# *Obstructive Sleep Apnea (OSA) Classification using Analysis of Breathing Sounds During Speech*

*Ruby M. Simply, Eliran Dafna, Yaniv Zigel*
Department of Biomedical Engineering, Faculty of Engineering Sciences,
Ben-Gurion University of the Negev
Beer-Sheva, Israel
simplyr@post.bgu.ac.il, elirandafna@gmail.com, yaniv@bgu.ac.il

***Abstract*— *Obstructive sleep apnea (OSA) is a sleep disorder in which pharyngeal collapse during sleep, causes a complete or partial airway obstruction. OSA is common and can have severe impacts, but often remains unrecognized. In this study, we propose a novel method which able to detect OSA subjects while they are awake, by analyzing breathing sounds during speech. The hypothesis is that OSA is associated with anatomical and functional abnormalities of the upper airway, which in turn, affect the acoustic parameters of a natural breathing sound during speech. The proposed OSA detector is a fully automated system, which consists of three consecutive steps including: 1) locating breathing sounds during continuous speech, 2) extracting acoustic features that quantify the breathing properties, and 3) OSA/non-OSA classification based on the detected breathing sounds. Based on breathing sounds analysis alone (90 male subjects; 72 for training, 18 for validation), our system yields an encouraging results (accuracy of 76.5%) showing the potential of speech analysis to detect OSA. Such a system can be integrated with other non-contact OSA detectors to provide a reliable and OSA syndrome-screening tool.*

***Keywords—Obstructive sleep apnea (OSA), speech signals, breath signals, signal processing, machine learning***

## I. INTRODUCTION

Obstructive sleep apnea (OSA) is defined as a repetitive collapse of the upper airways during sleep, characterized as a complete (apnea) or partial (hypopnea) collapse, while respiratory effort persists. OSA usually involves snoring and choking, and causes frequent awakenings, disrupted sleep, and excessive daytime sleepiness [1]. As the disorder progresses, sleepiness encroaches into all daily activities and can become disabling and dangerous. Accordingly, OSA has been shown to increase the risk of car accidents [1]. In addition, OSA can cause morning headaches, dry mouth and sore throat when awakening. When left untreated, OSA can even cause cardiovascular and neurocognitive disorders, such as heart failure, stroke, mood disorders, and hypertension [1].

OSA severity is defined by the number of obstructive apnea and hypopnea events per hour of sleep, which is known as the apnea-hypopnea index (AHI) [2].

Polysomnography (PSG) is a clinical procedure performed overnight and is currently accepted as a gold standard diagnostic assessment for OSA. The PSG monitors a number of body functions such as brain wave activity (EEG) and heart wave activity (ECG). The sensor that is usually used for detecting OSA is an oronasal thermal sensor, which detects both nasal and oral airflow [3]. Although PSG can provide an accurate assessment of the disorder, it is considered an expensive test for the health care system and inconvenient for patients. Moreover, due to the inconvenience caused to the patients, sleep habits are disrupted, leading to biased results. Hence, 75–80% of OSAs remain undiagnosed [4].

Earlier studies have already shown that OSA patients have anatomical and functional abnormalities of the upper airway that may affect their speech [5, 6]. Furthermore, in [7] it was noted that experienced speech pathologists were able to identify speech abnormalities in patients with OSA. As a result of those conclusions, a few studies were correspondingly made to try and find an objective acoustic analysis that could distinguish OSA from speech signals, rather than relying on a subjective assessment by an experienced listener [8-10].

The primary goal of the current study is to develop an alternative, non-invasive, and portable OSA monitoring tool that will be used to analyze the patient's speech while awake. This tool should estimate and predict OSA severity from a short speech prior to bedtime. Such a tool may increase accessibility, and decrease the percentage of undiagnosed cases.

