IOP Publishing Physiol. Meas. 41 (2020) 085006 https://doi.org/10.1088/1361-6579/abaa34

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

RECEIVED
4 January 2020

REVISED
23 July 2020

ACCEPTED FOR PUBLICATION
28 July 2020

PUBLISHED
4 September 2020

Amplitude spectrum trend-based feature for excitation location classification from snore sounds

Jingpeng Sun[1], **Xiyuan Hu**[2], **Chen Chen**[1], **Silong Peng**[1] **and Yan Ma**[3]

[1] Institute of Automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing 100190, People's Republic of China

[2] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China

[3] Center for Dynamical Biomarkers, Division of Interdisciplinary Medicine and Biotechnology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, United States of America

E-mail: xiyuan.hu@foxmail.com

Keywords: amplitude trend, signal decomposition, null space pursuit, snore classification, OSA

Abstract

Objective: Successful surgical treatment of obstructive sleep apnea (OSA) depends on the precise location of the vibrating tissue. Snoring is the main symptom of OSA and can be utilized to detect the active location of tissues. However, existing approaches are limited, owing to their inability to capture the characteristics of snoring produced from the upper airway. This paper proposes a new approach to better distinguish different snoring sounds that are generated from four different excitation locations. *Approach*: First, we propose a robust null space pursuit algorithm for extracting the trend from the amplitude spectrum of snoring. Second, a new feature from this extracted amplitude spectrum trend, which outperforms the Mel-frequency cepstral coefficient (MFCC) feature, is designed. Subsequently, the newly proposed feature, namely the trend-based MFCC (TCC), is reduced in dimensionality by using principal component analysis. Finally, a support vector machine is employed for the classification task. *Main results*: By using the TCC, the proposed approach achieves an unweighted average recall of 87.5% on the classification of four excitation locations on the public dataset Munich Passau Snore Sound Corpus. *Significance*: The TCC is a promising feature for capturing the characteristics of snoring. The proposed method can effectively perform snore classification and assist in accurate OSA diagnosis.

1. Introduction

Snoring bothers over 50% of the general population almost every night (Young *et al* 1997). It is due to the turbulent flow of air through the collapsed upper airway, causing a vibration of some tissues during sleep (Culebras 1996). Obstructive sleep apnea (OSA) is characterized by a repetitive partial or complete upper airway obstruction, which intermittently causes reductions in airflow (hypopneas) or cessations of breathing (apneas) (Farney *et al* 2011). In contrast, simple snoring cases exhibit no apnea or hypopnea events. Previous studies revealed that the incidence of OSA is estimated at 5% of the world's population, and a high percentage of patients (80%–90%) with moderate or severe OSA are believed to be undiagnosed (Finkel *et al* 2009). Without treatment, OSA can result in both acute (e.g. congestive heart failure, stroke, and even sudden death) and chronic (e.g. hypertension, cardiovascular diseases, and diabetes) conditions (Redline and Strohl 1998, Young *et al* 2002, Coccagna *et al* 2006, Tarasiuk *et al* 2006, Somers *et al* 2008). Surgical treatment is a common option to cure OSA. Unfortunately, several surgical treatments have failed to address the inability to precisely localize the origin tissue of the vibration. A targeted and less invasive surgical treatment is better suited, especially for severe OSA patients. Therefore, the accurate localization of the vibration or obstruction position is a necessary condition for a successful treatment.

Accordingly, drug-induced sleep endoscopy (DISE) has been established to address the localization problem (El Badawey *et al* 2003). DISE is an intranasally inspected procedure performed using a flexible

© 2020 Institute of Physics and Engineering in Medicine

nasopharyngoscope during artificial sleep, which is induced by applying narcotics to OSA patients. Although DISE is known as a powerful approach for identifying the location of the vibration and obstruction and studying the dynamic upper airway, it is straining, time consuming, and costly. Furthermore, it cannot reflect the natural sleep situation because it is performed during artificial sleep.

These disadvantages limit the application of DISE. Therefore, alternative methods should be developed to complement or replace DISE for identifying the location of the vibration and obstruction. Detecting snores via their acoustic characteristics has unique advantages, and snores can be recorded conveniently without attaching sensors to subjects. Therefore, such applications are preferred over other physiological signals, and the acoustic characteristics of snores have been of great interest in recent years. As mentioned, snoring is produced by the vibration of some tissues in the collapsed upper airway during sleep. The upper airway acts as an acoustic filter during the production of snoring sounds. Hence, changes in the structure or vibration location in the upper airway are generally revealed in the acoustical properties of snores.

Some acoustic features, such as the subband energy ratio and intensity, spectral, and pitch-related features have been proposed to extract relevant diagnostic information from the snoring sound for snore classification. For instance, average power and spectral entropy can reflect the occurrence of apnea events (Cavusoglu *et al* 2008, Azarbarzin and Moussavi 2013). Spectral analysis (Fiz *et al* 1996) or linear regression performed on the power spectrum density (Azarbarzin and Moussavi 2013) has been proposed to determine whether or not snores are caused by apnea. Hummel *et al* (2016) classified obstructive and central sleep apnea by using 16 acoustic features (such as periodicity or spectral centroid). Deep features (Wang *et al* 2018, Arsenali *et al* 2018, Lim *et al* 2019) have also been used to classify snoring sounds.

