Automatic Picking of Snore Events from Overnight Breath Sound Recordings*

Vinayak R Swarnkar, Udantha R Abeyratne, *Senior Member, IEEE*, Roneel V Sharan, *Member, IEEE*

***Abstract*— Snoring is one of the earliest symptoms of Obstructive Sleep Apnea (OSA). However, the unavailability of an objective snore definition is a major obstacle in developing automated snore analysis system for OSA screening. The objectives of this paper is to develop a method to identify and extract snore sounds from a continuous sound recording following an objective definition of snore that is independent of snore loudness. Nocturnal sounds from 34 subjects were recorded using a non-contact microphone and computerized data-acquisition system. Sound data were divided into non-overlapping training (n = 21) and testing (n = 13) datasets. Using training dataset an Artificial Neural Network (ANN) classifier were trained for snore and non-snore classification. Snore sounds were defined based on the key observation that sounds perceived as 'snores' by human are characterized by repetitive packets of energy that are responsible for creating the vibratory sound peculiar to snorers. On the testing dataset, the accuracy of ANN classifier ranged between 86 – 89%. Our results indicate that it is possible to define snoring using loudness independent, objective criteria, and develop automated snore identification and extraction algorithms.**

I. Introduction

Snoring is generally considered as a perceivable vibratory sound produced by anatomical structures in the upper airway during sleep. It is a breathing-related sound more common in the inspiratory phase than the expiratory phase. The prevalence of habitual snoring is estimated to be around 40% in adult men and 20% in adult women[1] while a much higher percentage of people snore on an occasional basis.

Snoring has known to cause social and marital problems in the society. However, it is more than a mere nuisance to bed partners; it is a cardinal symptoms[2] of a serious disease called Obstructive Sleep Apnea (OSA). In OSA complete or partial obstruction develops in upper-air passage during sleep. The narrowing of upper airways leads to the vibration of the soft tissues producing snore sounds. Patients with OSA often report years of snoring prior to the onset of OSA related symptoms [2].

A number of studies [3-7] have proposed algorithms to automatically detect snore sounds, leading to mixed performance outcomes. It is impossible to compare their results due to differences in instrumentation, classification methodology and variation in study designs and test populations. The method proposed in [6, 7] was developed and validated using the same set of subjects. The accuracy of the method proposed in [4] dropped with the increased severity of OSA. The method proposed in [3] achieved high accuracy, however, the validation dataset consisted only of selected snore sounds.

Moreover previous studies [3, 4, 6, 7] have failed to provide an objective definition for the 'snore' used in their work. Dafna et al [5] defined snore sounds as "a breathing sound that occurred during an inspiration with an intensity >20 dB". This is fundamentally a loudness-based definition in which the expiratory part of the breath was not included. In addition, this method could label the loud breathing sound as snoring, which are normally not considered as snoring. In this paper, we address these issues and in particular:

(i) Follow an objective definition of 'snore' independent of the loudness of the breathing sound. Our definition is based on the key observation that sounds perceived as 'snore' by humans are characterized by pseudo-repetitive packets of energy (see Fig. 1) that are responsible for the characteristic vibratory sounds. We call the distance between such packets the 'pitch' of snoring [8].

(ii) Propose an automated method to detect snore sounds from continuous overnight recording of breath sounds recorded using a non-contact microphone. Algorithms were tested on a clinical dataset of OSA and non-OSA patients.

II. Procedure for Paper Submission

A. Definition of snore

We define **'Breathe Episode'** as the sound originated from the patient from the start of an inspiration to the corresponding end of expiration'. This period of breathe episode may contain small segments of silence, between the end of inspiration and the start of expiration.

Snore Episode is defined as a **'Breathe Episode'** containing periodic packets of energy, even for a small portion of the episode. If the breathe episode did not have these periodic structures, anywhere during the duration of episode (inspiratory or expiratory part), that sound was called as a Pure Breath Episode. Fig. 1 (a) and Fig. 1 (b) shows sound segments containing a breath episode and a snore episode respectively.

