# Sleep Stage Classification by Non-contact Vital Signs Indices using Doppler Radar Sensors*

Masayuki Kagawa, *Member, IEEE*, Kazuki Suzumura and Takemi Matsui

*Abstract*— **Disturbed sleep has become more common in recent years. To improve the quality of sleep, undergoing sleep observation has gained interest as a means to resolve possible problems. In this paper, we evaluate a non-restrictive and non-contact method for classifying real-time sleep stages and report on its potential applications. The proposed system measures heart rate (HR), heart rate variability (HRV), body movements, and respiratory signals of a sleeping person using two 24-GHz microwave radars placed beneath the mattress. We introduce a method that dynamically selects the window width of the moving average filter to extract the pulse waves from the radar output signals. The Pearson correlation coefficient between two HR measurements derived from the radars overnight, and the reference polysomnography was the average of 88.3% and the correlation coefficient for HRV parameters was the average of 71.2%. For identifying wake and sleep periods, the body-movement index reached sensitivity of 76.0%, and a specificity of 77.0% with 10 participants. Low-frequency (LF) components of HRV and the LF/HF ratio had a high degree of contribution and differed significantly across the three sleep stages (REM, LIGHT, and DEEP; $p < 0.01$). In contrast, high-frequency (HF) components of HRV were not significantly different across the three sleep stages ($p > 0.05$). We applied a canonical discriminant analysis to identify wake or sleep periods and to classify the three sleep stages with leave-one-out cross validation. Classification accuracy was 66.4% for simply identifying wake and sleep, 57.1% for three stages (wake, REM, and NREM) and 34% for four stages (wake, REM, LIGHT, and DEEP). This is a novel system for measuring HRs, HRV, body movements, and respiratory intervals and for measuring high sensitivity pulse waves using two radar signals. It simplifies measurement of sleep stages and may be employed at nursing care facilities or by the general public to improve sleep quality.**

## I. Introduction

An increasing number of adults suffer from disturbed sleep, and sleep observation has gained interest as a means to solve this problem and improve sleep quality. Additionally, nursing and medical settings need to be able to classify sleep states in real-time without interference. Sleep can roughly be divided into two states: rapid eye movement (REM) and non-REM (NREM; N1, N2, N3, and N4) sleep, which alternate in cycles of about 90 min. Polysomnography (PSG) is considered the gold standard in sleep research and allows accurate assessment of most aspects of sleep. Unfortunately, it also requires at least one night in a specialized sleep lab. Expert technicians apply numerous sensors to the patient (possibly affecting sleep) and evaluate the collected data manually in 30 s intervals [1]. During a PSG test, maintaining a normal sleeping routine is difficult for a participant because many sensors are in contact with the body and the reliability and accuracy of sleep-stage classification are reduced.

As an alternative to PSG, attempts have been made to classify sleep stages using heart rate variability (HRV) derived from electrocardiograms [2–5]. In this paper, we propose a new non-restrictive and non-contact method for classifying real-time sleep stages that overcomes the problems inherent in PSG and contact HRV methods. To accurately extract heartbeat intervals, we propose using dynamic selection of window width (DSWW) with a moving average filter and dynamic channel selection (DCS) to select the best radar channel from four radar-output signals. DCS enabled us to accurately extract respiratory intervals (RIs) and heart beat intervals (RR-intervals). Non-contact vital sign detection was processed based on these two methods even though the quality of radar signals was not always optimal and measurement conditions varied during the night when participants changed their sleeping postures.

## II. Methods

### A. Measuring System and Data Collection

The non-contact classifying system for sleep stages is shown in Figure 1. Two Doppler radar sensors (24-GHz, NJR4262J; New Japan Radio, Tokyo, Japan) were installed beneath the mattress. The output of the radar was 40 mW EIRP and the incident power density on the body surface was $1.5 \times 10^{-2}$ mW/cm$^2$, which is much lower than the Japanese