As for the determination of the vibration location, Beeton *et al* (2007) discriminated palatal and nonpalatal snores by combining a two-means clustering method and the statistical dimensionless moment coefficients of skewness and kurtosis. Agrawal *et al* (2002) argued that frequency can be used to distinguish vibration locations, with the observation that palate snores occurred at 137 Hz, tongue snores occurred at a high frequency of 1243 Hz, and snores produced by the epiglottis and the tonsillar were characterized by 490 and 170 Hz, respectively. Many different schemes have been proposed (Friedman *et al* 2002, Iwanaga *et al* 2003, Abdullah *et al* 2003, Vicini *et al* 2012), and one of them is the velum-oropharyngeal-tongue-epiglottis (VOTE) (Kezirian *et al* 2011), which is a popular and widely used classification scheme that distinguishes the difference among four structures within the upper airway (figure 1). Qian *et al* (2017) introduced nine acoustic features to evaluate their effectiveness in capturing the structural characteristics of snoring generated by four tissues. They compared the performances of different feature sets and attempted to obtain the best classification performance by combining feature sets with several classifiers. Neural networks (Freitag *et al* 2017, Amiriparian *et al* 2017, Vesperini *et al* 2018, Schmitt and Schuller 2019, Zhang *et al* 2020) and some well-known classifiers (Rao *et al* 2017, Nwe *et al* 2017, Albornoz *et al* 2017, Demir *et al* 2018, Qian *et al* 2019), such as support vector machines (SVMs) (Rao *et al* 2017, Nwe *et al* 2017, Albornoz *et al* 2017, Demir *et al* 2018), random forest, and naive Bayes (Qian *et al* 2019), have been built for the classification of the excitation location. Amiriparian *et al* (2017) proposed a convolutional neural network (CNN), which was used to capture the characteristics of four types of snoring while taking spectral features as input. They achieved an unweighted average recall (UAR) of 67% on the test dataset. Similarly, Freitag *et al* (2017) proposed a CNN paradigm to classify snoring sounds. The difference between this study and the study by Amiriparian *et al* (2017) is that they adopted a hybrid 'end-to-evolution' approach by combining deep CNN and evolutionary feature selection. A UAR of 66.5% was obtained on the test dataset. Vesperini *et al* (2018) employed the deep scattering spectrum technique, multi-layer perceptron neural networks, and Gaussian mean supervectors for VOTE snore classification. As a result, a UAR of 74.19% was achieved on the test dataset. Schmitt and Schuller (2019) combined CNN and long short-term memory to build an end-to-end deep neural network classifier. They achieved a UAR of 67.0% on the test dataset. Concerning standard machine learning based approaches, Rao *et al* (2017) achieved a UAR of 49.58% on a development dataset by modeling the production process of snores from lungs to lips/nose with a dual source-filter model. Nwe *et al* (2017) used a fusion of CNN with random forest and SVMs, for a UAR of 51.7% on the test dataset. Demir *et al* (2018) and Albornoz *et al* (2017) employed an SVM classifier to classify acoustic features extracted from a spectrogram and spectral features, with UARs of 72.62% and 48.10% on the test and development datasets, respectively. Qian *et al* (2019) employed a bag of wavelet-based audio-words to represent features and obtained a UAR of 69.4% on the test dataset with a naive Bayes classifier.

This study aims to develop a classification algorithm that accurately distinguishes VOTE snoring by capturing the filter characteristics of the upper airway. We propose an improved signal decomposition algorithm to extract the trend of the amplitude spectrum by modeling the upper airway as a filter from the source-filter perspective. A new feature is obtained by performing cepstral analysis on the extracted trend, to capture the filter characteristics of the upper airway during snoring. The experimental results of our approach demonstrate satisfactory performance on a public test dataset.

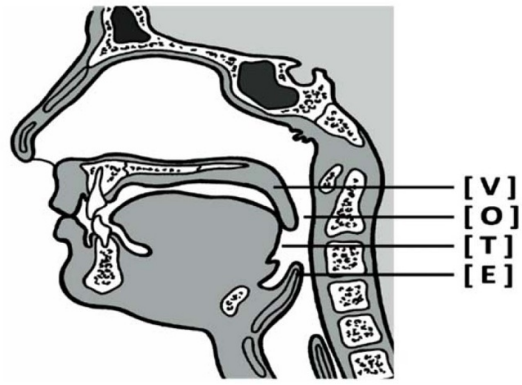

Figure 1. Diagram of the VOTE scheme in the upper airway. © [2017] IEEE. Reprinted, with permission, from Qian *et al* (2017).

Table 1. Number of snoring samples per class.

	Train	Development	Test	Total
V	168	161	155	484
O	76	75	65	216
T	8	15	16	39
E	30	32	27	89
Total	282	283	263	828

The remainder of this paper is organized as follows: section 2 describes the datasets; section 3 introduces the proposed algorithm; section 4 presents the experimental results; section 5 discusses these results; section 6 provides conclusions and an outlook for future work.

2. Dataset

The Munich Passau Snore Sound Corpus (MPSSC) (Janott *et al* 2018) was used in this study. The MPSSC is the first available public snoring dataset that focuses on the excitation location issue. This dataset contains 828 snoring episodes from 219 subjects who had undergone diagnostic DISE because of suspected OSA, obtained at three clinical centers between 2006 and 2015. Note that 205 of the 219 subjects were male and 14 were female. Their ages ranged from 24 to 78 years, with an average age of 49.8 years. These snoring sounds were recorded with different equipment (i.e. nasopharyngoscope, recording system, and microphone) at a same sample rate of 44 100 Hz and a 16 bit resolution. Each of them was labeled as V (velum), O (oropharyngeal), T (tongue), and E (epiglottis) according to the VOTE scheme and vibration tissue. The samples were normalized and resampled to 16 000 Hz. The duration of the snore samples varies from 0.73 to 2.75 s and the average duration is 1.46 s. Figure 1 shows the corresponding structures of VOTE. More specific anatomical information is presented below:

- Velum: velopharyngeal area;
- Oropharyngeal: oropharyngeal lateral walls;
- Tongue: anteroposterior tongue base;
- Epiglottis: epiglottis.

Figure 2 depicts waveform and spectrogram examples of VOTE snoring. The spectrograms illustrate that most of the snoring energy is in the low-frequency range, i.e. inspiring, which will be described in more detail in the next section.

The dataset was stratified into a train, development, and test subset. Table 1 presents the number of snoring samples per class.

