35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

OSA Severity Assessment based on Sleep Breathing Analysis using Ambient Microphone

E. Dafna, A. Tarasiuk, Y. Zigel, *member, IEEE*

Abstract — **In this paper, an audio-based system for severity estimation of obstructive sleep apnea (OSA) is proposed. The system estimates the apnea-hypopnea index (AHI), which is the average number of apneic events per hour of sleep. This system is based on a Gaussian mixture regression algorithm that was trained and validated on full-night audio recordings. Feature selection process using a genetic algorithm was applied to select the best features extracted from time and spectra domains. A total of 155 subjects, referred to in-laboratory polysomnography (PSG) study, were recruited. Using the PSG's AHI score as a gold-standard, the performances of the proposed system were evaluated using a Pearson correlation, AHI error, and diagnostic agreement methods. Correlation of *R*=0.89, AHI error of 7.35 events/hr, and diagnostic agreement of 77.3% were achieved, showing encouraging performances and a reliable non-contact alternative method for OSA severity estimation.**

***Keywords*: OSA, Snoring, Signal processing, GMR.**

I. Introduction

Obstructive sleep apnea (OSA) is a chronic disorder affecting 2% to 7% of adults and can lead to significant consequences, such as cardiovascular morbidity [1, 2]. OSA involves partial or complete collapse of the upper airway during sleep, often causing noisy breathing. The gold standard for OSA diagnosis is polysomnography study [3, 4]. The commonly used score to determine OSA severity is the apnea-hypopnea-index (AHI), which is calculated as the average number of apnea and hypopnea events per hour.

Snoring is the most common symptom of OSA, occurring in 70% to 95% of OSA patients [5]. Little is known about acoustic characteristics of snoring events in adults. Earlier studies investigated snoring sound intensity [6, 7], spectral [7, 8], and pitch-related [9, 10] features. Azarbarzin et al. [11] and Ben-Israel et al. [12] proposed methods involving analysis of sequential properties of snore variations during sleep time; this kind of analysis revealed that OSA patients have greater variances of snore properties. Maimon et al. [13] showed a positive correlation of *R=0.66* between AHI and snore intensity through the night in a study conducted on 1643 subjects. It is possible that snoring signals carry essential information able to discriminate between patients of different degrees of OSA severity. However, the majority of previous literature has not attempted to estimate the AHI.

In this study, we developed and validated a snore analysis algorithm enabling estimation of apnea hypopnea index (AHI_{EST}) based solely on full-night audio signals acquired by a non-contact microphone. Moreover, we explored thousands of features corresponding to *time* and *spectra* domains, and meticulously selected an optimal set of features by feature selection algorithm. The AHI_{EST} in this case was estimated using a Gaussian mixture regression (GMR) model. Using the PSG's AHI score as a gold-standard (AHI_{PSG}), the performances of the proposed system were evaluated using Pearson correlation, AHI error, and diagnostic agreement methods.

II. Methods

The OSA severity estimation system consists of two major phases – design phase for system training, and validation phase for system evaluation. The top-level design of this system is shown in Fig. 1. Our previously published snore detector [14] and sleep/wake detector [15] were used as the first steps of each phase. Once the snores are detected, an analysis of all snore events across the night is performed; various acoustic features are extracted and investigated. The OSA severity estimation (AHI_{EST}) is performed according to a fitting (regression) model – GMR.

