A NOVEL METHOD FOR OBSTRUCTIVE SLEEP APNEA SEVERITY ESTIMATION USING SPEECH SIGNALS

M. Kriboy[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, Beer-Sheva, Israel

mayakr@post.bgu.ac.il, tarasiuk@bgu.ac.il, yaniv@bgu.ac.il

ABSTRACT

Obstructive sleep apnea (OSA) is a prevalent sleep disorder associated with anatomical abnormalities of the upper airway. It is known that anatomic changes in the vocal tract affect the acoustic parameters of speech. We hypothesize that the speech signal contains valuable information that can be utilized for the assessment of OSA severity. We prospectively included 131 men with a variety of OSA severities; subjects were recorded immediately prior to polysomnography study while reading a one-minute speech protocol. Features from time and spectra domains were extracted, and a feature selection procedure was applied. Using a support vector regression (SVR), the proposed system estimates OSA severity, which is defined by the apnea-hypopnea index (AHI: the average number of apneic events per hour of sleep). Correlation of R=0.67, AHI error of 10.17 events/hr, and diagnostic agreement of 66.7% were achieved. This study provides the proof of concept that it is possible to estimate OSA severity by analyzing speech signals.

Index Terms— OSA, speech signal processing, SVR.

1. INTRODUCTION

Obstructive sleep apnea (OSA) is a common sleep disorder characterized by repeated obstructive cessation of breathing and snoring during sleep. This disorder can lead to fragmented sleep, excessive daytime sleepiness and cardiovascular morbidity. Epidemiological data estimate the prevalence of OSA in the adult population to be in the range of 3% to 7% [1]. However, most OSA patients remain undiagnosed [2]. Untreated OSA is a major health burden because of potential complications such as cardio-vascular disorders, cognitive impairment, diabetes, and depression [1].

The upper airway obstruction can be complete (apnea) or partial (hypopnea). OSA severity is typically measured using the apnea-hypopnea index (AHI), which is the average number of apneas and hypopneas during one hour of sleep; AHI of less than five is considered as normal [3]. Both anatomic and neuromuscular factors are involved in the upper airway obstruction in OSA [3]. Essentially, pharyngeal collapse can happen when the normal decrease in the pharyngeal dilator muscle tone at the onset of sleep is accompanied by a narrowed and/or highly compliant pharynx [4].

Currently, the gold standard diagnostic study for OSA is polysomnography (PSG). PSG study usually consists of recordings of various biological signals, including electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), pulse-oximetry, and airflow measurements. PSG is time consuming, expensive, and uncomfortable for the patient; therefore many patients remain undiagnosed [1]. These disadvantages led to seeking alternative approaches for OSA diagnosis.

OSA is associated with several anatomical and functional abnormalities of the upper airway, probably due to evolution of the speech production system [5-7]. In fact, in most OSA patients the proportions between the upper airway soft tissue mass and the space made by the bony structure of the upper airway are higher than normal [4]. It is well known that anatomic and functional changes in the vocal tract components affect the acoustic parameters of speech [8]. It was suggested that some acoustic speech features of patients with OSA syndrome may be distinct from those of non-OSA subjects [9]. A perceptual study [9] confirmed the clinical practice claim that some patients with OSA syndrome have unintelligible speech; it was found that patients with OSA suffer from resonance, phonation, and articulation anomalies. Robb et al. [10] compared OSA and non-OSA patients' vocal tract resonances of vowels and found differences in formant band-width and frequencies; Pozo et al. [11] found dissimilarity in the difference between second and third formant frequency in severe OSA patients and non-OSA patients. Goldshtein et al. [12] further investigated the effect of OSA on speech and a wider variety of acoustic features was examined.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

Based on the these findings, we hypothesize that some acoustic speech features of patients with OSA may be distinct from those of non-OSA patients and that we can utilize this fact to design an automated system that will estimate OSA severity.

Few studies have been dedicated to the task of OSA/non-OSA classification using speech signal analysis [12-14]. To our knowledge this is the first study that attempts to estimate OSA severity (i.e., AHI) solely based on speech signals analysis. In the current study we developed and validated a novel AHI estimation algorithm based on the support vector regression (SVR) model using a super-vector constructed from the means of an adapted Gaussian mixture model (GMM). The proposed system provides a robust, effective, and non-invasive method for the estimation OSA severity.

