Original Article

Prediction of Obstructive Sleep Apnea Based on Respiratory Sounds Recorded Between Sleep Onset and Sleep Offset

Jeong-Whun Kim[1] · Taehoon Kim[2] · Jaeyoung Shin[2] · Goun Choe[1] · Hyun Jung Lim[1] · Chae-Seo Rhee[1] · Kyogu Lee[2]
Sung-Woo Cho[1]

[1]*Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Bundang Hospital,*
Seoul National University College of Medicine, Seongnam;

[2]*Music and Audio Research Group, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Korea*

Objectives. To develop a simple algorithm for prescreening of obstructive sleep apnea (OSA) on the basis of respiratory sounds recorded during polysomnography during all sleep stages between sleep onset and offset.

Methods. Patients who underwent attended, in-laboratory, full-night polysomnography were included. For all patients, audio recordings were performed with an air-conduction microphone during polysomnography. Analyses included all sleep stages (i.e., N1, N2, N3, rapid eye movement, and waking). After noise reduction preprocessing, data were segmented into 5-s windows and sound features were extracted. Prediction models were established and validated with 10-fold cross-validation by using simple logistic regression. Binary classifications were separately conducted for three different threshold criteria at apnea hypopnea index (AHI) of 5, 15, or 30. Prediction model characteristics, including accuracy, sensitivity, specificity, positive predictive value (precision), negative predictive value, and area under the curve (AUC) of the receiver operating characteristic were computed.

Results. A total of 116 subjects were included; their mean age, body mass index, and AHI were 50.4 years, 25.5 kg/m^2, and 23.0/hr, respectively. A total of 508 sound features were extracted from respiratory sounds recorded throughout sleep. Accuracies of binary classifiers at AHIs of 5, 15, and 30 were 82.7%, 84.4%, and 85.3%, respectively. Prediction performances for the classifiers at AHIs of 5, 15, and 30 were AUC, 0.83, 0.901, and 0.91; sensitivity, 87.5%, 81.6%, and 60%; and specificity, 67.8%, 87.5%, and 94.1%. Respective precision values of the classifiers were 89.5%, 87.5%, and 78.2% for AHIs of 5, 15, and 30.

Conclusion. This study showed that our binary classifier predicted patients with AHI of ≥15 with sensitivity and specificity of >80% by using respiratory sounds during sleep. Since our prediction model included all sleep stage data, algorithms based on respiratory sounds may have a high value for prescreening OSA with mobile devices.

Keywords. Obstructive Sleep Apnea; Respiratory Sounds; Polysomnography; Machine Learning

INTRODUCTION

Obstructive sleep apnea (OSA) is closely related to important medical conditions such as hypertension, cardiovascular diseases, neurovascular diseases, and metabolic syndromes [1]. In addition, the prevalence of OSA reaches approximately 26% in middle-aged to older adults [2] and is increasing with the rate of obesity in the general population. However, OSA is frequently underrecognized and underdiagnosed, possibly because of the limited accessibility of diagnostic tests due to high costs and in-

sufficient test facilities for sleep studies.

The gold standard examination is attended, in-laboratory, full-night polysomnography with multichannel monitoring. Although it provides an exact set of sleep data including apnea hypopnea index (AHI), many subjects at risk of OSA cannot afford polysomnography. Thus, various portable or at-home sleep test devices have been developed and are currently in use. However, current tests with portable devices also incur high expenses when performed repeatedly [3].

Recently, multiple smartphone applications have been developed to prescreen snoring or OSA [4,5]. They can be easily downloaded and used in open application online markets. During sleep, various signals are continuously produced by the body that may be used for prediction of OSA severity; these signals include respiratory sounds, such as silent or loud breathing sounds, regular or irregular breathing sounds, snoring, gasping, and cessation of breathing sound. Respiratory sounds during sleep can be easily recorded by using microphones embedded in most currently available smartphones. However, their performance and accuracy have rarely been compared and tested through simultaneous studies with polysomnography.

In the present study, we extracted a large volume of audio features from sleep breathing sounds recorded during full-night polysomnography, in which respiratory sound signals were perfectly synchronized with other body signals in the polysomnographic data set. The current study was performed to develop algorithms for prescreening of OSA with a large set of audio features and evaluate their performances in comparison with the results based on polysomnography.

