Apnea and Hypopnea Events Classification Using Amplitude Spectrum Trend Feature of Snores*

Jingpeng Sun, Xiyuan Hu, Yingying Zhao, Shuchen Sun, Chen Chen and Silong Peng

Abstract— **Research on snores for Obstructive Sleep Apnea Syndrome (OSAS) diagnosis is a new trend in recent years. In this paper, we proposed a snore-based apnea and hypopnea events classification approach. Firstly, we define the snores after the apnea event and during the hypopnea event as apnea-event-snore (AES) and hypopnea-event-snore (HES), respectively. Then, we design a new feature from the trend of the amplitude spectrum of snores. The newly proposed feature can be viewed as an improvement of the Mel-frequency cepstral coefficient (MFCC) feature, which is well-known for speech recognition. The extracted features were fed to principle component analysis (PCA) for dimension reduction and support vector machine (SVM) for apnea and hypopnea events classification. The experimental results demonstrate the efficiency of the proposed algorithm in using snores to classify apnea and hypopnea events.**

I. INTRODUCTION

The obstructive sleep apnea syndrome (OSAS) was first recognized as a significant health problem in 1956 [1]. In clinical practice, OSAS is scored by Apnea-Hypopnea Index (AHI), the number of pathological respiratory events (apneas and hypopneas) per hour during sleep. AHI is obtained by Polysomnography (PSG), which is the golden standard diagnostic technique for OSAS. PSG requires individual to spend a whole night in a sleep laboratory and wear about 20 electrodes to record physiological signals (EEG, ECG, EMG, EOG, SpO2, etc). It is poorly tolerated, inconvenient and expensive. Furthermore, it may not truly reflect the patients' condition based on only one night record. Thus, an environment friendly, comfortable, convenient and low cost strategy for OSAS diagnosis is urgently required.

OSAS is caused by upper airway collapse during sleep and one of its cardinal symptoms is snoring. Many studies have shown that snoring carries information relating to the degree of collapse of the upper airway [2-5].

*This work was supported by the Natural Science Foundation of China (61571438).

Jingpeng Sun is with Institute of automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, 100190, P.R. China sunjingpeng2015@ia.ac.cn

Xiyuan Hu is with Institute of automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, 100190, P.R. China xiyuan.hu@ia.ac.cn

Yingying Zhao is with South Campus of Guanganmen Hospital, China Academy of Chinese Medical, Beijing, 102600, P.R. China zyy851221@126.com

Shuchen Sun is with Guanganmen Hospital, China Academy of Chinese Medical, Beijing, 100053, P.R. China sunsc0922@163.com

Chen Chen is with Institute of automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, 100190, P.R. China chen.chen@ia.ac.cn

Silong Peng is with Institute of automation, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, 100190, P.R. China silong.peng@ia.ac.cn

Considering the close linkage between snoring and OSAS, snore recording is non-contact, non-invasive and easy to operate. Snoring analysis has aroused great interest in recent years. Pevernagie et al. [2] discussed the acoustic properties of snore. Drawing inspiration from automatic speaker recognition (ASR), Hidden Markov models with Gaussian observation probability distributions have been applied to snorer group recognition [6, 7]. Artificial neural networks have also been used for snore detection and classification [8-11]. Although some of these models have achieved high-accuracy for some tasks, their assumption can not meet the clinical need. For example, In [12], their model performs well for the classification of the excitation location of snores, but the data they used can not reflect natural sleep. Because the corpus are uncut recordings from Drug Induced Sleep Endoscopy (DISE) examinations, this model might not be able to directly used in clinic.

