Biomarkers of sleep disorders

Antonio Vieira and **Isabel Trancoso** and **Alberto Abad**
Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

Abstract

This thesis takes on the work of Catarina Botelho in her Thesis "Speech as a Biomarker for Sleep Disorders and Sleep Deprivation", focusing on the automatic detection of Obstructive Sleep Apnea (OSA). This sleep disorder can cause fatal traffic accidents due to fatigue as well as work related accidents and so far the best diagnosis exam, the Polysomnography, is not practical and makes the patient uncomfortable.

Our objective was to improve on C. Botelho's work by adding to it cough and snore analysis to the sleep disorders detection and comprehension to all the other speech signals collected.

By evaluating the features of the subjects' speech signals with an SVM classifier we were able to reach a 91% accuracy rate with majority voting.

The results were obtained from two different corpora, the first one has 40 subjects and it was compiled from YouTube vlogs and the second one has 26 subjects and it was compiled through a JotForm survey.

Keywords: Osa, Sleep disorders, Speech, Cough, Computational diagnosis.

1 Introduction

This work focuses on sleep disorders, more specifically on the analysis of biomarkers for the detection of sleep disorders. It follows up on the work done by Catarina Botelho during her MSc thesis entitled "Speech as a Biomarker for Sleep Disorders and Sleep Deprivation" (which has also an associated paper: "Speech as a Biomarker for Obstructive Sleep Apnea Detection" [4]).

	Control	OSA
#F	12	6
#M	8	19
Age - F	22 $\pm$11	55 $\pm$9
Age - M	36 $\pm$10	53 $\pm$10

Table 1: Portuguese Sleep Disorders (PSD) Corpus.

The decrease in the muscle tone of the upper airway dilator muscle, excessive compliance of the pharyngeal wall and anatomical alterations of the respiratory tract are some of the causing factors of Obstructive Sleep Apnea. These factors also cause articulatory anomalies, phonation anomalies and resonance anomalies, that are correlated with speech and are able to be identified through speech analysis.

C. Botelho collected a speech dataset in which the subjects had to complete several tasks and their speech signals were recorded. However this initial corpus was not as balanced as desired, and could jeopardize the results, so a second corpus was formed in order to avoid that possibility. C. Botelho also collected a small In-the-Wild Obstructive Sleep Apnea (WOSA) Corpus with data obtained from audio in YouTube videos.

The best experimental results of C. Botelho's Thesis for OSA detection (TPR of 88% and TNR of 80%) were achieved with a feature set selected based on the literature and an ensemble of SVM, LDA and kNN classifiers with majority vote. These results were achieved in the PSD corpus, with 25 subjects suffering from OSA and 20 control subjects. The results were also validated with data acquired from YouTube.

	Control	OSA
#F	11	8
#M	11	11
Age - F	50 ±8	61 ±14
Age - M	43 ±10	55 ±10

Table 2: Portuguese Sleep Disorders balanced (PSD-b) Corpus.

1.1 Motivation

The previously mentioned Thesis talks about how sleep disorders, such as Obstructive Sleep Apnea (OSA) and Insomnia, cause sleep Deprivation on those who suffer from them. Consequently, theses subjects may suffer from fatigue, mood alteration, decreased work performance, traffic and accidents and also work accidents due to sleep deprivation. It also refers to the fact that Obstructive Sleep Apnea can cause Diabetes, reduce life quality and increase mortality and morbidity by cardiovascular diseases. The gold standard for the diagnosis of OSA is the Polysomnography (PSG) study. However, it is expensive and uncomfortable, due to the fact that subjects have to be attached to sleep monitoring equipment whilst trying to sleep [4].

1.2 Objectives

We expanded the corpus whilst keeping it balanced, focusing on the attainment of data related to other biomarkers such as simulated snoring and cough in order to assess their potential contribution to the detection of Obstructive Sleep Apnea.