2. METHODS

In order to create an automatic system for OSA severity estimation, speech signals of 131 male subjects (speakers) were analyzed. The signals underwent pre-processing; a voice activity detector was implemented and features were extracted. AHI estimation algorithm based on SVR was developed and validated using the hold-out method. Figure 1 presents a block diagram of the system.

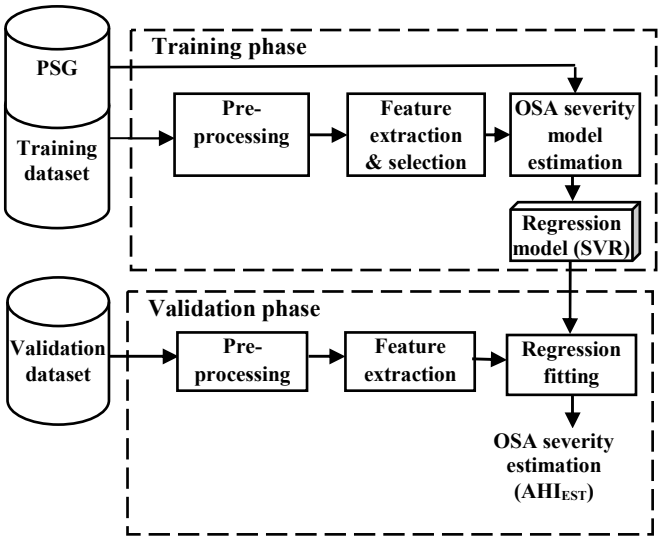

Figure 1: Block diagram of the OSA severity estimation system.

2.1. Pre-processing and feature extraction

Each digitized speech signal underwent a pre-processing procedure of DC removal, normalization, and pre-emphasizing. The signals were framed to 30 msec frames with 50% overlap; silence removal took place using a voice activity detector based on [15].

Using all the speech segments (i.e., voiced, unvoiced phonemes, and transition states), seventy-one features were extracted; the extracted features can be divided into four groups: time domain features, such as the log energy and its first and second derivatives; spectral features, such as 18 linear predictive coding coefficients (LPC) [8]; cepstrum domain features such as 16 mel-frequency cepstral coefficients (MFCC) and their first and second derivatives; and features for the detection of hyper-nasal speech, such as high and low LPC difference [16], and standard deviation of PSD [17]. Table 1 summarizes the extracted features.

Table 1: Extracted features.

Feature name	*No. of features*	*Feature symbol*
log Energy	1	E
Δ log Energy	1	ΔE
ΔΔ log Energy	1	ΔΔE
Mel frequency cepstral coefficient (MFCC)	16	c_1–c_{16}
ΔMFCC	16	Δc_1–Δc_{16}
ΔΔMFCC	16	$\Delta\Delta c_1$–$\Delta\Delta c_{16}$
Linear predictive coding (LPC)	18	a_1–a_{18}
High and low order LPC difference	1	LPCdiff
Standard deviation of PSD	1	PSD-STD
Total	71	

2.2. Feature selection

In order to avoid over-fitting and to find the most discriminative features, a feature selection procedure was performed using the sequential forward selection algorithm (SFS) [18]. Although a regression model-dependent scheme for feature selection might be more suitable for feature selection in this study, it is also very computationally expensive. Therefore, we have chosen to use a two-class classifier-dependent scheme.

A GMM-based classifier was designed using the training data. The GMM is defined as the weighted sum of *M* Gaussian component densities [19]:

$$p(\mathbf{x}\,|\,\omega)=\sum_{i=1}^{M} b_i \mathrm{g_i}\left(\mathbf{x}\,|\,\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right) \quad (1)$$

where ω is a given class, $\mathbf{x}$ is a d-dimensional data vector (feature vector), b_i is the weight of the i^{th} Gaussian, and g_i is the component density of the form:

$$\mathrm{g_i}(\mathbf{x}\,|\,\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)=\frac{1}{(2\pi)^{d/2}\left|\boldsymbol{\Sigma}_i\right|^{1/2}}\exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)}{2}\right\} \quad (2)$$

with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. These parameters are collectively represented as a model:

$$\omega=\left\{b_i,\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i\right\};\ i=1\ldots M \quad (3)$$

Two models were designed to represent the probability density of two groups: OSA (ω_O) and non-OSA (ω_{NO}). In this study we have used the area under (AU) the receiver operating characteristic (ROC) curve as the criterion for

feature selection. The AU was calculated via a *k*-fold cross-validation over the training dataset.

2.3. AHI estimation using a regression model

The OSA severity estimation system is based on the state of the art approach of support vector regression model using a super-vector constructed from the means of an adaptive GMM.

