https://doi.org/10.5664/jcsm.9094

JCSM | Journal of Clinical Sleep Medicine

SCIENTIFIC INVESTIGATIONS

Automatic classification of excitation location of snoring sounds

Jingpeng Sun, PhD[1,2,3]; Xiyuan Hu, PhD[4]; Silong Peng, PhD[1,2]; Chung-Kang Peng, PhD[3]; Yan Ma, MD, MPH[3]

[1]*Institute of Automation, Chinese Academy of Sciences, Beijing, People's Republic of China;* [2]*University of Chinese Academy of Sciences, Beijing, People's Republic of China;* [3]*Division of Interdisciplinary Medicine and Biotechnology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts;* [4]*School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, People's Republic of China*

Study Objectives: For surgical treatment of patients with obstructive sleep apnea-hypopnea syndrome, it is crucial to locate accurately the obstructive sites in the upper airway; however, noninvasive methods for locating the obstructive sites have not been well explored. Snoring, as the cardinal symptom of obstructive sleep apnea-hypopnea syndrome, should contain information that reflects the state of the upper airway. Through the classification of snores produced at four different locations, this study aimed to test the hypothesis that snores generated by various obstructive sites differ.

Methods: We trained and tested our model on a public data set that comprised 219 participants. For each snore episode, an acoustic and a physiological feature were extracted and concatenated, forming a 59-dimensional fusion feature. A principal component analysis and a support machine vector were used for dimensional reduction and snore classification. The performance of the proposed model was evaluated using several metrics: sensitivity, precision, specificity, area under the receiver operating characteristic curve, and *F1* score.

Results: The unweighted average values of sensitivity, precision, specificity, area under the curve, and *F1* were 86.36%, 89.09%, 96.4%, 87.9%, and 87.63%, respectively. The model achieved 98.04%, 80.56%, 72.73%, and 94.12% sensitivity for types V (velum), O (oropharyngeal), T (tongue), and E (epiglottis) snores.

Conclusions: The characteristics of snores are related to the state of the upper airway. The machine-learning–based model can be used to locate the vibration sites in the upper airway.

Keywords: machine learning; multiscale entropy; snore classification; obstructive sleep apnea hypopnea syndrome

Citation: Sun J, Hu X, Peng S, Peng C-K, Ma Y. Automatic classification of excitation location of snoring sounds. *J Clin Sleep Med*. 2021;17(5):1031–1038.

BRIEF SUMMARY

Current Knowledge/Study Rationale: Snores, produced by the vibration of collapsed structures of the upper airway during sleep, can be used to identify sites of snore generation. We aimed to explore distinct acoustic and physiological characteristics of different snore types and develop a machine-learning based approach to enable automatic classification of snoring sounds for identification of the obstruction site.

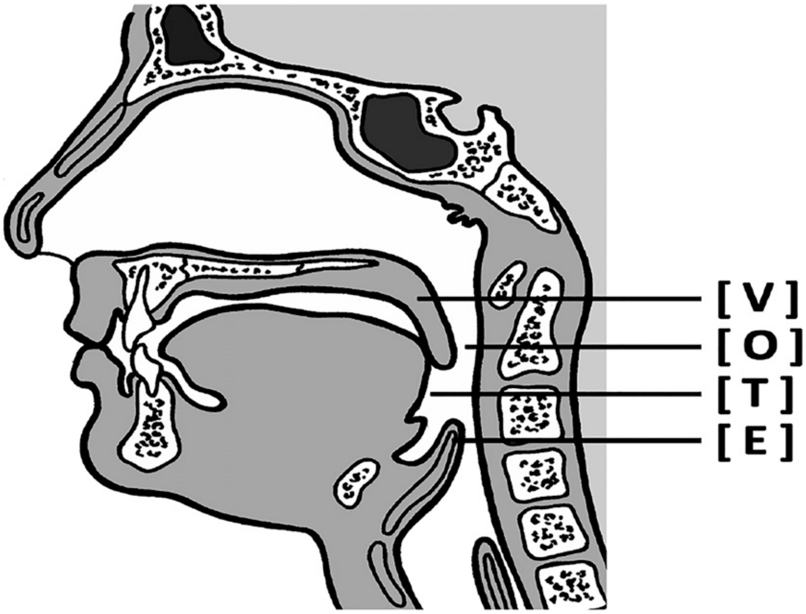
Study Impact: This is the first study to investigate both the acoustic characteristics of snores and their physiological characteristics. We found that snores generated by different sites have unique features, and snore sounds contain information that can be used to locate the vibration sites in the upper airway. Such approaches may enable automatic classification of snore types and assist in treatment decisions for patients with obstructive sleep apnea-hypopnea syndrome.

INTRODUCTION

Snoring is a respiratory noise produced by the vibration of collapsed structures of the upper airway during the inspiratory phase of breathing. The reported prevalence of snoring is 50% in adults and 3.2%–12.1% in children.[1–3] It is considered the earliest symptom of obstructive sleep apnea-hypopnea syndrome (OSAHS),[4] and it is closely related to complications associated with OSAHS, such as diabetes and cardiovascular diseases.[5] Continuous positive airway pressure (CPAP) is the most common treatment for OSAHS, but nonadherence to this treatment is a significant problem.[6] A common non-CPAP medical intervention is surgery, especially for patients with mild or moderate OSAHS.[7] It is noteworthy that many surgical options are available for different obstruction locations; for example, if the obstruction is at the palatal wall, the treatment can be an expanding operation on the side palatal wall, whereas if the obstruction is limited to the oropharynx/palate, uvulopalatopharyngoplasty is a better choice.[8,9] The low success rate of surgery makes this treatment controversial.[10,11] Careful patient selection is required in clinical practice based on the sites of obstruction in the upper airway, and knowledge of the precise location can improve the success rate of surgical treatment.[12,13] Therefore, precise assessment of the obstructive location in the upper airway is of the utmost importance, but the method of determination is still imprecise.