MATERIALS AND METHODS

Study participants and polysomnography

Patients with habitual snoring, with or without witnessed apnea, who underwent attended, in-laboratory, full-night polysomnography at a sleep center of a tertiary hospital between October 2013 and March 2014, were included in this study. Patients were excluded if they had central sleep apnea, neurological disorders, neuromuscular disease, heart failure, or any other critical medical condition. Polysomnography (Embla N 7000, Reykjavik, Ice-

land) included electroencephalography, electrooculography, chin and limb electromyography, electrocardiography, nasal pressure transducer, thermistor, chest and abdomen respiratory inductance plethysmography, and pulse oximetry. OSA and hypopnea were defined as previously described [6]: apnea was defined as cessation of airflow for at least 10 seconds, while hypopnea was defined as a >50% decrease in airflow for at least 10 seconds or a moderate reduction in airflow for at least 10 seconds associated with arousals or oxygen desaturation (<4%) [7]. AHI was defined as the total number of apneas and hypopneas per sleep hour. Written informed consent was obtained before enrollment and the study protocol was approved by the Institutional Review Board of Seoul National University Bundang Hospital (IRB No. B-1404/248-109).

Collection and preprocessing of sleep breathing sounds

For all patients, audio recordings were performed throughout the night of polysomnography by using an air-conduction microphone (SURP-102; Yianda Electronics, Shenzhen, China) linked to a video recorder, which was placed 1.7 m above the patient's bed, near the ceiling (Fig. 1). Respiratory sounds recorded
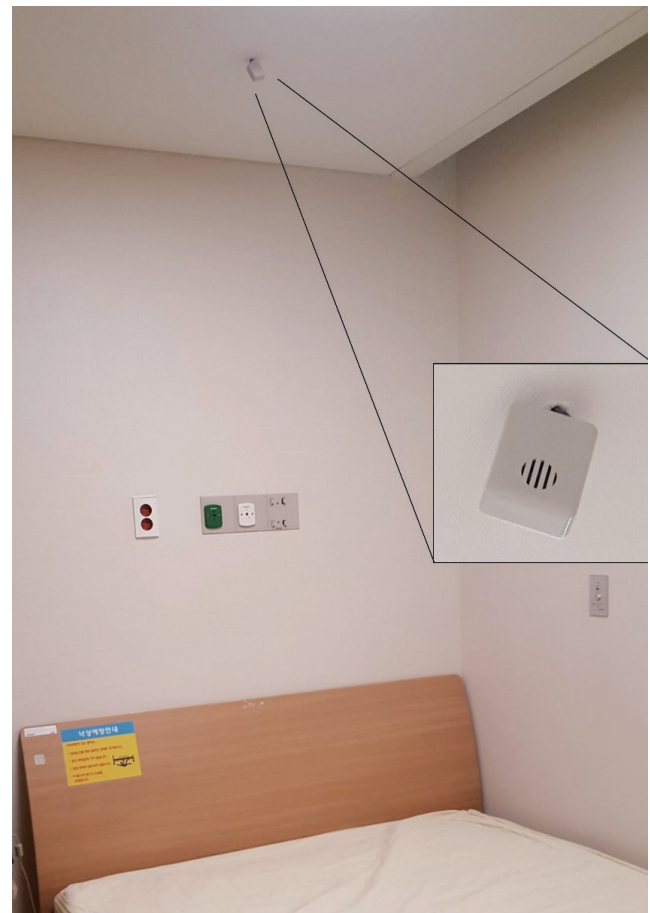
Fig. 1. A bed for polysomnography and a microphone (inset) on the ceiling.

HIGHLIGHTS

- Analyses included respiratory sounds from all sleep stages, including N1, N2, N3, rapid eye movement, and waking.
- We analyzed quite many patient sounds simultaneously recorded with polysomnography.
- Our binary classifier predicted patients with apnea hypopnea index of ≥15 with sensitivity and specificity of >80% by using respiratory sounds during sleep.

throughout the sleep time were used for analyses. Analyses were performed by using audio data from all sleep stages comprising stage N1 sleep, stage N2 sleep, stage N3 sleep, rapid eye movement sleep, and waking, from sleep onset to sleep offset. Audio data from each patient were converted into a wave format file with an 8-kHz sampling frequency by using FFmpeg, which is a free software for handling multimedia data [8]. Noise reduction was conducted for preprocessing with a spectral subtraction method [8].

