AUTOMATIC DETECTION OF SNORING EVENTS USING GAUSSIAN MIXTURE MODELS

E. Dafna,[1] A. Tarasiuk,[2] Y. Zigel[1]
[1]Department of Biomedical Engineering, Ben-Gurion University of the Negev, Beer–Sheva, Israel
[2]Sleep-Wake Disorders Unit, Soroka University Medical Center, and Department of Physiology, Faculty of Health Sciences, Ben-Gurion University of the Negev, Israel

Abstract: **In this work, an automatic snore detection system of acoustic snoring signals has been designed. Its purpose is to assist an alternative non-invasive method for diagnosing obstructive sleep apnea (OSA) based on acoustic signal processing. The detector is based on Gaussian mixture models that were trained and validated on full night acoustic signals that were recorded from a sleep laboratory, along with polysomnographic tests taken from patients with widely distributed severity of OSA. The snore detection system includes steps from noise reduction through event detection and all the way to snore identification.**

In order to analyze the performance of our proposed detector, a total of more than 80,000 acoustic episodes from 33 different OSA patients were manually segmented into snore and non-snore episodes; among the non-snore episodes we can find a variety of sleep related noises such as blanket and pillow murmurs, moaning, groaning, coughing, and talking. The validation dataset was recorded using two different audio recorders to ensure the robustness of the detector.

The events' total identification rate was 97.12% with 96.02% positive detection of snore as snore (sensitivity) and 97.90% detection of noise as noise (specificity).

Keywords: **Obstructive Sleep Apnea, Snore detection, GMM**

I. INTRODUCTION

Obstructive sleep apnea (OSA) is a common sleep related breathing disorder in which the upper airways (UA) are collapsed, causing rapid and shallow breathing (hypopnea) or even total prevention of inhalation for at least 10 seconds (apnea), causing suffocation and frequent arousal during sleep. The main consequences of OSA are daytime sleepiness and increased risk of severe cardiovascular diseases, resulting in high risk of strokes and even sudden death [1,2].

Today, the gold standard for OSA diagnosis is polysomnography (PSG) [3] study, which requires a whole night diagnosis at a sleep laboratory while the subject is connected to numerous sensors; this study is expensive and the waiting list is long; Moreover, during this procedure sleep conditions are unnatural; these issues lead to seeking alternative methods of OSA diagnosis.

Snoring is the most common symptom of OSA, occurring in 70% to 95% of patients [4]. Snoring is caused by the vibration of soft tissues due to turbulent airflow through a narrow oropharynx in the UA [5], such a narrow oropharynx is more common among patients with OSA than subjects without OSA [6]. Earlier studies [7,8] suggested that the snores may play a key-role in detecting and distinguishing between healthy (non-OSA) and OSA patients. Since snores can be recorded using a non-contact microphone in any place, even at patients' homes, natural sleep can be obtained, and snore event detection can be used as the first stage of OSA detection system using audio signals.

In recent years, several snore/non-snore classification techniques have been published. Duckitt et al. [9] proposed a classification method for snore/non-snore episodes using mel-frequency cepstral coefficients (MFCC) with hidden Markov model (HMM) and achieved a detection rate of 82%-89% (from 6 simple snorer subjects). Cavusoglu et al. [10] proposed a method using sub-band spectral energy distributions along with robust linear regression (RLR) and principal component analysis (PCA); their detection rate was around 90.2% (using 15 subjects for each design and validation, ~9000 simple & OSA snore episodes in total).

We propose a Gaussian mixture model (GMM)–based method for snore/non-snore detection that involves acoustic feature extraction from three different feature-space domains: time, energy, and frequency. The proposed system produces a detection rate of 97.12% for snores and noise, using a real OSA population that was referred to PSG study. The system is robust for variety of snores, regardless of the subjects' gender or their OSA severity.

II. METHODS

Thirty-three patients (over 18 years old) scheduled for the sleep laboratory, were recorded during one night with a digital audio recorder device (EDIROL R-4) using a directional microphone (RODE NTG-1) at a distance of 1 meter above the head level and stored along with the PSG signals; the acquired audio signals are digitized at a

Claudia Manfredi (edited by), *Models and analysis of vocal emissions for biomedical applications : 7th international workshop : August 25-27, 2011*, ISBN 978-88-6655-009-9 (print) ISBN 978-88-6655-011-2 (online)

sampling frequency of 44.1 kHz, PCM, 16 bits per sample.

The raw audio signal is processed using the proposed snore detection system which is shown in Fig. 1.