3. Method

The proposed method first uses a source-filter to model the sound propagation procedure in the upper airway, which reflects the physical mechanism of snoring. The spectrum trend of the snoring sounds is then extracted to capture the filter characteristics of the upper airway. Finally, the corresponding features are

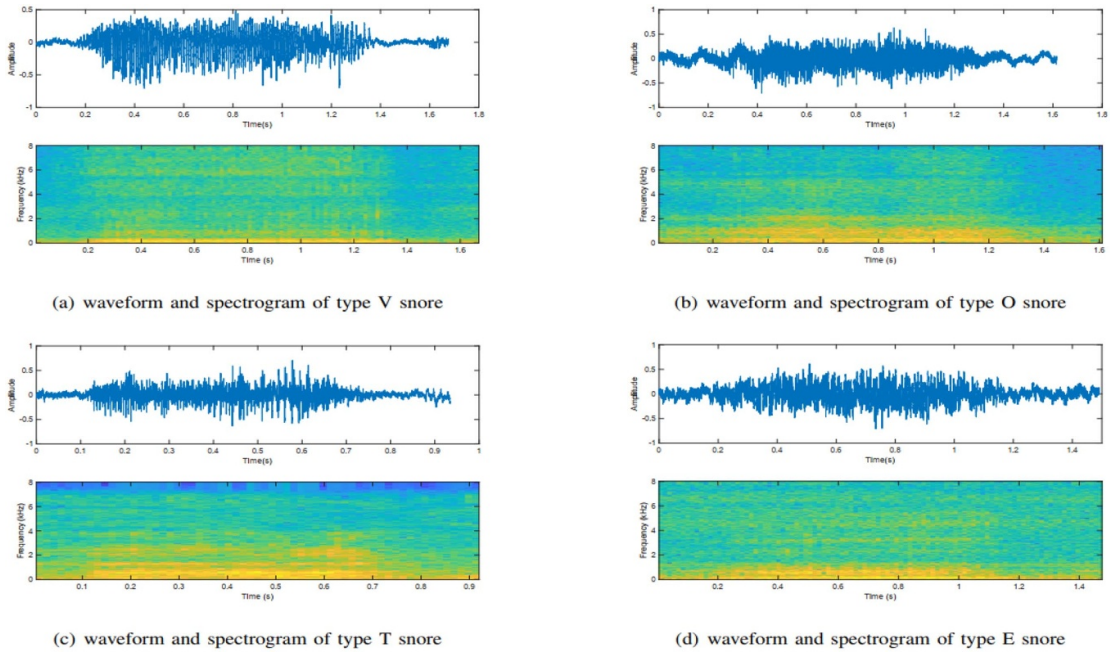

Figure 2. Examples of VOTE snore.

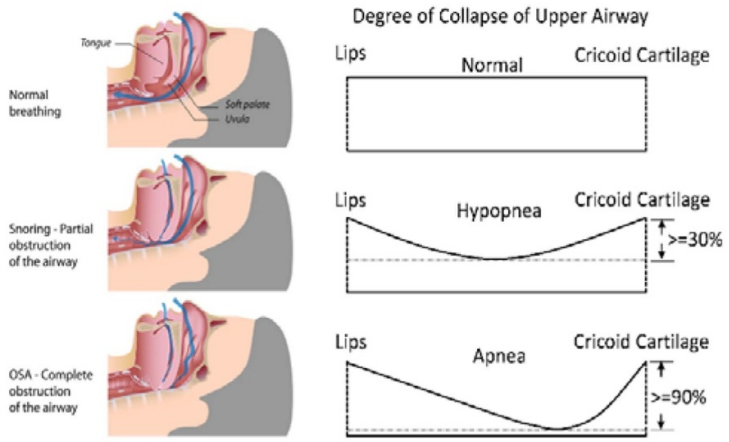

Figure 3. States of the upper airway during sleep (left) and the corresponding simplified tube models (right). From top to bottom: Normal, hypopnea, and apnea.

extracted from the spectrum trend, which provides a good indicator of the state of the upper airway for the snoring classification.

3.1. Source filter model

During inspiration, the collapsed upper airway constricts the path from the mouth to the cricoid cartilage. The pressure increases as the degree of collapse increases. At one point, the airflow that flows through the narrowed upper airway vibrates some of the soft tissues, producing a snore. Therefore, snoring varies with the condition of the upper airway, including both the location of the vibration and the degree of collapse. For example, as shown in figure 3, if the nasal airflow decreases by more than 30% from the prevent baseline (3% oxygen desaturation is needed) and lasts at least 10 s, a hypopnea event is happening. An apnea event is determined when the decrease drops by more than 90% and lasts more than 10 s (Berry *et al* 2018).

According to the acoustic theory for sound propagation, a precision snoring model should take many factors into account, such as the propagation of sound, shape variation of the upper airway, and loss of energy resulting from heat conduction and friction. While it is difficult to integrate all of the above factors into a model, some simplified models have been proposed, which can well approximate the snoring propagation process. In this paper, we assume that the upper airway can be reduced to a tube (figure 3).

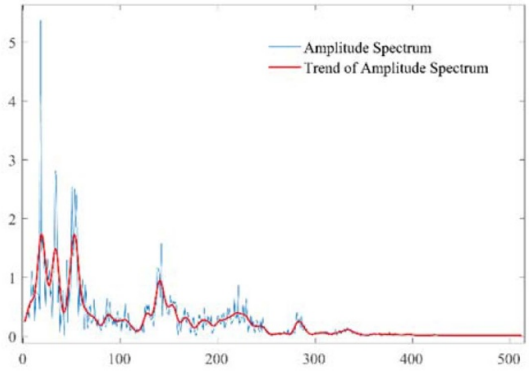

Figure 4. Illustration of the trend of the amplitude spectrum.

Therefore, the snoring sound is generated by the airflow passing through the vibrating part in the upper airway, and the airflow and the upper airway can be treated as a source and a time-varying filter, respectively. The following model can be used to represent such a source-filter model:

$$x(n) = e(n) * u(n), \tag{1}$$

where '*' represents the linear convolution symbol; n denotes the nth element of sequences; $x(n)$ is the snore; $e(n)$ denotes the source excitation; and $u(n)$ is the response of the upper airway.

Figure 4 shows the amplitude spectrum of snoring. The conventional formant analysis proves that formants reflect the acoustic resonance characteristics of the cavities (the human vocal tract or the upper airway). Formants are the frequencies at which peaks of the smoothed spectrum are observed. Therefore, the characteristics of formant may be reflected or included in a low frequency component of the amplitude spectrum. If we compare the spectrum to a signal, the formants can be considered as the 'low-frequency component' of the amplitude spectrum. A widely used representation of the 'low-frequency component' is the spectral envelope, which can be derived by computing the real cepstrum of a windowed short-time signal.

However, the spectral envelope approximates the 'low-frequency component' by performing filter banks on signals. Here, we directly used the smooth-varying trend of the amplitude spectrum to approximate its 'low-frequency component'. We extracted an adaptive and robust trend from various amplitude spectra by improving the null space pursuit (NSP) algorithm (Peng and Hwang 2010) with a new iterative algorithm and a hyperparameter updating strategy. Figure 4 shows an example of using the improved NSP algorithm to extract the 'low-frequency component' from an amplitude spectrum.

