SPEECH AS A BIOMARKER FOR OBSTRUCTIVE SLEEP APNEA DETECTION

M. Catarina Botelho[1], *Isabel Trancoso*[1], *Alberto Abad*[1], *Teresa Paiva*[2]

[1]INESC-ID/Instituto Superior Técnico, University of Lisbon, Portugal;
[2]CENC Sleep Medicine Center, Lisbon, Portugal

ABSTRACT

Obstructive sleep apnea (OSA) is a prevalent sleep disorder, responsible for a decrease of people's quality of life, and significant morbidity and mortality associated with hypertension and cardiovascular diseases. OSA is caused by anatomical and functional alterations in the upper airways, thus we hypothesize that the speech properties of OSA patients are altered, making it possible to detect OSA through voice analysis. To address this hypothesis, we collected speech recordings from 25 OSA subjects and 20 controls, designed a feature set, and compared different machine learning algorithms for binary classification. We achieved a True-Positive-Rate of 88% and a True-Negative-Rate of 80% with a majority vote ensemble of SVM, LDA and kNN classifiers. These results were validated with in-the-wild data acquired from *Youtube*. Moreover, the negative impact of sleep disorders on working memory was also shown by the results obtained in one of the recorded verbal tasks.

***Index Terms*—** Obstructive Sleep Apnea, Speech, Machine Learning, Cognitive Load

1. INTRODUCTION

Obstructive sleep apnea (OSA) is a sleep-concerned breathing disorder characterized by a complete stop (apnea) or decrease (hyperpnea) of the airflow, despite the continued or increased inspiratory efforts [1]. Patients with OSA report decreases in their quality of life, mood and personality changes, relationship discord associated with loud snoring [2], depression, cognitive impairment, and excessive daytime sleepiness [3]. OSA is also associated with diabetes [3] and significant morbidity and mortality associated mainly with hypertension and cardiovascular diseases [1, 4]. According to Senaratna et al. [5], the prevalence of OSA (more than 5 respiratory events/hour) ranges from 9% to 38%, with higher values in men and elderly groups.

The gold standard diagnosis of OSA is based on a polysomnography (PSG), an exam that requires that the patients spend a night connected to several electrodes, which not only is not the best indicator of the patients sleeping habits [2], but also it is time consuming and uncomfortable for the patient [6]. Thus, cost-effective, quicker and more comfortable alternatives to diagnose sleep disorders are necessary.

Many efforts have been made to find alternative approaches to diagnose and monitor OSA. Some are based on the idea that OSA patients have altered craniofacial structures, thus aiming to detect OSA from facial photographies [7, 8]; some rely on wearable devices for home monitoring of snoring [9, 10] or for measuring biosignals, such as air flow through temperature fluctuations [11], nasal air-pressure [11], actigraphy [11, 12], peripheral arterial pulsation, and arterial oxygen saturation [12]. In general, these alternative diagnosing and monitoring technologies present some limitations, for instance, some require overnight recordings, and some capture the data obtrusively. On the other hand, speech is known to carry much information about the speaker's gender, age, emotions, personality traits and health. If we consider that speech production is present in daily life situations, its acquisition can be unobtrusive, it provides quantitative data relatively quickly, it does not require expensive recording tools, and it is free from sensor calibration [4], speech appears as a potentially valuable biomarker for sleep disorders.

The use of speech as a biomarker for OSA has been previously explored in the literature based either on sustained vowel productions [13], on read speech [4, 14, 15, 16, 6, 17, 18], and on yes/no questions and lists of words [14, 15, 16].

In this work, we addressed OSA automatic detection, through the analysis of read and spontaneous speech in Portuguese, as well as in-the-wild data acquired form video blogs (vlogs). The remaining of this document is organized as follows. Section 2 describes how OSA pathophysiology may influence speech properties and makes a summary of previous works that address OSA detection through speech analysis. Section 3 characterizes the corpora used in our experimental work, the Portuguese Sleep Disorders (PSD) Corpus and the in-the-wild OSA (WOSA) corpus, as well as the design of the recording protocol. Section 4 describes the feature set we defined for OSA classification. Next, in section 5, we present our experiments and the obtained results. In section 6, we investigate the relationship between cognitive load and sleep disorders. Finally, section 7 presents conclusions and directions for future work.