B. Recording setup

The data acquisition environment for this work is the Sleep Diagnostic Laboratory of The Princess Alexandra Hospital, Australia. Our subject population includes individuals with symptoms such as daytime sleepiness, snoring, tiredness lethargy etc. and who are suspected of OSA. They were referred to the hospital for a routine PSG test. Breathing sounds of the patients were recorded using a high fidelity computerized data recording system consists of

*Research supported by Australian Research Council, grant number DP120100141.

Vinayak R. Swarnkar, Udantha R. Abeyratne, and Roneel V. Sharan are with the School of ITEE, The University of Queensland, Brisbane, QLD 4072, Australia, (e-mail: vinayak@itee.uq.edu.au, udantha@itee.uq.edu.au, r.sharan@uq.edu.au; phone: +61-7-33669063).

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

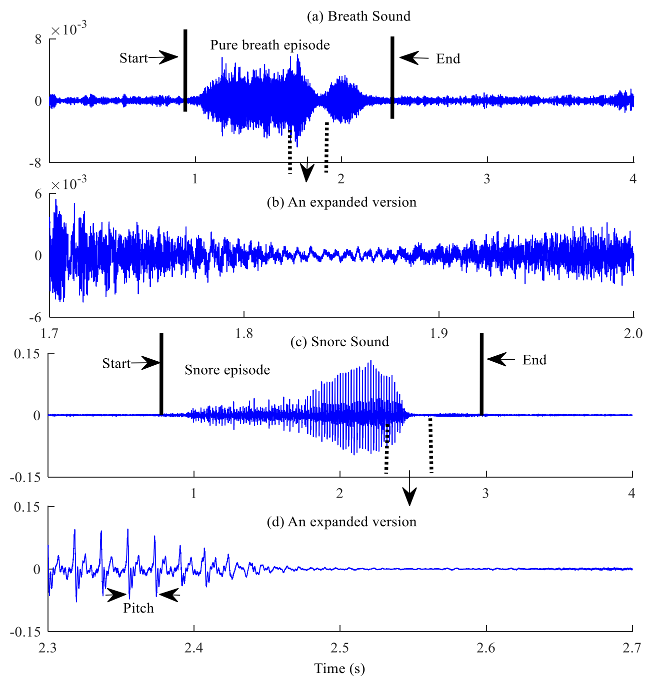

Figure 1. (a) a breath sound, (b) its expanded version (c) a snore sound segment (d) its expanded part illustrating the concept of pitch. Pitch value of the pure breath sound episode was 0 ms whereas snore episodes' pitch was detected as 18 ms. The mean loudness of pure breath episode was 57db (max = 64 db) and that for snore episodes was 85 db (max = 94 db).

a professional quality pre-amplifier, A/D converter unit (Model Mobile Pre-USB, M-Audio, California, USA) and a matched pair of low-noise microphones having cardioid directional beam pattern (Rode NT3). Sound data were recorded at a sampling rate of 44.1 K Hz with a 16-bit resolution. The nominal distance from the microphone to the face of the patients' was 50cm, but could vary between 40cm to 70cm due to patient movements. The study protocols were approved by the human ethic committee of the Princess Alexandra Hospital and the University of Queensland.

C. Sound data

In this study we analyzed nocturnal breathing sound data from N = 34 subjects. Length of sound data from each subject varied between 6 to 8 hours. We divided these subjects into two non-overlapping datasets, (i) Training dataset (D_{Tr}) and (ii) Testing dataset (D_{Ts}). The datasets D_{Tr} and D_{Ts} were mutually exclusive.