3.2. Trend extraction with the robust null space pursuit algorithm

The NSP algorithm (Peng and Hwang 2010) is an operator-based signal separation method that separates signals into several subcomponents by means of predefined operators. For example, given a linear operator Γ and a signal S, S_1, and R are determined such that

$$S = S_1 + R,$$

where R is the residual component and S_1 is in the null space of Γ, computed as follows:

$$\Gamma S_1 = 0 \Leftrightarrow \Gamma(S - R) = 0.$$

Mathematically, this can be modeled as

$$\min_R \|\Gamma(S-R)\|^2 \text{ s.t.} \|R\|^2 < \varepsilon. \tag{2}$$

To obtain R, Peng and Hwang (2010) proposed a regularization model:

$$\hat{R} = \arg\min_R \left\{ \|\Gamma_S(S-R)\|^2 + \lambda \left(\|D(R)\|^2 + \gamma \|S-R\|^2 \right) + F(\Gamma_S) \right\}. \tag{3}$$

The model is called the NSP algorithm. Γ_S is adaptively estimated from the signal S. D is an operator that regulates R. λ is a regularization parameter. γ is a leakage factor determining the amount of $S-R$ to be

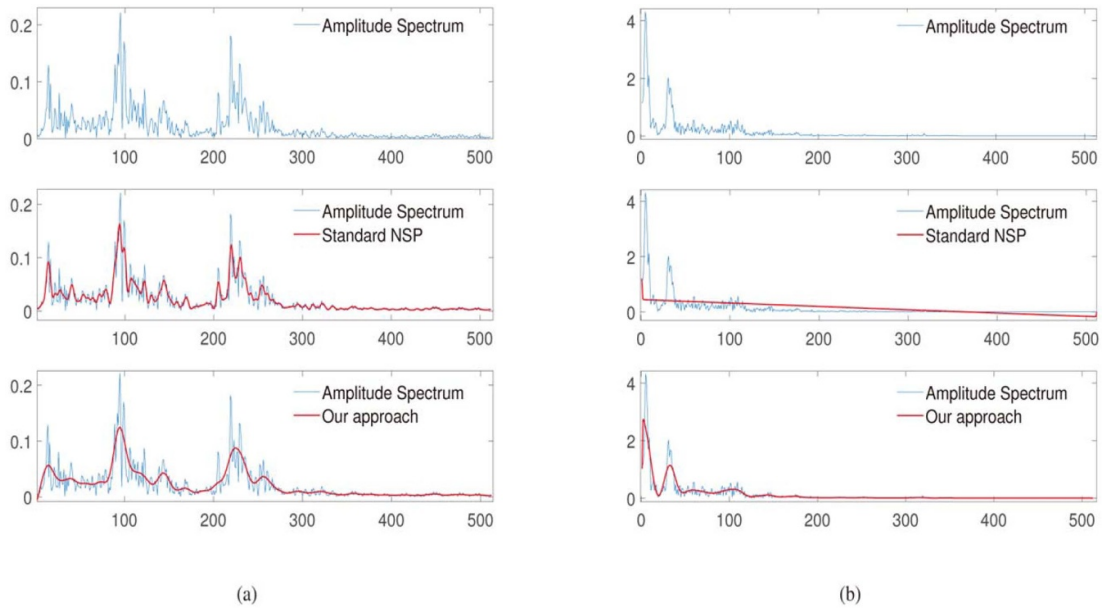

Figure 5. Trend extraction results using two different approaches: standard NSP in (Peng and Hwang 2010) and our approach. The blue line is the amplitude spectrum and the red line is the trend. (a) An illustrative case of both the standard NSP and our approach extracting the trend of a snore. (b) An illustrative case of the standard NSP failing to obtain the trend of a snore, while our approach maintains its effectiveness.

retained in the null space of Γ_S. The last term is the Lagrange term for the parameters of operator Γ_S. Parameters λ and γ are updated during the iteration. The iteration stop criterion is when R becomes stable.

However, the updating formulas of the hyperparameters λ and γ are closely related in the NSP algorithm (i.e. $\lambda^{(k+1)}$ depends on $\gamma^{(k)}$, and vice versa), which makes it difficult for R to obtain a desired stable solution in some real-life signals. As shown in the middle row in figure 5, the 'low-frequency component' extracted by the standard NSP algorithm either has some small oscillatory waves or is oversmoothed.

Here, we refer to $S-R$ as the trend representing the 'low-frequency component' of the input signal S. During the iterations, the trend changes from high- to low-frequency components via several mutations. The difference between two adjacent trends is oscillatory. When the difference reaches its first minimum, $S-R$, as intended, is the optimal trend.

To obtain a better solution of model (2), and to avoid introducing extra hyperparameters γ, we converted it into an unconstrained problem:

$$\min_R \|\Gamma(S-R)\|^2 + \lambda\|R\|^2, \tag{4}$$

where Γ is a discrete second-order difference operator, and $\lambda \geqslant 0$ is a termed regularization parameter. That is, for a signal s, performing Γ on the nth element of s means: $\Gamma s_n = s_n - 2s_{n-1} + s_{n-2}$.

In appendix A, we present the detailed process of solving R and λ. We name the algorithm the 'robust null space pursuit' (RNSP) and summarize it in algorithm 1.

Algorithm 1 Robust null space pursuit (RNSP)
Input S
Initialize R, Γ, $\lambda \leftarrow \|\Gamma\|^2$, $k \leftarrow 0$
repeat
$R^{(k+1)} \leftarrow \left(\Gamma^T\Gamma + \lambda I\right)^{-1}\left(\lambda R^{(k)} + \Gamma^T\Gamma S\right)$
$\lambda^{(k+1)} \leftarrow \frac{N_2}{N_1}\frac{\left\|\Gamma S - \Gamma R^{(k)}\right\|^2}{\left\|R^{(k)}\right\|^2}$
$k \leftarrow k+1$
until $R^{(k+1)} - R^{(k)}$ reaches its first minimum
return $S - R^{(k+1)}$

3.3. Feature extraction and classification

As a powerful and efficient feature, the Mel-frequency cepstral coefficient (MFCC) has dominated the field of speech recognition for a long time. Similar to the real cepstrum, the MFCC is defined as the real cepstrum of

a windowed short-time signal derived from the fast fourier transform (FFT) of the input signal, but with a nonlinear frequency scale (i.e. the window size varies nonlinearly), which can be viewed as an approximation of the auditory system behavior. The spectral envelope is representative of the upper airway, and the MFCC represents the envelope. Motivated by this, we fed the trend obtained by the improved NSP algorithm into the Mel-filter banks to obtain the feature coefficient. We called this the trend-based MFCC (TCC).