2. PATHOPHYSIOLOGY'S IMPACT ON SPEECH AND RELATED WORK

OSA pathophysiology can be explained by anatomical alterations in the upper respiractory tract, such as: retrognathia and high arched palate [2], elongated or excessive tissue of the soft palate, large tongue, swollen uvula, large tonsils and redundant pharyngeal mucosa [1]; excessive compliance of the pharyngeal wall; and decrease in upper airway dilator muscle tone [19], which may compromise its normal function of maintaining the pharyngeal lumen open during inspiration, contributing to the collapse of the upper airways. These alterations, which are responsible for snoring and airway narrowing, also cause modifications in the acoustic properties of voice [20]. Moreover, the recurrent laryngeal nerve, which is responsible for the enervation of dilator muscles of the upper airway, is also responsible for enervating the cricoarytenoid muscles, responsible for opening

This work was supported by national funds through *Fundação para a Ciência e a Tecnologia* (FCT) with reference UID/CEC/50021/2013. The authors would like to thank Joana Correia, for her support with the acquisition of the WOSA corpus, in the final part of C. Botelho's MSc. Thesis.

the vocal folds. This will also cause alterations to the speech signal. The main anomalies are [20]:

Articulatory anomalies, caused by hypotonus of the upper airway muscles or lack of regulated innervations to the breathing musculature, which cause neuromotor dysfunction [20]. This is associated with articulatory disorders, especially dysarthria, which affects the articulation of vowels and consonants, and causes slurred speech [4].

Phonation anomalies, consequence of a larynx inflammation caused by snoring, typical in OSA patients [4], which could affect the vocal folds.

Resonance anomalies can be explained by an abnormal coupling of the vocal tract with the nasal cavity - Pozo et al [4] suggested that OSA speech exhibits smaller differences between nasal and non nasal vowels than healthy speech.

Given these anomalies, several authors have addressed the automatic detection of OSA through voice analysis, making use of corpora in Spanish [4, 17, 13, 18] and Hebrew [14, 15, 16, 6]. The most common acoustic features in these works are Mel frequency cepstral coefficients (MFCC), linear prediction cepstral coefficients (LPCC), Energy, Harmonics-to-noise ratio (HNR), jitter and formants frequency and bandwidth. The most common classifying and regression methods are Gaussian Mixture Models (GMM) [4, 14, 15] , Linear Discriminant Analysis (LDA) [16, 17], k-Nearest Neighbors (kNN) [16, 13], Support Vector Machines (SVM) [13], Bayesian Classifiers [13], Neural networks ([13]), Adaboost [13], and Support Vector Regression (SVR) [6, 18].

Espinoza-Cuadros et al. [18] collected the largest corpus, having obtained worse results than the other works reviewed in both regression and classification tasks. This motivated them to make a careful review of previous works and on possible pitfalls that could be responsible for overoptimistic results. The authors pointed out three main pitfalls: small corpora and very often unbalanced in terms of classes, which is more prone to overfitting; presence of confounding variables such as gender, age and body mass index unevenly distributed between classes; and feature selection on high dimensionality feature spaces when little data is available, which is also most likely to cause data overfitting.

Goldshtein et al. [14] is the only work so far that made a separate analysis for female and male speakers, and they reported better results for female speakers. Elisha et al [15] observed that the phonemes carrying more distinguishing information were the vowel /a/ and the nasal phonemes (/m/ and /n/), which is consistent with the resonance anomalies previously described.

The works of Kriboy et al. [16] and Sol-Casals et al. [13] hypothesized that acoustic properties of speech that are altered by body position help distinguish between OSA and Non-OSA subjects. In fact, there is an increased frequency and severity of apneas in supine position, most likely due to unfavorable airway geometry, increase in collapsibility, gravity and inadequate dilator muscle compensation.