2 Background

2.1 Sleepiness and Sleep Disorders

Sleep is a cyclic process with four stages. One complete cycle lasts approximately 90 minutes, and one night of sleep includes 5 to 6 cycles. Sleep deprivation interferes with the maintenance of vital functions and results in increasing sleep propensity and destabilization during awake periods [16].

2.2 OSA

Obstructive sleep apnea (OSA) is a widespread but frequently unrecognized condition. It is triggered by pharyngeal collapse during sleep and marked by repeated awakenings, interrupted sleep and consequent extreme sleepiness. The link between OSA and asthma, metabolic syndrome, diabetes, heart failure and other disorders is acknowledged [10].

2.3 Neurocognitive somatic symptoms of OSA

People with OSA often have difficulties focusing and retaining focus for long periods. OSA can, primarily through intermittent hypoxia, promote cognitive impairment. Hypersomnolence can also play a role in the development of neurocognitive impairment due to sleep fragmentation [3].

2.4 Insomnia

Dissatisfaction with sleep related to trouble falling asleep or staying asleep or waking up too early is occurring on a weekly basis in about one third of adults. Prolonged sleeplessness, however, is frequently associated with severe anxiety, impaired activity during the day, or both. A diagnosis of insomnia disorder is necessary in such circumstances. Chronic insomnia is all related to declines in perceived fitness and quality of life, increases in occupational accidents and absenteeism, and even fatal injuries. Insomnia symptoms may also be an independent risk factor for suicide attempts and deaths from suicide, independent of depression [15].

3 Related Work

3.1 Speech Signals

The most widely used objective measure of OSA severity is the apnea-hypopnea index (AHI) This index reflects the average number of obstructive apnea and hypopnea events per hour of sleep. This study "Diagnosis of Obstructive Sleep Apnea Using Speech Signals From Awake Subjects" [12] reports an innovative system to identify OSA subjects while they are awake, not asleep, by extracting speech signals from subjects .

The three main extractions from the speech data base used in this work were Breathing Segments, Sustained Vowels and Continuous Speech Signal.

The system achieved an average accuracy of 77.14%, a sensitivity of 75%, and a specificity of 79%.

Another paper called "Prediction of Sleepiness Ratings from Voice by Man and Machine" [8] looks in more detail at the Interspeech 2019 computational paralinguistics challenge on the prediction of sleepiness ratings from speech using samples of the Düsseldorf Sleepy Language Corpus (DSLC). This challenge was notable because the performance of all entrants was uniformly poor, with even the winning system only achieving a correlation of r=0.37. The authors look at whether the task itself is achievable, and whether the corpus is suited to training a machine learning system for the task.

A listening experiment was performed using samples from the corpus and show that a group of human listeners can achieve a correlation of r=0.7 on this task, although this is mainly by classifying the recordings into one of three sleepiness groups. The corpus, because of its construction, confounds variation with sleepiness and variation with speaker identity, and this was the reason why machine learning systems failed to perform well:

- Through the analysis of the corpus itself in section 4, the authors have seen that a major problem is the confounding of speaker identity and sleepiness ratings in the corpus.

- Each corpus partition contains different speakers, and each speaker only produced a narrow range of sleepiness ratings. This makes it very hard to learn features of sleepiness from the training set without at the same time learning features of identity.

- When those features are exploited by the prediction model, they may work well to measure similarity between speakers in the test set to speakers in the training set, but it is not necessarily the case that those similar speakers have similar sleepiness ratings.

It could be concluded from this paper that sleepiness rating prediction from voice is not an impossible task, but that good performance requires more information about sleepy speech and its variability across listeners than is available in the DSLC corpus.

3.2 Deep Neural Networks in Speaker Recognition

A third paper, called "X-VECTORS: ROBUST DNN EMBEDDINGS FOR SPEAKER RECOGNITION" [13], has some useful information regarding speaker recognition systems using speech signals.

There were four recognition systems developed for this study, which consist of two i-vector baselines and the DNN x-vector system.