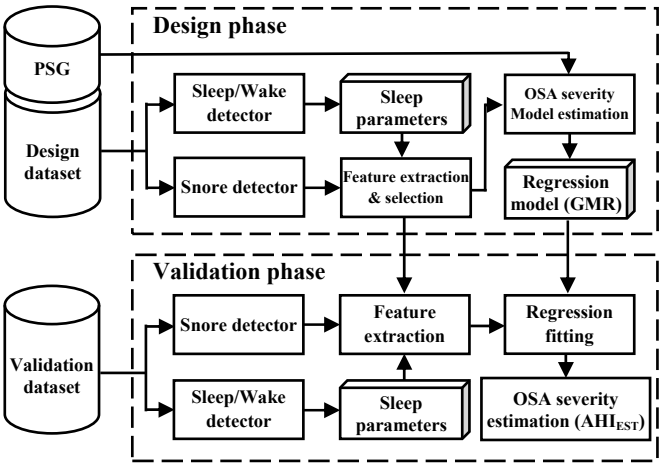

Figure 1. Block diagram of the audio-based OSA severity estimation system.

A. Experimental setup

One-hundred-fifty-five adult patients (subjects) who were referred to the sleep lab participated in this study. The subjects underwent PSG test and, simultaneously, the whole-

*This work was supported by the Israel Ministry of Industry and Trade, **The Kamin Program, award no.** 46168 **to YZ and AT.**

E. Dafna is with the Department of Biomedical Engineering, Ben-Gurion University of the Negev, Beer–Sheva, Israel (elirand@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, Beer-Sheva 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).

night audio signal was recorded using a digital audio recorder (sampling frequency of 16kHz, 16 bits per sample). The audio acquisition was performed using a non-contact microphone (Rode, NTG-1) that was connected to a digital audio recording device (Edirol, R4-pro) and placed 1m above the patient's bed. After PSG, a sleep expert scans the PSG signals and marks suspicious breathing obstructions as apnea and hypopnea events.

The OSA severity score, also known as AHI, can vary from expert to expert since human errors can occur. To get a more accurate and true result in this study, another sleep expert double checked the entire sleep time of the tested patient for errors. The Institutional Review Committee of the Soroka Medical Center approved the study protocol number 10141. Subjects' characteristics are presented in Table I.

TABLE I. SUBJECTS' CHARACTERISTICS

	System Design (n=80)	System Validation (n=75)	P value
Gender(M/F)	48/32	47/28	0.7334
Age (yr)	55.3 ± 14.8	55.4 ± 15.2	0.9670
(range)	(23-82)	(24-81)	
AHI (events/hr)	22.2 ± 18.2	22.5 ± 18.5	0.9191
(Range)	(1.0-82.2)	(0.5-87.4)	
BMI (kg/m^2)	32.0 ± 6.1	32.2 ± 6.1	0.8386
(range)	(16.8-39)	(17.2-38.6)	
ESS (score)	10.9±6.1	11.0±6.2	0.9195
Recorded length (min)	444.2 ± 64.2	452.3 ± 65.7	0.4389
Snores detected (#)	2775±1470	2821±1543	0.8495

The values are presented as mean ± SD corresponding to the relevant units. The *p* value is determined using unpaired t-test or χ^2.

In this case, the separation between the two groups (design and validation) was done randomly by maximizing the Bonferroni corrected *p* value when comparing the four parameters (AHI, BMI, Age, and Gender) in mean and standard deviation. No significant differences were found between system design (N = 80, m/f 48/32) and validation (N = 75, m/f 47/28) groups in age, BMI, snoring, ESS, AHI, associated morbidities, or tobacco smoking (Table I).

B. Snore detector

In this study we used our proposed snore detector [14]. The detector is based on signal enhancement and Gaussian mixture model fed by features extracted from time, energy, and spectra domains. The overall accuracy detection of our proposed system is > 97%.

C. Sleep/Wake detector

An additional module used in this study is the sleep quality estimation system [15]; this system automatically detects sleep/wake epochs from audio signal and calculates sleep quality parameters such as total sleep time. The overall accuracy detection of our proposed system for sleep/wake decision is about 82%. The total sleep time will help in normalizing some audio features that will be extracted later.

D. Feature extraction

One of the main challenges is to find how OSA severity is expressed through the snore characteristics. In order to do so, we would like to find appropriate features that represent the snore characteristics of a subject.