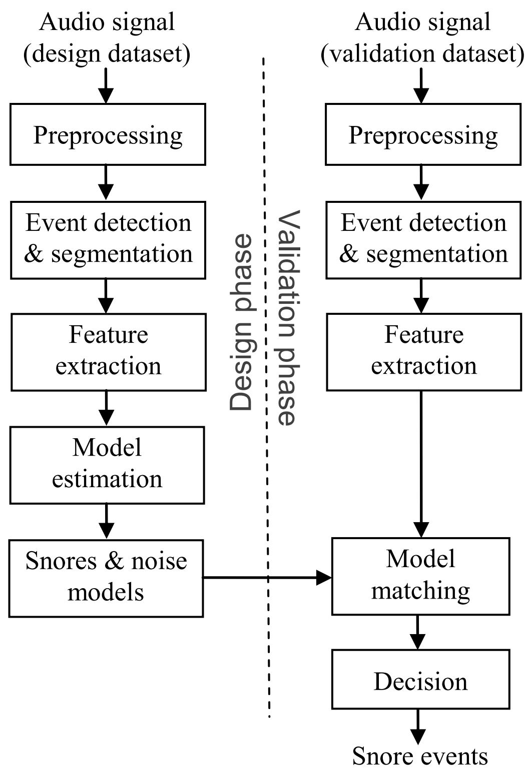


Fig. 1. – Block diagram of the snore detection system

Preprocessing

The signal is down-sampled to 16 kHz. A noise reduction (spectral subtraction) algorithm is applied on the full night audio based on the Wiener-filter, which is based on tracking a priori SNR using the decision-directed method proposed by Scalart et al. [11].

Event detection & segmentation

The audio events were automatically detected and segmented using an adaptive energy threshold.

In order to do so, the event detection module must include a few steps in order to achieve a potential snore event. At first the full night audio signal is analyzed in chucks of one-minute segments, for every segment an energy vector is calculated with a frame size of 60 ms and 75% overlap; the combined energy vectors are stored and will be used in the following steps. Later, a threshold is calculated using the estimated probability density function (pdf) of the energy values, where the first peak of the pdf is considered to be the background noise. With the completion of calculating every minute of the entire audio file, a smoothing technique based on median values is applied on the threshold vector to smooth outlier values.

With every energetic event that surpasses the relevant threshold, a boundary fine tuning technique is applied, based on the slopes of the linear regression fitting curves of 10 energetic samples (150 ms window) outside the event region on both sides; the event region is increased on each side as long as the slope does not change its sign.

Next is the fragmentation test – some of the events that are too close to each other (< 200 ms) are suspected to be a fragmented event (such as split snores); these events undergo a spectral similarity test of the 100 ms adjacent windows; in case of similarity the events are merged to form one event.

In order to improve detection efficiency, an event duration test is applied to remove unlikely snore events; only 200 ms to 3500 ms events are considered to be a potential snore event and sent to the snore classifier model.

Feature extraction

At this stage from each suspected event a 40-dimensional feature vector is calculated, consisting of three different sets of features:

1) Energy set

We included seven features such as skewness and kurtosis both for energy distribution in amplitude and in time, a normalized area beneath the energy envelope when a square shape represents one, a volume density rate [(max-min)/max] of the energy, and slope, which represents the slope from the beginning to the highest peak within the energy normalized duration.

2) Spectral-domain set

Twenty-seven features were included in this set. First we calculated 20 MFCCs for every 16 ms window (with a 50% overlapping) of the entire event; from that MFFC matrix, some of the features are extracted. We included the median (along time) of the first 16 of 20 coefficients from the MFCC matrix.

Two dynamic MFCC's distance (d_1 and d_2), which measure the MFCC's variance along time (1):

$$d_1 = \frac{1}{20}\sum_{k=1}^{20} VAR[MFCC(k,n)] \qquad (1)$$

Where the variance (VAR) of MFCC(k,n) is estimated along time n and another version of distance uses the derivatives of MFCCs (2):

$$d_2 = \frac{1}{20}\sum_{k=1}^{20} VAR\left[\frac{d}{d\tau} MFCC(k,n)\right] \qquad (2)$$

We also included a spectral flux, which was measured as the variance of the DFT along time (32 ms window duration, amplitude in dB).