Extraction of audio features

To extract audio features that describe the characteristics of respiratory sounds and can be used for prediction of OSA, full-length, de-noised audio data of each patient were segmented into 5-s window signals and a large set of audio features was extracted from those 5-s window signals by using jAudio, a Java-based audio feature extraction software. The audio feature extraction framework is summarized in Fig. 2.

Prediction of OSA

Prediction of OSA, based on breathing sounds during sleep time, was performed with a simple logistic regression model. To establish and validate prediction models, or classifiers, 10-fold cross-validation was used. Enrolled patients were randomly divided into 10 equal-sized subgroups. Of the 10 subgroups, a single subgroup was retained for validation of the prediction model; the remaining nine subgroups were used for training. The cross-validation process was then repeated 10 times (10 folds), with each of the 10 subgroups used once for validation. This gives 10 evaluation results, which are averaged. Then learning algorithm was then invoked a final (11th) time on the entire dataset to obtain the final model [9]. Model performance measures, such as accuracy, sensitivity, specificity, positive predictive value (PPV, precision), negative predictive value (NPV), and area under the curve (AUC) of the receiver operating characteristic, were computed for each fold. Binary classifications were conducted for three different threshold criteria at AHI of 5, 15, or 30; prediction models (classifiers) were separately established for each threshold. Machine learning was performed with a free software, Weka [9]. Other statistical analyses were performed by using IBM SPSS ver. 22.0 (IBM Corp., Armonk, NY, USA). Results are presented as mean±standard deviation.
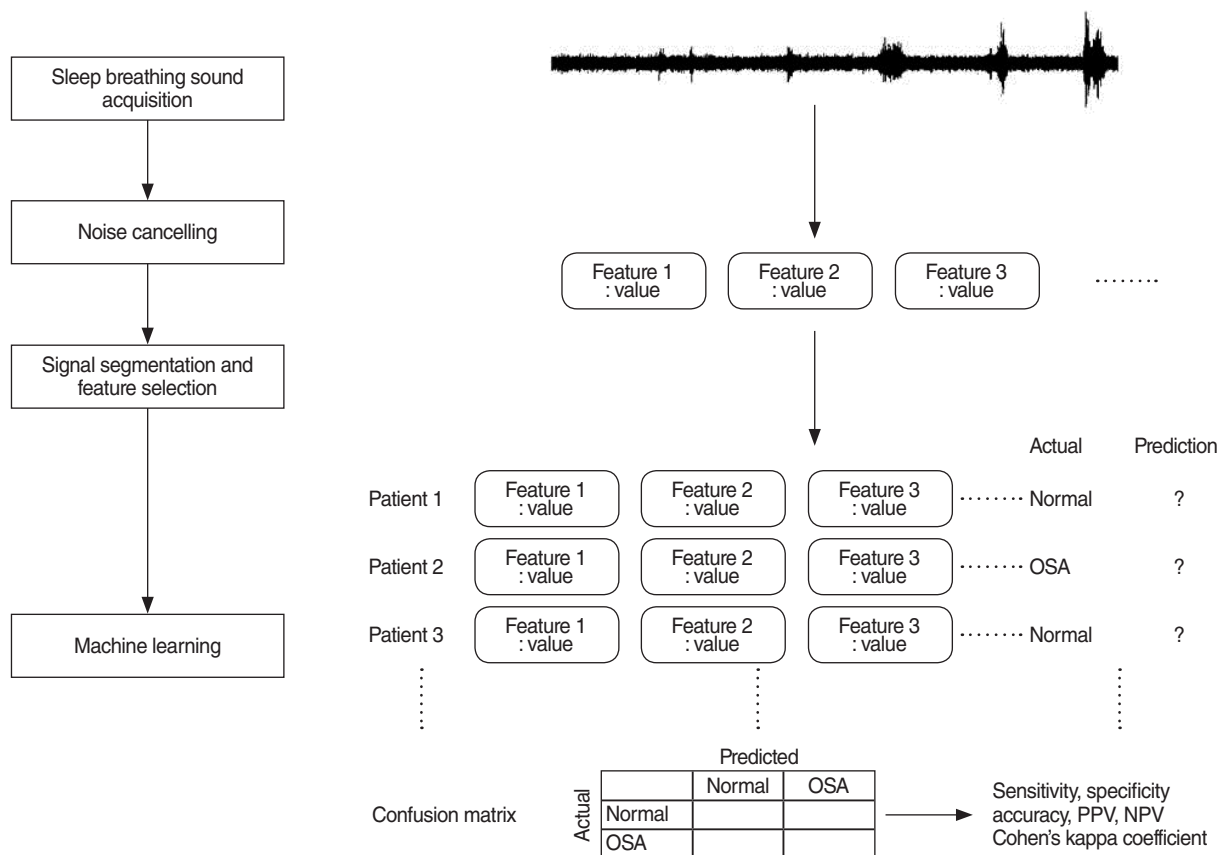