We extracted 127 snore features that have the potential to distinguish between healthy and OSA subjects, and hence may reflect OSA severity. Those features can be divided into two major feature categories corresponding to the *time* and *spectra* domains. Those feature categories are presented in Table II.

TABLE II. FEATURE CATEGORY

Feature category	Features Counts
I. Time related features	**25**
a) Periodicity features (Inter-events)	10
b) Duration and sample scattering (Intra-events)	4
c) Energy features (Intra-events)	11
II. Spectral related features	**102**
a) Spectral parameterization (Intra-events)	68
b) Bio-characteristic frequencies (Intra-events)	10
c) Dynamic frequencies features (Intra-events)	24

In order to estimate OSA severity from these snore features, a whole-night analysis is needed, therefore statistical distribution parameters such as moments (mean, variance, skewness, and kurtosis) and other parameters (min, max, median, and mode) were calculated from some of the features (Figure 2).

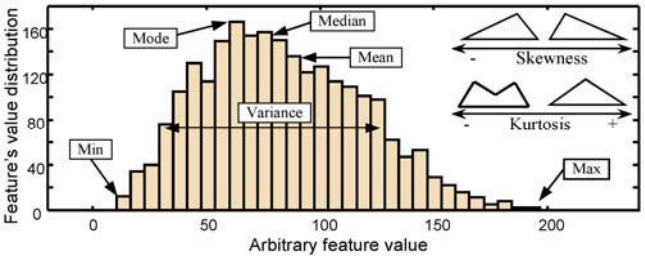

Figure 2. Visualization of eight parameters that were used to determine score distributions.

These statistical parameters were calculated from three types of time-windows: whole-night and two running windows (fixed-length window, and a group of snores window). In this way the number of features was increased.

1) Whole night snore distribution

Here we seek a parameterization for the distribution of each single snore feature (Fig. 2) among the entire signal – all detected snores. We also included multiplication or division of several feature pairs. The most intuitive example of that kind of combination is the "*number of apneas*" (counting silence segments within snore groups) divided by "*total sleep time*" (extracted using the sleep/wake detector).

2) Running windows

Here, we investigated the variability of a feature along time. For this purpose we divided the entire night into consecutive windows: either fixed length (1 minute) or group

of snores (separated by >1 min of silence). From each window that surpasses minimum snore counts, the statistical parameters were extracted. The next step is to apply the same parameterization technique among all the windows for every parameter.

Eventually, 2673 features were calculated using this parameterization technique of the different feature categorizations (Table II). Each feature was transformed into a linear scale, which is more appropriate for estimating AHI. This feature transformation was performed using a parabolic function (curve fitting) in the form of:

$$\tilde{x} = ax^2 + bx + c \tag{1}$$

where x represents the feature value before the transformation, and $\tilde{x}$ represents the transformed feature; a, b, and c represent the parabolic coefficients. Using this transformation, the feature scatter becomes more linear and can be modeled using relatively few regression parameters. An illustration of the parabolic transformation is shown in Fig. 3.

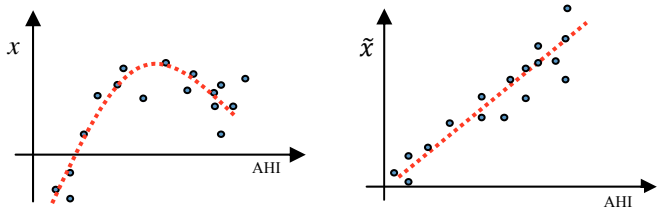

Figure 3. The left panel illustrates a feature's values distribution according to true AHI score. The right panel represents the corresponding values after the parabolic transformation. Note how the transformation improved the correlation to the true AHI and can be easily fitted using fewer Gaussians that are needed for the GMR.

E. Feature selection

Now, we have too many features – we cannot use all of them to estimate AHI; this is a classic case for using a feature selection process to reduce feature dimensionality, keeping only the most powerful OSA severity related features and avoiding over-fitting.

