Detection of Breathing Sounds During Sleep Using Non-Contact Audio Recordings

T. Rosenwein, E. Dafna, A. Tarasiuk, Y. Zigel

***Abstract*— Evaluation of respiratory activity during sleep is essential in order to reliably diagnose sleep disorder breathing (SDB); a condition associated with serious cardio-vascular morbidity and mortality. In the current study, we developed and validated a robust automatic breathing-sounds (i.e. inspiratory and expiratory sounds) detection system of audio signals acquired during sleep. Random forest classifier was trained and tested using inspiratory/expiratory/noise events (episodes), acquired from 84 subjects consecutively and prospectively referred to SDB diagnosis in sleep laboratory and in at-home environment. More than 560,000 events were analyzed, including a variety of recording devices and different environments. The system's overall accuracy rate is 88.8%, with accuracy rate of 91.2% and 83.6% in in-laboratory and at-home environments respectively, when classifying between inspiratory, expiratory, and noise classes.**

Here, we provide evidence that breathing-sounds can be reliably detected using non-contact audio technology in at-home environment. The proposed approach may improve our understanding of respiratory activity during sleep. This in return, will improve early SDB diagnosis and treatment.

***Keywords*: sleep disorder breathing, breathing-sounds detection, audio signal processing, random forest.**

I. Introduction

Sleep disordered breathing (SDB) is a prevalent disorder affecting 2% to 7% of adults. SDB is a condition characterized by repeated episodes of complete or partial collapse of the upper airways during sleep [1]. This disorder can cause snoring, excessive daytime sleepiness, cognitive dysfunction, severe cardiovascular diseases, and even death [2]. The gold standard for SDB diagnosis is polysomnography (PSG), in which patients are required to spend a night in a sleep laboratory connected to numerous sensors. PSG is expensive, unsuitable for mass screening, and may affect normal sleep [3]. Due to these disadvantages, acoustical analysis of respiratory activity during sleep has recently received considerable attention as potential alternative to PSG [4-8]. Such a non-contact acoustic technology may be applied even in the patients' home; this technology may offer as comfortable evaluation of respiratory activity in natural environment, resulting in unbiased results.

In recent years, several studies have explored snore detection. Karunajeewa et al. [5] proposed a three-class classification method of snoring (voiced non-silence), breathing (unvoiced non-silence), and silence; an accuracy of 90.74% was reported, and after noise reduction the classification accuracy improved to 96.8%. Yadollahi et al. [6] achieved accuracy of 95.7% when classifying into snore and breath classes. Azarbarzin et al. [7] classified sound episodes into snore and no-snore classes; accuracy of 93.1% was reported when using ambient microphone. All of the above studies included between 12 and 30 subjects with a relatively low number of episodes (up to 12,000 episodes).

In our previous study we developed a snore detection system [8]. By analyzing the false-detected events, we found that due to similar audio characteristics, most misclassifications were between inspiration and expiration events (episodes). Distinguishing between inspiratory and expiratory sound events will allow better estimation of breathing patterns. Moreover, several studies have found that expiratory events may hold hidden information for SDB diagnosis [9-11]. Thus, by using the additional expiratory information, SDB diagnosis and sleep quality assessment can be improved [12,13].

In the current study, we developed a breathing-sound detection system enabling to accurately differentiate between inspiratory and expiratory sound events. We propose a three-class classification algorithm that distinguishes between inspiration (snore), expiration, and noise classes, based on random forest classifier [14]. As far as we know, this is the first attempt to classify inspiratory and expiratory sounds from non-contact audio signals during sleep. The proposed system is potentially robust to variety of acoustic conditions i.e. different environments (in-laboratory and at-home), signal to noise ratios (SNR), and variety of inspiration, expiration and noise sounds.

II. Methods

An audio signals database was constructed from 84 adult patients (>18 years old) as shown in Table 1. A total of 57 patients were recorded at the Sleep-Wake Disorder Unit (Soroka University Medical Center) during the PSG examination. Fifty-two patients were recorded using a digital audio recording device (Edirol *R-4 Pro*) connected to a non-

*This work was supported in part by the Israeli Ministry of Industry and Trade, The Kamin Program, award no. 46168.

T. Rosenwein, E. Dafna are with the Department of Biomedical Engineering, Ben-Gurion University of the Negev, Beer–Sheva, Israel (rosenwei@post.bgu.ac.il).

A. Tarasiuk is with the Sleep-Wake Disorders Unit, Soroka University Medical Center, and Department of Physiology, Faculty of Health Sciences, Ben-Gurion University of the Negev, Israel (tarasiuk@bgu.ac.il).