Fig. 2. Study framework. Sound data were acquired, followed by noise cancelling, and feature selection. From these inputs and with labeled result from the polysomnography of the same patient, machine learning had been performed. OSA, obstructive sleep apnea; PPV, positive predictive value; NPV, negative predictive value.

Table 1. General and polysomnographic characteristics

Variable	AHI<5 (n=28)	5≤AHI<15 (n=28)	15≤AHI<30 (n=30)	AHI ≥30 (n=30)
Age (yr)	43.2	54.0	53.9	50.3
Male:female	10:18	18:10	24:6	26:4
Body mass index (kg/m^2)	23.1	24.6	26.9	27.3
AHI (/hr)	1.1	8.9	22.0	57.5

AHI, apnea hypopnea index.

Table 2. Features extracted from respiratory sounds during sleep

Name of feature	No. of derived features
Beat histogram	172
Area method of moments & derivative	100
Mel frequency cepstral coefficient & derivative	52
Linear predictive coding & derivative	40
Area method of moments of constant Q-based Mel frequency cepstral coefficients	20
Area method of moments of log of constant Q transform	20
Area method of moments of Mel frequency cepstral coefficients	20
Method of moments & derivative	20
Beat sum & derivative	4
Compactness & derivative	4
Fraction of low energy windows & derivative	4
Peak-based spectral smoothness & derivative	4
Relative difference function & derivative	4
Root mean square & derivative	4
Spectral centroid & derivative	4
Spectral flux & derivative	4
Spectral rolloff point & derivative	4
Spectral variability & derivative	4
Strength of strongest beat & derivative	4
Strongest beat & derivative	4
Strongest frequency via fast fourier transform maximum & derivative	4
Strongest frequency via spectral centroid & derivative	4
Strongest frequency via zero crossings & derivative	4
Zero crossings & derivative	4
Total number of features	508

RESULTS

General characteristics and polysomnographic findings

A total of 116 subjects (78 men and 38 women) with a mean age of 50.4±16.7 years were included in the present study. Their mean body mass index was 25.5±3.9 kg/m^2, and the mean AHI was 23.0±24.0/hr. The numbers of subjects with AHI <5, 5≤ AHI <15, 15≤ AHI <30, and AHI ≥30 were 28, 28, 30, and 30, respectively. Their characteristics, according to OSA severity, are summarized in Table 1. The mean total sleep time between sleep onset and offset was 369.7±104.8 minutes.

Characteristics of sound-associated features

The mean number of 5-s windows from the respiratory sounds recorded during total sleep time was 4,436.8±1,258.4 per patient. A total of 508 audio features were extracted from those windows using the jAudio software. The representative features were beat histogram, area method of moments, Mel frequency cepstral coefficient (MFCC), linear predictive coding, area method of moments of constant Q-based MFCCs, area method of moments of log of constant Q transform, area method of moments of MFCCs, and method of moments (Table 2).

Performance of binary classifiers for OSA

When the AHI threshold for binary classification was defined as 5, 15, and 30, the numbers of subjects whose OSA severity was accurately predicted was 96 (82.7%; Cohen's kappa coefficient (κ)=0.54, 95% confidence interval [CI], 0.36 to 0.72), 98 (84.4%; κ=0.69; 95% CI, 0.51 to 0.87), and 99 (85.3%; κ=0.59; 95% CI, 0.41 to 0.77), respectively, out of 116 subjects.