We used a Genetic algorithm [16] as a feature selection procedure. The criterion for feature selection was the AHI estimation error; this error was calculated as the mean absolute difference between AHI_{EST} and AHI_{PSG}. The AHI_{EST} was calculated using the GMR (see section F.)

F. AHI estimation using a regression model

In this study we used 2nd order GMR as a regression model to estimate the AHI_{EST}. The key idea of GMR is to construct a mixture of Gaussians for the joint density of the data (**x** ,y) [17]; In our case, **x** represents a feature vector (after feature selection), and y represents the suitable AHI_{PSG}. In the design phase, after Gaussian mixture model training, GMR parameters were estimated for each Gaussian (k = 1,…,K): mean vectors – $\boldsymbol{\mu}_x^k$, $\boldsymbol{\mu}_y^k$, covariance matrices – $\boldsymbol{\Sigma}_{xx}^k$, $\boldsymbol{\Sigma}_{yx}^k$, and the mixing weights, w_k [17]. In the validation phase, AHI_{EST} was calculated using the following equations:

$$AHI_{EST}(\mathbf{x}) = \sum_{k=1}^{K} w_k(\mathbf{x}) \cdot m_k(\mathbf{x}) \tag{2}$$

where **x** represents the new (tested) feature vector, and $m_k(\mathbf{x})$ represents the regression of the k^{th} Gaussian:

$$m_k(\mathbf{x}) = \boldsymbol{\mu}_y^k + \boldsymbol{\Sigma}_{yx}^k (\boldsymbol{\Sigma}_{xx}^k)^{-1} (\mathbf{x} - \boldsymbol{\mu}_x^k) \tag{3}$$

Note that in case of one Gaussian (K=1, w_1=1), eq. (3) is equal to a simple linear regression.

GMR is a complex classifier that can easily be affected by model over-fitting due to the numerous free variables such as mean vector and covariance matrix for each Gaussian. Since we performed feature transformation (see above), the complexity was reduced significantly.

III. RESULTS

One-hundred and fifty-five patients participated in the design and validation of this system (80 – design, 75 – validation). Table I summarizes the database used for the OSA recognition model.

Feature selection – Based on the outcome of the snore detector and the sleep/wake detector, features were extracted. Three features were automatically selected to estimate AHI (using the feature selection process described in section II.*E*): **1**) *The number of estimated apneas/total sleep time*, **2**) *Inter-snore homogeneity*:

$$Snore\ homogeneity = \underset{j}{\mathrm{var}}\left(\underset{i}{\mathrm{var}}\left(\underset{n}{Skew}\left(E_n\right)\right)\right) \tag{4}$$

where E_n represents the energy signal at frame index n, of snore index i within the j^{th} snore group. **3**) *The radius of the 3rd formant × energy ratio*. The *energy ratio* calculation:

$$Energy\ ratio = \frac{\sum_{n=1}^{10/Fr} E_{T_f+n}}{\sum_{n=1}^{10/Fr} E_{T_i-n}} \tag{5}$$

where T_i and T_f represent the initial and the final frame index of the tested snore respectively, and Fr is the frame rate (Fr = 0.015sec). In case of apnea event, this *energy ratio* produces extreme values (very small or very large). As mentioned, each feature value underwent a parabolic transformation in order to reduce the complexity of the GMR.

AHI estimation – AHI_{EST} was estimated by a GMR model, fed by the three features as independent variables. In order to avoid over-fitting, the GMR model included only two Gaussians. Figure 4 presents a scatter plot of AHI determined by PSG (AHI_{PSG}) versus estimated AHI (AHI_{EST}). For the design dataset, the correlation coefficient was R=0.875 (p < 0.001) and average absolute AHI error was 7.15 events/hr; the diagnostic agreement in that case was 81.25%. Using the same parameters of the GMR for the validation dataset, the correlation coefficient was R=0.892 (p < 0.001) and AHI error was 7.35 events/hr. The diagnostic agreement in that case was 77.33%. Examining the Bland-Altman plots (Fig. 5) comparing AHI_{PSG} versus AHI_{EST} showed no consistent bias, i.e., the mean difference was only 0.5 events/hr for system design and validation. The plots also show that the AHI_{EST} corresponded more closely to AHI_{PSG} when the mean AHI was <15 events/h in the validation study.