Y. Zigel is with the Department of Biomedical Engineering, Ben-Gurion University of the Negev, Beer–Sheva, Israel (corresponding author, phone: +972-8-6428372; fax: +972-8-6428371; e-mail: yaniv@bgu.ac.il).

contact, directional condenser microphone (RODE NTG-1) that was placed 1.0 m above the patient's head. In addition, 5 more patients were recorded using a handy audio recorder (Olympus LS-5) that was attached to the dresser behind the patient’s head. This type of recorder was also used for at-home recordings. Along with the acquired signals, we included the patient's information and the data recorded from the PSG for future analysis.

Twenty-seven patients were recorded at their own homes. A technician installed a WatchPat [15], which is a portable device for sleep study, on the patient’s wrist. A handy recorder was placed on the dresser beside the patient’s bed. Both the handy recorder and the WatchPat device were simultaneously operated. The in-laboratory and at-home acquired audio signals are digitized at a sampling frequency of 44.1 kHz, PCM, 16 bits per sample.

Using the hold out method, the classification procedure was divided into two phases: design (training) and validation (testing), as shown in the block diagram in Figure 1. The system consists of four main stages: pre-processing, event detection and segmentation, feature extraction and selection, and finally model matching into inspiration, expiration, or noise classes.

A. Pre-processing

The acquired audio signals were first down sampled to 16 KHz, and an adaptive noise reduction (spectral subtraction) algorithm was applied [16].

B. Event detection and segmentation

Adaptive energy threshold was used in order to automatically find suspected energetic audio events (sound segments). After finding these energetic events, segmentation procedure was performed in order to find the exact time boundaries of events [8].

Among the events, we manually labeled more than 560K events from both non-SDB and SDB snorers. The noise events were assembled from biological resources (such as talking, blanket/pillow noises, murmurs, and groaning) and third party resources (such as cell-phone/beeper rings, shutting doors, car horns, and barking dogs). It is important to mention that the at-home recordings have considerably lower SNR than in-laboratory recordings, with a wider variety of noises, such as TV/radio noise and analog clock noise, and the frequency of their appearance is higher. Although each respiratory cycle consists of inspiration and expiration, due to limited microphones’ sensitivities, more inspiration events were detected.

TABLE I. SUBJECT’S DATABASE INFORMATION.

	All	Design	Validation
# Subjects	84	35	49
Gender(M/F)	48/36	18/17	30/19
AGE - range (mean ± std)	19.1-82 53.8±13.9	29-82 55.5±12.6	19.1-81 52.4±14.9
AHI - range (mean ± std)	2-76.8 19.6±15.5	2-66.1 21.5±18.1	2-76.8 18.3±13.4
BMI - range (mean ± std)	16.8.9-53 31.2±5.1	24.5-38 30.7±3.3	16.8-53 31.7±6.1
Inspiration Events	188850	72228	116622
Expiration Events	132519	41314	91205
Noise Events	240461	93258	147203
Edirol *R-4*/LS-5	52/32	25/10	27/22

BMI – Body mass index, AHI – Apnea and hypopnea index.

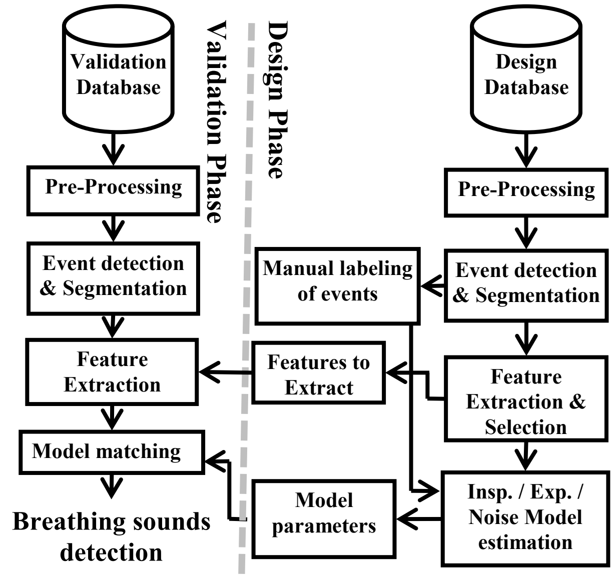

Figure 1: Block diagram of the proposed system.

C. Feature extraction & Selection

For each suspected audio event, a 351-dimensional feature vector based on energy, time, spectra, and high order spectra (HOS)-related domains was extracted, including all the 127 feature-pool described in [17]. In this work we added features from two main groups: time varying (TV) related and high order spectra (HOS) related.