The authors found that data augmentation is a strategy for improving their performance that is easily implemented and effective. They found that two standard i-vector baselines on SRE16 Cantonese were significantly outperformed by the x-vector system.

3.3 Simulated snoring

In this next paper, Michael Herzog et al.[7] carried out a study on simulated snoring and how it was able to predict the AHI.

From all the experiments that were conducted in this work, the following correlations were discovered:

- During simulated snoring, an increase in the AHI was associated with increased degree of dorsal movement of the tongue base.
- An increase in pharyngeal collapse at the tongue's base level was associated with a rise in the AHI.
- During simulated snoring, pharyngeal collapse at the velum level was not associated with a high overnight AHI.
- Mueller maneuver, Mallampati index, tonsil size, and dorsalization of the tongue base "static" exams do not correlate with the overnight AHI.

Michael Herzog et al. concluded that in awake patients with suspected OSA, a "dynamic" evaluation of the upper airway provided reliable prognostic data for the prediction of sleep disordered breathing. There is a significant correlation between the change of the lower airway during simulated snoring and the AHI.

3.4 Acoustic analysis of cough

William Thorpe et al.[14] conducted a study on the acoustic analysis of voluntary cough, which showed that it can be useful in the diagnosis of respiratory diseases.

In this work, it was concluded that cough sound analysis could be used in clinical practice to determine diagnoses of respiratory disease in children where clinical measurements would otherwise be difficult to obtain. Coughs may also be recorded at home when symptoms arise, allowing for later study and identification of cough related with asthma.

4 Proposed Solution

4.1 OSA detection using speech signals

This study focused on using just the speech biomarker as a possible single indicator of Obstructive Sleep Apnea.

4.1.1 Corpus

In the current thesis, the original in-the-wild OSA (WOSA) subset was expanded to more than twice the number of videos. On total, the WOSA corpus now includes speech recordings from 40 English-speaking subjects, 22 OSA patients, and 18 healthy controls.

Table 3: Corpus description.

	#male subjects	#female subjects
OSA	12	10
Controls	9	9

The OSA subjects reported in their vlogs that they have either ”obstructive sleep apnea” or ”weight-related sleep apnea”, implying that their condition is obstructive rather than central. Other variable to consider is the age range. The control subjects’ speech recordings were taken from a random selection of vlogs with unrelated topics. As a result, the corpus may be noisy.

4.1.2 Feature Extraction

We extracted the audio of these vlogs and each audio file was processed by a Voice Activity Detector (VAD).We then split each audio file into 4 second segments with an overlapping of 0.5 seconds for feature extraction. Three feature sets were extracted for every audio file, eGeMAPS, i-vectors and x-vectors. These feature sets are normally associated with speaker recognition, however, we believe that the presence of OSA symptoms can be associated with the speaker through these features.

eGeMAPS Features

The Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) contains 88 features: the arithmetic mean and coefficient of variation of 18 LLDs, 8 functionals applied to pitch and loudness, 4 statistics over the unvoiced segments, 6 temporal features, and 26 additional cepstral parameters and dynamic parameters. [6]

To create this feature set, we used the eGeMAPS configuration of OpenSMILE which is a free and open-source program for extracting characteristics from audio data.[2]

i-vectors and x-vectors

In order to extract i-vectors and x-vectors[13] from the audio files, we used pretrained models in kaldi toolkit [11].

4.1.3 Classifiers and hyperparameter choice

We chose to use a Support Vector Machine because it has been successfully used in previous works, namely Catarina Botelho’s work [4], being a frequently used classifier in scenarios with limited training data. We performed a grid search with leave-one-speaker-out cross validation in order to select the best SVM hyperparameters using the scikit-learn toolkit from python [1]. The following parameters were tested:

- kernel: linear, sigmoid and radial basis function (rbf).
- C: 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10, 100, 1000 and 10000.
- γ: 1e-5, 1e-4, 1e-3, 1e-2, 1e-1 and 1 (only varied for the rbf kernel).