Under this scheme a GMM universal background model ω^{UBM} (GMM-UBM) with M^{UBM} components is trained. The UBM is a large (high order) GMM trained to represent the speaker-independent distribution of features [20], while the GMM is adapted from the UBM to represent an individual speaker. A key method for combining the GMM-UBM approach with support vector machine (SVM) is to represent each speaker in a high dimension feature space using a super-vector obtained by concatenating the Gaussians' mean vectors of an adapted GMM [21].

The UBM database was constructed from 85 male speakers recorded using the same recording device as the database for this study, reading a different text protocol in Hebrew. The UBM speakers were not involved in any other aspect of this study. After the UBM was trained, a model was trained for each speaker (from the OSA database) separately - $\omega^{Adapted}$, by adapting the UBM model to the speaker's feature vectors. In this research a maximum a posteriori (MAP) estimation was used for the adaptation process [20].

From each speaker's model a super vector was formed by stacking the Gaussian mean vectors of the mixture components $\boldsymbol{\mu}_i^{Adapted}$. In addition, two more features were added to each speaker's super-vector: age and BMI. Finally, the super-vector is of the form:

$$\mathbf{x} = [\boldsymbol{\mu}_1^{Adapted}, \boldsymbol{\mu}_2^{Adapted} \ldots \boldsymbol{\mu}_{M^{UBM}}^{Adapted}, BMI, Age]^T \quad (4)$$

It is well known that age and BMI are somewhat correlated with OSA severity, so adding these features may assist in increasing the system's performance.

In order to estimate the AHI, a support vector regression model was used. The SVR generates input-output mapping functions from a set of labeled training observations [22]. Given a set of training observations and labels, $D = \{(\mathbf{x}_1, \gamma_1), \ldots (\mathbf{x}_N, \gamma_N)\}$, the model seeks a function $f(\mathbf{x})$ that has the smallest deviation from the actual labels γ_i for all the training data, and at the same time is as flat as possible [22]. SVR parameters such as error cost factor *C* and kernel parameters were calibrated using leave one out (LOO) cross-validation on the training data set. Different orders of polynomial kernel were tested in this stage.

For SVR, the score for a tested observation is calculated as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \gamma_i k(\mathbf{x}, \mathbf{z_i}) + b \quad (5)$$

where $f(\mathbf{x})$ is the test score function, γ_i are the labels, α_i, α_i^*, and b are the model parameters, $k(\mathbf{x},\mathbf{z}_i)$ is the kernel function, $\mathbf{z_i}$ are the support vectors, and $\mathbf{x}$ is the tested super-vector.

3. EXPERIMENTAL SETUP

The database for this research was constructed from speech signals of 131 male subjects who were referred to the Sleep-Wake Unit of Soroka University Medical Center for PSG study; in order to evaluate sleep-disordered breathing. To avoid over-fitting, the database was divided into two separate datasets: training (*n=80*) and validation (*n=51*). Each subject was recorded using a digital audio recorder (Handy recorder "H4" by "ZOOM") reading a one-minute text protocol in Hebrew that was designed to emphasize certain characteristics of speech. We recorded at a sampling rate of 44.1 kHz and downsampled to 16 kHz (16 bits/sample). The text protocol included sustained utterance of vowels; specific long sentences containing considerable amounts of nasals and vowels; yes or no questions; and a list of isolated words. Immediately after speech recording, each subject underwent complete PSG examination; after examination, the PSG signals were analyzed and scored and an AHI value was given by the sleep clinic's medical staff. Subjects' age, BMI, and AHI are summarized in Table 2.

Table 2: Subjects' characteristics.

	Number of subjects	*AHI [events/hr] (range)*	*BMI [kg/m^2] (range)*	*Age [yr] (range)*
Training dataset	80	21.7±18.9 (2-74.7)	29.5±5.3 (17-45)	52.6±14.9 (19.1-83.5)
Validation dataset	51	21.1±18.9 (0.7-108)	29.61±5.2 (20.1-47.6)	52.2±15.0 (22.4-82.6)

The values are presented as mean ± SD corresponding to the relevant units.

4. RESULTS AND DISSCUSSION

Performance evaluation was conducted using the hold-out method. The feature selection procedure resulted in a five-dimensional subset presented in Table 3. The selected features are a combination of LPC- and MFCC-based features; this result suggests that combining different types of features can reveal more discriminative information. Those features are associated with changes in vocal tract shape and its perceptual effect [8], which are apparent in OSA patients.