Training dataset – This dataset consisted of sound recordings from 21 subjects. The main criterion in designing a D_{Tr} is that it should contain the whole range of snore sounds and their variations as well as non-snore sounds expected in a practical setting. This way, the pattern classifier model can learn the characteristics of the variety of snore sounds and learn to differentiate them from the non-snore sounds. We designed D_{Tr} by randomly picking representative Snore Episodes and non-snore sounds from each subject in the training dataset. Non-snore sounds included Pure Breath episodes (PBE), biological sounds including speech, laughing, grunting, coughing, yawning, etc., and ambient noises such as footsteps, trolley movements, doors closing, bed creak, and duvet noise etc. It is essential to engineer the training set to avoid introducing unproductive biases to the training process. The subset of data obtained from the training set is called $D_{Tr}^{(s)}$.

Testing dataset – This dataset consisted of sound recordings from 13 subjects. While the algorithm designer is free to define the training set, the D_{Ts} cannot be anything other than the sound stream actually recorded in the hospital. In this dataset we tested our algorithms using two sound segments objectively taken from each subjects in D_{Ts}. The first segment was taken from 0^{th} minute (start of sound file) to 5^{th} minute of data. During the first segment, the patient is usually awake and active giving us an opportunity to collect a large number of non-snore sounds to test our method. The second segment was taken from start of the 1^{st} Snore Episode to the end of the 50^{th} Snore Episode. The second segment will consist of both snore sounds and non-snore sounds. The subset of data obtained from the testing set is called $D_{Ts}^{(s)}$.

D. Silence removal & manual labeling

The first step in snore identification algorithm is removing silence and identifying sound segments. For silence removal we implemented a simple algorithm proposed in [9]. This method is based on two simple audio features the signal energy and the spectral centroid.

Let P_{Tr} and P_{Ts} are the total number of sound episodes detected in $D_{Tr}^{(s)}$ and $D_{Ts}^{(s)}$ dataset respectively. All the sound episodes obtained were subjected to an episode duration test to filter out lengthy sound episodes, which are least likely to belong to snore class. We removed all the sound episodes which were > T s in duration. Optimal value for T was computed using snore episodes in $D_{Tr}^{(s)}$. Let Q_{Tr} and Q_{Ts} are the total sound episodes in $D_{Tr}^{(s)}$ and $D_{Ts}^{(s)}$ dataset respectively after removing episodes greater than > T seconds.

Manual labeling – All the sound episodes were manually labeled as either Snore Episodes or Non-snore Sounds by a human scorer by carefully listening to the sounds and simultaneously looking at the time scale and spectrogram.

(i) Any sound episodes originated from the patient and satisfying the snore definition given in section II(A) were labeled as **'Snore Episodes'**.
(ii) All the other sound episodes which does not satisfy the definition of 'snore' including sounds such as speech, laughter, cough, ambiguous noise, or pure-breathe sound were labeled as '**Non-snore sound'**.

This manual scoring is used as the reference standard against which results of automatic classification will be compared.

E. Feature extraction

We recognize that human speech and snore share many similarities in the mechanism of their generation [10]. Snore sounds are generated due to vibration of soft tissues in the upper airway that carries vital information about the changing state of upper airway. The upper airway acts as an acoustic filter during snore production similarly the vocal tract performs similar function in the case of speech. Considering the similarity of human speech production and snore sounds and rather similar anatomical structure involved in the production of two sounds, it is highly likely that acoustic features which have been successfully used in the speech analysis will also be useful in snore sound analysis. Based on this hypothesis following features were extracted from each sound episode: (i) Non-gaussianity Score, NGS – to measure the amount of deviation of a data segment from Gaussianity, (ii) Pitch – essentially refers to the fundamental frequency of the vocal cord, (iii) Log Energy – to capture loudness based characteristics of the snore, (iv) kurtosis – measures peakedness of the probability density distribution of a data, (v) Mel-frequency cepstral coefficients, MFCC – MFCC are widely used in speech recognition system and provides some resilience to the non-linguistic sources of variance in speech signal. We included 12 MFCC components. Further details of all the features we extracted could be found in [11, 12].