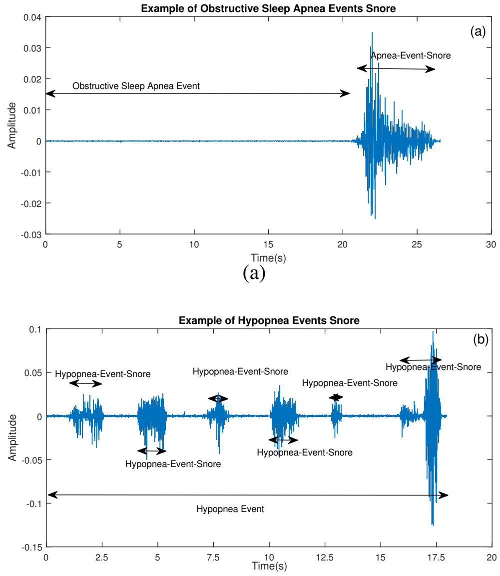
Hummel et al. [13] employed 16 acoustic features, including 10 features for identifying basic breath sound (inspiration, expiration and snoring) and 6 features designed based on sleep apnea pathophysiology, for classification of obstructive and central sleep apnea with linear support vector machine (SVM). It should be point out that their audio files were 2.5 to 8 minutes (mean duration was 5.3 ± 0.7 minutes) long that contained central or obstructive sleep apnea events, which might contain more than one events in an audio file leading to inaccurate AHI.

In this paper, to better capture the characteristics of repiratory events in natural sleep and more accurately calculated AHI. We define hypopnea-event-snore (HES) and apnea-event-snore (AES) as the snores during the hypopnea event and the first snore for a short time (within 5 seconds) after an apnea event, respectively. As shown in figure 1, (a) reprensents example of AES, (b) is a example of HES. Furthermore, we propose a novel feature from the trend of the amplitude spectrum of snores. Then the extracted features were fed to principle component analysis (PCA) and support vector machine (SVM) for apnea and hypopnea events classification. The experimental results demonstrate that our method performs well on both recall (sensitivity) and precision (positive predictive value) on the test set collected by ourselves.

II. METHOD

A. Feature extraction

According to the acoustic nature of snoring and the definition of respiratory events, an apnea event is defined as airflow drops by more than 90% of baseline at the mouth and



(a)

(b)

Fig. 1. Examples of (a) obstructive sleep apnea event snore and (b) hypopnea event snore. Apnea event snores were the first snore after an apnea event. Hypopnea event snores were snores during a hypopnea event.

nose for 10 seconds or more and at least 90% of the event's duration meets the amplitude reduction criteria. While with drops of airflow by more than 30% of baseline at least for 10 seconds and the desaturation of SpO2 is greater or equal to 4% from the pre-event baseline are called hypopnea event [14]. The human upper airway can be seen as a filter during the production of snoring sounds, which can be simulated by the following mathematical model:

$$s(n) = v(n) * u(n) \tag{1}$$

where n represents the $n-th$ point of each sequence, $s(n)$ denotes snore, $u(n)$ denotes the upper airway response and $v(n)$ is the source excitation sequence. The symbol '*' denotes the linear convolution operation.

The "low frequency" component of the amplitude spectrum is a good representation of the filtering characteristics of the human vocal tract. Because the spectral envelope of the amplitude spectrum is a rough representation of the "low frequency" component of the amplitude spectrum, it has been transplanted to the Mel-frequency cepstrum coefficients (MFCC) for speech recognition. The MFCC is derived by computing the real cepstrum of a windowed short-time signal derived from the Fast Fourier Transformation (FFT) of that signal. The difference from the real cepstrum is that a nonlinear frequency scale is used in Mel-spectrum. The nonlinear frequency scale approximates the behavior of the auditory system. Davis and Mermelstein [15] show that the Mel-frequency scale representation beneficial for speech recognition. In fact, MFCC is obtained by performing the cepstral analysis on Mel-spectrum for the purpose of characterizing the spectral envelope. Therefore, it can not give a visual representation of the envelope. Experiments reveal that the trend of the amplitude spectrum performs better than envelope in terms of the representation of the low frequency of the amplitude spectrum, thus, it is more effectively to describe the filter characteristics of upper airway.

