OBSTRUCTIVE SLEEP APNEA DETECTION 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) syndrome is a prevalent sleep disorder that affects 2% of women and 4% of men aged more than 50 years. OSA is associated with anatomical and functional abnormalities of the upper airway due to evolution of the speech production system. It is well known that anatomic and functional changes in the vocal tract components affect the acoustic parameters of speech; hence, our hypothesis is that there is a difference between speech characteristics of OSA-patients and non-OSA subjects, and that we can utilize this fact to design an automated system that will discriminate between the two groups. The database for this study consists of 103 male subjects recorded while reading a one-minute speech protocol in Hebrew, just before they underwent a full polysomnography examination; all the non-silence segments were used for the task of OSA/non-OSA classification in order to exploit all the information hidden in the speech signal (i.e., voiced, unvoiced phonemes, and transition states). Moreover, the system doesn't require an automatic or manual phoneme identification stage prior to the classification; that allows fast and low-complexity diagnosis of OSA patients. Results of 80.65% sensitivity and 80% specificity were achieved using the hold out validation method.**

Keywords**: obstructive sleep apnea, speech signal processing, speaker recognition.**

I. INTRODUCTION

Obstructive sleep apnea (OSA) syndrome is a prevalent sleep disorder in which complete or partial airway obstruction, caused by pharyngeal collapse during sleep, leads to choking, loud snoring, frequent awakenings, disrupted sleep, and excessive daytime somnolence. The obstruction of the airway can cause complete airflow absence (apnea) or reduction in the airflow (hypopnea). OSA syndrome is defined as five or more episodes of apnea or hypopnea per hour (apnea hypopnea index – AHI) with associated symptoms (e.g., excessive daytime sleepiness, snoring, fatigue) [1]. The main factors that encourage the upper airway to collapse are anatomical features and insufficient neuromuscular compensation during sleep [2].

Recent studies suggest that 4% of men and 2% of women aged more than 50 years suffer from symptomatic OSA [1]. With the increasing prevalence of obesity (which is most potent risk factor for OSA), the number of patients who suffer from OSA syndrome has significantly increased, and is expected to continue in the same direction in the future [1]. OSA is often associated with numerous complications such as cardio-vascular disorders, stroke, diabetes, and depression [3].

The gold standard diagnostic test for sleep apnea is polysomnography (PSG) during an entire night. The PSG usually consists of recordings of various biological signals, including electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG). PSG is time consuming, labor intensive, expensive, and uncomfortable for the patient; therefore many patients remain undiagnosed [3].

Fox [4] performed a perceptual study that confirmed the clinical practice claim that some patients with OSA syndrome have abnormality in voice emission that makes it difficult to understand their speech. Fox has found that patients with OSA suffer from resonance, phonation, and articulation anomalies. In addition, different studies have confirmed that OSA is associated with anatomical and functional abnormalities of the upper airway due to evolution of the speech production system [5,6,7]. 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 [2]. It is well known that anatomic and functional changes in the vocal tract components affect the acoustic parameters of speech; hence, it was suggested that some acoustic speech features of patients with OSA syndrome may be distinct from those of non-OSA subjects [6].

Our hypothesis is that there is a difference between the speech characteristics of OSA-patients and non-OSA subjects, and that we can utilize this fact to design an automated system that will discriminate between the two groups.

Previous studies with a similar hypothesis [8,9,10] were completely based on automatic and/or manual identification of phonemes prior to the classification stage. However, the segmentation procedure complicates the classification process, and might result in additive error whether it is done manually or automatically, especially when dealing with people who suffer from speech anomalies. Blanco [11] found that the performance of automatic phoneme identification system were significantly lower for subjects with OSA in comparison to healthy subjects. Moreover, previous studies have used only voiced phonemes, and ignored

the unvoiced speech segments that might contain valuable distinctive information as well.