F. Classification model training and testing

Let M_{Tr} be a feature matrix computed using sound episodes Q_{Tr} in $D_{Tr}^{(s)}$. Now, the problem in our hand is a typical binary classification problem. We have a training data in M_{Tr} and we have a class associated with each feature vector in M_{Tr}. In this paper we investigate the use of an Artificial Neural Network (ANN) [13] as the pattern classifier. Use of ANN as classifier is inspired by the human capability to precisely identify the snore sounds from the other background sounds regardless of their intensity or duration. Moreover, ANN has the advantage of classifying data using non-linear decision boundaries, based on a process of supervised learning with a set of given examples.

ANN we trained consisted of a feed-forward network with 1 input layer equal to the size of input feature vector, **H** hidden layers of **E** neurons and 1 output layer of 1 neuron. Both values for H and E were optimized using the training dataset. Hyperbolic tangent sigmoid transfer function was used in all the layers. Network was trained using Levenberg-Marquardt back-propagation training algorithm [13] and mean square error was used as the performance estimation function.

During the model training process, we adopted a k-fold cross validation technique (setting K=4). At the end of this process, we end up with four different ANN models. The performance of these classifiers were individually evaluated on sound episodes Q_{Ts} in testing set $D_{Ts}^{(s)}$. Automatic classification results were compared against manual classification.

III. RESULTS

In this paper we have 21 patients in training dataset (D_{Tr}), with male to female ratio of 10:11 and mean age 47±12 years and mean Respiratory Disturbance Index (RDI) = 32±32. In testing dataset, we have data from 13 patients with male to female ratio 6:7 and mean age 48±11 years and mean RDI = 34±34.

The length of the training sound data we created in section II (C) was 2 hours and 32 minutes. After removing silence in section II (D), from D_{Tr} a total of P_{Tr} = 1892 sound episodes were detected. Of P_{Tr} = 1892 sound episodes 1035 were Snore Episodes and 857 were non-snore sounds. Silence removal algorithm picked >99% of sound episodes in D_{Tr} indicating a high sensitivity of algorithm in picking sounds. Histogram distribution of the snore durations indicated that more than 99% of snore episodes have duration <4s. Therefore, in duration filter at section II (D), we set T = 4 seconds. After removing sound episodes > 4s, D_{Tr} had Q_{Tr} = 1815 sound episodes, with 1025 snore episodes and 790 non-snore sound episodes.

The overall length of the testing sound data, $D_{Ts}^{(s)}$, created in section II (C), was 2 hours and 20 minutes. It had 650 snore episodes from 13 subjects. After removing silence in section II (D), from $D_{Ts}^{(s)}$ a total of P_{Ts} = 1551 sound episodes were detected with 640 snore episodes. Ten snore episodes were missed by silence removal algorithm. After removing sound episodes > 4s, $D_{Ts}^{(s)}$ had Q_{Ts} = 1437 sound episodes, with 635 snore episodes and 802 non-snore sound episodes.

A. Cross-validation results

In section II (F) we trained ANN classifier following k-fold cross validation technique setting K=4 to separate 'snore sounds' from 'other sounds'. ANN architecture were optimized for (i) number of neurons **E** in hidden layer & (ii) for number of hidden layers **H**. To optimize for number of neurons, we use H = 1 and varied E from 2 to 50 and computed training and validation performances. At start ANN performances increases rapidly with increase in E. After E crosses 10 improvement is minimal and after E>15 no improvement can be seen in ANN performance. On contrary to this time taken to train an ANN is constant till E≤15 and starts increasing thereafter. Therefore, we set E = 15. Next using E = 15 we optimized for number of hidden layers by varying H from 1 to 10. There was no change in ANN performance with increase in H however time to train ANN does increases considerably. Therefore, we set H = 1.

The Table I shows k-fold cross-validation classification results for ANN (setting E = 15 and H =1). On validation dataset, ANN can identify snore sounds with mean sensitivity of 94±2% and specificity of 93±4%. The kappa value for ANN were 0.86±0.02, 'almost perfect agreement'[14].