The procedure for obtaining the TCC of signal $x(n)$ consists of five steps.

(1) Preprocessing: The following filter is used to perform pre-processing:

$$H(z) = 1 - az^{-1},$$

where a is a constant.

In our work, the snores were framed with a frame duration of 1024 points, and a frame hop of 512 points. The Hamming window was used to reduce frequency leakage.

(2) Computing the power spectral $X(k)$ of signal $x(n)$ by applying the FFT:

$$X(k) = |FFT(x(n))|.$$

(3) The trend of $X(k)$ denoted as $X_1(k)$ using the RNSP algorithm (as proposed in section 3.2).
(4) Computing the TCC by feeding the extracted trend $X_1(k)$ to the Mel-frequency filter banks and keeping the first 13 coefficients.
(5) Representing dynamic characteristics of a snore with differential and acceleration (first- and second-order difference of TCC) coefficients. Therefore, the differential coefficients are obtained as follows:

$$d_t = \frac{\sum_{j=1}^{J} j\left(\mathrm{TCC}_{t+j} - \mathrm{TCC}_{t-j}\right)}{2\sum_{j=1}^{J} j^2}, \tag{5}$$

where d_t denotes the differential coefficient, and TCC_{t+j} is the $(t+j)$ th element of TCC, $J = 2$. Similarly, we can obtain the acceleration coefficients.

Finally, a 39-dimensional feature vector (static, differential, and acceleration coefficients) was generated for each frame. However, the snore duration varied, resulting in the different dimensions of the snore feature matrix. We addressed this problem by averaging the features of all frames to derive the final 39-dimensional feature.

To classify the segments into four classes (i.e. 'V,' 'O,' 'T,' and 'E'), an 11-dimensional final feature was obtained by feeding the extracted 39-dimensional feature into a principal component analysis (PCA) for dimension reduction (90% energy was retained in this study), and the binary SVM classifiers were arranged in a one-against-one strategy. The theory of the SVM technique can be found in Vapnik (2013).

4. Results

We evaluated our method on the MPSSC dataset for snore classification. The dataset was divided into train, development, and test subsets (table 1). The SVM classifier equipped with a radial basis function (RBF) kernel was employed to classify the four categories of VOTE snore episodes, in which the RBF kernel parameters were optimized by grid search. The margin and scale of the RBF kernel were 0.0027 and 2, respectively. The final feature (an averaged 39-dimensional vector) of each snoring sound was fed into the PCA and SVM for dimensionality reduction and classification. Table 2 shows the comparison of the UAR values between our method and those of the reported previous studies on the MPSSC dataset. Our method outperformed those reported in previous studies with 18.15% and 13.31% UAR on the development and test set, respectively.

Figure 6 shows the confusion matrix of the RNSP in the VOTE classification. Among the four classes of snores, types V and E achieved the highest accuracy (i.e. above 90%), whereas type T achieved 74%. Interestingly, even though the amount of training data for both types T and E was small, type E had the highest accuracy, while type T had the lowest. We will discuss this result in detail in the next section.

In addition, to better observe the performance of the TCC and compare it with that of the MFCC, figure 7 shows the displayed sensitivity and specificity with the same classifier (i.e. SVM), performing on

Table 2. Comparison of classification results on MPSSC.

Methods	Development	Test
Amiriparian *et al* (2017)	44.8%	67.0%
Freitag *et al* (2017)	57.6%	66.5%
Vesperini *et al* (2018)	67.14	74.19%
Schmitt and Schuller (2019)	59.1%	67.0%
Rao *et al* (2017)	49.58%	-
Nwe *et al* (2017)	57.13%	51.7%
Demir *et al* (2018)	37.82%	72.62%
Albornoz *et al* (2017)	48.10%	-
Qian *et al* (2019)	35.0%	69.4%
Ours	**85.29%**	**87.5%**

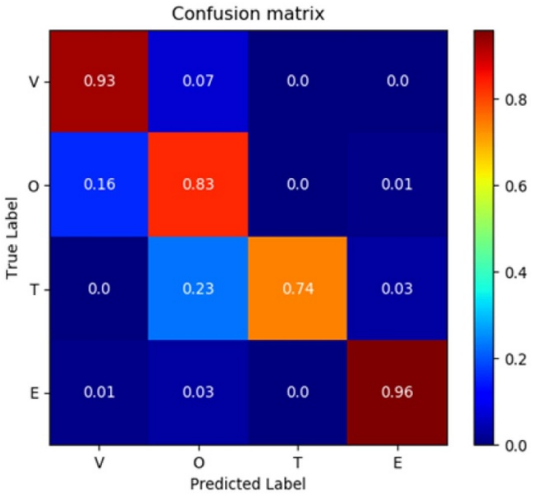

Figure 6. Confusion matrix of the four-category classification task.

various values of the operating point in the form of receiver operating characteristic (ROC) curves. The ROC curve of the TCC was consistently above that of the MFCC. Furthermore, the empirical bootstrap was used to obtain the two-sided 95% confidence interval for the area under the ROC (AUC) and the number of bootstrap replicates was 1000. The 95% confidence limits of the AUC of MFCC and TCC are [0.9795, 0.9803] and [0.9846, 0.9853], respectively. Considering that the AUC estimates for TCC and MFCC may fall into each other's confidence limits, the superiority of TCC makes more sense when 95% confidence limits are considered. Although the advantages of the TCC over the MFCC were not very large, these results indicated that our approach is comparable and superior by discovering the underlying relationship among different snores.

5. Discussion

We proposed a signal decomposition algorithm, called RNSP, and a novel feature based on the amplitude spectrum trend to classify snoring produced at four excitation locations. The results presented in section 4 clearly show that the proposed new feature allows for a better classification with the VOTE scheme. Furthermore, the performance of the new feature compares favorably to those of the other algorithms and outperforms the powerful feature MFCC with a promising improvement.