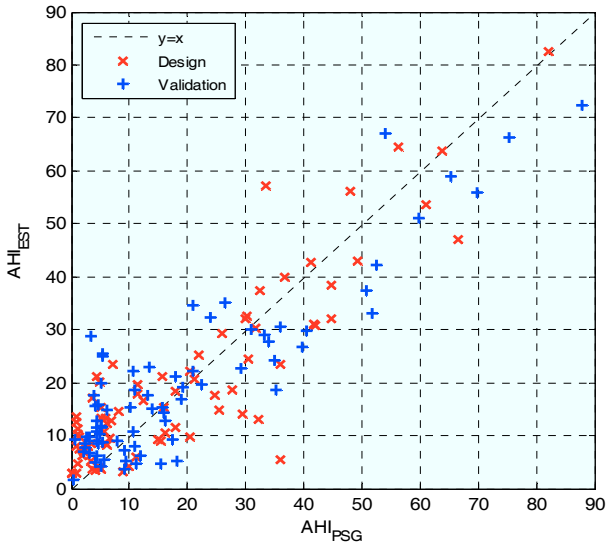

Figure 4. Regression performance of the proposed AHI estimation.

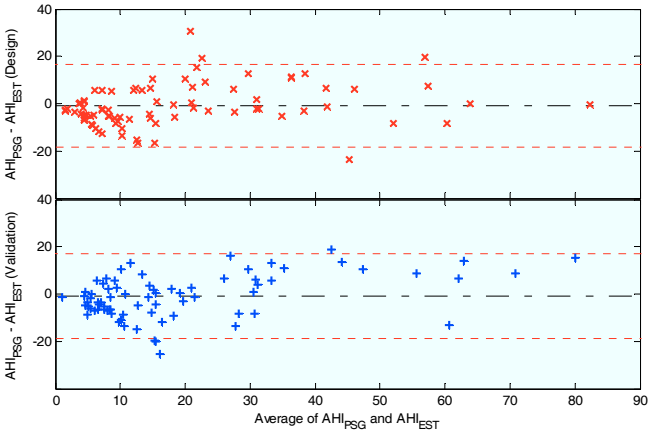

Figure 5. Bland-Altman plot of the estimated AHI derived from the design database (top panel) and from validation database (bottom panel). Lines indicate the average difference and the 2 standard deviations.

IV. Discussion and Conclusion

In this study, we explored OSA severity assessment via regression model; a huge pool (> 2600) of audio features was used, containing complementary domains such as time and spectra. Each feature underwent a parameterization process in order to assess its distribution function, and a parabolic transformation was used in order to help the regression step. Three features were selected from this feature pool using a feature selection technique.

Note that the three selected features are a combination of the time and spectra domains. The first selected feature extracts the apnea events in the audio signal and calculates the average number of apneas per hour – which is actually the definition of AHI. The sleep/wake detector [15] was an essential pre-step for this feature calculation and for some others. The second selected feature calculates the variance of the snores skewness based on the energy signal – which is a measure for inter-snore homogeneity – strengthening the theory that OSA snorers tend to have a greater variance of snores. The third selected feature can be seen as another version of apnea tracking using short-time energy ratio combined with some intra-snore spectral information.

Snores were detected from all the subjects (Table I). Generally, when the number of detected snores is higher, the estimated AHI is more accurate.

In comparison with earlier studies that estimate AHI using audio signals, our proposed method (R=0.892) is superior to others (R=0.842 in [12], R=0.66 in [13]).