Using the training data, GMM order of *M*=8 has been proven to be the most efficient. The fact that a relatively low order was the most efficient isn't surprising since we deal with a two-class classification problem. Higher order could

represent different sub-classes, such as accent and age, and therefore impair the results.

The diagnostic agreement [23] approach and correlation were used to assess the accuracy of our system in AHI prediction. Diagnostic agreement is defined when the estimated AHI and the true AHI are both above 40 (events/hr), or if the true AHI is less than 40 and the estimated AHI was within 10 events/hr from the true AHI. The motivation for using the diagnostic agreement is that small changes in AHI between our system and the PSG might be clinically unimportant. Figure 2 presents a scatter plot of AHI determined by PSG (AHI_{PSG}) versus estimated AHI (AHI_{Est}). The correlation coefficient was found to be R=0.67 (p<0.001), and average absolute AHI error was 10.17 events/hr; diagnostic agreement of 66.7% was achieved. The Bland-Altman plot (Figure 3) shows that the AHI_{Est} corresponds more closely to the AHI_{PSG} when the mean AHI was less than 35 events/hr. We can also see that there is no consistent bias, i.e., the mean difference AHI_{Est}-AHI_{PSG} was close to zero (-0.58 events/hr).

Table 3: Selected features.

Selection order	*Feature symbol*
1	a15
2	ΔΔc9
3	a17
4	ΔΔc12
5	c16

When examining Figure 2, one can see a trend toward a plateau at AHI>35 events/hr; it is noticeable that patients with high AHI are somewhat under-estimated. The figure also shows that only one patient with AHI>15 events/hr is missed; this result is encouraging, since the system is meant to be used as a tool for initial screening of OSA patients, where a low misdetection rate is very important, while under-estimation of severe patients, as long as they are diagnosed as severe, is tolerable.

5. CONCLUSIONS & FUTURE WORK

This study provides the proof of concept that it is possible to estimate OSA severity by analyzing speech signals. To our knowledge, this is the first study that estimates AHI using speech. Further studies are needed to explore the effect of age and accent (about 40% of our subjects speak Hebrew with an accent) on speech signal production of OSA patients. Since our data consists of patients referred to a sleep study for sleep disorder breathing evaluation, it appears that there is a shortage of negative control group; therefore, further validation in a broader population is required.

The proposed method can be used as a basis for future development of an initial screening tool of potential OSA patients. This type of a tool can significantly reduce the number of undiagnosed patients and reduce the number of patients referred unnecessarily to sleep clinics. The system is fully automated, and is based on speech signal recordings only; therefore it can allow effective, fast, patient-friendly, and low cost diagnosis of potential OSA patients.

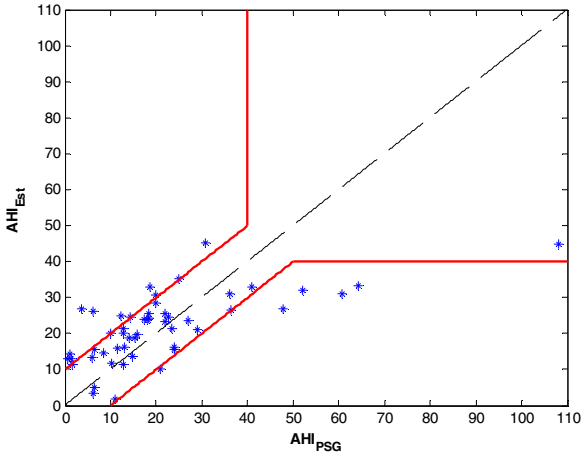

Figure 2: Estimation results, (estimated AHI VS. PSG diagnosis) the dashed line represents identity line, and solid lines represent diagnostic agreement boundaries.

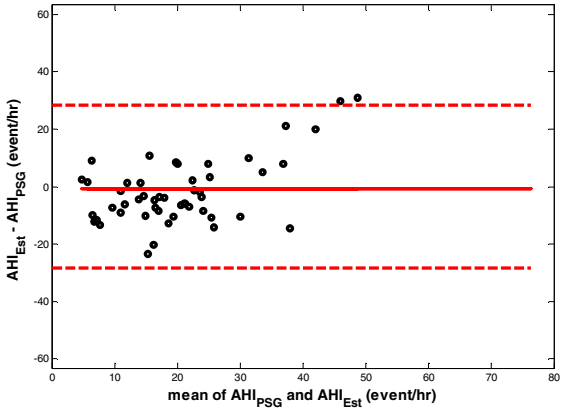

Figure 3: Altman-Bland plot. Solid line indicates the average difference and dashed lines indicate two standard deviations.