The current, broadly used approach to locate upper airway obstruction is drug-induced sleep endoscopy (DISE), which requires certified surgeons to perform endoscopy during a pharmacologically induced sleep. Among several descriptive classification systems for DISE results, the velum-oropharyngeal-tongue-epiglottis (VOTE) system classifies upper airway obstructions by using anatomical structures, as shown in **Figure 1**. It is widely used because of its simplicity and accuracy,[14] but it has some major limitations. First, it could potentially lead to artificial additional sites owing to oversensitivity of the

Figure 1—Diagram of the VOTE scheme in the upper airway.



In this figure, V, O, T, and E refer to four locations of the upper airway: velum (velum area), oropharyngeal (palatine tonsils and the lateral pharyngeal wall tissues), tongue (tongue base), and epiglottis. Reprinted, with permission, from Qian K et al. Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis. *IEEE Transactions on Biomedical Engineering*. 2017;64(8):1731–1741. https://doi.org/10.1109/TBME.2016.2619675.

observation, prolonged examination, or misunderstanding of the correlation between the DISE image and considerably decreased airflow.[7] Second, the procedure is performed during drug-induced sleep, and it is questionable whether the architecture of artificial sleep is comparable to that of natural sleep.[12] Third, the procedure requires special rooms, equipment, and trained personnel.[15]

There has been increasing demand for alternative approaches that are free of these limitations to complement or replace DISE for the identification of obstruction locations. Snoring, as a product of the vibration location in the upper airway, carries vital information about the condition of the upper airway; such information is simple to acquire and analyze during natural sleep.[16] Different vibration locations are correlated with the distinct acoustic characteristics of the snoring.[17]

Some attempts have been made using artificial neural networks to classify VOTE sites based on snoring[18–20]; however, these methods did not consider the physiological information contained in the snoring, and accuracy was thus not high. Therefore, the primary purpose of this study was to develop an enhanced algorithm to improve the accuracy of locating an obstructive site by capturing the information of the state embedded in snoring.

METHODS

Database

We used the Munich Passau Snore Sound Corpus (MPSSC) in this study.[17] Snores were derived from endoscopic recordings of DISE examinations and labeled as V, O, T, and E, corresponding to four locations of the upper airway: velum (velum area), oropharyngeal (palatine tonsils and the lateral pharyngeal wall tissues), tongue (tongue base), and epiglottis. This is a sample-oriented rather than individual-oriented database, which means that snores from the same individual can be belong to different classes. The snoring sounds were recorded at a sample rate of 44,100 Hz with different equipment (nasopharyngoscope, recording system, cell phone) at three clinical centers from 2006 to 2015.

Participants with snores originating from more than one location and samples that were corrupted by noise were excluded. To diversify the samples, a maximum of six snoring episodes of the same class were included per participant. The database comprised 219 participants whose average age was 49.8 (range = 24–78) years; male participants predominated (93.6%). A total of 828 snore episodes were captured with an average duration of 1.46 (0.73–2.75) seconds per episode. The number of episodes and average duration per category are shown in **Table 1**.

To avoid the bias of dividing the sample into a training set and a test set, given the imbalance of the sample, a hierarchical cross-validation strategy was used for training and testing. Specifically, 80% of samples from the four classes, proportional to the data set, were selected as the training set, and the remaining 20% was used as the test set. That is, a fivefold cross-validation strategy was used in our analysis, meaning that the described procedure was repeated five times. This secondary analysis was conducted on deidentified data and did not involve a research protocol requiring approval by an institutional review board or ethics committee.

Signal processing

Feature extraction

The process whereby sound is perceived by humans goes through three stages: *production* (the core reason why the sound

Table 1—Number of snore episodes per class.

Class		V	O	T	E
No. of patients		134	56	10	24
No. of episodes	Total number	484	216	39	89
	Mean (over patient)	3.61	3.86	3.90	3.71
Duration (s)	Mean (over episode)	1.5457	1.4688	1.2670	1.5202
	SD	.3478	.3108	.3716	.3629

E = epiglottis, O = oropharyngeal, SD = standard deviation, T = tongue, V = velum.

is different), *transmission* (medium sending sound from the source to the receiver), and *reception*. During sleep, snore sounds are generated by vibrations at different obstruction sites and through resonance of the upper airway. We assumed that the vibration processes at different sites have different physiological characteristics (such as frequency and complexity). Specifically, perception of the human ear is most accurate at low frequencies, and in this range, the perception and frequency are linearly related. The human auditory system perceives higher frequencies less accurately, and above a particular range, perception correlates logarithmically with frequency. Logarithmic representation means that the difference between high frequencies is compressed and hence not perceived as accurately. There are various psychoacoustic models of frequency perception scales. One common model is the mel scale. The mel frequency cepstral coefficient (MFCC) was designed to capture the perception characteristics using the mel-frequency scale. In addition, the degree of complexity of different physiological signals generated by different systems or by the same system during different conditions differs. Therefore, for each snore episode, two features were extracted to capture the characteristics of the production and reception processes: complexity (defined by multiscale entropy, MSE) of the zero-crossing rate (ZCR) and MFCC. Furthermore, principal component analysis (PCA) was used to capture the most significant components of the features.