3. CORPORA

This section describes the two corpora collected for this study: the Portuguese Sleep Disorders (PSD) Corpus and the in-the-wild OSA (WOSA) corpus.

3.1. Portuguese Sleep Disorders (PSD) Corpus

This corpus includes speech recordings of people suffering from sleep disorders (OSA and insomnia) and of volunteers self-diagnosed as not suffering from sleep disorders. The collection of speech from patients took place at Centro de Encefalografia e Neurofisiologia Clínica (CENC), during their first medical appointment. Patients were later diagnosed with OSA (25), insomnia (32) or both (3). The collection of speech from 20 volunteers took place at the Clinic and at INESC-ID facilities. All patients under 18 years old, or suffering from psychiatric comorbidities, or already under treatment, or unable to speak Portuguese were excluded from our study. The Ethics Committee of Instituto Superior Técnico approved the experimental protocol.

The recordings included four tasks: 1) reading the Portuguese version of the tale "The North Wind and the Sun", a phonetically rich text; 2) pronouncing elongated vowels /a/ and /i/, which are expected to reveal phonation anomalies and some degree of muscle fatigue; 3) a reading span task; and 4) free description of Vincent Van Gogh's painting "Bedroom in Arles".

The reading span task (task 3), based on [21], required participants to read out aloud 10 possibly illogical sentences, and to classify them as logical or illogical. At the same time, letters were being displayed in between the sentences, that the subjects had to memorize and later identify. This task was designed to enable the assessment of working memory, with the number of correctly recalled letters, in the correct order (score) [22]. This exercise was chosen to allow for a cross study between sleep disorders, interpreted here as signs of fatigue, and cognitive load assessment through speech analysis.

Table 1 (a) and b)) shows the number of subjects in the PSD corpus, the number of audio files per task, per subject, and the average duration of the audio files per task. This is, to the best of our knowledge, the first speech corpus for sleep disorders fully spoken is Portuguese, although three of the speakers were Brazilian, two were Spanish and one was Angolan. It is also a unique sleep disorder corpus, due to the inclusion of spontaneous speech and the relationship of one of the tasks with cognitive load assessment.

Table 1: PSD corpus characterization. #F and #M refer to the number of female and male speakers, respectively.

(a)

Class	#F	#M	Age (mean $\pm std$)
Control	12	8	34 $\pm$11
OSA	6	19	53 $\pm$10
Insomnia	21	11	50 $\pm$14

(b)

Task	#Audio Files	Duration (s) (mean $\pm std$)
1	1	33.8 $\pm$5.6
2	2	3.5 $\pm$1.3
3	10	4.3 $\pm$1.4
4	1	52.1 $\pm$31.7

3.2. In-the-wild OSA (WOSA) Corpus

Following the idea presented by Correia et al. in [23], we created a second corpus obtained from *vlogs* available at *youtube.com*. This pilot corpus currently includes recordings from 16 English speaking subjects, 8 OSA and 8 controls. In each class, half of the subjets were male and half female. From each *vlog*, 12 segments were extracted, with average duration of 7.2 s, in order to allow a more fair comparison with the PSD corpus.

From the 8 OSA subjects, 6 were under treatment of continuous positive airway pressure during sleep, 1 used an oral appliance and 1 was not under treatment. The control subjects were randomly selected from *vlog* featuring unrelated topics.

Each vlog included (or not) a self claim of suffering from OSA, which was manually verified. In spite of the noisy labels, this small set was intended, on one hand, to enable the validation of the results obtained with the laboratory corpus, and on the other hand to serve as a proof-of-concept on using in-the-wild data to diagnose OSA, and to classify OSA speech in subjects undertaking treatment.

4. FEATURE EXTRACTION

4.1. Knowledge-based feature characterization

Based on the literature review and on the expected anomalies described in Section 2, we represent each audio file as a vector of 111 features. This feature set includes 25 features common to the eGeMAPS [24]: mean and standard deviation (std) of the frequency and bandwidth of formant 1, 2 and 3; mean and std of HNR; mean and std of jitter; mean, std, and percentile 20, 50, and 100 of F0; and mean and std of all frames and of only voiced frames of Spectral Flux. The set also includes 2 extra features: the differences between the frequencies of second and first formants, and third and second formants. Moreover, we added the mean of the 12 MFCC plus their first and second order derivatives, and 48 LPCC.

