Bio-Medical Materials and Engineering 26 (2015) S787–S793
DOI 10.3233/BME-151370
IOS Press

Efficient snoring and breathing detection based on sub-band spectral statistics

Xiang Sun [a], Jin Young Kim [a,*], Yonggwan Won [a], Jung-Ja Kim [b] and Kyung-Ah Kim [c]
[a] *Department of ECE, College of Engineering, Chonnam National University, Gwangju, 500-757, South Korea*
[b] *Department of Biomedical Engineering, College of Engineering, Chonbuk National University, Jeonju, South Korea*
[c] *Department of Biomedical Engineering, College of Medicine, Chonbuk National University, Jeonju, South Korea*

Abstract. Snoring, a common symptom in the general population may indicate the presence of obstructive sleep apnea (OSA). In order to detect snoring events in sleep sound recordings, a novel method was proposed in this paper. The proposed method operates by analyzing the acoustic characteristics of the snoring sounds. Based on these acoustic properties, the feature vectors are obtained using the mean and standard deviation of the sub-band spectral energy. A support vector machine is then applied to perform the frame-based classification procedure. This method was demonstrated experimentally to be effective for snoring detection. The database for detection included full-night audio recordings from four individuals who acknowledged having snoring habits. The performance of the proposed method was evaluated by classifying different events (snoring, breathing and silence) from the sleep sound recordings and comparing the classification against ground truth. The proposed algorithm was able to achieve an accuracy of 99.61% for detecting snoring events, 99.16% for breathing, and 99.55% for silence.

Keywords: Snoring, acoustic analysis, obstructive sleep apnea, sub-band spectra

1. Introduction

Obstructive sleep apnea (OSA) is a common breathing disorder which is caused by obstruction of the upper airway [1]. It is characterized by repetitive pauses in breathing during sleep, has a number of symptoms, including repetitive pauses in breathing during sleep, and is usually associated with Hypoxemia Syndrome, which can result in tiredness during the daytime, increased risk of stroke or cardiovascular disease, and even sudden apnea [2, 3]. Moreover, without identification, these symptoms may last for years, during which time the individual may become conditioned to the sleepiness and fatigue associated with sleep disturbance. Recent studies found that OSA is usually accompanied with heavy snoring, which was clinically indicated to be one of the most common symptoms, occurring in 70% to 95% of cases [4]. Earlier studies [5] indicated that the snoring may

* Address for correspondence: Jin Young Kim, Department of ECE, College of Engineering, Chonnam National University, Gwangju, 500-757, South Korea. Tel.: (+82)625301757; Fax: (+82)625301759; E-mail: beyond@jnu.ac.kr.

play a key-role in the diagnosis and differentiation between healthy individuals and OSA patients. Although, whole night polysomnography (PSG) can be used for analysis of the acoustic characteristics of a patient's sleep habits [6], this method requires a full-night diagnosis with the individual connected to a number of facilities in the diagnosis room. However, it is both difficult and expensive to implement for PSG for every patient. Therefore, a fast and efficient method to classify snoring and non-snoring noises is urgently needed.

Several related studies have produced techniques that exhibit good performance for snoring classification. Duckitt, Tuomi, and Niesler [7] applied a speech recognition technique in which the Mel-frequency cepstral coefficients (MFCC) features are extracted and then classified using a hidden Markov model (HMM), achieving a snoring classification rate of 89%. Cavusoglu *et al.* [8] developed a method that applies principal component analysis (PCA) to acquire 2-dimensional primary components from 15-dimensional sub-band spectral energy vectors, after which robust linear regression (RLR) is used for the classification. The classification accuracy of this method was 90.2% (using 15 subjects for design and testing each). Dafna *et al.* [9] later proposed a method for classification of snoring events based on a Gaussian mixture model (GMM), which involves 40-dimensional feature vectors using MFCC, time domain, and energy features. This method exhibited a detection rate of 98.1% for snoring and non-snoring sounds.

In the present study, we proposed a frame-based classification method using a set of simple and principal features. The feature employed is the normalized mean and standard deviation of the sub-band spectral energy, which shows an apparent distinction between snoring and other classes of events. We then adopted a support vector machine (SVM) for classification of each frame, and achieved an accuracy rate of 99.61% for snoring, 99.16% for breathing, and 99.55% for silence.