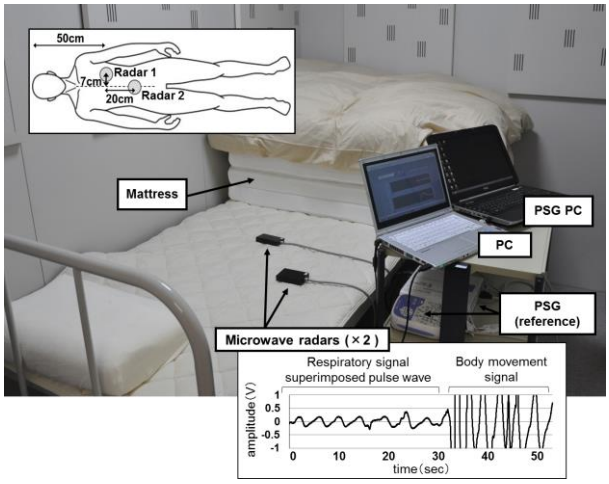


Figure 1. Setup of the sleep-stage classification system and signals obtained from microwave radars

*This research was supported in part by the Japan Society for the Promotion of Science KAKENHI (Grant Number 26350507).

M. Kagawa is with the Faculty of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, Japan (phone: +81-42-585-8467; fax: +81-42-585-8467; e-mail: kagawa.masa2@key.ocn.ne.jp)

K. Suzumura is with the Graduate School of System Design, Tokyo Metropolitan University (e-mail: kazuki.s1223@gmail.com)

T. Matsui is with the Graduate School of System Design, Tokyo Metropolitan University (e-mail: tmatsui@tmu.ac.jp)

safety guidelines for protection from harmful electromagnetic waves (1 mW/cm$^2$). Sample output signals are shown in Figure 1. Radar output signals are composed of respiratory movements, pulse waves, and body movements, and the figure shows how the respiratory and heartbeat signals change drastically with body movements. The Doppler radar provided two orthonormal baseband signals (in-phase and quadrature baseband-outputs). We introduced the dynamic channel selection (DCS) technique to select the best radar channel with the least number of irregular RIs and RR-intervals from four radar output signals.

Ten healthy male volunteers from Tokyo Metropolitan University (aged 22.7 ± 0.9 years) participated in the experiment. Each individual participated in one night of recordings. Multi-channel complete PSG recordings (PSG1100; Nihon Kohden, Tokyo, Japan) were performed for every subject. After the PSG measurement, expert personnel evaluated the participant's sleep, and stages (wake, REM, N1, N2, N3 and N4) were classified in 30 s epochs according to the American Academy of Sleep Medicine rules [6–8]. All recording durations were 8 hours, lasting from 22:00 to 06:00 the next morning. The research protocol was approved by The Tokyo Metropolitan University Ethics Committee, and all participants gave their written informed consent.

The segment size for the feature vectors to be classified was set to 60 s, which is a tradeoff between feature quality and time resolution. Segments that are too short can lead to unreliable features, particularly in RI-fluctuation indices because breathing signals have slow rhythms. We used the following rule to transform the hypnogram values from the 30-s PSG segments to 60-s segment values: when two different sleep stages existed within a 60-s epoch, the epoch was defined as the lighter stage.

### *B. Detection of pulse waves from radar output signals*

Detecting pulse waves from radar output signals is not easy. It demands extracting a small signal (pulse waves) in the background of large noise (respiratory movements and body movements). We used a moving-average filter and a background-subtraction technique to detect the pulse waves from the radar output signals (Fig. 2). Consider the radar-output data series, S sampled at 100 Hz. $\mathrm{S}(n)$ can be expressed as equation (1).

$$\mathrm{S}(n) = \mathrm{H}(n) + \mathrm{R}(n) + \mathrm{B}(n) + \mathrm{W}(n), \qquad (1)$$

where $\mathrm{H}(n)$ are the pulse waves, $\mathrm{R}(n)$ are the respiratory movements, $\mathrm{B}(n)$are the body movements, and $\mathrm{W}(n)$is the white noise at the estimated point $n$. In spite of its simplicity, the moving-average filter is optimal for this task: a low pass filter that removes high frequency components while retaining a sharp step response. The output of the symmetric moving average filter with the input of $\mathrm{S}(n)$ is given by