The formant frequencies are expected to reveal resonance anomalies, and jitter and HNR are expected to reveal the articulatory anomalies. Formant bandwidths are expected to be altered due to the altered properties of the soft tissue (pharingeal mucosa and soft tissue). The spectral flux and F0-related features were included in this feature set because they proved to be relevant for sleepiness detection [25]. LPCC and MFCC were chosen because of their relationship with the vocal tract shape.

4.2. Automatic feature selection

From this original feature set, henceforth abbreviated as OFS, we derived two other subsets, in an effort to avoid high dimensionality problems given the reduced size of the corpus:

- **Random Forest feature selection (RF)**: 5 most relevant features according to the RF ranking: mean of $\Delta\Delta$MFCC[11], mean of MFCC[12], std of F0, mean of $\Delta\Delta$MFCC[12], and percentile 20.0 of F0.
- **Mann-Whitney U test ranking (M-W)**: 17 features with lower p-value according to the Mann-Whitney U test.

The number of features considered for the RF and M-W feature sets corresponds to a maximization of classification results obtained with SVM, linear kernel, using the PSD corpus. Hence, they may lead to overoptimistic results in that data set, a fact that may be revealed when worse results are obtained for other corpora.

5. EXPERIMENTAL RESULTS

This section reports on OSA binary classification experiments performed on: a) the PSD corpus, b) the subset of male speakers in this corpus (PSD-m), c) the WOSA corpus, and d) a combination of PSD+WOSA corpora. Notice that in the case of PSD data, the analysis of the elongated vowels was excluded, since their acoustic features differ significantly from the features extracted for the remaining data tasks, and the number of recorded vowels was too small to allow for a separate analysis. Additionally, the classification of insomnia patients was not addressed either.

Table 2 summarizes the main results achieved. The table reports results at subject level, in which each segment prediction is combined using a majority vote strategy. The full set of experiments involved the comparison of five classifiers: SVM, kNN, LDA, Naïve Bayes and Random Forest. All classifiers were tested using the three feature sets, with leave-one-speaker-out cross validation. We also tested the combination of different classifiers, using both majority vote, and a weighted approach - quadratic best-worst weighted vote [26]. However, for sake of clarity, we report results only for the three best combinations of feature sets and classifiers according to PSD results (first row of Table 2). The second line of Table 2 shows the results obtained for the PSD-m subset. This subset was considered in order to verify whether the gender imbalance between classes was affecting the results. The third line shows corresponding results for the WOSA corpus. The last line shows results of training with a combination of PSD+WOSA, but testing only on speakers that belong to the WOSA corpus.

The best performing single classifier using the PSD corpus was SVM, both with the original and the RF feature sets. However, the combination by majority vote of SVM, kNN and LDA with the original feature set allowed an improvement of TNR to 80%. The RF feature set yielded the largest TPR in the PSD corpus (92%), which is essential for medical applications, but the correspondent TNR is rather low (65%). In fact, the RF feature set does not achieve satisfactory results in either the PSD-m subset, or the WOSA, or PSD+WOSA corpora. These observations suggest that the RF feature set was overfitting the PSD data.

On the other hand, OFS is able to characterize well other data sets, especially with the combination of SVM, LDA and kNN. In fact, this model applied to the subset PSD-m achieves TPR of 94.7% which represents an improvement, when compared to the whole PSD corpus. The TNR is much lower, though. This may be related to the fact that the subset PSD-m only includes 8 control subjects, compared to 19 OSA subjects.

Comparing the results obtained with PSD+WOSA with the WOSA and PSD alone, we observe that the combination of the two data sets produces worse results, especially in terms of TNR. This may be due to the fact that this is a cross language experiment and involves different types of verbal tasks.

The fact that we were able to accurately detect 87.5% of the OSA patients using the WOSA corpus suggest that the different types of treatments undertaken by the subjects from the WOSA corpus, do not annihilate the speech alterations typical of OSA speech.