B. Testing results

Training using k-fold cross validation, resulted in four

TABLE I. CLASSIFICATION RESULTS ON TRAINING DATASET USING K-FOLD CROSS VALIDATION. RESULTS ARE MEAN ± STANDARD DEVIATION VALUES.

	Sensitivity	Specificity	Accuracy	PPV	NPV	Kappa
Training	98±1	95±0	97±1	96±0	97±2	0.93±0.01
Validation	95±2	91±4	93±2	93±3	93±1	0.86±0.04

TABLE II. CLASSIFICATION RESULTS ON TESTING DATASET, $D_{Ts}^{(s)}$.

	Sensitivity	Specificity	Accuracy	PPV	NPV
K=1	87	89	88	88	88
K=2	90	88	89	87	91
K=3	84	87	86	85	86
K=4	82	89	86	87	85

sets of ANN models. Each of these models were tested on a testing dataset $D_{Ts}^{(s)}$.

As an example, Fig. 2 shows the results of classifier in identifying snore sounds from continuous audio data. Audio data in this Fig. 2 is from the testing dataset. According to this figure, silence removal algorithm picked the sound episodes with a very high accuracy. In Fig.2 example, the ANN classified these sound segments into ‘snore’ and ‘other sounds’ with 100% accuracy. Table II shows classification results of ANN classifiers on complete testing dataset.

IV. Discussion and Conclusion

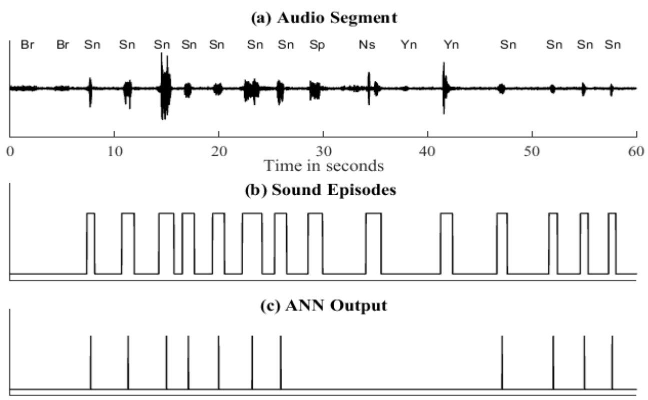

Figure 2. Automatic identification of snore episodes from a continuous audio segment of 1-minute duration from the testing dataset. (a) Audio signal (b) Identified sound episodes after removing silence; (c) Classification of sound episodes into ‘Snore’ or ‘non-snore sound’ using ANN classifier. Br – Breathe; Sn – Snore; Sp – Speech; Ns – Noise; Yn – Yawn; FP – False Positive; FN – False negative.

In this paper, we developed and tested a new algorithm to identify snore sounds from the continuous nocturnal sound recording. Classification algorithms were developed using breathing sound data from 21 subjects, following a k-fold cross validation technique. Algorithm were further tested on sound data from 13 subjects, different from training dataset. On testing dataset, the ANN classifier achieved mean sensitivity and specificity of 87.5±3.5% and 88±1% respectively.

In the past research has shown that accuracy of snore detection algorithms drops in patients with higher RDI[4, 15]. To test this hypothesis, we divided our subject population in testing dataset into two groups, groups 1 with subjects RDI<15 and group 2 with RDI >15. Then we computed the performance of the classifiers separately on these groups. The Fig. 3 shows the boxplot for ANN classifiers, illustrating model performance on subjects with RDI<15 and RDI>15. One-way analysis of variance test revealed no significance difference in model performance within the two groups, p>0.9.