Signal transformation and complexity measurement

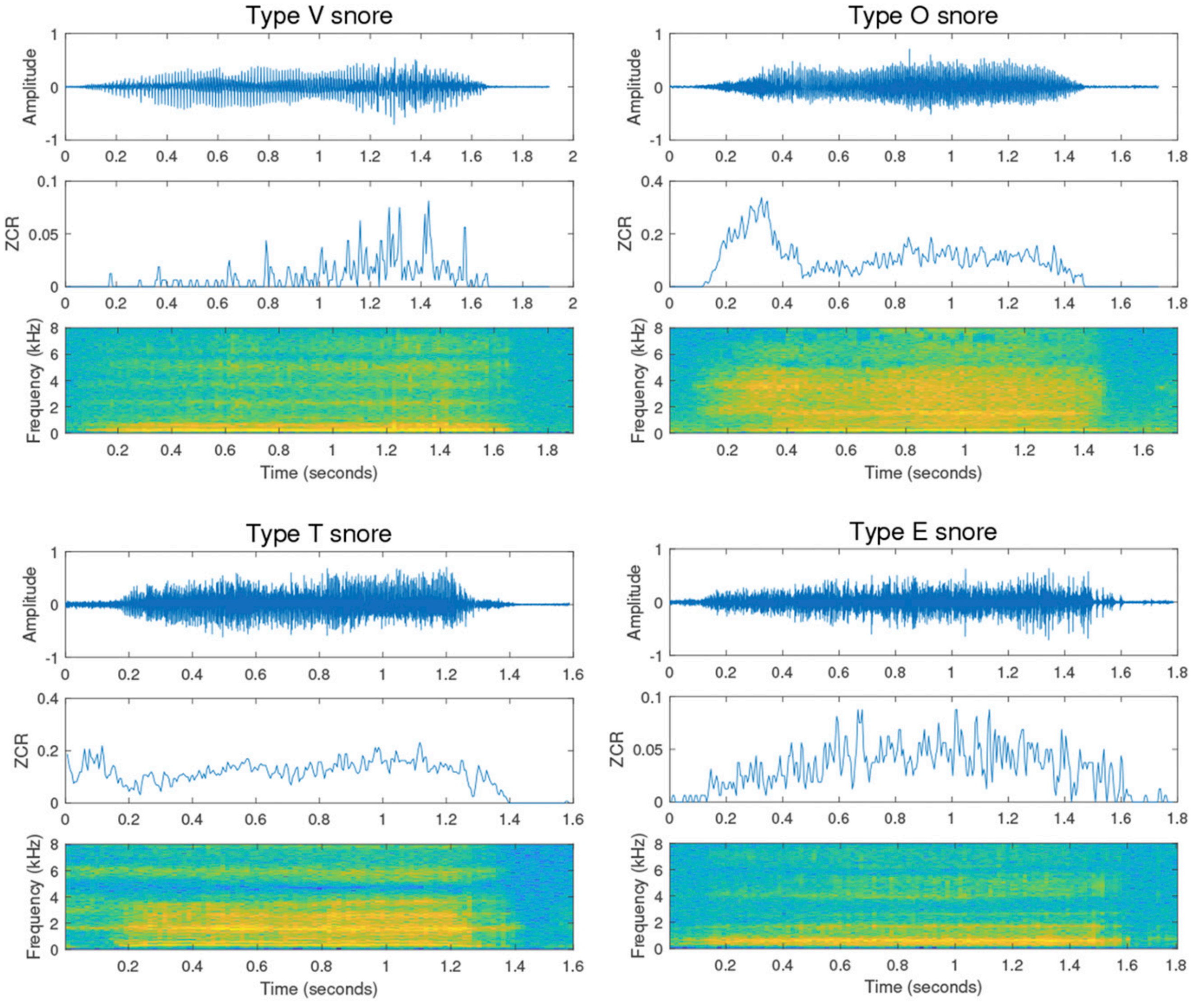
The ZCR is a commonly used technique in sound classification. ZCR transformation can provide an estimation of the frequency content of a signal in a time-domain at low computational cost[21]; thus, it is often used as a part of the front-end processing. ZCR is defined as the number of time-domain zero-crossings within a defined region of a signal, divided by the number of samples of that region. It represents the smoothness of the signal and is an indicator of the frequency (examples are shown in **Figure 2**). Compared with the raw snoring sounds, the ZCR-transformed signals contain more meaningful information. As different vibrational tissues can be considered different physiological systems, the snoring sounds produced by them contain their characteristics (vibration frequency, amplitude, degree of upper airway obstruction, and so forth). Therefore, both the original snoring sounds and the ZCR-transformed signals remain nonstationary and nonlinear in nature. To capture the characteristics of snoring sounds, we applied MSE as the measurement of complexity.