5.1. Phoneme relevance for OSA detection

Table 3 represents a comparative analysis of the results obtained in each of the three verbal tasks included in the recording protocol that were considered. For task 3, the analysis is done separately for each read sentence.

Portuguese is a language that uses many diphthongs and nasal sounds (/m/, /n/, /ɲ/, /ĩ/, /ẽ/, /ɐ̃/, /õ/, /ũ/, /j̃/ and /w̃/). The inventory of nasal sounds is even wider in some regions of the country. In Table 3, one can observe that the sentence for which the classification

Table 2: Results achieved with the three best models: RF features and SVM classifier; OFS features and SVM classifier; and OFS features and the majority vote fusion of SVM, LDA and kNN, for the four data sets studied. TPR stands for True Positive Rate, TNR for True Negative Rate and WA for Weighted Accuracy.

	RF features; SVM			OFS features; SVM			OFS features; SVM+LDA+kNN		
	TPR (%)	TNR (%)	WA (%)	TPR (%)	TNR (%)	WA (%)	TPR (%)	TNR (%)	WA (%)
PSD	92.00	65.00	80.00	88.00	75.00	82.22	88.00	80.00	84.44
PSD-m	100.00	50.00	85.00	89.47	62.50	81.48	94.74	62.50	85.19
WOSA	12.20	37.50	25.00	75.00	87.50	81.25	75.00	87.50	81.25
PSD+WOSA	50.00	25.00	37.50	75.00	62.50	68.75	75.00	62.50	68.75

Table 3: Comparison of the classification results per verbal task with the number of nasal phonemes and diphthongs.TPR stands for True Positive Rate, TNR for True Negative Rate and WA for Weighted Accuracy.

Task	Nasal phonemes (%)	Diphthongs (%)	Performance TPR (%)	TNR (%)	WA (%)
1	12.6	6.4	84.0	70.0	78.8
3.1	13.5	5.7	84.0	65.0	75.6
3.2	**25.0**	**10.0**	**92.0**	**75.0**	**84.4**
3.3	18.8	6.3	72.0	65.0	68.9
3.4	6.5	6.5	84.0	70.0	77.8
3.5	12.1	5.1	92.0	65.0	80.0
3.6	8.9	4.4	80.0	85.0	82.2
3.7	14.0	7.0	84.0	75.0	80.0
3.8	14.3	7.1	84.0	75.0	80.0
3.9	11.5	1.9	92.0	65.0	80.0
3.10	16.0	4.0	88.0	60.0	75.6
4	-	-	92.0	55.0	75.6

results were better corresponds to the sentence with higher relative frequency of nasal phonemes and diphthongs. This represents an evidence for the presence of resonance anomalies related to the abnormal coupling of the vocal tract with the nasal cavity in OSA subjects, and for the presence of phonation anomalies, related to a possible inflammation of the larynx, caused by snoring. These findings are also consistent with [15], which reports that nasal phonemes carry more distinguishing information for OSA detection.

We also observe that task 4 (spontaneous speech), despite achieving a very high TRP (92%), achieves a rather low TNR (55%), comparable to chance. This may be due to the fact that non-speech segments were only removed in the beginning and in the end of each file. In fact, the duration of interpausal units in this task may also be an important cue to cognitive load analysis.

6. SLEEP DISORDERS' IMPACT ON WORKING MEMORY

Baddeley (1992) defined working memory as a brain system that provides temporary storage of information, as well as processing of that information [27]. The shared resources for storage and processing in the working memory involve a trade off between the complexity of processing a given task and short-term memory [28]. Cognitive load refers to the demands placed on a person's working memory by (a) the main task that he or she is performing, (b) any other task(s) he or she was performing concurrently, and (c) distracting aspects of the situation in which he or she finds himself or herself [29].

We compared the score obtained by control subjects and sleep disordered subjects (for this analysis, we consider both insomnia and OSA patients) in task 3 of the PSDC corpus, and observed that control subjects obtained a mean score of 7.7/10, insomnia patients obtained 5.3/10 and OSA patients obtained 4.8/10. This suggests that sleep disorders impair working memory, which is coherent to the findings of previous works that use functional brain imaging to associate sleep deprivation and impaired working memory [30][31].