Figure 6 illustrates that the algorithm has a higher accuracy for types V (93%) and E (96%), compared with types O (83%) and T (74%). Meanwhile, type V can be easily distinguished from type T or E, and type E can be distinguished from type V or O. The misclassification rates for such types were very low. However, the difference between types T and O or types O and V can be subtle sometimes, which may result in a wrong classification. This subtleness is mainly due to the physiological position of V, O, T, and E . V and E are located at both ends of the most collapsible part of the upper airway, while O and T are located between V and E, which leads to types V and E being easier to classify and types O and T being more difficult to classify. This is the reason why types T and E have almost the same number of samples but achieved entirely different classification accuracies.

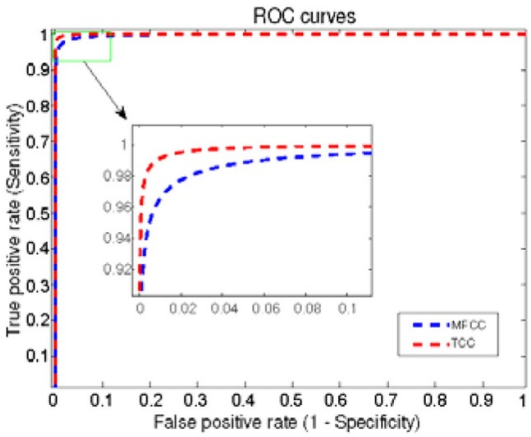

Figure 7. ROC curves for TCC and MFCC on the test set.

Other features for snore classification have been proposed in the existing literature. Azarbarzin and Moussavi (2013) extracted the average power, zero-crossing rate, frequency of the spectral peak with the lowest frequency (F_0), frequency of the peak with maximum power (F_p), and spectral entropy from snoring, to classify non-apneic, hypopneic, and post-apneic snoring. They obtained 92.9% sensitivity, 100% specificity, and 96.4% accuracy in their experiment.

Meanwhile, Cavusoglu *et al* (2008) used four features, namely snoring episode power durations, snoring episode separations, average snoring episode power, and the short time coefficient of variation sequences, to explore the possibility of distinguishing among simple snorers and OSA patients. However, their experiment only showed that these features have the potential to classify snoring, but do not have a quantitative indicator.

Ng *et al* (2008) argued that the formant frequencies of snoring contain essential information and are representative of the physical frequency transfer function of the upper airway. Their study achieved 88% sensitivity and 82% specificity. Qian *et al* (2017) proposed a feature set, called the wavelet energy feature, based on the wavelet transform theory. They compared several commonly used acoustic features on the classification of snore sounds according to the VOTE scheme. This feature set achieved a UAR of 78% with the best combination of features (crest factor, formants, MFCCs, etc) and classifiers (k-NN, LDA, SVM, etc). The highest accuracy was 89% for type T. The lowest accuracy was 63.9% for type O.

Compared with those studies, the RNSP algorithm proposed in this study achieved the highest UAR. However, a limitation of our study is that the number of samples was low, and no augmenting strategy was employed to balance the number of samples of different labels.

Therefore, a better performance can be expected when more recordings are available.

6. Conclusions and future work

We proposed herein a novel feature based on the RNSP algorithm to classify the snoring generated in different upper airway conditions, in which the TCC was effective in classifying different snoring sounds. In future, we will detect and classify respiratory events, considering that the snoring sounds generated during different respiratory events vary.

Acknowledgments

The National Natural Science Foundation of China supported this research through the research project ‘Research on Operator-based Robust Adaptive Signal Separation Algorithm and its Applications’ with Grant No. 61571438.

We would like to thank the anonymous reviewers for their valuable and insightful comments.

Appendices

Appendix A: Trend extraction with the RNSP algorithm

For the unconstrained problem:

$$\min_R \|\Gamma(S-R)\|^2 + \lambda\|R\|^2, \tag{A1}$$

where Γ is a discrete second-order difference operator, and $\lambda \geqslant 0$ is a termed regularization parameter. That is, for a signal s, performing Γ on the nth element of s means: $\Gamma s_n = s_n - 2s_{n-1} + s_{n-2}$. R can be obtained by solving

$$\frac{\partial\left(\|\Gamma(S-R)\|^2 + \lambda\|R\|^2\right)}{\partial R^T} = 0.$$

That is,

$$\Gamma^T\Gamma R - \Gamma^T S + \lambda R = 0.$$

Thus, we have

$$R = \left(\Gamma^T\Gamma + \lambda I\right)^{-1}\Gamma^T S. \tag{A2}$$

We obtain smoother solutions using (A2) with iterated Tikhonov regularization (Neumaier 1998) and initializing $R^{(0)} = 0$. R can then be updated using the following iterative formula:

$$R^{(k+1)} = \left(\Gamma^T\Gamma + \lambda I\right)^{-1}\left(\lambda R^{(k)} + \Gamma^T\Gamma S\right). \tag{A3}$$

After obtaining R, the next step is to compute λ which can be derived in a Bayesian framework. We assume that each sample in the residual signal R follows a Gaussian distribution with zero mean and variance σ_R^2, and the samples are independent:

$$p(R) = \mathrm{N}(0, \sigma_R^2), \tag{A4}$$

with (A1),

$$p(\Gamma S|\Gamma, R) = \mathrm{N}(\Gamma R, \sigma_{\Gamma S}^2), \tag{A5}$$

where $\sigma_{\Gamma S}^2$ represents the variance of a conditional probability density function. The l_2 regularization formulation in (A1) is equivalent to using a maximum *a posteriori* (MAP) formation with Gaussian prior (A4) and (A5):

$$\begin{aligned} &\arg\max_R p(R|\Gamma S) \\ &= \arg\max_R \frac{p(\Gamma S, R)}{p(\Gamma S)} \\ &= \arg\max_R (\ln p(\Gamma S|R) + lnp(R)). \end{aligned} \tag{A6}$$

