TÉCNICO LISBOA

Biomarkers of sleep disorders

António Moura Sarmento de Magalhães Vieira

Thesis to obtain the Master of Science Degree in

Computer Science and Engineering

Supervisor(s): Prof. Isabel Maria Martins Trancoso
Prof. Alberto Abad Gareta

Examination Committee

Chairperson: Prof. Manuel Fernando Cabido Peres Lopes
Supervisor: Prof. Alberto Abad Gareta
Member of the Committee: Prof. Rui Miguel Carrasqueiro Henriques

November 2021

Dedicated to my mother, my father and my grandmother.

Acknowledgments

I would like to thank Professor Isabel and Professor Alberto, who helped me make this thesis and help me have a better understanding of speech processing since I had no prior knowledge in that matter. I would also like to thank Doctor Teresa Paiva who helped me build the JotForm Corpus by disclosing the JotForm survey to the CENC patients.

Resumo

Esta proposta de tese retoma o trabalho de Catarina Botelho “Speech as a Biomarker for Sleep Disorders and Sleep Deprivation” e complementa-o adicionando análise de tosse e ressonar à detecção e diagnóstico de distúrbios do sono para todos os outros sinais de fala recolhidos. Esse distúrbio do sono (Apneia Obstrutiva do Sono) pode causar acidentes de trânsito fatais por fadiga e também acidentes de trabalho e até o momento o melhor exame de diagnóstico, a polissonografia, não é prático e é desconfortável para o paciente.

Ao avaliar as características dos sinais de fala das pessoas testadas com um SVM, fomos capazes de alcançar uma taxa de precisão de 91% por votação.

Os resultados foram obtidos a partir de dois corpora distintos, o primeiro com 40 indivíduos e foi compilado a partir de vlogs do YouTube e o segundo com 26 indivíduos e foi compilado através de um questionário JotForm.

Palavras-chave: Apneia Obstrutiva do Sono, insónia, distúrbios do sono, fala, tosse, diagnóstico automático.

Abstract

This thesis proposal takes on the work of Catarina Botelho in her Thesis ”Speech as a Biomarker for Sleep Disorders and Sleep Deprivation”. 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 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, insomnia, sleep disorders, speech, cough, computational diagnosis.

Contents

List of Tables

List of Figures

Chapter 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” [1]).

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.

Some papers used as research material for this Thesis are: ”Speech dysfunction of obstructive sleep apnea: A discriminant analysis of its descriptors” by A, Fox et al. [2], ”Detection of severe obstructive sleep apnea through voice analysis” by J. Solé-Casals et al. [3] and “Automatic detection of obstructive sleep apnea using speech signal analysis” by O. Elisha et al. [4] .

C. Botelho collected a speech dataset in which the subjects had to complete several tasks. The first task was to read a small tale, the Portuguese version of ”the North Wind and the Sun”. This fable is famous by its use in phonetic descriptions of languages as an illustration of spoken language. It is recommended by the IPA for the purpose of eliciting all phonemic contrasts that occur in English when conducting tests by foreign users or of regional usage. Whilst translating it to the Portuguese language, it was kept in mind that the phonetics of this language was highlighted. The second task was to record elongated vowels spoken by the subjects. The third task was a reading span task, were the subjects were recorded reading some sentences and between sentences they were shown a letter of the alphabet to memorize, at the end of the task, the subjects’ memory was recorded according to the capability of memorizing the interpolated letters. In the fourth and last task, subjects had to describe a sleep related image. Most of the subjects that suffered from sleep disorders revealed negative feelings towards the image.

The initial corpus was composed of 12 female subjects and 8 male subjects for the control group and 6 female subject and 19 male subject for the Obstructive Sleep Apnea confirmed group **(Table 1)**.

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

Table 1.1: Portuguese Sleep Disorders (PSD) Corpus.

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

Table 1.2: Portuguese Sleep Disorders balanced (PSD-b) Corpus.