We repeated the score comparison across the age ranges with available data, in intervals of ten years, and we observed that the mean score of control subjects is higher than that of sleep disordered subjects, for all age ranges.

Naturally, one should be cautious when drawing conclusions, because these score differences can also be influenced by other factors, for which we do not hold information.

7. CONCLUSIONS

This study addresses OSA detection with both read and spontaneous speech collected in a sleep clinic, and also using in-the-wild data.

The feature set we designed for OSA detection is able to achieve very promising results, when compared to previous works, although the corpora were very different, both in terms of language and type of materials collected (read speech, yes/no questions, words, and sustained vowels in all other works reviewed).

We found evidence for the resonance and phonation anomalies foreseen by the work of Fox [20] with the comparative analysis of performance versus the relative frequency of nasal phonemes and diphthongs.

Although the results obtained are very promising, the major limitation is the reduced size of the corpora under study.

The results obtained with the vlog corpus were particularly interesting, as they involved cross language experiments, as well as experiments with subjects undertaking different types of treatment.

The need for manual verification of the OSA claim in the WOSA corpus was the main factor preventing its expansion. In fact, our initial query retrieved vlogs from subjects with central sleep apnea as well as vlogs from subjects with obstructive sleep apnea. We plan to address this issue in the near future using a bag-of-words model applied to the vlog transcription [23].

Future work will also further explore the spontaneous speech subset of the PSD corpus, in particular, the intra-speech inhalation sounds during which the entire vocal tract is exposed to enable maximum intake of air. These breath sounds have been shown to carry information about the speakers identity [32], and may also reveal

some of the anomalies caused by OSA.

We also observed an impairment of the working memory related to the presence of sleep disorders. We hope this evidence contributes to raise general awareness to the importance of sleep health.

References

[1] J. Arnold, M. Sunilkumar, and V. Krishna et al., "Obstructive sleep apnea," *Journal of pharmacy & bioallied sciences*, vol. 9, no. Suppl 1, pp. S26, 2017.

[2] T. Paiva, M. Andersen, and S. Tufik, "Sono e a medicina do sono. 1ª edição," 2014.

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

[4] R. Pozo, J. Murillo, and L. Gómez et al., "Assessment of severe apnoea through voice analysis, automatic speech, and speaker recognition techniques," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, pp. 982531, 2009.

[5] C. Senaratna, J. Perret, and C. Lodge et al., "Prevalence of obstructive sleep apnea in the general population: a systematic review," *Sleep Medicine Reviews*, vol. 34, pp. 70–81, 2017.

[6] M. Kriboy, A. Tarasiuk, and Y. Zigel, "A novel method for obstructive sleep apnea severity estimation using speech signals," in *ICASSP*. IEEE, 2014.

[7] A. Balaei, K. Sutherland, and P. Cistulli et al., "Automatic detection of obstructive sleep apnea using facial images," in *ISBI*. IEEE, 2017.

[8] H. Nosrati, N. Sadr, and P. de Chazal, "Apnoea-hypopnoea index estimation using craniofacial photographic measurements," in *CinC*. IEEE, 2016.

[9] B. Calabrese, F. Pucci, and M. Sturniolo et al., "Automatic detection of obstructive sleep apnea syndrome based on snore signals.," in *MAVEBA*, 2009, pp. 185–188.

[10] C.-M. Cheng, Y.-L. Hsu, and C.-M. Young et al., "Development of a portable device for telemonitoring of snoring and obstructive sleep apnea syndrome symptoms," *Telemedicine and e-Health*, vol. 14, no. 1, pp. 55–68, 2008.

[11] R. Puri, A. Athanassiadis, and N. Gill et al., "Design and preliminary evaluation of a wearable device for mass-screening of sleep apnea," in *EMBC*. IEEE, 2016.