Section 2 describes the feature extraction, including overall structure of the system in Sub-section 2.1, acoustic analysis of the database in Sub-section 2.2, and feature extraction in Sub-section 2.3. In Section 3, details of the database and the results of the experiment are presented. Finally, the conclusion is given in Section 4.

2. Methods

2.1. Overall structure

The raw sleep recordings are processed following the proposed detection system shown in Figure 1. The overall system structure is composed of the following steps.

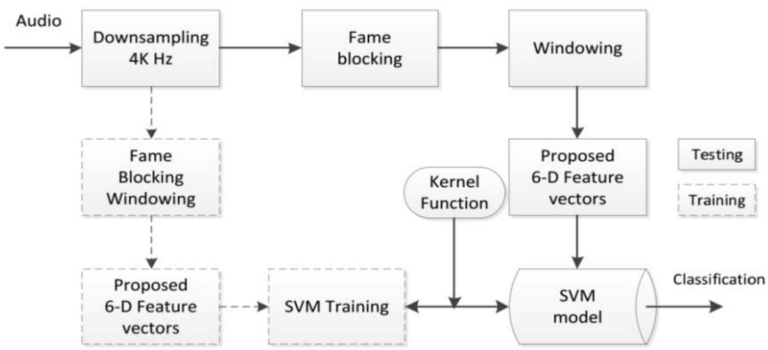
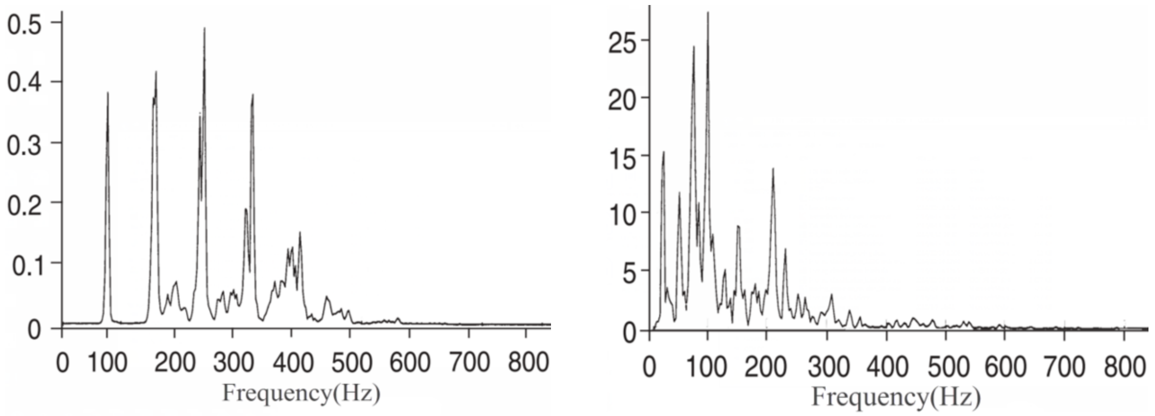


Fig. 1. Block diagram of the snoring classification system.



(a). Power spectrum of the first type of snoring. There is a fundamental frequency with clearly identified harmonics.

(b). Power spectrum of the second type of snoring. Maximum energy is around the fundamental frequency, with a lot of peaks in the range from 50Hz to 300Hz.

Fig. 2. Analysis of two types of snoring spectra.

Firstly, breathing sound signals are segmented at a constant size by framing, and the spectral features are obtained from the segmented signals. In the training stage, the SVM classifier processes features that have been tagged as silence, snoring, inspiration, and expiration, and the SVM training produces support vectors for each class. During the testing stage, the SVM classifier classifies the input feature vectors into the different breathing modes.

2.2. Acoustic analysis

Snoring samples are observed in the frequency domain. Beck *et al.* [10] previously demonstrated that snoring and breathing usually exhibit a strong energy output in the range from 64 Hz to 800 Hz. Herein, we investigated the spectral signals in that band and found the following additional distinctive characteristics.