Snore is one of the easiest biological signals available to develop a simple, non-contact technology to assess sleep disorders such as OSA. The basic requirement of such technology is identification of snore sounds from nocturnal continuous audio recording. The proposed algorithm can act as a front-end application for any snore based OSA diagnosis/screening systems. It can also be used to compute commonly used clinical indicators such as snore density and

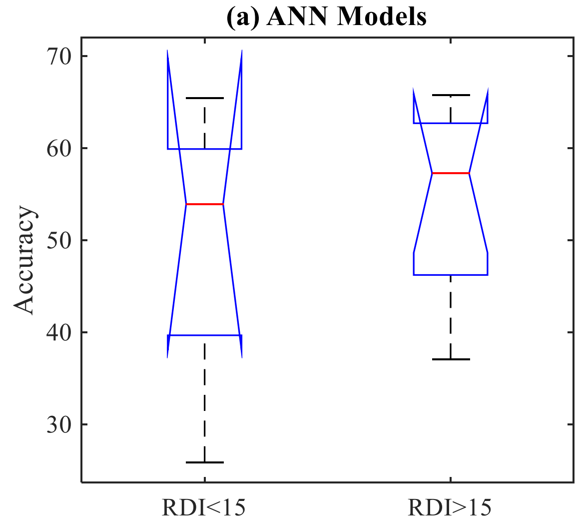

Figure 3: Boxplot showing performance of the classification models in snore identification on patients with RDI < 15 (4 subjects) and RDI > 15 (9 Subjects).

snore intensity from the whole night sound data and study snore demography with respect to sleep position, sleep stage, male/female differences in OSA patients.

References

1. Hoffstein, V., *Apnea and snoring: state of the art and future directions.* Acta oto-rhino-laryngologica Belgica, 2001. **56**(2): p. 205-236.
2. Lugaresi, E., et al., *Staging of heavy snorers' disease. A proposal.* Bulletin europeen de physiopathologie respiratoire, 1982. **19**(6): p. 590-594.
3. Azarbarzin, A. and Z. Moussavi, *Automatic and unsupervised snore sound extraction from respiratory sound signals.* Biomedical Engineering, IEEE Transactions on, 2011. **58**(5): p. 1156-1162.
4. Cavusoglu, M., et al., *An efficient method for snore/nonsnore classification of sleep sounds.* Physiological measurement, 2007. **28**(8): p. 841.
5. Dafna, E., A. Tarasiuk, and Y. Zigel, *Automatic Detection of Whole Night Snoring Events Using Non-Contact Microphone.* PloS one, 2013. **8**(12): p. e84139.
6. Duckitt, W., S. Tuomi, and T. Niesler, *Automatic detection, segmentation and assessment of snoring from ambient acoustic data.* Physiological measurement, 2006. **27**(10): p. 1047.
7. Jané, R., et al. *Automatic detection of snoring signals: validation with simple snorers and OSAS patients*. in *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE*. 2000. Chicago IL: IEEE.
8. Abeyratne, U.R., A.S. Wakwella, and C. Hukins, *Pitch jump probability measures for the analysis of snoring sounds in apnea.* Physiological measurement, 2005. **26**(5): p. 779.
9. Giannakopoulos, T. and A. Pikrakis, *Introduction to Audio Analysis: A MATLAB Approach*. 2014, Academic Press: San Diego, USA.
10. Coleman Jr, J.A., *Pathophysiology of Snoring and Obstructive Sleep Apnea.* Snoring and obstructive sleep apnea, 2003: p. 19.
11. Abeyratne, U., et al., *Obstructive sleep apnea screening by integrating snore feature classes.* Physiological Measurement, 2013. **34**(2): p. 99.
12. Abeyratne, U.R., et al., *Cough Sound Analysis Can Rapidly Diagnose Childhood Pneumonia.* Annals of biomedical engineering, 2013. **41**(11): p. 2448-2462.
13. Duda, R.O., P.E. Hart, and D.G. Stork, *Pattern classification*. 2012: John Wiley & Sons.
14. Viera, A.J. and J.M. Garrett, *Understanding interobserver agreement: the kappa statistic.* Fam Med, 2005. **37**(5): p. 360-363.
15. Azarbarzin, A. and Z. Moussavi, *Relationship Between Obstructive Sleep Apnea And Snoring Type.* Am J Respir Crit Care Med, 2012. **185**: p. A6433.