The MSE algorithm is based on the application of approximate entropy[22] or sample entropy[23] for different scales of the same process. Approximate entropy and sample entropy are measures of regularity (orderliness) in serial data. The largest values of entropy correspond to completely random sequences; however, an increase in entropy may not imply a higher physiological complexity for physiological signals. For example, we can obtain a new signal with a higher entropy value by shuffling a snoring sound, but the new signal does not contain any meaningful physiological information because single-scale–based entropy cannot reveal the long-range correlation on a multiple temporal scale. Therefore, MSE has been developed to quantify the complexity of a physiological signal over different time scales. It has been widely used in the field of signal analysis, especially for the analysis of physiological signals.[24–31] In our analysis, we calculated entropies on 20 scales of ZCR-transformed signals as the measurement of complexity, and we compared the complexity levels of the four snore types.

Mel-frequency cepstral coefficients

MFCCs are the most widely used acoustic features in speech recognition and other audio-based research tasks.[32–35] It is an acoustic approach that uses not only the production phase but also the reception phase. Specifically, it is representative of the vocal tract by computing the envelope of the time-power spectrum of the speech signal. On the other hand, according to psychophysical studies, human auditory system perception of a sound's frequency follows an overall nonlinear scale, which is spaced linearly at frequencies below 1 KHz and logarithmically above 1 KHz. MFCC characterizes the human ear perception of frequency by using a set of nonlinear spaced triangular bandpass filters. Considering the similarities between snoring production and speech generation, here the upper airway acts like the vocal tract. We believe the MFCC can capture differences in the acoustic characteristics of snoring sounds generated by different vibration locations. It is worth noting that because the duration of snores is different, the total number of MFCCs is different. We used the first 13 coefficients by taking into consideration the fact that most of the signal feature is compacted in the first few coefficients owing to the properties of the cosine transform. Furthermore, its delta (first-order difference) and acceleration (second-order difference) features related to the change in the characteristics of snores over time were added. Finally, a total of 39 features (13 MFCC, 13 Δ, and 13 acceleration) were extracted from each snoring sound.

Figure 2—Examples of snore sound signals and their transformation.



For each snore type, the upper panel shows the raw audio waveform of a snore episode, the middle panel shows the corresponding ZCR-transformed signal, and the lower panel shows the spectrogram of snore episode. E = epiglottis, O = oropharyngeal, T = tongue, V = velum, ZCR = zero crossing rate.

PCA

PCA is used abundantly in many fields because it is an unsupervised, simple approach to reduce the redundant information from original data sets by performing linear transformation techniques. PCA provides a way to reveal the hidden crucial information that often underlies the original data. PCA aims to remove redundant information by linearly combining the original features and retaining the effective information of the data with as few dimensions as possible. In our study, we used PCA to extract the most significant components of features and reduce the computation for dimensionality reduction.

Classification

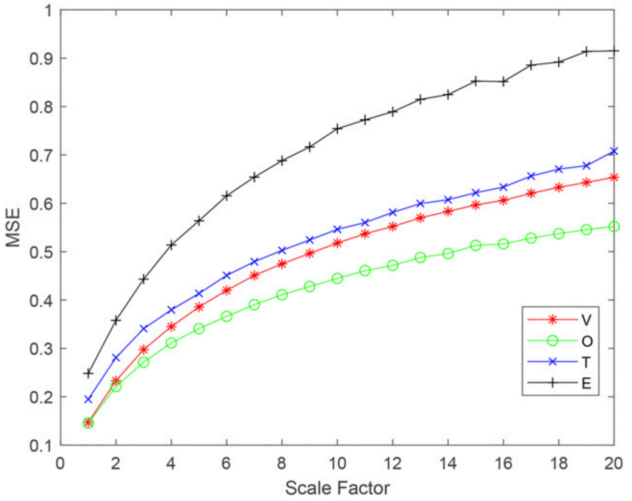
Machine learning is a proven potent method in classification tasks. One of the most powerful yet simple machine-learning approaches is a support vector machine. It aims to find a hyperplane that separates the test data set into a discrete, predefined number of classes in a manner consistent with the training examples.

It has also been used increasingly in the past few years in physiological signal analysis as it can provide a systematized architecture for analyzing and extracting important information from complex data.[36] Hence, a support vector machine–based machine-learning model may be promising for the classification of snoring sounds. In our study, we selected 80% of the MPSSC data set randomly for training; the remainder were used for testing. This process was repeated 10 times, and the average accuracy was calculated as the final accuracy.

Statistical analysis

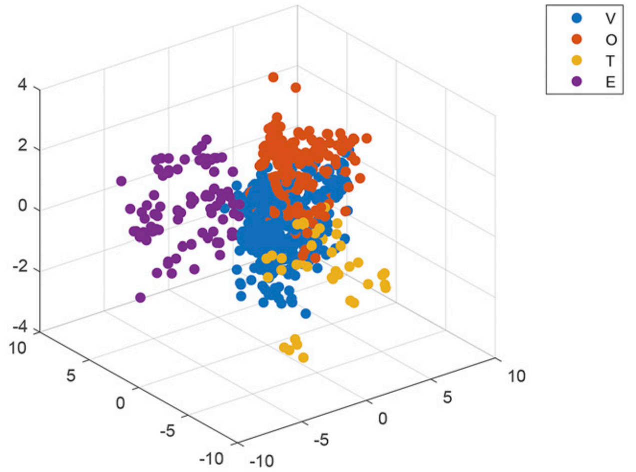
Signal processing and statistical analyses were performed using MATLAB 2017b (MathWorks, Inc., Natick, MA). Descriptive statistics were reported with mean and standard deviation for

Figure 3—Complexity level of ZCR-transformed snoring sound.