In [10], subjects were recorded in a sitting position while emitting a series of sustained phonemes. Afterwards, the subjects were recorded reading a one-minute speech protocol spoken in Hebrew [11]. In a recently performed study [12], a system was built using a fusion of three different subsystems for speech signal analysis, combined with the subject's BMI and age. These subsystems included the exploitation of short-term and long-term acoustic features of continuous speech, as well as features of sustained vowels, in order to extract the relevant information from the speech signals.

The structural and physiological changes that are associated with OSA, such as narrower and more collapsible pharynx, were expected to affect the breathing sounds within a speech signal. Breathing is the process that moves air in and out of the lungs, to allow the diffusion of oxygen and carbon dioxide. Some notable differences between the breathing sounds of people with different degrees of OSA severity during

The Israel Science Foundation (ISF) supported this work: award number 1403/15.

ISBN 978-90-827970-1-5 © EURASIP 2018

wakefulness have been shown in [13]. The research analyzed only the tracheal breath sounds at medium flow rate (without any talking). The novelty of this study is in its use of the spontaneous breathing sounds within a continuous speech signal to estimate OSA in order to, eventually, integrate this approach into spontaneous speech.

## II. METHODS

### A. Experimental Setup

The database for this study comprised three subsets and includes 100 male speech signals in total. The first subset includes 66 subjects who were referred to the Sleep-Wake Unit (Soroka University Medical Center) for a PSG study in order to evaluate their sleep disorders. The second subset consists of 24 subjects who were sent by the Sleep-Wake Unit to have at-home OSA evaluation using WatchPAT 200 (Itamar Medical Ltd., Israel) [14]. The third subset consists of 10 students from Ben-Gurion University of the Negev who volunteered to participate in the research using WatchPAT evaluation. Table 1 shows subjects' characteristics.

Each of the 100 participants was recorded using a digital audio recorder (Zoom H4 handy recorder) while reading a text protocol in Hebrew. The digital signals were recorded using a sampling rate of 44.1 kHz (PCM, 16 bits/sample). Afterwards, each subject underwent an overnight sleep study using PSG or WatchPAT to assess the ground truth AHI score.

The Institutional Review Committee of Soroka Medical Center (Helsinki Committee) and the Human Subjects Research Committee of Ben Gurion University approved the study protocol.

### B. Breathing Detection

After receiving a recording of the speech signals, an analysis was performed to extract only breathing sounds. In order to maintain uniformity among breathing sounds across participants, four identical sentences from the speaking protocol were chosen.

The algorithm in this section is a modified version of the detector presented in [15], and was adjusted to better capture breathing sounds during speech. A block diagram of the breathing detection system is given in Fig. 1 and detailed in the following steps:

**Pre-processing:** Audio signal of speech was first divided into 10 msec sub-frames (hamming window, no overlaps) and hence defined the temporal resolution of the detector. Let denote these sub-frames as $\mathbf{x}_i$. Each sub-frame was then underwent a DC removal and pre-emphasis filtering (transfer function: $H(Z)=1-0.95Z^{-1}$).

TABLE 1. Subjects' characteristics.

| Database | Diagnosis | Subjects | AHI | Age | BMI |
|---|---|---|---|---|---|
| **Training** | Healthy | 38 | 4.78±2.49 | 38.16±15.44 | 28±4.94 |
| | OSA | 34 | 29.97±18.36 | 52.60±13.81 | 30.05±4.54 |
| **Testing** | Healthy | 9 | 4.26±2.32 | 36±13.24 | 26.67±4.52 |
| | OSA | 9 | 25.98±15.95 | 57.93±12.1 | 29.98±4.25 |

Values are presented as mean ± std.
* Ten subjects were excluded, see Result section B for more information.

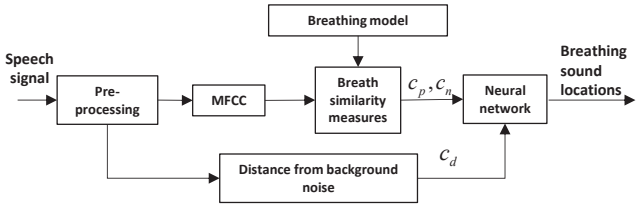


*Figure 1. Block diagram of the breathing detection system.*

To automatically detect breathing sounds from a continuous speech signal, it was first necessary to define the breathing sounds manually. Manual segmentation was used as the true label and was used to train a breathing model. In order to make the technique simple, the breathing sounds were to be the same length; hence a 120 msec segment from the center of each breathing sound was used. This will be referred to as the "breathing segment". This length was chosen because it was the shortest breathing sound that was segmented.