4.1.4 Classification results

The accuracy was measured using majority voting, meaning that a speaker is considered to be correctly classified if more than 50% of the respective files are correctly classified. With that in mind, Table 4 shows the best hyperparameters found by the grid search for each feature set. The best results are in Table 5.

Table 4: Best SVM hyperparameters for each feature set.

Feature set	kernel	C	γ
eGeMAPS	sigmoid	0.1	1e-02
i-vectors	rbf	100	1e-04
x-vectors	linear	0.01	1e-05

Table 5: Best SVM results for each feature set.

Feature set	Accuracy (% by majority voting)
eGeMAPS	61.29
i-vectors	74.19
x-vectors	70.97

4.2 OSA detection using different acoustic biomarkers

This next experiment focused on using four different acoustic biomarkers as possible indicators of Obstructive Sleep Apnea.

4.2.1 Corpus

The corpus for this experiment was compiled using data obtained from a JotForm [9] survey which was

Table 6: JotForm Corpus description.

	#male subjects	#female subjects	Age					
			21 - 30	31 - 40	41 - 50	51 - 60	61 - 70	71+
OSA	11	2	1	2	1	3	5	1
Controls	11	2	1	2	1	3	5	1

disclosed to the public and to the patients of the CENC [5] (Centro de Electroencefalografia e Neurofisiologia Clínica), which was important to gather medically diagnosed OSA patients.

Four types of acoustic signals were requested in the survey:

- Cough: where the subjects recorded themselves coughing.
- Snore: where the subjects faked snoring while recording.
- Sustained Vowel: where subjects recorded themselves while making the sound of the vowel "a" as long as they could.
- Read and Spontaneous Speech: where subjects recorded themselves reading a short tale and describing an image.

Since 13 subjects were OSA patients, we were able to create a corpus with 26 subjects, using those patients and by building a healthy control group with similar characteristics and the same number of subjects as well.

4.2.2 Feature Extraction

The procedure from the previous experiment was repeated on this corpus.

4.2.3 Cough Biomarker experiment

The first audio files that were analysed were the recording of the subjects' coughs. Since OSA is deeply related with the upper airways, cough analysis can be beneficial in its detection.

The accuracy was measured using majority voting, in the same manner as the previous speech study. Table 16 shows the best hyperparameters found by the grid search for each feature set extracted from the cough recordings and the best results are in Table 17.

Table 7: Best SVM hyperparameters for each feature set from the cough subset.

Feature set	kernel	C	γ
eGeMAPS	sigmoid	10	1e-02
i-vectors	sigmoid	10000	1
x-vectors	linear	0.01	1e-05

Table 8: Best SVM results for each feature set from the cough subset.

Feature set	Accuracy (% by majority voting)
eGeMAPS	91.67
i-vectors	79.17
x-vectors	62.5

4.2.4 Snoring Biomarker experiment

The same procedure as before was applied.

We decided to analyse it in the same context as in these experiments.

Table 18 shows the best hyperparameters found by the SVM grid search and the best results are in Table 19.

Table 9: Best SVM hyperparameters for each feature set from the snore subset.

Feature set	kernel	C	γ
eGeMAPS	sigmoid	1000	1e-04
i-vectors	sigmoid	10	1
x-vectors	linear	100	1e-03

Table 10: Best SVM results for each feature set from the snore subset.

Feature set	Accuracy (% by majority voting)
eGeMAPS	52.0
i-vectors	52.0
x-vectors	60.0

4.2.5 Sustained Vowel experiment

The same experiment was done on the sustained vowel audios.The Table 20 shows the best hyperparameters found by the SVM grid search and the best results are in Table 21.

Table 11: Best SVM hyperparameters for each feature set from the sustained vowel subset.

Feature set	kernel	C	γ
eGeMAPS	sigmoid	10000	1e-03
i-vectors	sigmoid	1	1
x-vectors	sigmoid	10	1e-03

Table 12: Best SVM results for each feature set from the sustained vowel subset.