Multiscale entropy (MSE) of ZCR-transformed snoring sound time series. Entropy of type E snores is significantly higher than the other three snore types (post hoc test indicated $P < .05$). E = epiglottis, O = oropharyngeal, T = tongue, V = velum, ZCR = zero crossing rate.

Figure 4—Plot of three-dimensional feature extracted from the originally high-dimensional acoustic signals.



The final three-dimensional features extracted from the 59-dimensional acoustic feature. Each axis represents a new combined feature, and each point represents one snore episode. The features of the four types of snores have distinctly different distribution spaces, which makes it easier for SVM classifier to classify them. Each point represents one snore episode. SMV = support machine vector.

continuous data (eg, duration of snore episodes) and number (count) for categorical data. Comparisons of continuous variables between snore types (eg, complexity of ZCR-transformed signals) were assessed by analysis of variance. A P value < .05 was considered statistically significant. Accuracy was calculated to evaluate the overall performance of the proposed method. For each snore type, the sensitivity (true positive rate), specificity (true negative rate), area under the receiver operating characteristics curve, precision, and *F1* score were also investigated to evaluate the robustness of our method. The precision (also called the positive predictive value) is the fraction of relevant instances among the retrieved instance, that is:

$$precision = \frac{number\ of\ true\ positive}{number\ of\ true\ positive + number\ of\ false\ positive}$$

For conveying the balance between sensitivity and precision, the *F1* score was used, defined as follows:

$$F1 = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity}$$

A confusion matrix was used to present the results in a clean and unambiguous way.

RESULTS

Complexity features of the ZCR-transformed audio signals

Complexity levels of the ZCR-transformed snoring sounds of the four snore types were compared (**Figure 3**). Overall, the complexity (indicated by MSE values) increased dramatically from small scales to larger scales. Differences in complexity were more distinguishable at larger scales (eg, entropies on scale 3 to scale 20). The entropy of type E snores was significantly higher than that of the other three snore types ($P < .05$).

Reduced feature space of snores

From MSE and MFCC analyses, a total of 59 dimensional features were obtained for each snore type. The dimensionality of the features (20 entropies from MSE analysis and 39 parameters from MFCC analysis) was then analyzed by PCA. By calculating the variance contribution rate of new (linearly combined) features with the variance of the original data, we found that three-dimensional subspace is sufficient for this classification task. Therefore, the final three-dimensional features were obtained by projecting the 59 features into a three-dimensional principal component subspace. **Figure 4** demonstrates the distribution of the reduced acoustic feature space of the snores.

Accuracy of the analysis

The performance of our proposed approach yielded an overall accuracy of 92.17%, with an unweighted average sensitivity of .8636 and an unweighted average precision of .8909 (**Table 2**). Among the four snore types, specificity (.9933) was highest in type E snores. The area under the receiver operating characteristic curve values of snore types V, O, T, and E were .9982, .8321, .7507, and .9349, respectively. The *F1*-score of 0.8763 was obtained as the harmonic mean of the sensitivity and precision. *F1* scores were highest in type V snores and lowest in type T snores (**Table 2**). **Figure 5** shows the confusion matrix of the proposed method in VOTE classification. Among the

Table 2—Statistical results of experiment.

	Sensitivity	Precision	Specificity	AUC	*F1*
V	.9804	.9434	.9063	.9982	.9615
O	.8056	.8788	.9692	.8321	.8406
T	.7273	.8000	.9871	.7507	.7619
E	.9412	.9412	.9933	.9349	.9412
Unweighted average	.8636	.8909	.9640	.8790	.8763

AUC = area under the receiver operating characteristic curve, E = epiglottis, *F1* = the harmonic mean of precision and sensitivity, O = oropharyngeal, T = tongue, V = velum.

four snore types, types V and E showed the highest sensitivity (> 90%), whereas the sensitivity of type T was 72.73%.

DISCUSSION

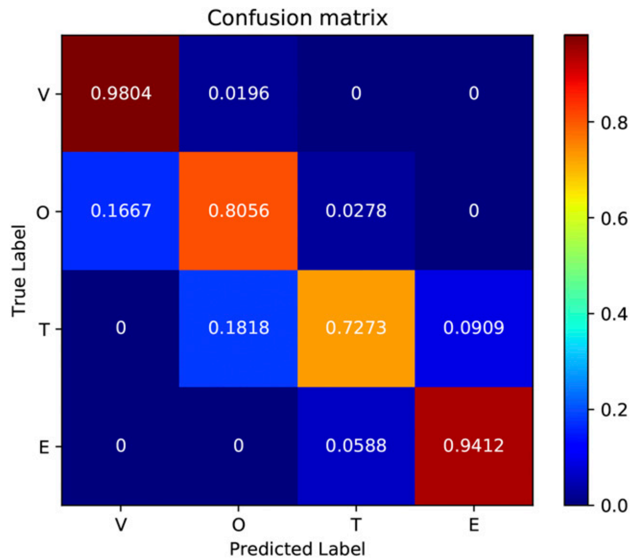
In this study, we developed a machine-learning model that used the acoustic and physiological features of snores to classify snores produced at different vibration locations in the upper airway. Our proposed approach was able to capture automatically the complexity nature of snore sounds generated from different sites in the upper airway. The overall performance of our model showed an unweighted average area under the receiver operating characteristic curve of 87.9% on the MPSSC data set. We believe our approach can contribute to the automatic classification of vibration sites from the snore sound signal.