**Feature extraction:** Three features were extracted from each sub-frame. These features were specifically designed to measure either the distance of the sub-frame to (pre-defined) breathing model (two features in total), or the distance to the background noise template (one feature).

To be consistent with the length of the breathing segment, for the first two features, a "tested frame" was defined as 12 sub-frames (120 msec). For each breathing segment and tested frame, a MFCC matrix was obtained in which the columns were the sub-frame numbers, and the rows were the coefficient numbers. The MFCC matrix of a tested frame was denoted by $\mathbf{Y}$. The breathing model was estimated by calculating MFCC's mean ($\mathbf{T}$) and variance ($\mathbf{V}$) across all breathing segments. Then, the normalized difference matrix ($\mathbf{D}$) between the breathing model and the tested frame, was computed, according to the following equation:

$$\mathbf{D} = \frac{\mathbf{Y} - \mathbf{T}}{\mathbf{V}} \quad (1)$$

Eventually, the first similarity measure ($c_p$) which was used as the first feature, was computed according to the following equation:

$$c_p = \left( \sum_{j=1}^{n} \sum_{c=1}^{Nc} | \mathbf{D}_{jc} |^2 \right)^{-1} \quad (2)$$

where $n$ is the number of frames and $N_c$ is the number of MFCCs computed for each frame. Due to the inverse operator in the equation, the more similar the tested frame was to the breathing model, the higher this feature.

In addition, the singular value decomposition (SVD) on the concatenation MFCC matrices of the breathing segments was

ISBN 978-90-827970-1-5 © EURASIP 2018

computed. Then, the normalized singular vector ($\mathbf{sv}$) corresponding to the largest singular value was derived.

The second similarity measure ($c_n$), which was used as the second feature, was computed by taking the sum of the inner products between the singular vector and the columns of the MFCC matrix of the tested frame, according to the following equation:

$$c_n = \sum_{j=1}^{n} < \mathbf{sv}, \mathbf{Y}_j > \quad (3)$$

Since the singular vector was expected to capture the most essential features of the breath sound, $c_n$ was expected to be high when the frame contains information of breathing sounds.

The third feature was the spectrum distance from the background noise ($c_{d_i}$). In order to increase the separation between the breathing sounds and the background noise, a spectrum distance between the sub-frame and the background noise of the continuous speech signal was performed. The spectrum of each sub-frame was calculated using: $\mathbf{s}_i = \ln(|FFT(\mathbf{x}_i)|)$. In order to get the background noise representation for each utterance, the energy for each sub-frame was calculated and the lower percentile was considered as the background noise frames. The mean vector of those frames' spectrum was defined as the background noise spectrum ($\mathbf{n}$). Then the mean square error (MSE) of each sub-frame from the background noise representation was calculated. Because each speech signal had different background noise, we normalized each frame value by the 80th percentile of the values in the signal ($p_{80}(c_d)$), formulated as follows:

$$c_{d_i} = \frac{E[(\mathbf{n} - \mathbf{s}_i)^2]}{p_{80}(c_d)} \quad (4)$$

**Classification:** Each sub-frame got a classification decision, defining whether it was a breathing frame or not, i.e., frame by frame decision. In the training phase, the three features, i.e., the similarity measures and the distance from background noise, were fed into a feedforward neural network (NN) along with a manual labeling for breathing/non-breathing. This NN contained 1 hidden layer with 10 hyperbolic tangent sigmoid neurons and output layer of softmax function for two classes (breathing/non-breathing). A cross-entropy loss function with weighted errors was chosen for the training procedure, in order to overcome the inequality proportions of frames:

$$Weight_i = \frac{N}{\sum_{j=1}^{N} \mathrm{Bool}\left(Frame_j = Frame_i\right)}, \quad (5)$$
$$Frame_i \in \{Breathing, Non-breathing\}$$

where $N$ is the total number of sub-frames for the design dataset, and Bool is the Boolean operator resulting in "1" if the statement is true.