The time varying subgroup consists of parametric models for power spectrum density (PSD) estimation, affected by the vocal tract’s shape; this subgroup allows us to investigate the dynamics of the signal, i.e., to characterize the way the statistical properties of the signal changes over time [18]. Since the sounds of inspiration and expiration are different, we expect that different features values will be achieved. On the entire energetic event, 60 time varying linear predictive coding (TVLPC) coefficients were extracted using 3 Chebychev base functions and model order of 20. Eighty reflection coefficients were extracted using the time varying lattice (TVLATTICE) network [19] with 4 Chebychev base functions and 20-section lattice model.

The HOS are spectral components of higher moments. The bi-spectrum $B(f_1, f_2)$ of a signal is the Fourier transform (FT) of the signal’s third order statistics. The bi-spectrum is a complex-valued function of two frequencies and exhibits symmetry; hence it was computed in the non-redundant region, termed Ω [20]. Each energetic event was framed into frames of 100 msec with 30% overlap. The bi-spectrum was computed using 512 DFT samples from an average taken over all frames. In order to emphasize the

differences, the features were extracted once as is, and using log transform. Eighty four features were extracted as follows:
Sum – the sum of absolute value of the bi-spectrum in the non-redundant region.
Projection ratio – taking the mean value of an axis (f_1 or f_2) causes a reduction of the bi-spectrum's dimensions to a vector $\bar{B}(f_2)$ or $\bar{B}(f_1)$, respectively. On this vector, 5 sub-band ratios are calculated [21].
10 sub-bands 90% – For each sub-band on each projection, the max value of the vector $\bar{B}(f_2)$ (or $\bar{B}(f_1)$) is calculated. The value of the element that has 90% of this value is saved.
12 sub-bands half point – a variation on [22] was calculated for each projection's sub-band and on the entire projection.
5 radial sub-bands (RSB) – The sum of the absolute value of the bi-spectrum in equal 5 radial sub bands were calculated according to (1):

$$RSB_{1:5} = \frac{\sum_{[r_{i-1},r_i]} |B(f_1,f_2)|}{\sum_{\Omega} |B(f_1,f_2)|}, \quad (1)$$

where $[r_{i-1},r_i]$ is the region inside Ω that is trapped between the i^{th} and the $i^{th}+1$ radius.

Gray level co-occurrence matrix (GLCM) $p(i,j)$, which calculates how often a pixel with gray level *i* value occurs adjacent to a pixel with value *j* [23], was computed on the log of the high order spectra matrix using 4 different orientations: horizontal, lateral, and diagonal (down left or up right). Using the statistics of the GLCM, several features were extracted:
Homogeneity – a measure of the closeness of the distribution of elements to the diagonal:

$$Hom = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}. \quad (2)$$

Correlation – how correlated a pixel is to its neighbors across all GLCM.

$$Cor = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)\,p(i,j)}{\sigma_i \sigma_j}, \quad (3)$$

where μ_x and σ_x are the mean and standard deviation values of variable x over the GLCM matrix.

After extracting all features, a feature normalization procedure was conducted to obtain unity standard deviation and zero mean, using the design database distribution.

Due to high correlation between features, the 150 most significant features were selected according to the design dataset, using random forest's Gini importance measure [24].

D. Model estimation

In this stage we used the design database in order to estimate the model using random forest classifier. Random forest is an ensemble of B classification trees $\{T_1(X),\ldots,T_B(X)\}$, where $X=\{x_1,\ldots,x_d\}$ is a *d*-dimensional feature vector of a suspected energetic event. Each tree returns its prediction $\{\hat{Y}_1 = T_1(X),\ldots,\hat{Y}_B = T_B(X)\}$, where $\hat{Y}_b; b=1,\ldots,B$ is the prediction made by the b^{th} tree. The majority vote of all trees determines the final prediction $\hat{Y}$ [25]. At every tree, each node is split using the best among a m_{try} subset of features randomly chosen out of the d features, where $m_{try} < d$. During the training procedure m_{try} was chosen empirically as well as the number of trees, B according to the error on the design database.

E. Model matching and breathing sound detection

After the forest is grown, the independent test database is passed down through the forest. For each event, the percentage of trees that voted for each class is returned, and the majority vote determines the final decision: inspiration, expiration, or noise:

$$\hat{Y} = \arg\max_{i=1,2,3}(s_i);\quad s_i = \frac{1}{B}\sum_{b=1}^{B} I(\hat{Y}_b = i), \quad (4)$$

where I is the indicator function that returns 1 if $\hat{Y}_b$, the b^{th} tree's decision, is equal to *i*, the i^{th} class label, and returns 0 otherwise.