When the AHI threshold criterion was 5, 88 subjects had mild to severe OSA with AHI ≥5. The sensitivity and specificity of the prediction model (classifier) at AHI of 5 were 87.5% (95% CI, 78.3% to 93.3%) and 67.8% (95% CI, 47.6% to 83.4%), respectively. The PPV (precision) and NPV were 89.5% (95% CI, 80.6% to 94.8%) and 63.3% (95% CI, 46.9% to 79.5%), respectively. When the AHI threshold was 15, 60 subjects had moderate to severe OSA with AHI ≥15, based on polysomnography. The sensitivity and specificity of the binary classifier at AHI of 15 were 81.6% (95% CI, 69.1% to 90.1%) and 87.5% (95% CI, 75.3% to 94.4%), respectively. The PPV and NPV were 87.5% (95% CI, 75.3% to 94.4%) and 81.6% (95% CI,

69.1% to 90.1%), respectively. When the AHI threshold was 30, 30 patients had polysomnographic severe OSA with AHI ≥30. The sensitivity and specificity of the binary classifier at AHI of 30 was 60% (95% CI, 40.7% to 76.8%) and 94.1% (95% CI, 86.3% to 97.8%), respectively. The PPV and NPV were 78.2% (95% CI, 55.8% to 91.7%) and 87% (95% CI, 78.2% to 92.9%), respectively. Average AUCs were 0.83, 0.901, and 0.91, when the prediction models of binary classifications were tested for different cut-off values at AHI of 5, 15, and 30, respectively. The results are summarized in Fig. 3.

DISCUSSION

The purpose of this study was to analyze all respiratory sounds occurring between sleep onset and offset based on polysomnography and then predict OSA by using respiratory sounds from patients during sleep time. More than 500 variables, or features, that were extracted from the breathing sounds were used to develop a classifier (i.e., an OSA-predicting machine learning algorithm).

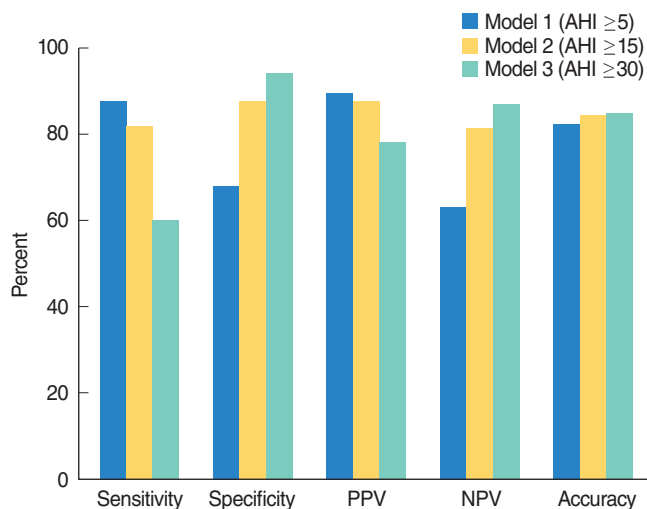We evaluated binary classifiers that were trained to classify

Fig. 3. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of binary classifiers at apnea hypopnea index (AHI) of 5, 15, and 30 for prescreening of obstructive sleep apnea.

patients into two groups. The binary classifier for AHI of 5 had the lowest specificity, while the classifier for AHI of 30 had the lowest sensitivity and lowest precision (PPV); accuracies of the three classifiers were similar. Cohen's kappa coefficient, which measures interrater agreement for categorical values, was highest when the cutoff value of AHI was set at 15. Therefore, the best performance of the OSA prediction algorithm with recorded respiratory sounds during sleeping time was observed when AHI of 15 was the classification cutoff.

Thus far, few studies have been performed with sounds that were generated during sleep, especially involving machine learning techniques; most studies have been conducted with a small number of patients [10-13]. Previous studies mainly focused on analyzing the characteristics of snoring, or distinguishing snoring from non-snoring sounds. One study reported that the classification performance of human visual scorers and a machine learning algorithm was similar in terms of automatically identifying snoring; this study used a support vector machine, which is a machine learning technique [14]. There have also been studies that used analysis of snoring sounds to predict the anatomical location in the upper airway where the snoring sounds were generated [15-17]. However, analyses in these studies were limited to only a short period of time, drug-induced sleep. Another study used neural network analysis to separate snoring from non-snoring segments; this study also analyzed only breathing sounds from the trachea, and was performed in very few patients (<10) [18]. Another study of the trachea respiratory sound was performed in 147 awake patients; it reported that prediction based on breathing sounds was superior to that based on anthropometric features in predicting polysomnographic AHI of ≥10 [19]. However, it differs from our study in that sounds were recorded when the patients were awake. A further study