$$\mathrm{MA}(i) = \frac{1}{N}\sum_{j=-m}^{j=m} \mathrm{S}(i+j),\ N = 2m+1, \qquad (2)$$

where $N$ is the number of points in the average, denoted as a window width, given $i$ as a measuring-point index. The cutoff frequency of this filter decreases with the width of the window. For example, if one wants a 0.55 Hz cutoff frequency to separate the respiratory movement signals from the pulse waves, then the moving average window width $N$ should be 81 (810 ms). Similarly, the window width will be 41 (410 ms), 21 (210 ms), or 11 (110 ms) when making a cutoff frequency of 1.1 Hz, 2.2 Hz, or 4.4 Hz, respectively. If the window width $N$ is an appropriate value (e.g., 81, 41, 21, or 11), the moving-average filter is an exceptionally good low pass filter and eliminates high frequency movements $\mathrm{H}(i)$ and $\mathrm{W}(i)$. The filter output signal $\mathrm{MA}(i)$ can then be approximated as

$$\mathrm{MA}(i) \cong \mathrm{R}(i) + \mathrm{B}(i). \qquad (3)$$

We subtracted the filter output signal $\mathrm{MA}(i)$ from the original radar output $\mathrm{S}(i)$ to extract the pulse waves as follows:

$$\mathrm{S}(i) - \mathrm{MA}(i) \cong \mathrm{H}(i) + \mathrm{W}(i) \cong \mathrm{H}(i). \qquad (4)$$

The conditions for catching the pulse waves via radar always change according to body position and posture in bed. Each minute, we dynamically selected the window width (DSWW) from four values (81, 41, 21, or 11) such that it provided the least numbers of irregular periods in detected RR-intervals (Fig. 2). In Figure 2b, posture-dependent small amplitude R-peaks and a lower signal to noise ratio were observed. We averaged the ten newest RR-intervals and defined irregular intervals as those that were < 83% or > 166% of this average [9]. R-peaks were detected using an adaptive-threshold algorithm. Heart rates (HRs) were calculated from the average of the ten newest RR-intervals.

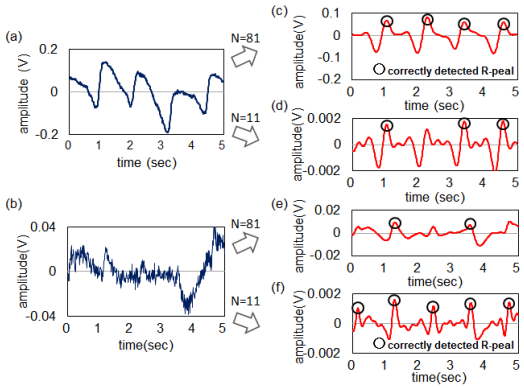


Figure 2. Dynamic selection of the most suitable window width. (a-b) Radar output signals and (c-f) extracted pulse waves. (c) and ( f) are examples of the best choice for the window width.

### *C. HRV calculation from radar pulse waves*

HRV spectral analysis was applied using the autoregressive method on a series of 120 RR-intervals obtained from radar pulse waves. We determined the low-frequency (LF) components (0.04–0.15 Hz), which were linked to sympathetic modulation with some parasympathetic influence,

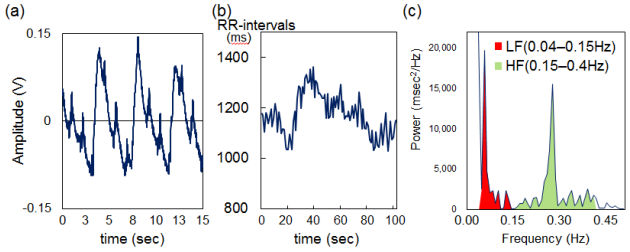


Figure 3. HRV calculation from radar pulse waves. (a) Radar output signal, (b) RR-interval series with spline interpolation at the sampling frequency of 100 Hz and (c) HRV frequency domain parameters LF and HF.

and the high-frequency (HF) components (0.15–0.4 Hz), which reflected parasympathetic activity (Fig. 3). HRV is highly sensitive to artifact and errors. To ensure accurate results, we therefore introduced the DSWW method to manage artifact and RR-interval errors appropriately prior to performing spectral analysis.