III. Results and Discussion

The system was designed and validated using two different recording devices in different environmental conditions. In this way, over-fitting of the system to the microphone's specs as well as to the room parameters is reduced.

The number of trees in the random forest classifier was empirically determined to be $B = 400$ and m_{try} was set to 13

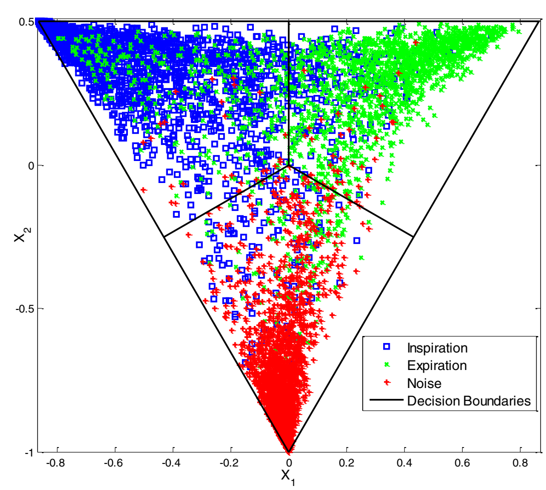

Figure 2: Projection of the events' scores (entire validation dataset) and the decision boundaries. X_1, X_2 are the 2-dimensional projection axes of the 3-dimensional decision score distribution space. As a result of the dependency between decision scores s_i, the 3-dimensional graph can be projected into 2-dimensions. For a clear display, one out of every 50 events was randomly selected due to the amount of data.

TABLE II. CONFUSION MATRIX OF ALL VALIDATION DATA.

True Label \ Classified as	Inspiration	Expiration	Noise
Inspiration	**87.4%**	8.3%	4.4%
Expiration	16.2%	**76.3%**	7.5%
Noise	0.8%	1.3%	**97.9%**

TABLE III. CLASSIFICATION RESULTS: ACCURACY

Subset category \ # of Classes	3 class: Insp.\ Exp.\ Noise	2 class: Insp. & Exp.\ Noise
In-laboratory	91.23%	96.71%
At-Home	83.61%	93.65%
All (In- laboratory and at-home)	88.84%	95.75%

according to the error on the design-set. Figure 2 shows the distribution of the projected events' scores; we can see a good separation between the three classes. Table 2 shows the confusion matrix of the validation results. The diagonal represent the true detection rate of each class. Table 3 summarizes the accuracy measure when isolating the place of recording, i.e., at-home and in-laboratory, and when observing the 3-class and 2-class (breathing vs. noise) classification problems. Although at-home recordings are characterized by low SNR, different recording devices, and environmental conditions, the system's performance did not deteriorate drastically comparing to in-laboratory recordings (see Table 3). In order to eliminate the influence of the type of recording device, a comparison between the performances of the LS-5 and the Edirol *R-4 Pro* (in-laboratory validation dataset) was conducted; no significant differences were observed.

IV. CONCLUSIONS AND FUTURE WORK

This large study included signals from different environments in at-home and in-laboratory conditions. We propose automatic breathing sounds classification system that is robust to both recording device and SNR conditions. This system may be used for further development of SDB detection and diagnosis algorithms that are based on breathing sounds analysis [12], suggesting a better examination of the entire respiratory cycle using breath-by-breath analysis. Currently, we are seeking new features that will better-discriminate inspiration and expiration events, in order to decrease our false detection rates.

ACKNOWLEDGMENT

We would like to thank Mrs. Bruria Freidman from the Sleep Wake Disorder Unit of Soroka University Medical Center, for her support and collaboration.

REFERENCES

[1] P.E. Peppard, E. Paul, T. Young, M. Palta, J. Skatrud, "Prospective Study of the Association between Sleep-Disordered Breathing and Hypertension," *N Engl J*, vol. 342, no. 19, pp.1378-1384, 2000.

[2] N.M. Punjabi, "The epidemiology of adult obstructive sleep apnea," *Proc. of the American Thoracic Society*, vol. 5, pp. 136, 2008.

[3] E. Osman, J. Osborne, P. Hill and B. Lee, "Snoring assessment: Do home studies and hospital studies give different results?" *Clinical Otolaryngology & Allied Sciences*, vol. 23, pp. 524-527, 1998.