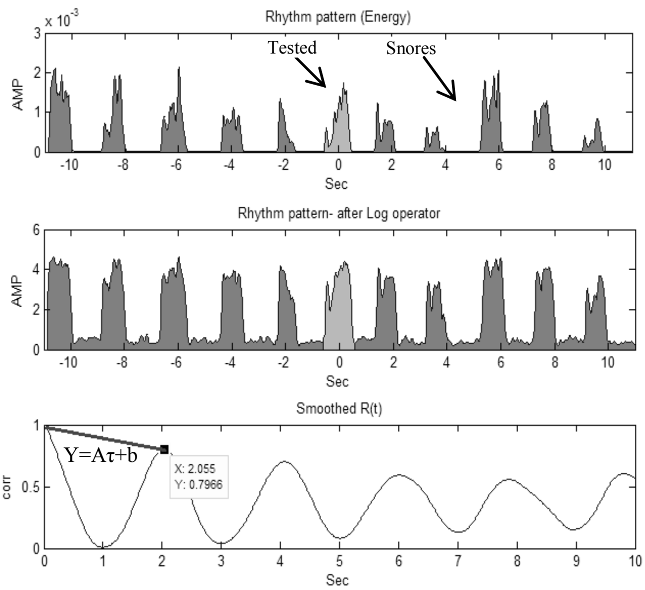
A four sub-band frequencies distribution with a bandwidth of 2 kHz each – but only the first three was taken, frequency centroid, and the difference between the centroid of the initial half episode and the second half episode over time.

Pitch related features were added as well, such as pitch, pitch strength, and pitch density [8].

3) Time-domain set

Six features were included in this set such as episode duration, zero-crossing rate, rhythm period, and period strength; For the two last features we seek for a snoring pattern (evenly repeated events) which is calculated via autocorrelation of a 20 sec interval which includes the energy signal of the event surroundings; the rhythm period is the location (in time) of the first $R(\tau)$ peak. When the rhythm strength is measured as the product of the peak value of the first $R(\tau)$ and the variance between the $R(\tau)$ curve and the $A\tau+b$ linear regression fitting of $R(0)$ to that peak, the more "delta" shape the $R(\tau)$ the greater the error and therefore the greater the strength of the rhythm. A demonstration of snores rhythm is shown in Fig. 2; while this feature alone cannot be relied on, with the addition of the rest of the features, its contribution is a major addition in discrimination between snores and uncorrelated noises, but cannot be relied on alone.

We also added the ratio between forward and backward rhythm periods of the adjacent events as well as the strength of the ratio measured as the root square of the product between forward and backward strengths.



Demonstration of snores rhythm

Fig. 2. – The upper figure represents the energetic pattern of snores when the middle event is the tested snore. To emphasize the rhythm, a log operator and rescaling is applied to the energetic segment as shown in the middle figure. At the bottom is the autocorrelation R(t) of the segment.

Model estimation

In the design phase, two GMM-based models are estimated, one for snore events and one for non-snore events (model order 7 for snores and order 32 for non-snores) using feature vectors from the first 20 patients (design data). For this process, manually labeled events were used.

Model matching

Using the estimated models and the feature vectors, calculated from each event from the validation dataset (13 patients), a general classification decision (snore/non-snore) is performed using log-likelihood ratio (LLR) scores.

Decision

An adaptive LLR threshold is calculated using all the scores to assemble a pdf in order to find a minimum between the bi-modal Gaussian densities.

Among the events we recorded 35000+ noise events and 46000+ snores – both from simple and OSA snorers; the noise events were assembled from breathing, talking, blanket noises, and other non-snore events.

The tested group (for validation) contained 13 subjects, three of whom were recorded using another portable hand recorder (Olympus SL5, sampling frequency of 44.1 kHz, PCM, 16 bits per sample) located on the dresser beside the pillow in order to see if different recording devices and microphones can be used, although it was not included in the training (design) process.

III. RESULTS

The experiment was conducted using the database that is shown in Table 1.

Table 1 –Subjects' database information

	All	System Design	System Validation
# Subjects	33	20	13
Gender(M/F)	18/15	9/11	9/4
AGE - range	25-82	37-82	25-81
(mean ± std)	51.9±12.4	54.0±10.9	48.1±14.5
AHI - range	2.2-64.9	2.2-64.9	5.9-47.9
(mean ± std)	18.3±15.3	16.9±16.6	20.7±13.1
BMI - range	22.9-39.1	26.4-38	22.9-39.1
(mean ± std)	29.5±6.9	28.7±7.5	31.0±5.8
Snores (M/F)	23125/22109	16127/17502	6998/4607
Noise Events (M/F)	21486/13685	8457/9971	13029/3714
Device (RODE NTG-1/LS-5)	30/3	20/0	10/3

The design set for the model estimation has almost an equal number of men and women and a wide range of OSA severity – AHI of 2–65. For the validation set we included also three recordings from a handy recorder as shown in Table 1; we deliberately validated episodes recorded from a handy recorder in order to eliminate over fitting of the signals to our microphone's specs. The overall detection rate was 97.12% with 96.02% for snores and 97.90% for noise, with a confusion matrix as shown in Table 2.