[12] S. Pittman, N. Ayas, and M. MacDonald et al., "Using a wrist-worn device based on peripheral arterial tonometry to diagnose obstructive sleep apnea: in-laboratory and ambulatory validation," *Sleep*, vol. 27, no. 5, pp. 923–933, 2004.

[13] J. Solé-Casals, C. Munteanu, and O. Martín et al., "Detection of severe obstructive sleep apnea through voice analysis," *Applied Soft Computing*, vol. 23, pp. 346–354, 2014.

[14] E. Goldshtein, A. Tarasiuk, and Y. Zigel, "Automatic detection of obstructive sleep apnea using speech signals," *IEEE Transactions on biomedical engineering*, vol. 58, no. 5, pp. 1373–1382, 2011.

[15] O. Elisha, A. Tarasiuk, and Y. Zigel, "Automatic detection of obstructive sleep apnea using speech signal analysis," in *Afeka-AVIOS Speech Processing Conference 2012*, 2012.

[16] M. Kriboy, A. Tarasiuk, and Y. Zigel, "Detection of obstructive sleep apnea in awake subjects by exploiting body posture effects on the speech signal," in *EMBC*. IEEE, 2014.

[17] A. Benavides, R. Pozo, and D. Toledano et al., "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.

[18] F. Espinoza-Cuadros, R. Pozo, D. Toledano, J. Alcázar-Ramírez, E. Gonzalo, and L. Hernandez-Gomez et al., "Reviewing the connection between speech and obstructive sleep apnea," *Biomedical engineering online*, vol. 15, no. 1, pp. 20, 2016.

[19] M. Blumen, F. Chabolle, and E. Rabischong et al., "Dilator muscles of the pharynx and their implication in the sleep apnea syndrome of obstructive type. review of the literature," in *Annales d'oto-laryngologie et de chirurgie cervico faciale: bulletin de la Societe d'oto-laryngologie des hopitaux de Paris*, 1998, vol. 115, pp. 73–84.

[20] A. Fox, P. Monoson, and D. Morgan, "Speech dysfunction of obstructive sleep apnea: A discriminant analysis of its descriptors," *Chest*, vol. 96, no. 3, pp. 589–595, 1989.

[21] B. Schuller, S. Steidl, and A. Batliner et al., "The interspeech 2014 computational paralinguistics challenge: Cognitive & physical load," in *INTERSPEECH*, 2014.

[22] N. Unsworth, R. Heitz, and J. Schrock et al., "An automated version of the operation span task," *Behavior research methods*, vol. 37, no. 3, pp. 498–505, 2005.

[23] J. Correia, B. Raj, and I. Trancoso et al., "Mining multimodal repositories for speech affecting diseases," 2018.

[24] F. Eyben, K. Scherer, and B. Schuller et al., "The Geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.

[25] C. Botelho, "Speech as a biomarker for sleep disorders and sleep deprivation," M.S. thesis, Instituto Superior Tcnico, University of Lisbon, 2018.

[26] F. Moreno-Seco, J. Inesta, and P. De León et al., "Comparison of classifier fusion methods for classification in pattern recognition tasks," in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2006, pp. 705–713.

[27] A. Baddeley, "Working memory," *Science*, vol. 255, no. 5044, pp. 556–559, 1992.

[28] P. Barrouillet, S. Bernardin, and S. Portrat et al., "Time and cognitive load in working memory.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 33, no. 3, pp. 570, 2007.

[29] A. Berthold and A. Jameson, "Interpreting symptoms of cognitive load in speech input," in *UM99 User Modeling*, pp. 235–244. Springer, 1999.

[30] M. Chee and W. Choo, "Functional imaging of working memory after 24 hr of total sleep deprivation," *Journal of Neuroscience*, vol. 24, no. 19, pp. 4560–4567, 2004.

[31] R. Thomas, B. Rosen, and C. Stern et al., "Functional imaging of working memory in obstructive sleep-disordered breathing," *Journal of Applied Physiology*, vol. 98, no. 6, pp. 2226–2234,

2005.

[32] W. Zhao, Y. Gao, and R. Singh, "Speaker identification from the sound of the human breath," *arXiv preprint arXiv:1712.00171*, 2017.