Feature set	Accuracy (% by majority voting)
eGeMAPS	72.0
i-vectors	68.0
x-vectors	72.0

Table 15: Male only JotForm Corpus description.

	#male subjects	#female subjects	Age					
			21 - 30	31 - 40	41 - 50	51 - 60	61 - 70	71+
OSA	11	0	1	2	1	3	4	0
Controls	11	0	1	2	1	3	4	0

4.2.6 Speech experiment

Finally, in order to compare this corpus to the previous in-the-wild one, speech samples of the subjects went through the same experimental process. The Table 22 shows the best hyperparameters found by the SVM grid search and the best results are in Table 23.

Table 13: Best SVM hyperparameters for each feature set from the speech subset.

Feature set	kernel	C	γ
eGeMAPS	sigmoid	1000	1e-04
i-vectors	sigmoid	0.1	1
x-vectors	sigmoid	10000	1e-03

Table 14: Best SVM results for each feature set from the speech subset.

Feature set	Accuracy (% by majority voting)
eGeMAPS	64.0
i-vectors	84.0
x-vectors	84.0

4.3 OSA detection using different biomarkers: Male only corpus

This experiment focused on using the same four different biomarkers as before to assess them as probable cues for Obstructive Sleep Apnea detection using a man only corpus.

4.3.1 Corpus

The corpus for this experiment is the same as the JotForm corpus described before, however it only contains male subjects. Since the previous JotForm corpus was not balanced gender wise containing only 4 female subjects, using only the male subjects might be useful to see if the models behave differently or if the results maintain. The same procedure as before was performed on this corpus.

4.3.2 Experimental results

Table 16: Best SVM hyperparameters for each feature set from the cough subset, male only corpus.

Feature set	kernel	C	γ
eGeMAPS	rbf	10000	1e-04
i-vectors	sigmoid	10	1
x-vectors	sigmoid	1	1

Table 17: Best SVM results for each feature set from the cough subset, male only corpus.

Feature set	Accuracy (% by majority voting)
eGeMAPS	77.27
i-vectors	72.72
x-vectors	63.63

Table 18: Best SVM hyperparameters for each feature set from the snore subset, male only corpus.

Feature set	kernel	C	γ
eGeMAPS	sigmoid	1	1e-05
i-vectors	sigmoid	10000	1e-02
x-vectors	rbf	100	1e-04

Table 19: Best SVM results for each feature set from the snore subset, male only corpus.

Feature set	Accuracy (% by majority voting)
eGeMAPS	54.54
i-vectors	54.54
x-vectors	72.72

Table 20: Best SVM hyperparameters for each feature set from the sustained vowel subset, male only corpus.

Feature set	kernel	C	γ
eGeMAPS	sigmoid	10000	1e-03
i-vectors	sigmoid	1	1
x-vectors	sigmoid	1000	1e-02

Table 21: Best SVM results for each feature set from the sustained vowel subset, male only corpus.

Feature set	Accuracy (% by majority voting)
eGeMAPS	77.27
i-vectors	81.82
x-vectors	68.18

Table 23: Best SVM results for each feature set from the speech subset, male only corpus.

Feature set	Accuracy (% by majority voting)
eGeMAPS	59.09
i-vectors	81.81
x-vectors	77.27

Table 22: Best SVM hyperparameters for each feature set from the speech subset, male only corpus.

Feature set	kernel	C	γ
eGeMAPS	rbf	10	1e-01
i-vectors	sigmoid	0.1	1
x-vectors	sigmoid	1	1e-03

4.4 Result discussion

For the first experiment, the best result is 74.19% of accuracy with majority voting.This result was obtained by evaluating the i-vectors extracted from the corpus data.

The best result overall is 91.67% and it was obtained from the cough subset in the JotForm corpus however, the best result obtained from the same subset in the male-only JotForm corpus is 77.27%. This might be due to the lack of data in the JotForm corpus and its trimming in order to form male-only JotForm corpus.