Since extracting trend from an amplitude spectrum is a signal separation problem, we use null space pursuit (NSP) algorithm to extract it. The NSP approach [16][1], uses an adaptive operator Γ_S to decompose a signal S into two subcomponents: U and R ($S = U + R$). It can be formulated as an optimization problem:

$$\min_R \left\{ \|\Gamma_S(S-R)\|^2 + \lambda_1(\|R\|^2 + \gamma\|S-R\|^2) + F(\Gamma_S) \right\}, \tag{2}$$

where Γ_S is adaptively estimated from the signal S, λ_1 is regularization parameter, γ is leakage factor, and the last term is the Lagrange term for the parameters of the operator Γ_S. Minimizing the term $||\Gamma_S(S-R)||^2$ ensures that $S-R$ is in the null space of the operator Γ_S. The advantages of NSP are that the design of the operator Γ_S can be customized to the target signal S, and that the operator's parameters and the Lagrangian multipliers can be adaptively estimated [16]. More details about the NSP algorithm is available in [16] on signal decomposition. Figure 2 shows an example of extracting the trend from an amplitude spectrum.

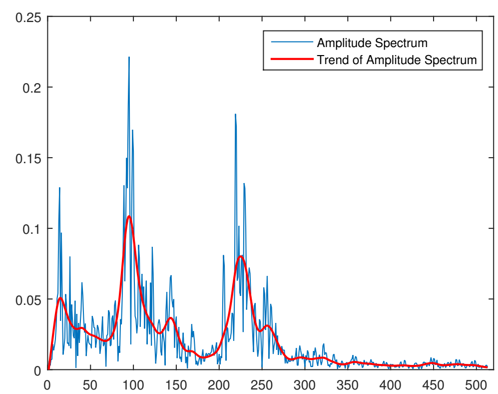


Fig. 2. Illustration of the trend of amplitude spectrum.

B. Null space pursuit coefficient (NSPC) Algorithm

The proposed NSPC algorithm consists seven steps as follows:

1) Pre-process: Pre-emphasis technique is used in order to compensate the high frequency loss in upper airway. Pre-emphasis filter is generally set to

$$H(z) = 1 - az^{-1}, \tag{3}$$

where a is a const.

The snores are framed into 1024 points frames, with 50% overlap between frames. Each frame is windowed with

[1]Source codes of NSP can be downloaded from:`http://mda.ia.ac/people/huxy/`

hamming window for reducing leakage in the frequency domain

2) The spectrum $S(k)$ of each frame $s(n)$ is obtained by applying the discrete fourier transform (DFT):

$$S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{-2\pi jkn/N}, 1 \leq k \leq K \qquad (4)$$

where N is 1024, $h(n)$ is a N sample long hamming window, and K is the length of the DFT.

3) The power spectral $E(k)$ of $s(n)$ is computed by:

$$E_i(k) = |S_i(k)|. \qquad (5)$$

4) The trend of $E(k)$, denoted as T_k, is extracted using NSP operator:

$$T_i(k) = NSP(E_i(k)), \qquad (6)$$

5) After applying the Mel-frequency filter banks to the extracted trend $E(k)$, the first 13 coefficients are kept and denoted as null space pursuit coefficient (NSPC).

6) Differential (first-order difference) and acceleration (second-order difference) coefficients are used to represent the dynamic information of snore. Differential coefficients is obtained due to the following equation:

$$d_t = \frac{\sum_{j=1}^{J} j(NSPC_{t+j} - NSPC_{t-j})}{2\sum_{j=1}^{J} j^2}, \qquad (7)$$

where d_t is the differential coefficient, $NSPC_{t+j}$ denotes the $(t+j)$-th element of NSPC, J denotes 2. Acceleration coefficients can be obtained in the same way by applying the equation (7) to differential coefficients.

At last, a 39 long feature vector is obtained for each frame. Due to the different length of snores, the number of frames are not the same. We solve this problem by averaging frames of each snore as its final feature.