[4] E. Dafna, A. Tarasiuk and Y. Zigel, "Automatic detection of snoring events using Gaussian Mixture Models," in *Proc. MAVEBA*, Florence, 2011.

[5] A.S. Karunajeewa, U.R. Abeyratne and C. Hukins, "Silence–breathing–snore classification from snore-related sounds," *Physiol. Meas.,* vol. 29*,* pp. 227, 2008.

[6] A. Yadollahi and Z. Moussavi, "Automatic breath and snore sounds classification from tracheal and ambient sounds recordings," *Med. Eng. Phys.*, vol. 32, pp. 985-990, 2010.

[7] A. Azarbarzin and Z.M. Moussavi, "Automatic and unsupervised snore sound extraction from respiratory sound signals," *IEEE Trans. Biomedical Eng.*, vol. 58, pp. 1156-1162, 2011.

[8] E. Dafna, A. Tarasiuk and Y. Zigel, "Automatic detection of whole night snoring events using non-contact microphone," *PLoS One*, vol. 8, pp. e84139, 2013.

[9] O. Resta, P. Guido, V. Picca, F. Scarpelli and M. Foschino, "The role of the expiratory phase in obstructive sleep apnoea," *Respir. Med*., vol. 93, pp. 190-195, 1999.

[10] M.E. Haupt, D.M. Goodman and S.H. Sheldon, "Sleep Related Expiratory Obstructive Apnea in Children," *Journal of Clinical Sleep Medicine (JCSM)*, vol. 8, pp. 673, 2012.

[11] M.H. Sanders and N. Kern, "Obstructive sleep apnea treated by independently adjusted inspiratory and expiratory positive airway pressures via nasal mask. Physiologic and clinical implications," *CHEST*, vol. 98, pp. 317-324, 1990.

[12] N. Ben-Israel, A. Tarasiuk, Y. Zigel, "Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults," *Sleep*, vol. 35, pp. 1299, 2012.

[13] E. Dafna, A. Tarasiuk and Y. Zigel, "Sleep-quality assessment from full night audio recordings of sleep apnea patients," in *EMBC*, San-Diego, pp. 3660-3663, 2012.

[14] L. Breiman, "Random forests," *Mach. Learning*, vol. 45, pp. 5-32, 2001.

[15] T. Penzel, K. Kesper, I. Pinnow, H.F. Becker and C. Vogelmeier, "Peripheral arterial tonometry, oximetry and actigraphy for ambulatory recording of sleep apnea," *Physiol. Meas.*, vol. 25, pp. 1025, 2004.

[16] P. Scalart, "Speech enhancement based on a priori signal to noise estimation," in *ICASSP 1996*, *Conf. Proc.*, Atlanta, pp. 629-632, 1996.

[17] E. Dafna, A. Tarasiuk and Y. Zigel, "OSA severity assessment based on sleep breathing analysis using ambient microphone," in *Engineering in Medicine and Biology Society (EMBC),* Osaka, pp. 2044-2047, 2013.

[18] Z. Babin, F.A. Flomen and I.D. Shallom, "Incorporation of time varying AR modelling in speech recognition system based on dynamic programming," in *Electrical and Electronics Engineers in Israel Proc.*, Tel-Aviv, pp. 289-292, 1991.

[19] Y. Grenier, "Time varying lattices and autoregressive models: Parameter estimation," in *ICASSP 1982, Conf. Proc.,* Paris*,* pp. 1337-1340, 1982.

[20] C.L. Nikias and J.M. Mendel, "Signal processing with higher-order spectra," *IEEE Signal Processing Magazine*, vol. 10, pp. 10-37, 1993.

[21] A.S. Karunajeewa, U.R. Abeyratne and C. Hukins, "Multi-feature snore sound analysis in obstructive sleep apnea–hypopnea syndrome," *Physiol. Meas.,* vol. 32, p. 83, 2011.

[22] I. Romero and L. Serrano, "ECG frequency domain features extraction: A new characteristic for arrhythmias classification," in *Engineering in Eng. Medicine and Biology Society (EMBS) Proc.*, Istanbul, vol. 2, pp. 2006-2008, 2001.

[23] R.M. Haralick, K. Shanmugam and I.H. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man and Cybernetics*, vol. 3, pp. 610-621, 1973.

[24] B.H. Menze et al., "A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC Bioinformatics*, vol. 10, pp. 213, 2009.

[25] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan and B.P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modelling," *Chemical Information and Modelling*, vol. 43, pp. 1947-1958, 2003.