5 Conclusion

This thesis focused on the automatic detection of obstructive sleep apnea using different acoustic biomarkers.

The best classification results for the in-the-wild corpus were obtained by extracting the i-vectors embedding with a value of 74.19% of accuracy with majority voting.

The second experiment obtained the highest result of this work, 91.67% of accuracy with majority voting.

The third experiment obtained as a best result 81.82% of accuracy with majority voting.

5.1 Limitations

The main limitation in this work was the lack of data in all the corpora and the uneven number of male subjects compared to female subjects.

Furthermore, the control subjects were chosen at random from the available healthy subjects who did not necessarily have the risk characteristics and were not subjected to a PSG test. As a result, we must accept the likelihood of a noisy data set.

Other impediment was the current global pandemic situation that prevented us from collecting more information about different biomarkers.

5.2 Future Work

We suggest that further experiments varying biomarkers should be carried out to see if there is a practicable method of diagnosing obstructive sleep apnea through different biomarkers. It is also important to increase all all corpora size in order to have more reliable results in all experiments.

References

[1] Scikit-learn python, https://scikit-learn.org/stable/ (July 2021).

[2] Audeering, Opensmile, https://www.audeering.com/opensmile/ (July 2021).

[3] D. W. Beebe, L. Groesz, C. Wells, A. Nichols, K. McGee, The neuropsychological effects of obstructive sleep apnea: a meta-analysis of norm-referenced and case-controlled data, Sleep 26 (3) (2003) 298–307.

[4] M. C. Botelho, I. Trancoso, A. Abad, T. Paiva, Speech as a biomarker for obstructive sleep apnea detection, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 5851–5855.

[5] CENC, Centro de electroencefalografia e neurofisiologia clínica, https://www.cenc.pt/ (2021).

[6] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, K. P. Truong, The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing, IEEE Transactions on Affective Computing 7 (2) (2016) 190–202.

[7] M. Herzog, T. Metz, A. Schmidt, T. Bremert, B. Venohr, W. Hosemann, H. Kaftan, The Prognostic Value of Simulated Snoring in Awake Patients With Suspected Sleep-Disordered Breathing: Introduction of a New Technique of Examination, Sleep 29 (11) (2006) 1456–1462.
URL https://doi.org/10.1093/sleep/29.11.1456

[8] M. Huckvale, A. Beke, M. Ikushima, Prediction of sleepiness ratings from voice by man and machine, in: Interspeech 2020, ISCA, 2020.
URL https://doi.org/10.21437/interspeech.2020-1601

[9] JotForm, Jotform surveys, https://www.jotform.com/ (2021).

[10] M. R. Mannarino, F. Di Filippo, M. Pirro, Obstructive sleep apnea syndrome, European journal of internal medicine 23 (7) (2012) 586–593.

[11] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, et al., The kaldi speech recognition toolkit, in: IEEE 2011 workshop on automatic speech recognition and understanding, no. CONF, IEEE Signal Processing Society, 2011.

[12] R. M. Simply, E. Dafna, Y. Zigel, Diagnosis of obstructive sleep apnea using speech signals from awake subjects, IEEE Journal of Selected Topics in Signal Processing 14 (2) (2020) 251–260.

[13] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, S. Khudanpur, X-vectors: Robust dnn embeddings for speaker recognition, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5329–5333.

[14] W. Thorpe, M. Kurver, G. King, C. Salome, Acoustic analysis of cough, in: The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001, 2001, pp. 391–394.

[15] J. W. Winkelman, Insomnia Disorder, New England Journal of Medicine 373 (15) (2015) 1437–1444.
URL http://www.nejm.org/doi/10.1056/NEJMcp1412740

[16] M. R. Zielinski, J. T. McKenna, R. W. McCarley, Functions and mechanisms of sleep, AIMS neuroscience 3 (1) (2016) 67.