**Post-processing:** The assumption is that the shortest breathing sound was 120 msec. For that reason, after getting the decision for each sub-frame, a 12th-order one-dimensional median filter on the decision vector was applied, in order to smooth the results.

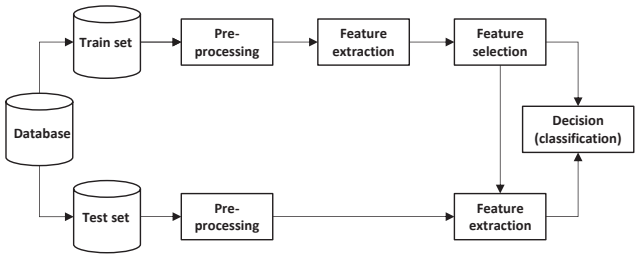


*Figure 2. Block diagram of the OSA classification system.*

### C. OSA Classifications

After locating the breathing sounds (section B), an OSA estimation was performed using only the isolated breathing sounds. To avoid over-fitting, the database was divided into two separate datasets: a training set (80%) and a test set (20%). A block diagram of the developed system is given in Fig. 2 and detailed in the following steps:

**Pre-processing:** Each breathing sound was framed into 30 msec frames with an overlap of 75% (10 msec frame rate) after the DC removal step.

**Feature extraction:** To investigate features suitable for classification between OSA and non-OSA, features such as different energy manipulations and MFCC were calculated. The feature list is given in Table 2. Since the number of breathing sounds is generally different between subjects, we apply several mathematical and statistical calculations over the detected breathing sounds including: mean (average of each feature), median, 90th percentile, std, skewness, kurtosis, and the features of the most energetic breathing sound. Eventually, each subject had a feature vector with 104 features.

**Feature selection and classification:** We used feature selection in order to choose the best discriminative features (among the extracted 104 features) and to avoid overfitting. In order to choose the features that identify the components of the audio signal that are good for estimating OSA, we used the forward feature selection algorithm [16] with 5-fold cross validation on the training set. The support vector machine (SVM) model with second order polynomial kernel was used

TABLE 2. Extracted features
(before adding statistical calculations).

| # | Feature name | Number of features |
|---|---|---|
| 1 | Average energy normalized with the speech energy | 1 |
| 2 | Average energy normalized with the background noise energy | 1 |
| 3 | Average zero crossing rates (ZCR) | 1 |
| 4 | Kurtosis | 1 |
| 5 | Mel-frequency cepstral coefficients (MFCC) | 12 |
| 6 | Pitch peak | 1 |

ISBN 978-90-827970-1-5 © EURASIP 2018

(both for the 5-fold cross validation and the final model), while the performance criterion determining the quality of the features was the average accuracy for all of the 5 models.

## III. Results & Discussion

### A. Breathing Detection

Performance evaluation was conducted using the leave one out (LOO) method. As described, the system has three features that were extracted in order to distinguish breathing and non-breathing sounds. A graphical example of a typical speech signal, as well as the features, is given in Fig.3.

As one can see from Fig 3, the breathing sounds have higher breathing similarity values and lower values of spectral distance from the background noise, as opposed to the speech sections, as expected.

This study's goal was to estimate OSA using breathing sounds taken from a continuous speech signal, meaning that if a non-breathing sound was mistakenly considered as a breathing (false positive), it could cause inaccurate results. Accordingly, the false positive rate should be as low as possible (high specificity is needed) even if it comes at the expense of a lower sensitivity.