Previous efforts have been made to apply machine-learning techniques to classify snores. Using the same MPSSC data set containing data for the four snore types, Amiriparian et al[18] proposed a convolutional neural network (CNN) that was used to capture the characteristics of four types of snoring while using spectral features as input. They achieved an unweighted average sensitivity of 67% on the test data set. Similarly, Freitag et al[19] proposed a CNN paradigm to classify snoring sounds. Their study differed from that of Amiriparian et al[18] in that they adopted a hybrid "end-to-evolution" approach by combining deep CNN and evolutionary feature selection. They obtained an unweighted average sensitivity of 66.5% on the test data set. Vesperini et al[20] used deep scattering spectrum, multilayer perceptron neural networks, and Gaussian mean supervectors for VOTE snore classification. As a result, an unweighted average sensitivity of 74.19% was achieved on the test data set. Consistent with our study, all these studies considered the acoustic information of snore; however, the physiological information contained in the snores was overlooked in these previous studies. Our study is the first to distinguish snores generated at different obstructed parts of the upper airway by combining acoustic and physiological information, which allows our model a more efficient recognition ability to achieve better classification performance than prior studies.

Our results confirm that vibration sites are correlated with distinct characteristics of the snoring event. Snores originating from the epiglottis showed the highest MSE, whereas the MSE of oropharyngeal snores was the lowest. In clinical practice, the cardinal difficulty in improving the success rates of surgical treatment for patients with OSAHS is to target accurately the excitation location in the upper airway. For example, it has been proven that treatments for patients whose obstructive site is located at the soft palate are more successful[37,38] and that treatments for snores produced by the posterior pharyngeal walls or the tongue base have less effect.[39] In contrast, for simple palatal snorers, treatments that target mainly the hypopharyngeal area might not be a wise choice.[40,41] Therefore, it is helpful for clinical interventions to understand the mechanism of obstruction and obstructive sites in the upper airway. The test results showed that the model we presented is superior to those proposed in existing studies for the classification of snores generated at the four excitation locations, which indicates that it can be used to locate the generation location of snores. Furthermore, in contrast to DISE, it is more amenable to mass screening because it is noninvasive and inexpensive.

This study has some limitations. First, the whole airway cannot be visualized at once, which makes breath-to-breath labeling difficult. This was a secondary analysis, and our data were from an existing data set in which all snoring samples were

Figure 5—The confusion matrix of SVM classifier.



E, epiglottis, O = oropharyngeal, SMV = support machine vector, T = tongue, V = velum.

limited to a single site. Therefore, many important sources of obstruction are absent from this study and need further investigation. Second, some possible scenarios were not included in the original data set, for example, nonvibration obstruction sounds due to turbulent flow and release of pressure, collapse with complete silence, and snores generated from multisites. Therefore, the accuracy of our approach in the real world may be affected if such sources are the dominant reasons for snore sounds. According to the VOTE classification scheme,[14] for a simple snorer, snoring usually occurs during a stage of partial narrowing without complete collapse. Therefore, the samples in the MPSSC data set are limited to partial narrowing according to the VOTE classification.[17] For this primary snoring, snore sounds can be used to locate the site of vibration for targeted therapy. The existing data set does not include apneic snore sounds; therefore, to what extent the excitation location of snore sounds and the obstruction sites in apnea correlate remains unexplored. We encourage future studies to address this gap with other available data sets. Then, for type V snores, more targeted information, and specific classification of the site of collapse are needed to make treatment decisions because, even though they can capture some velopharyngeal obstruction information to some extent, the degree of obstruction and configuration of obstruction are not contained. Finally, the samples from patients with different classes of snores were unbalanced, with type V (obstruction at velum area) and type O (oropharyngeal obstruction) snores being dominant and type T (obstruction at tongue base) and type E (obstruction at epiglottis) snores much less sampled; however, the sample represents real-world distributions, and the ratio is consistent with clinical results and earlier findings of the distribution of snore types.[39,42] Furthermore, it is noteworthy that even though the number of training data for both types T and E was small, type E had the second highest sensitivity and type T the lowest, suggesting that the sample size does not determine the overall performance of the proposed technique. Future studies are encouraged to include data with multilevel obstructions.

Clinical implications

Assessment of the site of snore generation is a major challenge to decision makers for directing clinical treatment of snoring and OSAHS. Reports that can truly reflect the status of the upper airway are not available based solely on DISE examinations (or other similar evaluation techniques). Although DISE is widely used, it cannot reflect the natural sleep situation because it is based on artificial sleep, and only a part of the upper airway can be visualized at one time. The need for a more convenient and powerful approach to identify the site of snore generation is of major importance to the field of sleep-disordered breathing. Here, we propose a novel and valid approach to systematically locate the site of snore generation from a single snore sound. This algorithm can capture the difference of snores generated by different parts of the upper airway through acoustic and physiological characteristics of snore sounds. Although our analysis is based on a simplified classification system named the VOTE scheme, it can be used to assist in the treatment selection, especially for those with mild OSAHS because single-level obstruction is more common in mild OSAHS patients, whereas in severe OSAHS, a multilevel obstruction is more characteristic.[43–45] Furthermore, our method may enable a deeper understanding of clinical application value of snoring sounds. This study shows that locating vibrating sites acoustically has great application potential because snoring sounds contain recognizable physiological information that can reflect the state of the upper airway. Further studies are needed to reinforce our findings by applying our model to more specific and detailed larger database.