12. REFERENCES

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

[2] K. P. Pang and D. J. Terris, "Screening for obstructive sleep apnea: an evidence-based analysis," *American journal of otolaryngology,* vol. 27, pp. 112-118, 2006.

[3] M. R. Mannarino, F. Di Filippo, and M. Pirro, "Obstructive sleep apnea syndrome," *European Journal of Internal Medicine,* vol. 23, pp. 586-593, 2012.

[4] C. M. Ryan and T. D. Bradley, "Pathogenesis of obstructive sleep apnea," *Journal of Applied Physiology,* vol. 99, pp. 2440-2450, 2005.

[5] T. M. Davidson, "The Great Leap Forward: the anatomic basis for the acquisition of speech and obstructive sleep apnea," *Sleep medicine,* vol. 4, pp. 185-194, 2003.

[6] T. M. Davidson, J. Sedgh, D. Tran, and C. J. Stepnowsky Jr, "The anatomic basis for the acquisition of speech and obstructive sleep apnea: evidence from cephalometric analysis supports The Great Leap Forward hypothesis," *Sleep medicine,* vol. 6, pp. 497-505, 2005.

[7] Y. Finkelstein, D. Wexler, E. Horowitz, G. Berger, A. Nachmani, M. Shapiro-Feinberg*, et al.*, "Frontal and lateral cephalometry in patients with sleep-disordered breathing," *The Laryngoscope,* vol. 111, pp. 634-641, 2001.

[8] J. R. Deller, J. G. Proakis, and J. H. Hansen, *Discrete-time processing of speech signals*. New York: IEEE Press, 2000.

[9] A. W. Fox, P. K. Monoson, and C. D. Morgan, "Speech dysfunction of obstructive sleep apnea," *Chest,* vol. 96, pp. 589-95, 1989.

[10] M. Robb, J. Yates, and E. Morgan, "Vocal tract resonance characteristics of adults with obstructive sleep apnea," *Acta oto-laryngologica,* vol. 117, pp. 760-763, 1997.

[11] F. P. Rubén, B. M. Jose Luis, H. G. Luis, L. G. Eduardo, A. R. José, and T. Doroteo T, "Assessment of severe apnoea through voice analysis, automatic speech, and speaker recognition techniques," *EURASIP Journal on Advances in Signal Processing,* vol. 2009, 2009.

[12] E. Goldshtein, A. Tarasiuk, and Y. Zigel, "Automatic detection of obstructive sleep apnea using speech signals," *Biomedical Engineering, IEEE Transactions on,* vol. 58, pp. 1373-1382, 2011.

[13] O. Elisha, A. Tarasiuk, and Y. Zigel, "Detection of obstructive sleep apnea using speech signal analysis," in *Models and analysis of vocal emissions for biomedical applications Conf.*, Florence, 2011, pp. 13-16.

[14] M. Kriboy, A. Tarasiuk, and Y. Zigel, "Obstructive sleep apnea detection using speech signals," in *Afeka-AVIOS Speech processing Conf.*, Tel-Aviv, 2013, pp. 1-4.

[15] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *Signal Processing Letters, IEEE,* vol. 6, pp. 1-3, 1999.

[16] D. K. Rah, Y. I. Ko, C. Lee, and D. W. Kim, "A noninvasive estimation of hypernasality using a linear predictive model," *Annals of biomedical Engineering,* vol. 29, pp. 587-594, 2001.

[17] T. Pruthi and C. Y. Espy-Wilson, "Acoustic parameters for the automatic detection of vowel nasalization," in *INTERSPEECH*, 2007, pp. 1925-1928.

[18] H. Liu and H. Motoda, *Feature selection for knowledge discovery and data mining*. Massachusetts: Springer, 1998.

[19] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech and Audio Processing,* vol. 3, pp. 72-83, 1995.

[20] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital signal processing,* vol. 10, pp. 19-41, 2000.

[21] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *Signal Processing IEEE Lett.,* vol. 13, pp. 308-311, 2006.

[22] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing,* vol. 14, pp. 199-222, 2004.

[23] D. P. White, T. J. Gibb, J. M. Wall, and P. R. Westbrook, "Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home," *Sleep,* vol. 18, pp. 115-126, 1995.