The breathing detector performance evaluation can be seen in Table 3. As can be seen, the sensitivity of the algorithm is lower than the specificity. As mentioned, high specificity is more essential because it is preferable to lose information and ignore some of the breathing sounds, than to include any non-relevant information.

### B. OSA Classifications

In the OSA classification procedure, the features are mathematical and statistical measures that rely on the assumption that each subject has more than one breathing sound; therefore, subjects with less than two detected breathing sounds were excluded. An additional reason for removing the data that included only one breathing sound was that it did not have enough information in it. For these reasons, data from 10 subjects were excluded and the results refer to the 90 remaining

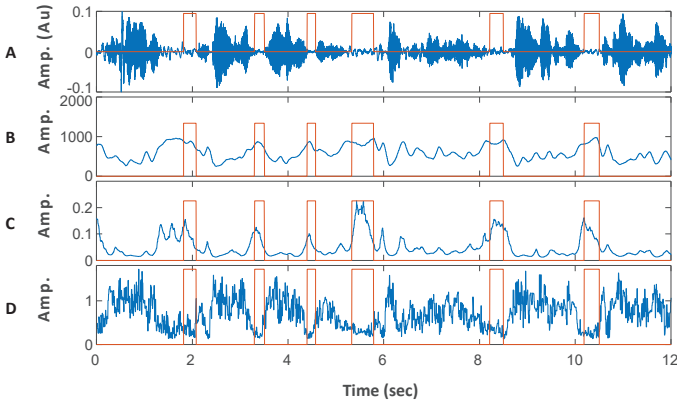


*Figure 3. Different signals/features (blue) and true labels (orange) – breathing sounds (higher value) and non-breathing sounds (lower value). A: Original signal. B: First similarity measure ($c_p$). C: second similarity measure ($c_n$). D: spectrum distance from the background noise ($c_d$).*

TABLE 3. Breathing detection performance evaluations.

| Sensitivity [%] | Specificity [%] | Cohen's kappa |
|---|---|---|
| 59.7 | 93.7 | 0.5 |

subjects: 43 of them were OSA patients while the remaining 47 were non-OSA subjects. To validate the system, the database was divided into two groups, the train and the test (see Table 1).

The feature selection used the average accuracy results of 5-fold cross-validations that were only taken from the training set in order to choose the best features. After 1000 iterations, it was found that three features were chosen much more frequently than the others. These selected features are: kurtosis of the average energy, median of the seventh MFCC, and the std of average ZCR.

Using only these three features, a new model was built using the features from the entire training set and was validated on the test set. The classification performance (Confusion matrix) can be seen in Table 4.

As shown, the sensitivity of the classification system is 55%, and the specificity is 100%. The average accuracy of the test set was 76.5%. In order to evaluate overtraining of the model, the average accuracy of the training set was calculated and found to be 76.0% as well.

## IV. Conclusions & Future Work

In this study, a new method has been proposed that uses breathing sound analysis for OSA severity prediction during wakefulness. The structural and physiological changes associated with OSA were expected to differentiate the breathing sounds in the people with OSA compared to healthy people.

The first step was to automatically detect breathing sounds within a continuous speech signal. The critical restriction was to give the OSA estimation system only the relevant information and not to consider a non-breathing sounds as a breathing sounds. Hence, a specificity of 93.7% and sensitivity of 59.7% were selected.

After detecting all of the breathing sounds, an OSA classifier system was designed (OSA/non-OSA). Three features representing statistical measures on MFCC, energy, and ZCR were chosen for estimating the existence of OSA, and obtained 76.5% accuracy.

Accordingly, it was concluded that the analysis of breathing sounds within speech signals of awake subjects could assist in

TABLE 4. Classification performance using the selected features.

| | | **True Labels** | |
|---|---|---|---|
| | | **OSA** | **Non-OSA** |
| **System Output** | **OSA** | 55% | 0% |
| | **Non-OSA** | 45% | 100% |

*Cohen's kappa coefficient is 0.54

ISBN 978-90-827970-1-5 © EURASIP 2018

the assessment of OSA. These conclusions are encouraging and pave the way for a simple, non-invasive, and inexpensive screening tool for the people suspected of having OSA.