### *D. Body Movement Index and Standard deviations of RIs*

We determined a body movement index (BI) to identify wake and sleep periods [10]. As seen in Figure 4, BI, expressing the size of the body movements, was not influenced by the distance between the radar and the body.

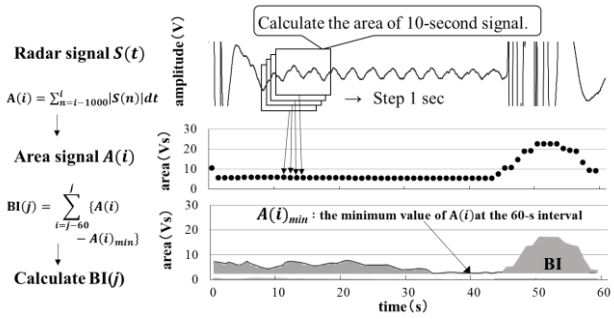


Figure 4. Body movement index (BI) calculation

We used the standard deviation of the RIs (SDR) as an RI fluctuation index in the time domain. Let the respiratory interval series derived from radar output signals, $R(i)$ $(i = 1,2,\cdots, N)$. The standard deviation σof RIs is defined as

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(R(i) - \tilde{R}\right)^2}, \qquad (5)$$

where $\tilde{R}$ is the average of the data-set $R(i)$. In general, SDR is expected to decrease with depth of NREM sleep [10].

## III. Results

While the average Pearson correlation coefficient between the HRs derived from radar overnight and the reference PSG was 88.3% (n = 10), the average correlation coefficient for HRV parameters (LF and HF) was lower at 71.2%. We selected six indices to classify sleep stages: LF, HF, LF/HF ratio, HR, BI, and SDR. The indices were newly derived each minute from radar output signals. First, we extracted the features for which the indices contributed to the identification of wake and sleep periods and to the classification of sleep stages. Then, we performed a canonical discriminant analysis that uses the four indices.

### *A. Identification of wake versus sleep by BI and HR*

Figure 5a is a scatter plot of boy movement index (BI) vs. the heart rates (HRs) from participant t2. Each plot of waking and sleep was defined by PSG. As the plot shows, BI can be used to identify wake and sleep states. Figure 5b shows the ROC curve. We obtained a BI cutoff value of 15.6 [Vs$^2$] for wake and sleep and an area under the curve (AUC) of 0.90. both BI and HR differed significantly between wake and sleep (n = 10, $p < 0.001$), but SDR did not (3 participants). The average AUC for BI was 0.81, while that for HR was 0.69. For identifying wake and sleep periods, the BI index reached a sensitivity of 76.0% and a specificity of 77.0% with 10

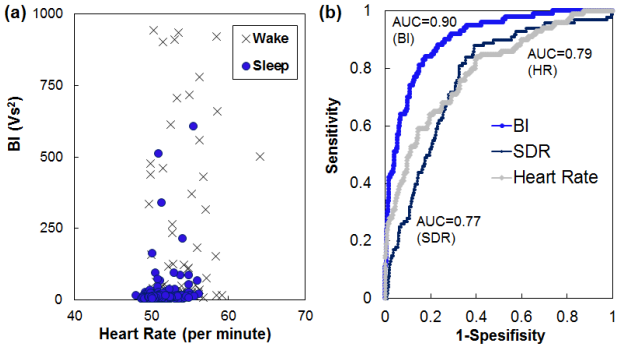


Figure 5. (a) Scatter plot of BI and HR. (b) ROC curve for discriminating subject's wake/sleep with BI, SDR and HR (Participant t2).

participants. The discrepancies occurred because people are sometimes motionless while awake. These results indicate that BI is a better identifier than HR or SDR for classifying wake and sleep.

### *B. Identification of REM, LIGHT, and DEEP by LF/HF*

Figure 6 is a scatter plot of the BI vs. LF/HF ratio from participant t2. As the plot shows, while LF/HF can be used to identify REM, LIGHT (NREM 1 and 2), and DEEP (NREM 3 and 4) sleep, identifying wake and sleep states using only LF/HF was difficult. Data were similar across the other 9 participants.