In this study we have designed a system that uses potential patients' speech recordings to automatically diagnose OSA, which is not dependent on phoneme segmentation and recognition. The classification system uses all the non-silence segments of the patient's speech signal, and therefore better exploits the hidden information in the speech signal and reveals the vocal tract's dynamics. Furthermore, analysis of all the non-silence segments allows us to investigate phonation, resonance, and articulation differences between OSA and non-OSA subjects. We designed a simple Gaussian mixture model (GMM)-based classifier, using mel frequency cepstrum coefficients (MFCC), and MFCC's first and second derivatives only; in order to avoid over-fitting and to find the most discriminative features a feature selection procedure was performed. Our goal is to use this system as a diagnostic tool for the early detection of OSA.

II. METHODOLGY

In order to create an automatic classification system for OSA diagnosis, speech signals of 103 subjects were analyzed. The signals underwent pre-processing; a voice activity detector was implemented and features were extracted. A GMM classifier was trained and validated using the hold out method. Figure 1 presents a block diagram of the system.

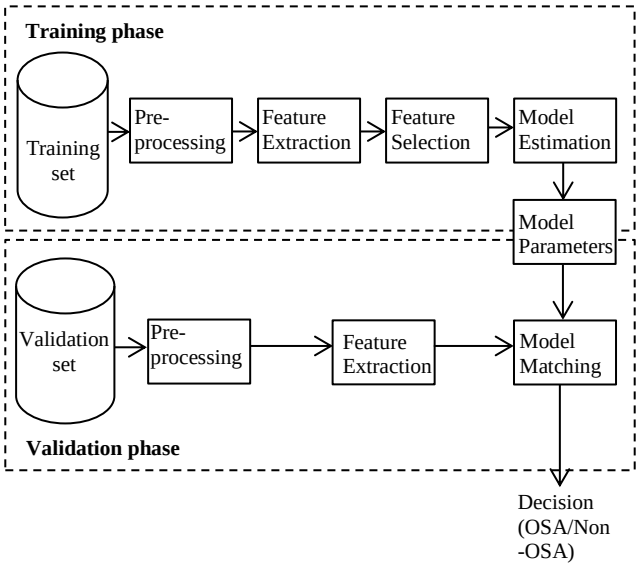

Figure 1 - System block diagram.

2.1 Pre-processing and feature extraction

Each digitized speech signal underwent a pre-processing procedure of DC removal, and pre-emphasizing; silence removal took place using a voice activity detector based on [12]. The signals were framed to 30 msec frames with 50% overlap. Forty-eight features were extracted: 16 MFCC, and their first and second derivatives, the ΔMFCC and the ΔΔMFCC, respectively.

2.2 Model estimation

Since the cepstrum's density has the benefit of being well modeled by a linear combination of Gaussian densities [13], a GMM classifier was implemented in this stage. The GMM is defined as the weighted sum of *M* Gaussian component densities [14].

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

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:

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

with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\sum_i$. In this work, the classifier was trained on a subset of features selected via a feature selection procedure. Two models were designed to represent the probability density of each group: OSA (ω_{O}) and non-OSA (ω_{NO}).

2.3 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 [15]. In this work 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 data set.

2.4 Validation

After the model parameters were estimated in the training phase, validation procedure was conducted to evaluate the system's performance. Each subject of the validation data was tested over a non-OSA model and an OSA model and scored using log-likelihood ratio:

$$\Lambda=\frac{1}{N}\left(\sum_{j=1}^{N}\log\left(\mathrm{p}\left(\mathbf{x}_j\,|\,\omega_O\right)\right)-\sum_{j=1}^{N}\log\left(\mathrm{p}\left(\mathbf{x}_j\,|\,\omega_{NO}\right)\right)\right) \qquad (3)$$

where $\mathrm{p}(\mathbf{x}_j|\omega_{\mathrm{O}})$ and $\mathrm{p}(\mathbf{x}_j|\omega_{\mathrm{NO}})$ are the likelihood probabilities of the j^{th} feature vector $\mathbf{x}_j$, given the model for OSA patients and non-OSA subjects, respectively, and N is the number of frames per subject.