Substituting Gaussian prior equations (A4) and (A5) into (A6), we obtain

$$\begin{aligned} &\arg\max_R p(R|\Gamma S) \\ &= \arg\max_R \left(-\frac{(\Gamma S - \Gamma R)^T(\Gamma S - \Gamma R)}{2\sigma_{\Gamma S}^2} - \frac{R^T R}{2\sigma_R^2}\right) \\ &= \arg\min_R \left(\|\Gamma S - \Gamma R\|^2 + \frac{\sigma_{\Gamma S}^2}{\sigma_R^2}\|R\|^2\right). \end{aligned} \tag{A7}$$

Comparing (A7) to (A1), we have

$$\lambda = \frac{\sigma_{\Gamma S}^2}{\sigma_R^2}. \tag{A8}$$

Using maximum likelihood estimation, we obtain

$$\hat{\sigma}_{\Gamma S}^2 = \frac{1}{N_1}\|\Gamma S - \Gamma R\|^2 \tag{A9}$$

and

$$\hat{\sigma}_{R}^2 = \frac{1}{N_2}\|R\|^2, \tag{A10}$$

where N_1 and N_2 are the lengths of signals ΓS and R, respectively.

Substituting equations (A9) and (A10) into (A8), we obtain

$$\lambda = \frac{N_2}{N_1}\frac{\|\Gamma S - \Gamma R\|^2}{\|R\|^2}. \tag{A11}$$

Therefore, the iterative formulation of λ is

$$\lambda^{(k+1)} = \frac{N_2}{N_1}\frac{\left\|\Gamma S - \Gamma R^{(k)}\right\|^2}{\left\|R^{(k)}\right\|^2}. \tag{A12}$$

In summary, we can perform alternate optimization with the help of (A3) and (A12). The iteration process will be stopped when $R^{(k+1)} - R^{(k)}$ reaches its first minimum. The trend $S - R^{(k+1)}$ can be obtained.

Appendix B: Spectrum trend extraction with loess regression (local nonparametric regression)

The loess regression is a nonparametric approach that was designed by Cleveland and Devlin (1988) to extract a smooth curve for a given signal. It is a local smoothing-based regression method. Specifically, a low-degree polynomial is used to fit a subset which was determined by a sliding window with a certain size (bandwidth) for extracting the smooth curve.

The smooth curve is a kind of trend; therefore, we compared the performance of this trend obtained by loess with the performance of our RNSP method. According to the definition of loess, its use is dependent on the bandwidth. Cross-validation was performed to find an optimal bandwidth parameter. We evaluated bandwidths from 0.01 to 0.5 in increments of 0.01. The experimental results found that loess achieved the highest UAR when the bandwidth was 0.02, which is the optimal bandwidth size. The examples of the trend extracted by loess with different bandwidths are shown in figure B1.

The loess method achieved a maximum UAR of 57.4% on the test dataset. After exploring the performance of loess with 10-fold cross validation, a higher UAR of 61.5% was obtained. However, comparing the performance of loess with RNSP, which achieved a UAR of 87.5% on the test dataset, we conclude that our RNSP method is more robust and has a higher performance than the loess method.

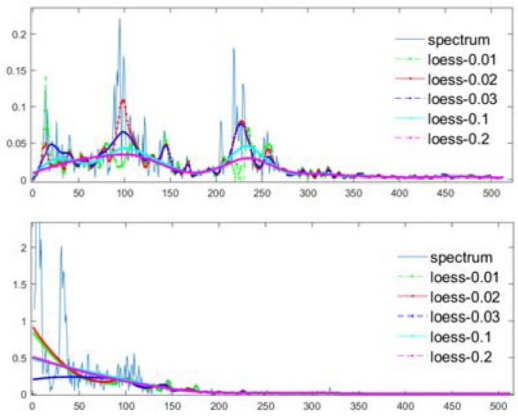

Figure B1. Trend extraction results using the loess method. Here, loess-0.01 refers to a bandwidth of 0.01. The remaining bandwidth parameters are not shown, for better visualization.

References

Abdullah V J, Wing Y K and van Hasselt C A 2003 Video sleep nasendoscopy: the Hong Kong experience *Otolaryngol. Clin. North Am.* **36** 461-71, vi

Agrawal S, Stone P, Mcguinness K, Morris J and Camilleri A E 2002 Sound frequency analysis and the site of snoring in natural and induced sleep *Clin. Otolaryngol Allied Sci.* **27** 162–6

Albornoz E M, Bugnon L A and Martínez C E 2017 Snore recognition using a reduced set of spectral features *XVII Workshop on Information Processing and Control (RPIC)* (IEEE) (https://doi.org/10.23919/rpic.2017.8214357)

Amiriparian S, Gerczuk M, Ottl S, Cummins N, Freitag M, Pugachevskiy S, Baird A and Schuller B W 2017 Snore sound classification using image-based deep spectrum features *INTERSPEECH 2017* (https://doi.org/10.21437/interspeech.2017-434)

Arsenali B, van Dijk J, Ouweltjes O, den Brinker B, Pevernagie D, Krijn R, van Gilst M and Overeem S 2018 Recurrent neural network for classification of snoring and non-snoring sound events *40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) (https://doi.org/10.1109/embc.2018.8512251)

Azarbarzin A and Moussavi Z 2013 Snoring sounds variability as a signature of obstructive sleep apnea *Med. Eng. Phys.* **35** 479–85

Beeton R J, Wells I, Ebden P, Whittet H B and Clarke J 2007 Snore site discrimination using statistical moments of free field snoring sounds recorded during sleep nasendoscopy *Physiol. Meas.* **28** 1225–36

Berry R B, Albertario C L, Harding S M, Lloyd R M, Plante D T, Quan S F, Troester M M and Vaughn B V 2018 *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* (Darien, IL: American Academy of Sleep Medicine)

Cavusoglu M, Ciloglu T, Serinagaoglu Y, Kamasak M, Erogul O and Akcam T 2008 Investigation of sequential properties of snoring episodes for obstructive sleep apnoea identification *Physiol. Meas.* **29** 879–98

Cleveland W S and Devlin S J 1988 Locally weighted regression: an approach to regression analysis by local fitting *J. Am. Stat. Assoc.* **83** 596–610

Coccagna G, Pollini A and Provini F 2006 Cardiovascular disorders and obstructive sleep apnea syndrome *Clin. Exp. Hypertens.* **28** 217–24

Culebras A 1996 *Clinical Handbook of Sleep Disorders* (Boston, MA: Butterworth-Heinemann Medical)

