection ratio (dashed line in Fig. 2). Fig. 2 shows the tradeoff relationship between correlation and detection ratio. We arbitrarily selected a threshold of 4.12, which gives a high correlation of 0.9, while giving a fair detection ratio of 44%.

The above algorithm results in a sequence of estimated respiratory rates for every timestep. Although we will evaluate the validity of this algorithm in a later section using a new set of test cases, it is worth discussing the characteristics of the estimated respiration rates here, as it will serve as the basis of the REM detection algorithm described next. Fig. 3 shows two examples of estimated respiratory rates for two subjects, one with detection ratio of 60% (Fig. 3(a)), and the other with 10% (Fig. 3(b)). In the former, the estimated respiratory rate closely follows the global trend of the respiratory rate (slowly decreasing from sleep onset to wake), while also capturing the local features like the surges caused by REM sleep. In the latter case, although the few detected values are accurate, the overall trend is difficult to deduce. In fact, we find that the detection ratio for the 24 individuals in the training case is fairly low (average: 49.5%, std.dev: 14.0%). This low detection ratio is something we need to consider when designing the REM detection algorithm.

C. REM Sleep Detection

We now propose a method for detecting REM sleep, using the respiratory rate data obtained as per the previous section. The method in this section depends on past findings that normal adults breathe more rapidly during REM sleep compared to during Non-REM sleep [21]. The occasional surges of respiration that occur several times a night is clearly recognizable in Fig. 3. Our aim is to detect sections of REM sleep by finding these surges in respiratory rate.

However, two characteristics of our respiratory rate data make finding these surges challenging:

1. Respiratory rates have a global decreasing trend from sleep onset to wake. Thus, separating REM from Non-REM is not simply a matter of setting a REM cutoff threshold.
2. Our detected respiratory rate are sparse, so any methods that rely on finding clear peaks (like finding changepoints in the differential) are likely to fail.

We propose a method that first constructs a respiratory rate baseline (the trend found after smoothing out local features), and then uses the baseline to normalize the respiratory rates to make the surges stand out. We describe the method in more detail. Fig. 4 shows intermediate results from each step of the process. The actual sleep stages are shown in Fig. 4(a).

Firstly, we estimate the respiratory rate for every minute of sleep using the algorithm described before, and smooth it out with quadratic interpolation (Fig. 4(b)). Next, we calculate the baseline by interpolating a quadratic function through the respiratory rates (Fig. 4(c)). Notice that the calculated baseline follows the global trend, ignoring local features, as desired. We normalize each respiratory rate by dividing by the baseline (Fig. 4(d)), making the individual REM surges stand out from the rest. By extracting sections where the normalized respiratory rate is continuously above 1, we identify individual sections of REM candidates. Fig. 4(e) shows that wherever respiration has been detected during REM, a REM candidate has also been successfully identified.

rem sleep detection
Thus, we use classification analysis to separate the valid and invalid candidates. For each candidate section, we calculate features like the maximum divergence and duration. Using an SVM [22] classifier trained on these features for whether the section intersects with a REM segment or not, we classify and find all REM candidates which are likely to be intersecting a REM segment. For every 1 minute epoch, we classify all epoches that are contained in a predicted REM segment as “REM epoch”, and all others as “unknown”. (Note that we cannot conclude that the epoches not intersecting a REM segment are Non-REM, since it may just be that respiration was not detected during a REM segment.)

III. RESULTS

We used data from the 10 new subjects for evaluation of the proposed algorithms.

A. Evaluation of Respiratory Rate Estimation

For two of the subjects (ids 33 and 34), we could not obtain the ground respiratory data due to technical difficulties, so they were not used in this evaluation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Percentage of detected time</th>
<th>Sleep Duration [min]</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>98.2%</td>
<td>53.4%</td>
</tr>
<tr>
<td>26</td>
<td>99.2%</td>
<td>26.8%</td>
</tr>
<tr>
<td>27</td>
<td>99.7%</td>
<td>62.5%</td>
</tr>
<tr>
<td>28</td>
<td>100.0%</td>
<td>82.6%</td>
</tr>
<tr>
<td>29</td>
<td>99.8%</td>
<td>40.8%</td>
</tr>
<tr>
<td>30</td>
<td>99.8%</td>
<td>22.1%</td>
</tr>
<tr>
<td>31</td>
<td>99.1%</td>
<td>37.8%</td>
</tr>
<tr>
<td>32</td>
<td>99.9%</td>
<td>26.9%</td>
</tr>
<tr>
<td>33</td>
<td>-</td>
<td>40.7%</td>
</tr>
<tr>
<td>34</td>
<td>-</td>
<td>48.9%</td>
</tr>
</tbody>
</table>