III. Experimental Framework

The database for this research was constructed from speech signals of 103 male subjects, who were referred to a sleep clinic. 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 stuff. Subjects' age, BMI, and AHI are summarized in Table 1. In order to avoid over-fitting the database was divided into two separate data sets: design (*n=62*) and validation (*n=41*).

Table 1- The subjects' information.

	Number of subjects	*AHI average ± STD*	*BMI average ± STD*	*Age average ± STD*
Design				
OSA	47	27.1±19.4	30.5±5.5	57.0±13.3
Non-OSA	15	4.4±2.2	27.6±3.8	43.4±16.1
Total	62	21.6±19.5	29.8±5.3	53.7±15.1
Validation				
OSA	31	27.0±19.3	30.4±5.6	56.8±12.6
Non-OSA	10	4.4±2.2	27.3±3.5	43.3±16.2
Total	41	21.5±19.4	29.6±5.3	53.6±14.6

IV. Results and discussion

The feature selection procedure resulted in a four-dimensional feature subset presented in Table 2. Three out of four selected features belong to the ΔΔMFCC set; this result supports our hypothesis that using all the speech (non-silence) segments exploits the information that is hidden in the signal's dynamic, which expresses the transition states of the vocal tract. GMM order of 7 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 two classes' classification problem; higher order could represent different sub-classes, such as age and accent, and therefore impair the results.

Table 2 – Selected features.

Selection order	Feature symbol
1	c3
2	ΔΔc12
3	ΔΔc14
4	ΔΔc16

The selected features: the third MFCC, and the second derivatives of the 12^{th}, 14^{th}, and 16^{th}.

System performances were evaluated using the validation data. Table 3 presents the classification system's performance and Figure 2 presents the ROC curve. Average correct classification is 80%.

Table 3 – Classification results of the suggested system.

	true label O	true label nO
classified as O	**80.65%**	20%
classified as nO	19.35%	**80%**

O indicates OSA, nO indicates non-OSA.

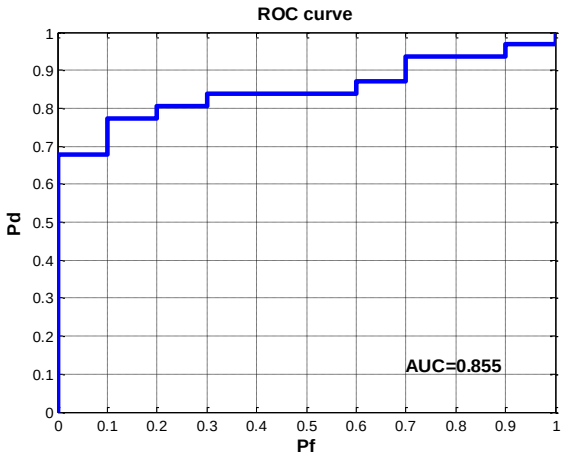

Figure 2 - Performance results evaluated over the validation data.

In [10] a classification system for the detection of severe OSA (i.e., AHI>30) was suggested; the study database consisted of 40 severe-OSA patients and 40 non-OSA subjects. The classification system was mainly based on vowels and nasal segments that were automatically segmented from continuous speech prior to the classification stage. 85% sensitivity and 77.5% specificity were achieved. Although we didn't exclude patients from our research and dealt with the entire range of OSA severity (i.e., entire range of AHI) the performances are quite similar.