Demir F, Sengur A, Cummins N, Amiriparian S and Schuller B 2018 Low level texture features for snore sound discrimination *40th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) (https://doi.org/10.1109/embc.2018.8512459)

El Badawey M R, Mckee G, Heggie N, Marshall H and Wilson J A 2003 Predictive value of sleep nasendoscopy in the management of habitual snorers *Ann. Otol. Rhinol. Laryngol.* **112** 40–44

Farney R J, Walker B S, Farney R M, Snow G L and Walker J M 2011 The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index *J. Clin. Sleep Med.* **7** 459-65B

Finkel K J *et al* 2009 Prevalence of undiagnosed obstructive sleep apnea among adult surgical patients in an academic medical center *Sleep Med.* **10** 753–8

Fiz J A, Abad J, Jane R, Riera M, Mananas M A, Caminal P, Rodenstein D and Morera J 1996 Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea *Eur. Respir. J.* **9** 2365–70

Freitag M, Amiriparian S, Cummins N, Gerczuk M and Schuller B W 2017 An 'end-to-evolution' hybrid approach for snore sound classification *INTERSPEECH 2017* (https://doi.org/10.21437/interspeech.2017-173)

Friedman M, Ibrahim H and Bass L 2002 Clinical staging for sleep-disordered breathing *Otolaryngol Head Neck Surg.* **127** 13–21

Hummel R, Bradley T D, Packer D and Alshaer H 2016 Distinguishing obstructive from central sleep apneas and hypopneas using linear SVM and acoustic features *38th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) (https://doi.org/10.1109/EMBC.2016.7591174)

Iwanaga K, Hasegawa K, Shibata N, Kawakatsu K, Akita Y, Suzuki K, Yagisawa M and Nishimura T 2003 Endoscopic examination of obstructive sleep apnea syndrome patients during drug-induced sleep *Acta Otolaryngol. Suppl.* **550** 36–40

Janott C *et al* 2018 Snoring classified: the Munich-Passau snore sound corpus *Comput. Biol. Med.* **94** 106–18

Kezirian E J, Hohenhorst W and de Vries N 2011 Drug-induced sleep endoscopy: the VOTE classification *Eur. Arch. Otorhinolaryngol* **268** 1233–6

Lim S J, Jang S J, Lim J Y and Ko J H 2019 Classification of snoring sound based on a recurrent neural network *Expert Syst. Appl.* **123** 237–45

Neumaier A 1998 Solving ill-conditioned and singular linear systems: A tutorial on regularization *SIAM Rev.* **40** 636–66

Ng A K, Koh T S, Baey E, Lee T H, Abeyratne U R and Puvanendran K 2008 Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Med.* **9** 894–8

Nwe T L, Tran H D and Ma B 2017 An integrated solution for snoring sound classification using Bhattacharyya distance based GMM supervectors with SVM, feature selection with random forest and spectrogram with CNN *INTERSPEECH 2017* (https://doi.org/10.21437/Interspeech.2017-1794)

Peng S and Hwang W-L 2010 Null space pursuit: an operator-based approach to adaptive signal separation *IEEE Trans. Signal Process.* **58** 2475–83

Qian K, Janott C, Pandit V, Zhang Z, Heiser C, Hohenhorst W, Herzog M, Hemmert W and Schuller B 2017 Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis *IEEE Trans. Biomed. Eng.* **64** 1731–41

Qian K, Schmitt M, Janott C, Zhang Z, Heiser C, Hohenhorst W, Herzog M, Hemmert W and Schuller B 2019 A bag of wavelet features for snore sound classification *Ann. Biomed. Eng.* **47** 1000–11

Rao M A, Yadav S and Ghosh P K 2017 A dual source-filter model of snore audio for snorer group classification *INTERSPEECH 2017* (https://doi.org/10.21437/interspeech.2017-1211)

Redline S and Strohl K P 1998 Recognition and consequences of obstructive sleep apnea hypopnea syndrome *Clin. Chest Med.* **19** 1–19

Schmitt M and Schuller B 2019 End-to-end audio classification with small datasets–making it work *27th European Signal Processing Conf. (EUSIPCO)* (IEEE) (https://doi.org/10.23919/eusipco.2019.8902712)

Somers V K *et al* 2008 Sleep apnea and cardiovascular disease: an American Heart Association/American College of Cardiology Foundation Scientific Statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Council, and Council on Cardiovascular Nursing *Circulation* **118** 1080–1111

Tarasiuk A, Greenberg-Dotan S, Simon T, Tal A, Oksenberg A and Reuveni H 2006 Low socioeconomic status is a risk factor for cardiovascular disease among adult obstructive sleep apnea syndrome patients requiring treatment *Chest* **130** 766–73

Vapnik V 2013 *The Nature of Statistical Learning Theory* (New York: Springer Science & Business Media)

Vesperini F, Galli A, Gabrielli L, Principi E and Squartini S 2018 Snore sounds excitation localization by using scattering transform and deep neural networks *Int. Joint Conf. on Neural Networks (IJCNN)* (IEEE) (https://doi.org/10.1109/ijcnn.2018.8489576)

Vicini C, De Vito A, Benazzo M, Frassineti S, Campanini A, Frasconi P and Mira E 2012 The nose oropharynx hypopharynx and larynx (NOHL) classification: a new system of diagnostic standardized examination for OSAHS patients *Eur. Arch. Otorhinolaryngol* **269** 1297–300

Wang J, Strömfeli H and Schuller B W 2018 A CNN-GRU approach to capture time-frequency pattern interdependence for snore sound classification *26th European Signal Processing Conf. (EUSIPCO)* (IEEE) (https://doi.org/10.23919/eusipco.2018.8553521)

Young T, Finn L and Kim H 1997 Nasal obstruction as a risk factor for sleep-disordered breathing. The university of wisconsin sleep and respiratory research group *J. Allergy Clin. Immunol.* **99** S757–62

Young T, Peppard P E and Gottlieb D J 2002 Epidemiology of obstructive sleep apnea: a population health perspective *Am. J. Respir. Crit. Care Med.* **165** 1217–39

Zhang Z, Han J, Qian K, Janott C, Guo Y and Schuller B 2020 Snore-GANs: improving automatic snore sound classification with synthesized data *IEEE J. Biomed. Health Inform.* **24** 300–10