Avg(Sd) 99.5% (0.6%) 44.3% (18.5%) 648.2 (58.0)

Comparing the estimated and actual rates for each minute, we found high correlation (R=0.88, p<0.001). 86% of the points were within 1 count of the actual rate. The mean absolute error (MAE) was 0.52 counts/min, comparable to that of the ground data.

One characteristic of our proposed method, as mentioned earlier, is its sparse output. Table I shows for each subject, the percentage of time during sleep with detected respiratory rate for both the ground data and the estimated data. The detection ratio was 44.3%±18.5% (avg±sd), a fairly low figure with large inter-subject variance.

B. Evaluation of REM Sleep Detection

The SVM classifier used in this evaluation was trained using the training set. We evaluate our algorithm using two indices: positive predictive value (PPV) and sensitivity. PPV is the percentage of epoches that were claimed to be REM by the algorithm that were actually correct. Sensitivity is the percentage of actual REM epoches that were identified by the algorithm. The other indices often employed that evaluate the accuracy of the negative claims (e.g. specificity, negative predictive value, etc) are vacuous, since our algorithm does not make negative claims; when tagging an epoch as “undefined”, it is not claiming that the epoch is “not REM”.

Table II shows the per-subject result of REM sleep detection. For 4 of the subjects (ids 26, 30, 33, 34), no REM epoches were claimed by the algorithm, thus the PPV could not be defined. The PPV for all subjects together is 64.5%, which is much higher than the chance level (20.8%). Individually, 4 of the subjects had high PPV of above 60%, and the other 2 subjects had low PPV below 30%. Note that for every subject that had a well-defined PPV, the value was higher than the subjects’ REM ratio (i.e the chance level). Of the epoches misclassified as REM, 11.3% were actually Wake. 24.2% were NREM1 or NREM2. No NREM3 or 4 epoches were misclassified as REM.

The sensitivity was very low for most subjects. The overall sensitivity was 11.0%, but 7 of the subjects had sensitivity less than 10%.

IV. DISCUSSION

The objective of this study was to propose and evaluate methods for estimating respiratory rate and detecting REM, both solely from actigraphic data.

We found that our respiratory rate estimation algorithm was accurate, achieving an high correlation (R=0.88) and low MAE (0.52 counts/min). However, it had a low detection ratio (44.3%). This is understandable, since our algorithm depends on the actigraphic device being near the chest during sleep by chance. While use of more sensitive accelerometers may improve the detection ratio, we think that sparseness in the output is an issue inherent in this method.

In this study, we assumed that the strongest peak within the 0.1~1Hz band represents respiration. Our evaluations suggest that this assumption is acceptable in the setting of our study. However, there may be free running conditions where this may not hold, such as in a moving vehicle. Future work should validate our algorithm in a wider variety of settings.

Our proposed REM detection algorithm had high PPV (64.5%), but low sensitivity (11.0%). Although some subjects had very low PPV, our algorithm still operated...
successfully for all subjects, with accuracy above chance level. In fact, PPV of 64.5% is higher than any other reported values for non-invasive REM detection methods. Furthermore, all misclassifications occurred between the Wake and NREM1~2 stages. By using methods from past studies [10]-[13] that promise to distinguish Wake, we may be able to improve our PPV by first detecting the Wake stage.