7) Final feature is obtained by applying PCA on the original 39 dimensional feature vector and then fed to SVM for classification.

III. Experiments and results

A. Data acquisition and labelling

Since the study of snoring is still under development, there are no open access dataset yet. According to the experiment design we have built up a snore dataset, and will make it public soon.

The data collection was conducted in the Sleep Center of South Campus of Guang'anmen Hospital, China Academy of Chinese Medical. The study is approved by the ethical committee of Guang'anmen Hospital, China Academy of Chinese Medical. Simultaneous with PSG, an audio of overnight breathing sound per patient was recorded by an iPhone 4. The phone was placed on the table 50cm away left or right side of the head of the patient, and the sampling frequency was set as 44.1kHz with 16-Bit resolution. The PSGs were scored by sleep technician according to the AASM score manual [17]. According to the scored PSG, snores during the hypopnea events (HES) or the first snore for a short time (within 5 seconds) after an apnea event (AES) were extracted manually. A sleep expert listened to each segment, identified and labelled each snore as HES or AES manually.

We collected and annotated a dataset of 4062 snore episodes from 14 men patients. The length of snore episodes vary from about 0.2 seconds to 5 seconds. The demographic information of patients is shown in Table I.

TABLE I
THE DEMOGRAPHIC INFORMATION OF THE 14 PATIENTS

	Mean	Std	Range
Age ($years$)	39.93	10.32	26-65
AHI ($events/h$)	51.52	18.29	28.5-79.4

B. Experiment

We investigated the performance of the proposed method for snore classification. SVM was employed to classify snore episodes into categories of HES or AES. Radial basis function (RBF) kernel was used. Parameters of kernel were obtained by grid search. Original snores were segmented into frames, and each frame contains 1024 points (50% overlap). For each snore episode, a 39 dimensional vector (static coefficients, differential coefficients and acceleration coefficients) was extracted as its feature, which was fed to SVM for classification.

The classification process consisted three steps. First, we computed the classification hyperplane from the training set of snores; then the snore in test set to be classified was projected into the subspace spanned by PCA; finally, the new snore was classified by the hyperplane.

The system performance was compared with the MFCC method, the experimental results are shown in figure 3. Forty percent, 50 percent, 60 percent, 70 percent, 80 percent, and 90 percent snore episodes were randomly selected from dataset for training and the remained were for testing. No matter how many samples were trained, our method always performs better than MFCC. With the number of training samples increase, our method achieves higher accuracy, however, the performance of MFCC starts to decrease at 80 percent (training samples). It indicates that our method is more robust than MFCC, since our method discovers the underlying relationship among different snores.

The more detailed comparison of recall rate, precison, F_1, accuracy between NSPC and MFCC are given in Table II. From which, we can find that our NSPC feature outperforms the MFCC feature for snore classification.

IV. Conclusion and Future work

we propose an approach based on NSP to extract trend of amplitude spectrum of snores as feature to distinguish apnea events and hypopnea events based on HES and AES. Experimental results have shown the effectiveness of our

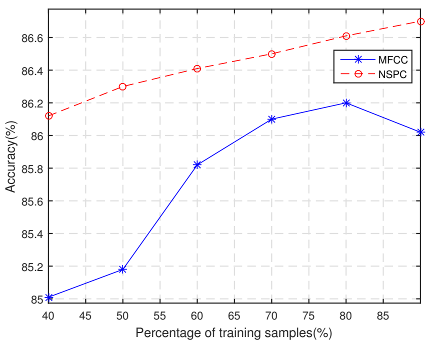


Fig. 3. Comparision of classification accuracy of NSPC and MFCC versus percentage of train samples.