1) Two types of snoring spectra were found. The first type, simple snoring signals, showed a power spectrum with a harmonic structure and a fundamental frequency ranging from 110-190 Hz [11], as shown in Figure 2(a).
2) The second type of snoring spectrum was characterized by a lot of low frequency peaks in the range of 50-300 Hz, with the rest of the energy of the snoring spectrum scattered around this range without clearly identifiable harmonics, as shown in Figure 2(b).
3) In the range from 300 Hz to 550 Hz, harmonic components in snoring signals were mixed with breathing noise of nearly equal strength.
4) In the range from 550 Hz to 800 Hz, breathing noises were more dominant than harmonic components, even though HCs were shown in the case of snoring.
5) Breathing sounds without snoring had a strong power output in the range from 300 Hz to 800 Hz.
6) Expiration while breathing produced a somewhat strong and short burst noise, different from that of inspiration.

The above observations indicate that the critical frequencies are 50 Hz, 300 Hz, 550 Hz and 800 Hz. That is, the band from 50 Hz to 800 Hz is of interest, and can be divided into three sub-bands according to the critical frequencies. For the features, the kurtosis and tonal properties of the snoring signals need to be considered. The tonal signals had spectral peaks and showed high kurtosis in the frequency domain. However, as the calculation of kurtosis is time-consuming, the standard deviation feature was adopted for the breathing and snoring classification.

2.3. Feature extraction

For the proposed method, a 6-dimensional feature vector was produced by computing the normalized means and standard deviations of 3 sub-bands at 50–300 Hz, 50–550 Hz, and 50–800 Hz for each frame. In addition, experiments were also carried out with non-overlapped sub-bands for comparison with the proposed method. Specifically, the sub-bands were distributed as 50 Hz to 300 Hz, 300 Hz to 550 Hz, ad 550 Hz to 800 Hz. Experimentally, the overlapped sub-bands showed better performance due to the similarity between expiration and snoring sounds in the sub-band from 550 to 800 Hz [12]. The frame size is shaped by a 128 ms Hamming window with a 50% overlap. Denoting a finite sub-band set ฿, the feature components at the jth frame are extracted as given below.

$$\mu_i = \int_{\mathfrak{B}_i} |F(j,f)|^2 \, df \tag{1}$$

$$\sigma_i = \sqrt{\int_{\mathfrak{B}_i} (|F(j,f)|^2 - \mu_i)^2 \, df} \tag{2}$$

Where ฿ $= \{[50{,}300], [50{,}550], [50{,}800]\}$, $\mathfrak{B}_i \in$ ฿, and $\mathfrak{B} = [50, 2000]$, and i is the number of sub-bands. Then, the normalized features are defined using

$$\mu_i^* = \mu_i M^{-1}, \;\; M = \int_{\mathfrak{B}} |F(j,f)|^2 df \tag{3}$$

$$\sigma_i^* = \sigma_i \mathcal{S}^{-1}, \;\; \mathcal{S} = \sqrt{\int_{\mathfrak{B}} (|F(j,f)|^2 - M)^2 \, df} \tag{4}$$

$F(j,f)$ is the fast Fourier transform in the sub-band $\mathfrak{B}_i$ at the jth frame. A median filter is applied after computing the energy of each frame. Finally, the 6-dimensional feature vector is applied to the SVM classifier for training.

3. Experimental results

A Zoom H2n portable recorder with a frequency response of 50-20000 Hz was used to record 8 hours of overnight sleep from each of four subjects with a sampling rate of 44.1 kHz, 16 bits per sample. Approximately 20% of the database was recorded using other devices. A down-sampling process was then adopted to change the sampling rate to 4 kHz. All of the subjects had acknowledged having snoring habits, and provided approximately 32 hours of recordings. The database contained sufficient snoring episodes for the analysis and experiments. The sounds present were categorized into four classes: expiration, inspiration, snoring, and silence. In order to determine the feasibility of using the proposed method to detect snoring and apnea, only the primary classes in the database were selected.

Table 1
Information about experimental datasets

	Total length	Subject A	Subject B	Subject C	Subject D
Expiration	295s(4613frames)	79s (1226frames)	83s (1239frames)	76s (1141frames)	67s (1007frames)
Inspiration	448s(6880frames)	102s(1597frames)	100s(1565frames)	125s(1957frames)	113s(1761frames)
Silence	288s(4504frames)	80s (1198frames)	68s(1070frames)	74s (1160frames)	69s (1076frames)
Snore	369s(5512frames)	95s (1487frames)	83s (1235frames)	89s (1397frames)	93s (1393frames)