In future research, a superior OSA detection system can be achieved by integrating breathing information with additional types of speech features such as freely spoken speech. The effect of body posture, and the hoarseness effect, should be addressed as well. In addition, to give these results better statistical validation, the database would need to be expanded.

## ACKNOWLEDGMENT

We would like to thank Prof. Ariel Tarasiuk and Mrs. Bruria Friedman from the Sleep Wake Disorder Unit of Soroka University Medical Center, for their support and collaboration.

## REFERENCES

[1] M. R. Mannarino, F. Di Filippo, and M. Pirro, "Obstructive sleep apnea syndrome," *European journal of internal medicine,* vol. 23, no. 7, pp. 586-593, 2012.

[2] N. M. Punjabi, "The epidemiology of adult obstructive sleep apnea," *Proceedings of the American Thoracic Society,* vol. 5, no. 2, pp. 136-143, 2008.

[3] R. B. Berry *et al.*, "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine," *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine,* vol. 8, no. 5, p. 597, 2012.

[4] V. Kapur, K. P. Strohl, S. Redline, C. Iber, G. O'connor, and J. Nieto, "Underdiagnosis of sleep apnea syndrome in US communities," *Sleep and Breathing,* vol. 6, no. 02, pp. 049-054, 2002.

[5] T. M. Davidson, J. Sedgh, D. Tran, and C. J. Stepnowsky, "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.

[6] 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.

[7] P. K. Monoson and A. W. Fox, "Preliminary observation of speech disorder in obstructive and mixed sleep apnea," *Chest,* vol. 92, no. 4, pp. 670-675, 1987.

[8] M. Kriboy, A. Tarasiuk, and Y. Zigel, "A novel method for obstructive sleep apnea severity estimation using speech signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014, pp. 3606-3610: IEEE.

[9] R. F. Pozo, J. L. B. Murillo, L. H. Gómez, E. L. Gonzalo, J. A. Ramírez, and D. T. Toledano, "Assessment of severe apnoea through voice analysis, automatic speech, and speaker recognition techniques," *EURASIP Journal on Advances in Signal Processing,* vol. 2009, no. 1, p. 982531, 2009.

[10] A. M. Benavides, R. F. Pozo, D. T. Toledano, J. L. B. Murillo, E. L. Gonzalo, and L. H. Gómez, "Analysis of voice features related to obstructive sleep apnoea and their application in diagnosis support," *Computer Speech & Language,* vol. 28, no. 2, pp. 434-452, 2014.

[11] M. Kriboy, A. Tarasiuk, and Y. Zigel, "Obstructive sleep apnea detection using speech signals," in *Proceedings of the annual conference of the Afeka-AVIOS in Speech Processing*, 2013, pp. 1-5.

[12] D. Ben-Or, E. Dafna, A. Tarasiuk, and Y. Zigel, "Obstructive sleep apnea severity estimation: Fusion of speech-based systems," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 3207-3210: IEEE.

[13] A. Montazeri, E. Giannouli, and Z. Moussavi, "Assessment of obstructive sleep apnea and its severity during wakefulness," *Annals of biomedical engineering,* vol. 40, no. 4, pp. 916-924, 2012.

[14] T. Ceylan, H. Fırat, G. Kuran, S. Ardıç, E. Bilgin, and F. Çelenk, "Quick diagnosis in obstructive sleep apnea syndrome: WatchPAT-200," *Iranian Red Crescent Medical Journal,* vol. 14, no. 8, p. 475, 2012.

[15] D. Ruinskiy and Y. Lavner, "An effective algorithm for automatic detection and exact demarcation of breath sounds in speech and song signals," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 15, no. 3, pp. 838-850, 2007.

[16] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature selection for SVMs," in *Advances in neural information processing systems*, 2001, pp. 668-674.

ISBN 978-90-827970-1-5 © EURASIP 2018