CONCLUSIONS

In conclusion, snores can reveal the condition of the upper airway. The proposed model provides an efficient tool to classify the snores produced in different upper airway states. Future studies are needed to determine the feasibility of applying this model in clinics and general populations.

ABBREVIATIONS

CNN, convolutional neural network
CPAP, continuous positive airways pressure
DISE, drug-induced sleep endoscopy
MFCC, Mel frequency cepstral coefficient
MPSSC, Munich Passau Snore Sound Corpus
MSE, multiscale entropy
OSAHS, obstructive sleep apnea-hypopnea syndrome
PCA, principal component analysis
VOTE, Velum-Oropharyngeal-Tongue-Epiglottis
ZCR, zero crossing rate

REFERENCES

1. Lugaresi E, Cirignotta F, Coccagna G, Piana C. Some epidemiological data on snoring and cardiocirculatory disturbances. *Sleep*. 1980;3(3-4):221–224.
2. Norton PG, Dunn EV. Snoring as a risk factor for disease: an epidemiological survey. *Br Med J (Clin Res Ed)*. 1985;291(6496):630–632.
3. Ersu R, Arman AR, Save D, et al. Prevalence of snoring and symptoms of sleep-disordered breathing in primary school children in Istanbul. *Chest*. 2004;126(1): 19–24.
4. Hoffstein V, Mateika S, Nash S. Comparing perceptions and measurements of snoring. *Sleep*. 1996;19(10):783–789.
5. Lee GS, Lee LA, Wang CY, et al. The frequency and energy of snoring sounds are associated with common carotid artery intima-media thickness in obstructive sleep apnea patients. *Sci Rep*. 2016;6(1):30559.
6. Sawyer AM, Gooneratne NS, Marcus CL, Ofer D, Richards KC, Weaver TE. A systematic review of CPAP adherence across age groups: clinical and empiric insights for developing CPAP adherence interventions. *Sleep Med Rev*. 2011; 15(6):343–356.
7. Blumen MB, Latournerie V, Bequignon E, Guillere L, Chabolle F. Are the obstruction sites visualized on drug-induced sleep endoscopy reliable? *Sleep Breath*. 2015;19(3):1021–1026.
8. Jung AR, Koh TK, Kim SJ, Lee KH, Cho JS, Kim SW. Comparison of level and degree of upper airway obstruction by Müller's maneuver and drug-induced sleep endoscopy in obstructive sleep apnea patients. *Auris Nasus Larynx*. 2017;44(5): 571–575.