Table 2
Results using the proposed feature

	Expiration	Inspiration	Silence	Snore
Expiration	**82.95**%	15.61%	0.89%	0.55%
Inspiration	11.05%	**85.52**%	0.21%	0.21%
Silence	0.02%	0.41%	**99.55**%	0.02%
Snore	0.03%	0.35%	0.01%	**99.61**%

Table 3
Results using the MFCC feature

	Expiration	Inspiration	Silence	Snore
Expiration	**78.21**%	16.18%	0.04%	5.57%
Inspiration	2.19%	**94.72**%	0.09%	3.01%
Silence	0.00%	46.42%	**53.49**%	0.09%
Snore	3.99%	0.02%	0.00%	**95.99**%

In order to gain adequate results from the data of only four subjects, we developed and tested our system using four independent experiments. Two main steps were employed to obtain the experimental data:

1) For each of the 8-hour recordings, episodes of 4 kinds of events (expiration, inspiration, silence and snoring) were randomly extracted and manually annotated.
2) In each of the four experiments, the annotated data from each subject was considered to be the testing data, while the data from the other three subjects were used for training.

The proposed approach was conducted using the datasets shown in Table 1. The data length for each class is given in the table, and the length in each cell was converted into frames using a frame size of 128 ms, with a 50% overlap. The four experiments were performed independently. In addition, the training and testing data for each experiment were non-overlapped.

The 6-dimensional feature vectors for the experimental datasets in Table 1 were produced and applied to the SVM for training and classification. The results were then compared with the ground truth for calculation of the accuracy. The confusion matrix shown in Table 2 indicates the integrated accuracy of the four experimental results. In this study, another set of feature Mel-frequency cepstrum coefficients (MFCC) was also implemented for comparison with the proposed method. MFCC has been recommended as the best features for audio event detection [13]. Therefore, several related works have used MFCC, and were able to obtain good performance. The spectral features are 42-dimensional MFCC, including log energy feature, 0th cepstrum coefficient, delta, and delta-delta coefficients. The features were also calculated every 64 ms with a 128 ms Hamming window using the same datasets, after which they were applied to the SVM classifier. The results using the MFCC features are shown in Table 3.

The performance for both kinds of feature showed that most errors occurred between expiration and inspiration. However, since expiration and inspiration can both be considered as breathing, the proposed feature achieved 99.16% performance for breath classification, while the accuracies for silence and snoring were 99.55% and 99.61%, respectively. Moreover, it can be seen from Tables 2 and 3 that the proposed feature outperformed MFCC for snoring classification. Specifically, the MFCC-based classification rates were 95.67% for breathing, 53.49% for silence, and 95.99% for snoring.

The detection performances using several episodes in our database, based on the frame-based classification results, are shown in Figure 3.

In addition, the running time for each method was also measured for comparison of the efficiencies using six samples with different data lengths. The results are shown in Figure 4.

4. Discussion and conclusion

In this study, we proposed a simple and efficient method to detect snoring, wherein the 6-dimensional sub-band standard deviations and mean power with normalization were considered as the features. The performance of the proposed system was found to be comparable to that of an

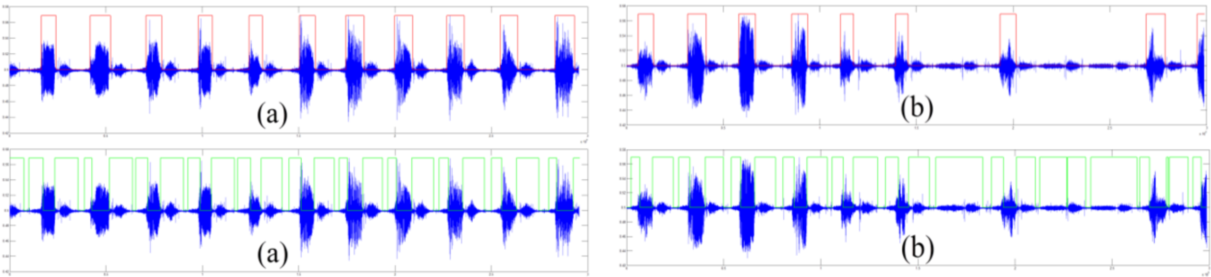


Fig. 3. Snoring and breathing episodes detection using two sequences of signal (a) and (b). The upper figures show the performances of snoring detection, the lower figures show the breathing detection.