Table 2 – Classification results

True label \ Class. As	Snore	Noise
Snore	**96.02%**	3.98%
Noise	2.10%	**97.90%**

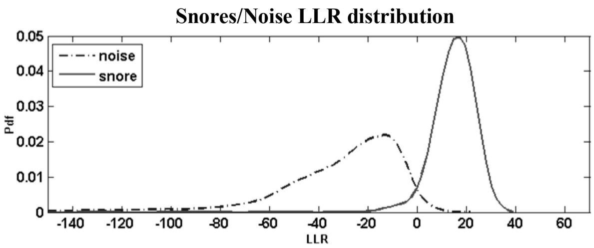


Fig. 3. – Log likelihood ratio (LLR) scores of snore and noise events

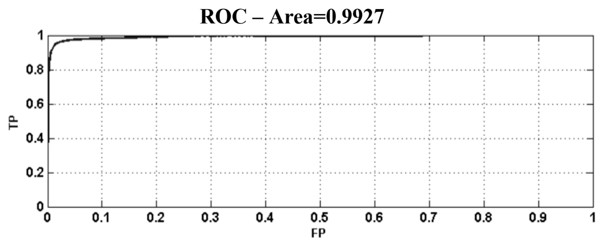


Fig. 4.– ROC curve – detection rate (True positive, TP) vs. (False Positive, FP) of snores

IV. DISCUSSION

The experiment was conducted on a total of 80,000+ audio events (snores and noise), which were extracted from 33 subjects as shown in Table 1, simple and OSA snorers, men and women – ensuring the robustness of the classifier. Furthermore, the validation group was assembled from signals that were recorded from different devices and even from different angles and distances, ensuring that the classification algorithm is robust to microphone type and angle; this implies that a small high-quality audio recording device can be used for home recordings, keeping the natural sleep of the patient.

According to the ROC curve in Fig. 4 and the LLR distribution in Fig. 3, we noticed that most of the errors were caused by a lack of distinction between smooth snores and breaths.

V. CONCLUSION

This paper proposed a snore detection system. The performance of the system is very encouraging – the detection rate is superior to earlier reported papers [9,10] and the system is ready for the next step – classification of OSA patients using snore analysis.

REFERENCES

[1] H.K. Yaggi, J. Concato, W.N. Kernan, J.H. Lichtman, L.M. Brass, V. Mohsenin, "Obstructive sleep apnea as a risk factor for stroke and death", N Engl J Med, vol. 353, pp. 2034-2041, 2005.

[2] A. Tarasiuk, S.G. Dotan, T. Simon, T. Tal, A. Oksenberg, H. Reuveni, "Low socioeconomic status is a risk factor for cardiovascular disease among adult obstructive sleep apnea patients requiring treatment", Chest, vol. 130, pp. 766-773, 2006.

[3] W. Flemons, "Obstructive sleep apnea", N Engl J Med, vol.347, pp. 498-504, 2002.

[4] M. Partinen, T. Telakivi, "Epidemiology of obstructive sleep apnea syndrome", Sleep, Vol.15, pp. S1-S4, 1992.

[5] V. Hoffstein, "Snoring", Chest, vol. 109, pp. 201-222, 1996.

[6] A. Malhotra, D.P. White, "Obstructive sleep apnea", Lancet, vol. 360, pp.237-245, 2002.

[7] A.K. Ng, T.S. Koh, E. Baey, T.H. Lee, U.R. Abeyratne, K. Puvanendran, "Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?", Sleep Medicine, vol. 9, issue. 8, pp. 894-898, 2008.

[8] N. Ben-Israel, A. Tarasiuk, Y. Zigel, "Nocturnal sound analysis for the diagnosis of obstructive sleep apnea", Conf Proc IEEE Eng Med Biol Soc., pp. 6146-6149, 2010.

[9] W. Duckitt, S. Tuomi, T Niesler, "Automatic detection, segmentation and assessment of snoring from ambient acoustic data", Physiol Meas, vol. 27, pp. 1047-1056, 2006.

[10] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu, T. Akcam, "An efficient method for snore/nonsnore classification of sleep sounds", Physiol Meas, vol. 28, pp. 841-853, 2007.

[11] P. Scalart, J.V. Filho, "Speech enhancement based on a priori signal to noise estimation", Conf Proc IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp.629-632, 1996.