TABLE II
THE DETAILS OF THE RESULTS OF EXPERIMENT

	NSPC		MFCC	
	HES	AES	HES	AES
Recall	85.39	88.74	84.56	87.42
Precision	88.51	85.82	87.25	84.87
F_1	86.89	87.29	85.93	86.18
Accuracy	87.05		85.97	

method on both recall and precision. To the best of our knowledge, this is the first devoted work on snore classification which explicitly considers the difference between hypopnea-events-snore and apnea-events-snore based solely on a snore episode. In the future, we will explore the intrinsic components of two types snore episode with signal separation.

ACKNOWLEDGMENT

We gratefully acknowledge the staff of the Sleep Center of South Campus of Guang'anmen Hospital, China Academy of Chinese Medical, especially thank Na Wang, Xinran Wang, Na Liang and Yingying Zhao for the valuable assistance with medical data acquisition and labelling.

REFERENCES

[1] C. S. Burwell, E. D. Robin, R. D. Whaley, and A. G. Bickelmann, "Extreme obesity associated with alveolar hypoventilation - a pickwickian syndrome," *Revista Medica De Chile*, vol. 20, no. 4, p. 570, 1954.

[2] P. D, A. RM, and D. M. M, "The acoustics of snoring." *Sleep medicine reviews*, vol. 14, no. 2, pp. 131–144, 2010.

[3] R. L. Wilkins, S. J. Krider, and R. L. Sheldon, *Clinical assessment in respiratory care*. Elsevier/Mosby,, 2014.

[4] G. Pillar, T. Etzioni, and N. Katz, *Acoustic snoring analysis can provide important information in OSA diagnosis.* Cambridge University Press,, 2012.

[5] F. Dalmasso and R. Prota, "Snoring: analysis, measurement, clinical implications and applications," *European Respiratory Journal*, vol. 9, no. 1, p. 146, 1996.

[6] D. L. Herath, U. R. Abeyratne, and C. Hukins, "Hmm-based snorer group recognition for sleep apnea diagnosis." in *Engineering in Medicine and Biology Society*, 2013, p. 3961.

[7] W. D. Duckitt, S. K. Tuomi, and T. R. Niesler, "Automatic detection, segmentation and assessment of snoring from ambient acoustic data," *Physiological Measurement*, vol. 27, no. 10, p. 1047, 2006.

[8] L. N. Tan and Y. Won, "Sleep snoring detection using multi-layer neural networks," *Bio-medical materials and engineering*, vol. 26, no. s1, pp. S1749–S1755, 2015.

[9] L. N. Tan and Y. Won, "Performance comparison of multilayer neural networks for sleep snoring detection," in *International Conference on It Convergence and Security*, 2015, pp. 1–3.

[10] J. M. Perezmacias, S. Adavanne, J. Viik, A. Varri, S. L. Himanen, and M. Tenhunen, "Assessment of support vector machines and convolutional neural networks to detect snoring using emfit mattress." *Conf Proc IEEE Eng Med Biol Soc*, pp. 2883–2886, 2017.

[11] M. Freitag, S. Amiriparian, N. Cummins, M. Gerczuk, and B. Schuller, "An 'end-to-evolution' hybrid approach for snore sound classification," in *INTERSPEECH*, 2017.

[12] K. Qian, C. Janott, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "Classification of the excitation location of snore sounds in the upper airway by acoustic multi-feature analysis," *IEEE transactions on bio-medical engineering*, vol. 64, no. 8, pp. 1731–1741, 2016.

[13] R. Hummel, T. D. Bradley, D. Packer, and H. Alshaer, "Distinguishing obstructive from central sleep apneas and hypopneas using linear svm and acoustic features." in *Engineering in Medicine and Biology Society*, 2016, p. 2236.

[14] R. B. Berry, *Fundamentals of Sleep Medicine*. Elsevier/Saunders,, 2012.

[15] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *Readings in Speech Recognition*, vol. 28, no. 4, pp. 65–74, 1990.

[16] S. Peng and W. L. Hwang, "Null space pursuit: An operator-based approach to adaptive signal separation," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2475–2483, 2010.

[17] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. Vaughn, "The aasm manual for the scoring of sleep and associated events," *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 2012.