In summary, a new method of estimating OSA severity was proposed and validated using 155 subjects (80 – design, 75 – validation). The method estimates AHI using GMR. Our results indicate that a diagnosis of OSA can be performed via a non-invasive, convenient, and inexpensive screening tool.

Acknowledgment

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

References

[1] N. M. Punjabi, "The epidemiology of adult obstructive sleep apnea," *Proc Am Thorac Soc*, vol. 5, pp. 136-143, 2008.

[2] A. Tarasiuk, S. Greenberg-Dotan, T. Simon, A. 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] H. Reuveni, E. Schweitzer, A. Tarasiuk, "A cost-effectiveness analysis of alternative at-home or in-laboratory technologies for the diagnosis of obstructive sleep apnea syndrome," *Med Decis Making*, vol. 21, pp. 451-458, 2001.

[4] W. W. Flemons, J. A. Rowely, W. M. Anderson, R. D. McEvoy, "Home diagnosis of sleep apnea: a systematic review of literature," *Chest*, vol. 124(4), pp. 1543-1579, 2003.

[5] V. Hoffstein, S. Mateika, D. Anderson, "Snoring: is it in the ear of the beholder?," *Sleep*, vol. 17, pp. 522-526, 1994.

[6] D. L. Van Brunt, K. L. Lichstein, S. L. Noe, R. N. Aguillard, K. W. Lester, "Intensity pattern of snoring sounds as a predictor for sleep-disordered breathing," *Sleep*, vol.20, pp.1151-1156, 1997.

[7] J. A. Fiz, R. Jané, J. Solà-Soler, J. Abad, M. A. García, J. Morera, "Continuous analysis and monitoring of snores and their relationship to the apnea-hypopnea index," *Laryngoscope*, vol. 120, pp. 854-862, 2010.

[8] A. K. Ng, T. S. Koh, U. Abeyratne, K. Puvanendran, "Investigation of obstructive sleep apnea using nonlinear mode interactions in nonstationary snore signals," *Ann Biomed Eng*, vol. 37, pp. 1796-1806, 2009.

[9] J. Solà-Soler, R. Jane, J. A. Fiz, J. Morrera, "Pitch analysis in snoring signals from simple snorers and patients with obstructive sleep apnea," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2, pp. 1527-1528, 2002.

[10] A. S. Karunajeewa, U. R. Abeyratne, C. Hukins, "Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome," *Physiol Meas*, vol. 32(1), pp. 83-97, 2011.

[11] A. Azarbarzin, Z. Moussavi, "Snoring sounds variability as a signature of obstructive sleep apnea," *Medical Engineering & Physics*, in press, 2012.

[12] N. Ben-Israel, A. Tarasiuk, Y. Zigel, "Obstructive Apnea Hypopnea Index Estimation by Analysis of Nocturnal Snoring Signals in Adults," *Sleep*, vol. 35, 1299-1305, 2012.

[13] N. Maimon, P. J. Hanly, "Does Snoring Intensity Correlate with the Severity of Obstructive Sleep Apnea?," *J Clin Sleep Med*, vol. 6(5), pp. 475-478, 2010.

[14] E. Dafna, A. Tarasiuk, Y. Zigel, "Automatic detection of snoring events using Gaussian mixture models," *Proc 7th international workshop of models and analysis of vocal emissions for biomedical applications, MAVEBA*, 2011.

[15] E. Dafna, A. Tarasiuk, Y. Zigel, "Sleep-quality assessment from full night audio recordings of sleep apnea patients," *34th Annual International IEEE EMBC Conference, IEEE EMBC*, 2012.

[16] D. Goldberg, "Genetic algorithms in search, optimization and machine learning," 1st ed. Boston: Addison-Wesley Professional, 1989.

[17] S. Calinon, Robot programming by demonstration: A probabilistic approach. CRC Press, 2009, pp. 38-44.