attempted to distinguish apneas from hypopneas by using audio signals. When all 2,015 apneas or hypopneas were analyzed, the accuracy was approximately 84.7%, which suggests the possibility of estimating apnea or hypopnea separately. However, considering that automatic detection of apneas and hypopneas has not been validated to estimate AHI as a measurement of OSA severity extracted by human scorers, further studies are needed for automation [20]. We have previously conducted studies to predict the severity of OSA by analyzing breathing sounds during sleep with machine and deep learning methods [8,21]. Although our previous studies used respiratory sounds during sleeping time, they were limited in that only N2 and N3 stages were included for analysis. Therefore, the present study is distinguished from our previous studies in that we tested prediction algorithm models in all sleep stages, including N1, N2, N3, rapid eye movement, and waking stages, between sleep onset and offset. Classifiers should be as simple and accurate as possible in predicting OSA. For this purpose, it is important to use the sounds of all sleep stages without any identification and/or separation of sleep stages.

The development of OSA prediction algorithms may be important in many respects. First, if predictive models are proven to have a certain level of performance, more patients will have the opportunity to prescreen their own OSA. In practice, considering the limitations of conventional polysomnography, many patients with OSA are underdiagnosed, despite its high prevalence (26% in middle-aged men) [22]. Individuals with repetitive snoring, daytime sleepiness, or poor sleep quality without definite OSA must now travel to the hospital for polysomnography. Regarding ambulatory home monitoring devices, there is a limitation (similar to polysomnography) in that symptomatic patients must first visit hospitals equipped with the appropriate devices. Second, the prediction algorithm tested in our study is an important contribution because most people currently use smartphones. Notably, smartphones have a sound recording function, which may enable prescreening of OSA [23]. Third, this type of prediction algorithm can be used for repetitive tests without economic or spatial burden. Therefore, given the night-to-night variability of OSA [24], it is useful to evaluate the average sleep state for an extended period of time, and to regularly monitor the effects of operation or intraoral devices. Fourth, this research may motivate more OSA patients to visit hospitals, thereby increasing the likelihood that they will undergo polysomnography and start treatment at the appropriate time. Since the risk of multi-organ diseases, such as cardiovascular, neurovascular, and metabolic diseases, may be reduced by early treatment of OSA, their social and economic burdens may decrease [25].

Our research also has several limitations. To be utilized in personal mobile devices, it is necessary to use the sound recorded from smartphones. While we have been able to synchronize polysomnography to determine sleep onset and offset, these

may be incorrectly determined when measured at home with a smartphone. Since the algorithm that we tested was a binary classifier, we could not distinguish the degree of OSA severity in detail. Overall performance was acceptable; however, it should be further increased. There may be several reasons for this, including inter-person variability of breathing sounds during sleep. The performance of the classifier may be improved by including more patients in future analyses. In addition, it is possible to develop algorithms that can predict apnea or hypopnea, as well as the severity of OSA. This study was based on a single session of polysomnography in each patient; therefore, it is also necessary to evaluate night-to-night variability of breathing sounds during sleep. Further, sound acquisition from the microphones of commercial smartphones, placed at the bedside, which is closer to real-world usage, must be also validated.

In conclusion, our binary classifier predicted patients with AHI of ≥15 with sensitivity and specificity of >80% by using respiratory sounds during sleep. Our study has strength in that it used data recorded in synchronization with polysomnography during actual sleep time. In addition, since our prediction model included all sleep stage data between sleep onset and offset, prediction classifiers based on respiratory sounds may have high future value as prescreening algorithms for OSA that can be used in personal mobile devices. Although our study used a large volume of patient data to diagnose OSA based on respiratory sounds, further studies with additional patients are needed to improve the performance of prediction algorithms.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