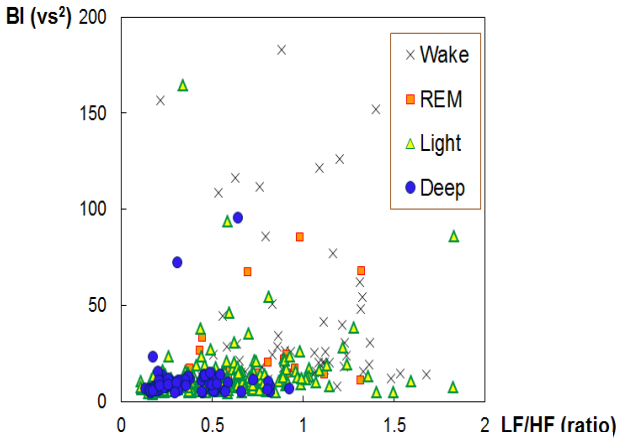


Figure 6. Scatter plot of BI and LF/HF ratio (Participant t2)

### *C. Analysis of variance for LF/HF and LF*

A one-way analysis of variance (ANOVA) was performed on the LF/HF ratio index for wake, REM, LIGHT, and DEEP stages. Post-hoc analysis using Bonferroni correction was performed as a follow-up. One-minute LF/HF had a high degree of contribution and showed a significant difference ($p < 0.01$) among the three sleep stages (REM, LIGHT, DEEP) (Fig 7). While LF showed a significant difference ($p < 0.01$) among the three sleep stages, HF did not ($p > 0.05$).

### *D. Canonical discriminant analysis*

By applying a canonical discriminant analysis (CDA) that separates the hyperspace created by the dependent parameters with hyperplanes, we were able to classify sleep stages by LF/HF ratio, LF, HF, and HRs derived from radar output signals. The highest linear discriminant function derived from CDA was

$$y = -0.60x_1 + 2.55x_2 + 1.07x_3 + 3.80x_4 - 2.11, \quad (6)$$

where $x_1$: HF, $x_2$: LF, $x_3$: LF/HF and $x_4$: HRs that were normalized to the range of 0–1. From equation (6), we know that HR contributed the most, followed by LF and LF/HF. The performance of the canonical discriminant analysis to identify wake and sleep periods and to classify the three sleep stages was evaluated in each leave-one-out (LOO) cross-validation loop. Table 1 shows the accuracy of this method compared with PSG. The accuracy of identifying wake and sleep was 66.4%. The accuracy of classifying three stages (wake, REM and NREM) was 57.1% and that of four stages (wake, REM, LIGHT and DEEP) was 34.0%.

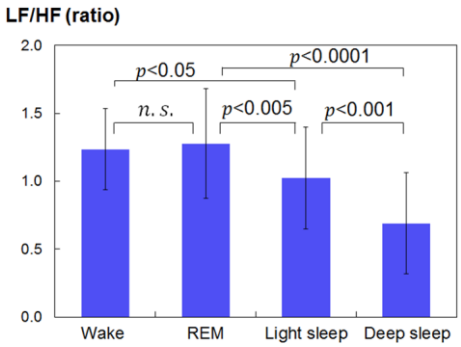


Figure 7. Analysis of variance and the Bonferroni test for LF/HF derived from radars.

TABLE 1 CONCORDANCE RATE BETWEEN RADAR AND POLYSOMNOGRAPHY

| Subject | Accuracy (%) | | |
|---|---|---|---|
| | 2 stages | 3 stages | 4 stages |
| t1 | 76.0 | 74.0 | 26.6 |
| t2 | 80.0 | 80.0 | **47.2** |
| t3 | 58.6 | 47.2 | 41.6 |
| t4 | 33.0 | 26.1 | 16.8 |
| t5 | 70.5 | 64.5 | **46.4** |
| t6 | 59.9 | 59.1 | 41.2 |
| t7 | 85.7 | 52.6 | 26.4 |
| t8 | 71.7 | 61.4 | 38.8 |
| t9 | 74.4 | 60.8 | 26.4 |
| t10 | 53.9 | 45.4 | 28.3 |
| Mean | 66.4 | 57.1 | 34.0 |