In a previous study [16] a different database was used (97 males) for diagnosing OSA, while using only 7 phonemes (\a\, \e\, \i\, \o\, \u\, \m\, \n\). Seven independent GMMs were designed for the OSA/non-OSA classification problem, and a fusion process was performed to combine the scores of these classifiers. The system was tested using two modes that differ in the way the phonemes were segmented: manually or automatically. In the manual segmentation mode 86.2% sensitivity and 75% specificity were achieved, while in the automatic segmentation mode 82.4% sensitivity and 80% specificity were achieved. When using the same algorithm presented in [16] on the database for this study, the results are 67.74% sensitivity and 80% specificity for the manual segmentation mode, while for

the automatic segmentation mode 64.52% sensitivity and 80% specificity were achieved. This decrease in results may be due to the fact that the database for the current study is more balanced (in terms of AHI, BMI, age, etc.) and includes more subjects. Comparing the results, one can see that the presented system outperforms the system that was suggested in [16]. Moreover, since the presented system doesn't require automatic and/or manual identification of phonemes prior to the classification stage, or the fusion of several classifiers, it has much lower complexity.

V. Conclusions

In this paper an automatic system for diagnosing OSA from speech was presented. The proposed method can be used for initial screening of potential OSA patients, and bring earlier diagnosis and treatment. Moreover, it can significantly reduce the number of patients referred unnecessarily to sleep clinics.

The system is fully automated with low complexity, and is based on speech signal recordings only; therefore it allows effective, fast, patient-friendly, and low cost diagnosis of potential OSA patients.

VI. References

[1] M. R. Mannarino, F. D. Filippo and M. Piroo, "Obstructive sleep apnea syndrome," *European Journal of Internal Medicine*, vol. 7, pp. 586-593, 2012.

[2] C. M. Ryan and T. D. Bradley, "Pathogenesis of obstructive sleep apnea," *Journal of Applied Physiology*, vol. 99, no. 6, pp. 2440-2450, 2005.

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

[4] A. W. Fox, P. K. Monoson, and C. Morgan, "Speech dysfunction of obstructive sleep apnea," *Chest*, vol. 96, no. 3, pp. 589-595, 1989.

[5] T. M. Davidson, "The great leap forward: the anatomic basis for the acquisition of speech and obstructive sleep apnea," *Sleep Medicine*, vol. 4, no. 3, pp. 185-194, 2003.

[6] T. M. Davidson and J. Sedgh, "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, no. 6, pp. 497-505, 2005.

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

[8] E. Goldshtein, A. Tarasiuk, and Y. Zigel, "Automatic detection of obstructive sleep apnea using speech signals," *IEEE Trans. on Biomedical Eng*., Vol. 58, pp. 1373-82, 2011.

[9] O. Elisha, Y. Zigel, and A. Tarasiuk. "Automatic Detection of Obstructive Sleep Apnea using Speech Signal Analysis" *Afeka-AVIOS Speech Processing Conference.* 2012.

[10] R. F. Pozo, J. L. B. Murillo, L. H. Gómez, E. L. Gonzalo, J. A. Ramírez, and D. T. Toledano, "Assessment of severe apnea through voice analysis, automatic speech, and speaker recognition techniques," *EURASIP Journal on Advances in Signal Processing*, 2009.

[11] J. L. Blanco, R. Fernandez, E. Lopez, and L. A. Hernandez, "Exploring differences between phonetic classes in Sleep Apnoea Syndrome Patients using automatic speech processing techniques," *The Phonetician*, vol. 97/98, pp. 36-56, 2011.

[12] J. Sohn, N. Kim, W. Sung, "A statistical model-based voice activity detection," *Signal Processing Letters, IEEE*, vol. 6, no. 1, pp. 1-3, 1999.

[13] J.P Campbell, "Speaker recognition: a tutorial," Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462, 1997.

[14] 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, no. 1, pp. 72-83, 1995.

[15] S. Theodoridis and K. Koutroumbas, *Pattern recognition*, 4th ed. Burlington: Academic Press, 1999.

[16] O. Elisha, “Obstructive sleep apnea detection and AHI estimation using speech signal analysis,” M.S. thesis, Dept. Biomed Eng., Ben-Gurion Univ., Beer-Sheva, Israel, 2012.