REFERENCES

1. Xie C, Zhu R, Tian Y, Wang K. Association of obstructive sleep apnoea with the risk of vascular outcomes and all-cause mortality: a meta-analysis. BMJ Open. 2017 Dec;7(12):e013983.
2. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. Am J Epidemiol. 2013 May;177(9):1006-14.
3. Ahmed M, Patel NP, Rosen I. Portable monitors in the diagnosis of obstructive sleep apnea. Chest. 2007 Nov;132(5):1672-7.
4. Ong AA, Gillespie MB. Overview of smartphone applications for sleep analysis. World J Otorhinolaryngol Head Neck Surg. 2016 Mar;2(1):45-9.
5. Camacho M, Robertson M, Abdullatif J, Certal V, Kram YA, Ruoff CM, et al. Smartphone apps for snoring. J Laryngol Otol. 2015 Oct; 129(10):974-9.
6. Jung HJ, Wee JH, Rhee CS, Kim JW. Full-night measurement of level of obstruction in sleep apnea utilizing continuous manometry. Laryngoscope. 2017 Dec;127(12):2897-902.
7. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep. 1999 Aug;22(5):667-89.
8. Kim J, Kim T, Lee D, Kim JW, Lee K. Exploiting temporal and nonstationary features in breathing sound analysis for multiple obstructive sleep apnea severity classification. Biomed Eng Online. 2017 Jan;16(1):6.
9. Frank E, Hall MA, Witten IH. The Weka Workbench. Online appendix for "data mining: practical machine learning tools and techniques". 4th ed. Burlington (NJ): Morgan Kaufmann; 2016.
10. Emoto T, Kashihara M, Abeyratne UR, Kawata I, Jinnouchi O, Akutagawa M, et al. Signal shape feature for automatic snore and breathing sounds classification. Physiol Meas. 2014 Dec;35(12):2489-99.
11. Sun X, Kim JY, Won Y, Kim JJ, Kim KA. Efficient snoring and breathing detection based on sub-band spectral statistics. Biomed Mater Eng. 2015;26 Suppl 1:S787-93.
12. Nguyen TL, Won Y. Sleep snoring detection using multi-layer neural networks. Biomed Mater Eng. 2015;26 Suppl 1:S1749-55.
13. Janott C, Schuller B, Heiser C. Acoustic information in snoring noises. HNO. 2017 Feb;65(2):107-16.
14. Samuelsson LB, Rangarajan AA, Shimada K, Krafty RT, Buysse DJ, Strollo PJ, et al. Support vector machines for automated snoring detection: proof-of-concept. Sleep Breath. 2017 Mar;21(1):119-33.
15. Koo SK, Kwon SB, Kim YJ, Moon JI, Kim YJ, Jung SH. Acoustic analysis of snoring sounds recorded with a smartphone according to obstruction site in OSAS patients. Eur Arch Otorhinolaryngol. 2017 Mar;274(3):1735-40.
16. Qian K, Janott C, Pandit V, Zhang Z, Heiser C, Hohenhorst W, et al. Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis. IEEE Trans Biomed Eng. 2017 Aug;64(8):1731-41.
17. Janott C, Schmitt M, Zhang Y, Qian K, Pandit V, Zhang Z, et al. Snoring classified: The Munich-Passau Snore Sound Corpus. Comput Biol Med. 2018 Mar;94:106-18.
18. Shokrollahi M, Saha S, Hadi P, Rudzicz F, Yadollahi A. Snoring sound classification from respiratory signal. Conf Proc IEEE Eng Med Biol Soc. 2016 Aug;2016:3215-8.
19. Elwali A, Moussavi Z. Obstructive sleep apnea screening and airway structure characterization during wakefulness using tracheal breathing sounds. Ann Biomed Eng. 2017 Mar;45(3):839-50.
20. Halevi M, Dafna E, Tarasiuk A, Zigel Y. Can we discriminate between apnea and hypopnea using audio signals? Conf Proc IEEE Eng Med Biol Soc. 2016 Aug;2016:3211-4.
21. Kim T, Kim JW, Lee K. Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques. Biomed Eng Online. 2018 Feb;17(1):16.
22. Erdenebayar U, Park JU, Jeong P, Lee KJ. Obstructive sleep apnea screening using a piezo-electric sensor. J Korean Med Sci. 2017 Jun; 32(6):893-9.
23. Saeb S, Cybulski TR, Schueller SM, Kording KP, Mohr DC. Scalable passive sleep monitoring using mobile phones: opportunities and obstacles. J Med Internet Res. 2017 Apr;19(4):e118.
24. White LH, Lyons OD, Yadollahi A, Ryan CM, Bradley TD. Night-to-

night variability in obstructive sleep apnea severity: relationship to overnight rostral fluid shift. J Clin Sleep Med. 2015 Jan;11(2):149-56.

25. Deacon NL, Jen R, Li Y, Malhotra A. Treatment of obstructive sleep apnea: prospects for personalized combined modality therapy. Ann Am Thorac Soc. 2016 Jan;13(1):101-8.