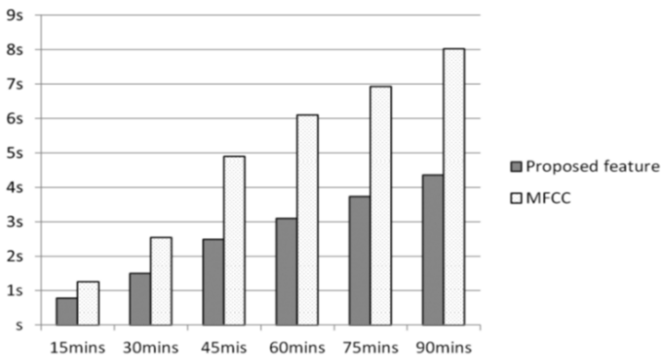


Fig. 4. The running time for the MFCC and our proposed method.

MFCC-based system, which was previously recommended as the best feature. Moreover, the database was recorded from different devices, even at various distances and angles, ensuring that the proposed algorithm is robust to microphone type and recording environment. In addition, different sampling rates were also tested in our experiments, revealing that the performances using 4 kHz were similar with those using higher sampling rates, but the calculation times were reduced. This implies that recorders of normal quality can be used for home recordings, allowing the natural sleep of the patients to be maintained during testing. The advantages of using the proposed system are that less calculation time is needed, and that it is also convenient to implement for bedside devices installed in a patient's home, or smartphone applications.

Future work includes the post-processing of snoring episode detection, identification of physiological sources of snoring, and a binary classification system to differentiate between simple snoring and the snoring from OSA patients to guide automatic analysis and diagnosis. Moreover, other kinds of feature extraction methods will also be explored using a more comprehensive database.

References

[1] M.R. Mannarino, F.D. Filippo and M. Pirro, Obstructive sleep apnea syndrome, European Journal of Internal Medicine **7** (2012), 586–593.
[2] A. Tarasiuk, S.G. Dotan, T. Simon, T. Tal, A. Oksenberg and H. Reuveni, Low socioeconomic status is a risk factor for cardiovascular disease among adult obstructive sleep apnea syndrome patients requiring treatment, Chest **130** (2006), 766–773.
[3] H.K. Yaggi, J. Concato, W.N. Kernan, J.H. Lichtman, L.M. Brass and V. Mohsenin, Obstructive sleep apnea as a risk factor for stroke and death, New England Journal of Medicine **353** (2005), 2034–2041.
[4] M. Partinen and T. Telakivi, Epidemiology of obstructive sleep apnea syndrome, Sleep, **15** (1992), S1–S4.
[5] N. Ben-Israel, A. Tarasiuk and Y. Zigel, Nocturnal sound analysis for the diagnosis of obstructive sleep apnea, Proceeding of International Conference of IEEE Engineering in Medicine and Biology Society, 2010, 6146–6149.
[6] R. Agarwal and J. Gotman, Digital tools in polysomnography, Jouranl of Clinical Neurophysiology **19** (2002), 136–143.
[7] W.D. Duckitt, S.K. Tuomi and T.R. Niesler, Automatic detection, segmentation and assessment of snoring from ambient acoustic data, Physiological Measurements **27** (2006), 1047–1056.
[8] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu and T. Akcam, An efficient method for snore/nonsnore classification of sleep sounds, Physiological Measurements **28** (2007), 841–853.
[9] E. Dafna, A. Tarasiuk and Y. Zigel, Automatic detection of whole night snoring Events using non-contact microphone, PLoS ONE 8 (2013), e84139.
[10] R. Beck, M. Odeh, A. Oliven and N. Gavriely, The acoustic properties of snores, European Respiratory Journal **8** (1995), 2120–2218.
[11] F. Dalmasso, G. Benedetto, R. Pogolotti, G. Righini and R. Spagnolo, Digital processing of snoring sounds, European Respiratory Journal **3** (1990), 528s–532s.
[12] X. Sun, J.Y. Kim and Y.G. Won, Simple and efficient spectral features for breathing and snoring sound classification, Journal of Korean Institute of Information Technology **12** (2014), 69–75.
[13] T. Kinnunen and H. Li, An overview of text-independent speaker recognition: From features to supervectors, Speech Communication **52** (2010), 12–40.