## IV. DISCUSSION

This is the first study that systematically compared HRV (LF, HF, and LF/HF), HRs, BI, and SDR derived from radar sensors in attempts to classify sleep stages [11 and 12]. Other methods using ballistocardiography, respiratory inductive plethysmography, or radar have been reported to reach an accuracy of 75 ± 10% for classification of three stages, which is better than our method [12]. The accuracy for our classification of four stages was less than 40% (Table 1). This might be a limitation of classification using only HRV indices with a CDA approach. Adding other parameters such as BI and SDR, and a hierarchical decision approach such as a decision tree will be necessary to solve this problem.

## V. CONCLUSION

We conducted an extensive study in which we used feature extraction to classify sleep stages. A system of two radars can be used to classify sleep stages and has the advantage of measuring body movements, body surface movements induced by respiration, and pulse waves with high sensitivity without needing bodily contact. Future work will investigate how a decision-tree system including BI and SDR indices can included to better classifying sleep stages. We think that the proposed method will enable easy measurement of sleep stages at nursing care facilities and in the general public, and can be employed to improve the quality of sleep.

## ACKNOWLEDGMENT

The authors would like to thank Dr. H. Tojima at Tokyo Rosai Hospital, for his generous support of this research and clinical advice.

## REFERENCES

[1] T. Willemen, D. Van Deun, V. Verhaert, M. Vandekerckhove, V. Exadaktylos, J. Verbraecken, S. Van Huffel, B. Haex, J. Vander Sloten, "An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification," *IEEE J Biomed Health Inform*, vol. 18, no. 2, pp. 661–669, 2014.

[2] T. Penzel, J.W. Kantelhardt, L.Grote, J.H. Peter, and A. Bunde, "Comparison of Detrended Fluctuation Analysis and Spectral Analysis for Heart Rate Variability in Sleep and Sleep Apnea," *IEEE Trans Biomed Eng*, vol. 50(10), pp. 1143–1151, 2003.

[3] AY. Schumann, RP. Bartsch, T. Penzel,PCh Ivanov, JW. Kantelhardt, "Aging effects on cardiac and respiratory dynamics in healthy subjects across sleep stages," *Sleep*, 33(7), pp. 943–955, Jul 2010.

[4] M. Kagawa, Y. Yoshida, M. Kubota, A. Kurita, and T. Matsui, "Non-contact heart rate monitoring method for elderly people in bed with random body motions using 24 GHz dual radars located beneath the mattress in clinical settings," *Journal of Medical Engineering & Technology*, 36(7-8), pp. 344–350, 2012.

[5] M. Peltokangas, J. Verho, A. Vehkaoja, "Night-time EKG and HRV monitoring with bed sheet integrated textile electrodes," *IEEE Trans Inf Technol Biomed*, 16(5):935-42, September 2012.

[6] A. Rechtschaffen and A. Kales, A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects. Washington, DC: US Public Health Service, US Government Printing Office, 1968.

[7] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. F. Quan, The AASM Manual for the Scoring of Sleep and Associated Events. Westchester,IL,USA: American Academy of Sleep Medicine, 2007.

[8] D. Moser, P. Anderer, G. Gruber, et al., "Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters," *Sleep*, 32(2), pp.139–149. Feb 2009.

[9] J. Pan, WJ. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transaction on Biomedical Engineering*, 32(3), 230–236, March 1985.

[10] M. Kagawa, N. Sasaki, K. Suzumura and T. Matsui, "Sleep stage classification by body movement index and respiratory interval indices using multiple radar sensors," in *IEEE EMBC*, Milan, August 2015, pp. 7606–7609.

[11] PI. Terrill, SJ. Wilson, S. Suresh, DM. Cooper, C. Dakin, "Attractor structure discriminates sleep states: recurrence plot analysis applied to infant breathing patterns," *IEEE Trans Biomed Eng*, 57(5), 1108–1116, May 2010.

[12] A. Tataraidze and L. Anishchenko, "Sleep stage classification based on bioradiolocation signals," in *IEEE EMBC*, Milan, August 2015, pp. 362–365.