9. Riley RW, Powell NB, Guilleminault C. Obstructive sleep apnea syndrome: a review of 306 consecutively treated surgical patients. *Otolaryngol Head Neck Surg*. 1993;108(2):117–125.
10. Carvalho B, Hsia J, Capasso R. Surgical therapy of obstructive sleep apnea: a review. *Neurotherapeutics*. 2012;9(4):710–716.
11. Choi JH, Lee JY, Cha J, Kim K, Hong SN, Lee SH. Predictive models of objective oropharyngeal OSA surgery outcomes: success rate and AHI reduction ratio. *PLoS One*. 2017;12(9):e0185201.
12. De Vito A, Carrasco Llatas M, Vanni A, et al. European position paper on drug-induced sedation endoscopy (DISE). *Sleep Breath*. 2014;18(3):453–465.
13. Hessel NS, de Vries N. Results of uvulopalatopharyngoplasty after diagnostic workup with polysomnography and sleep endoscopy: a report of 136 snoring patients. *Eur Arch Otorhinolaryngol*. 2003;260(2):91–95.
14. Kezirian EJ, Hohenhorst W, de Vries N. Drug-induced sleep endoscopy: the VOTE classification. *Eur Arch Otorhinolaryngol*. 2011;268(8):1233–1236.
15. Askar SM, Quriba AS, Hassan EM, Awad AM. Positional awake endoscopy vs DISE in assessment of OSA: a comparative study. *Laryngoscope*. 2020;130: 2269–2274.
16. Emoto T, Abeyratne UR, Akutagawa M, Nagashino H, Kinouchi Y. Feature extraction for snore sound via neural network processing. *Annu Int Conf IEEE Eng Med Biol Soc*. 2007;2007:5477–5480.
17. Janott C, Schmitt M, Zhang Y, et al. Snoring classified: the Munich-Passau Snore Sound Corpus. *Comput Biol Med*. 2018;94:106–118.
18. Amiriparian SGM, Gerczuk M, Ottl S, et al. Snore sound classification using image-based deep spectrum features. Presented at: *INTERSPEECH* 2017, 18th Annual Conference of the International Speech Communication Association; Stockholm, Sweden; August 2017; pp 3512–3516.
19. Freitag M. Amir S, Cummins N, Gerczuk M, Schuller B. An 'end-to-evolution' hybrid approach for snore sound classification. INTERSPEECH; 2017: 3507–3511.
20. Vesperini F, Galli A, Gabrielli L, Principi E, Squartini S. Snore sounds excitation localization by using scattering transform and deep neural networks. Paper presented at: 2018 International Joint Conference on Neural Networks (IJCNN); July 8–13, 2018; Rio de Janeiro, Brazil.
21. Bhattacharjee M, Prasanna S, Guha P. Time-frequency audio features for speech-music classification. Preprint posted online November 3, 2018. arXiv: 1811.01222.
22. Pincus SM. Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA*. 1991;88(6):2297–2301.
23. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*. 2000; 278(6):H2039–H2049.
24. Liau B-Y, Wu FL, Lung C-W, Zhang X, Wang X, Jan Y-K. Complexity-based measures of postural sway during walking at different speeds and durations using multiscale entropy. *Entropy*. 2019;21(11):1128.
25. Ma Y, Wu CW, Peng CK, et al. Complexity-based measures of heart rate dynamics in older adults following long- and short-term tai chi training: cross-sectional and randomized trial studies. *Sci Rep*. 2019;9(1):7500.
26. Hou F, Yu Z, Peng CK, Yang A, Wu C, Ma Y. Complexity of wake electroencephalography correlates with slow wave activity after sleep onset. *Front Neurosci*. 2018;12:809.
27. Humeau-Heurtier A. The multiscale entropy algorithm and its variants: a review. *Entropy*. 2015;17(5):3110–3123.
28. Ma Y, Shi W, Peng CK, Yang AC. Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches. *Sleep Med Rev*. 2018;37:85–93.
29. Shi W, Shang P, Ma Y, Sun S, Yeh C-H. A comparison study on stages of sleep: quantifying multiscale complexity using higher moments on coarse-graining. *Commun Nonlinear Sci Numer Simul*. 2017;44:292–303.
30. Ma Y, Tseng PH, Ahn A, et al. Cardiac autonomic alteration and metabolic syndrome: an ambulatory ECG-based study in a general population. *Sci Rep*. 2017;7(1):44363.
31. Yin Y, Peng C-K, Hou F, et al. The application of multiscale joint permutation entropy on multichannel sleep electroencephalography. *AIP Adv*. 2019;9(12):125214.
32. Abreu Araujo F, Riou M, Torrejon J, et al. Role of non-linear data processing on speech recognition task in the framework of reservoir computing. *Sci Rep*. 2020; 10(1):328.
33. Anumanchipalli GK, Chartier J, Chang EF. Speech synthesis from neural decoding of spoken sentences. *Nature*. 2019;568(7753):493–498.
34. Tokozume Y. *Learning environmental sounds with end-to-end convolutional neural network*. Presented at: the International Conference on Acoustics, Speech, and Signal Processing; March 2017; New Orleans, LA, pp 2721–2725.
35. Trigeorgis G, Nicolaou MA, Schuller BW, Zafeiriou S. Deep canonical time warping for simultaneous alignment and representation learning of sequences. *IEEE Trans Pattern Anal Mach Intell*. 2018;40(5):1128–1138.
36. Tomar D. A survey on data mining approaches for healthcare. *Int J Bio-SciBio-Technol*. 2013;5(5):241–266.
37. Iwanaga K, Hasegawa K, Shibata N, et al. Endoscopic examination of obstructive sleep apnea syndrome patients during drug-induced sleep. *Acta Otolaryngol Suppl*. 2003;440:36–40.
38. Hessel NS, Vries N. Increase of the apnoea-hypopnoea index after uvulopalatopharyngoplasty: analysis of failure. *Clin Otolaryngol Allied Sci*. 2004; 29(6):682–685.
39. Hessel NS, de Vries N. Diagnostic work-up of socially unacceptable snoring. II. Sleep endoscopy. *Eur Arch Otorhinolaryngol*. 2002;259(3):158–161.
40. Soares D, Sinawe H, Folbe AJ, et al. Lateral oropharyngeal wall and supraglottic airway collapse associated with failure in sleep apnea surgery. *Laryngoscope*. 2012;122(2):473–479.
41. den Herder C, Kox D, van Tinteren H, de Vries N. Bipolar radiofrequency induced thermotherapy of the tongue base: its complications, acceptance and effectiveness under local anesthesia. *Eur Arch Otorhinolaryngol*. 2006;263(11): 1031–1040.
42. Fiz JA, Jane R. Snoring analysis. A complex question. *J Sleep Disord Treat Care*. 2012;1(1):1–3.
43. Hohenhorst W, Hortscht M, Grünwald S, and Lamprecht J. Verteilung der vibrations-und obstruktionslokalisationen unter propofolschlaf. *Laryngo-Rhino-Otologie*. 2004;83(2):23–10.
44. Steinhart H, Kuhn-Lohmann J, Gewalt K, Constantinidis J, Mertzlufft F, Iro H. Upper airway collapsibility in habitual snorers and sleep apneics: evaluation with drug-induced sleep endoscopy. *Acta Otolaryngol*. 2000;120(8):990–994.
45. Hohenhorst W, Ravesloot M, Kezirian E, De Vries N. Drug-induced sleep endoscopy in adults with sleep-disordered breathing: technique and the VOTE classification system. *Oper Tech Otolaryngol--Head Neck Surg*. 2012; 23(1):11–18.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication June 19, 2020
Submitted in final revised form January 5, 2021
Accepted for publication January 6, 2021

Address correspondence to: Xiyuan Hu, PhD, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, People's Republic of China; Email: xiyuan.hu@foxmail.com and Yan Ma, MD, MPH, Division of Interdisciplinary Medicine and Biotechnology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA; Email: dr.yan.ma@gmail.com

DISCLOSURE STATEMENT

All authors have seen and approved the manuscript. Work for this study was performed at three clinical centers. The authors report no conflicts of interest.