A major limitation of our algorithm is the low sensitivity. One cause is the low detection ratio of respiration: all subjects with low detection ratio below 40% also had low REM sensitivity below 10%. However, some subjects had low REM sensitivity despite having high respiration detection ratio. We show two example outputs of our algorithms (Fig. 5) to explain how we could potentially improve performance. (1)(3-5)(7, 8)(11) are cases where the actual REM epoches have been correctly identified. In (2), the REM segment is missed despite the respiration surge being detected. This may be rectified by improving the calculation of the global trend. (6)(9)(13) are cases where respiration had not been detected in the first place. (10) was not classified as REM despite having a REM candidate associated with it. This is an error of the first place. All misclassifications occurred between the Wake and NREM1~2 stages. By using methods from past studies [10]-[13] that promise to distinguish Wake, we may be able to improve our PPV by first detecting the Wake stage.

Figure 5. Two example outputs of REM detection: (a) Subject 29 with high PPV and (b) Subject 25 with low PPV.

While various alterations in the detection method as discussed above may help in improving the performance, we are reluctant to believe that the issue of low specificity will be overcome entirely. This may be problematic for applications requiring precise, sensitive measurements, like the diagnosis of sleep disorders. For these purposes, other devices like mattress-type sensors or portable EEG headsets may be more suitable. However, we believe that the algorithm has the potential to be useful in many other applications that require monitoring of long term trends to a wider target, considering that actigraphs are becoming prevalent not only in the sleep research field, but also in consumer healthcare. We believe that future work will find appropriate applications for our sparse but long-term recordings of REM.

Although we have only validated our algorithms on a single variant of actigraphy, the device we used had specifications that were comparable to many of the common variants, so we speculate that applying our algorithms to other devices is simply a matter of retraining the parameters.

One limitation of this study is in our subject selection, which were all young healthy adults. Our results are therefore very preliminary, since we are unable to say at this point whether our method would achieve similar results with older or younger subjects, or with those having sleep disorders. Specific cases like SAS and PLMS could produce movement artifacts that interfere with our respiratory rate estimation. Such subjects may not display the same characteristics of respiratory rate during REM. Until our method can be verified against such subjects, we remain skeptical about the immediate clinical use of our algorithms.

V. CONCLUSION

We proposed an algorithm for predicting respiratory rate solely from actigraphic data, and then used these data to identify REM sleep segments. The REM classification accuracy had good PPV (64.5%), higher than previous non-invasive methods. We believe that its advantage of allowing long-term measurements with little stress gives it the potential to be useful in a variety of applications that previously could not have been possible.

Our study is important not only in the novel methods proposed, but also in demonstrating that actigraphy, which has for long been considered as merely a simple way of logging the sleep/wake cycle, has a far greater potential as a useful tool for sleep quality evaluation. We hope that future work in this area will improve on the methods presented in this article and use actigraphy for forming a more detailed understanding of human sleep.

ACKNOWLEDGMENT

The authors thank the Hyogo Rehabilitation Center Central Hospital, and the lab team for the collection of sleep data.

This work was funded fully by Hitachi Ltd.

REFERENCES


Ken Kawamoto received the B.S and M.CompSci degrees in computer science from the University of Oxford, Oxford, UK in 2007 and 2008. He joined the Central Research Laboratory, Hitachi Ltd., Tokyo, Japan in 2008, and currently works as a senior researcher in the Social Information Systems Research Department. His research interests include human activity sensing and simulations.

Hiroyuki Kuriyama received B.S in biomedical engineering from Hokkaido University in 1990. He joined Hitachi Ltd., Tokyo, Japan in 1990, and currently works as a senior researcher in the Central Research Laboratory. His research interests include sensor based behavior science and healthcare informatics.

Seiki Tajima received the M.D. from the Saga University School of Medicine, Saga, Japan in 1998. He has practiced and studied chronic fatigue pathology and sleep disorders at Fatigue clinical center, Osaka city university school of medicine (Osaka, 2005-2009), Kansas university of welfare sciences (Kashiwara, 2006-2009). Hyogo children's sleep and development medical research center (Kobe, 2009-present). He has specialized in practicing and researching the effect of sleep on fatigue recovery and the pathology of pediatric chronic fatigue syndrome using physiological data, e.g. actigraphy, PSG, EEG, and heart rate variability.