However this was an unbalanced corpus and would jeopardize the results, so a second corpus was formed in order to avoid that possibility. It was composed of 11 female and 11 male subjects for the control group (some of which were suffering with Insomnia, which could also improve differentiation between the two sleep disorders) and 9 female and 11 female subjects for the OSA group **(Table 2)**.

C. Botelho also collected a small In-the-Wild Obstructive Sleep Apnea (WOSA) Corpus with data obtained from audio in YouTube videos. This Corpus is composed of 4 male and 4 female subject in each group (Control and OSA) **(table 3)**.

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 we selected based on the literature and an ensemble of SVM, LDA and kNN with majority vote. These results were achieved in PSD corpus, with 25 subjects suffering from OSA and 20 control subjects. Since the corpus is small, the results were validated with data acquired from youtube, like previously mentioned. The PSD-b corpus achieved TPR of 80% and TNR of 72.72%.

Class	#Female subjects	#Male subjects	#Subjects under CPAP treatment	#Subjects using oral appliances	Subject not under treatment
Control	4	4	-	-	8
OSA	4	4	6	1	1

Table 1.3: in-the-Wild Sleep Apnea (WOSA) Corpus.

1.1 Motivation

The previously mentioned Thesis talks about how sleep disorder, 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 [1].

Some facts presented in C. Botelho's work are: between 230,000 and 345,000 people are expected to be killed in traffic accidents due to fatigue, one third of adults suffer of inadequate sleep, 9% to 38% of the adult population suffer from Obstructive Sleep Apnea, Sleep related traffic accidents have an injury severity level similar to alcohol intoxication related traffic accidents, 46% of Obstructive Sleep Apnea couples sleep in separate rooms, among others [1].

1.2 Objectives

We took the already developed work and the useful contents of C. Botelho's work and built on them, mainly by expanding 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.

Chapter 2

Background

2.1 Sleepiness and Sleep Disorders

Sleep is a naturally reversible state that contributes to the optimal development of physical and mental health, being regulated by cellular and molecular mechanisms [5]. Sleep is a cyclic process with four stages: the first three stages are associated with non-rapid eye movement (NREM) sleep, which is characterized by deeper sleep, regular breathing rate and slow electroencephalogram waves; the fourth stage is the rapid eye movement (REM) sleep, which is associated with variable breathing rate, brain waves that mimic the awake state and intense dreaming. One complete cycle lasts approximately 90 minutes, and one night of sleep includes 5 to 6 cycles. Generally, in the beginning of the night, the human sleep is deeper with short REM stages. Later in the night the REM stages become longer [5]. Sleep is essential for many vital tasks, including energy conservation, brain waste clearance, modulation of immune responses, performance, cognition, vigilance and psychological state [5]. Sleep deprivation thus interferes with the maintenance of vital functions and results in an increasing sleep propensity and destabilization during awake periods which have a negative impact in mood, cognitive performance and motor function. Certain neurocognitive domains, such as working memory, executive attention and divergent higher cognitive functions are specially sensitive to sleep loss [6]. Sleep deprivation may be caused by inadequate sleeping habits associated with work or life-style sleep restrictions [7], and by sleep disorders, such as insomnia, sleep apnea and restless legs syndrome [6]. It causes, in a first instance, a decrease in well-being and performance, daytime fatigue and sleepiness, and in long term, it is associated with cardiovascular disease, obesity, diabetes, hypertension, inflammation, anxiety, depression and premature death [8].

2.2 OSA

Obstructive sleep apnea (OSA) syndrome is a common sleep disorder in which snoring, choking, repeated awakenings, and prolonged daytime sleepiness are the main symptoms. Such sleepiness can also become debilitating and dangerous and has been related to an increase in the risk of accidents

involving vehicles. Dry mouth, morning headaches, and sore throat may also be caused by OSA upon waking. [9]

The condition of obstructive sleep apnea (OSA) is a widespread but frequently unrecognized condition triggered by pharyngeal collapse **(Figure 2.1)** during sleep and marked by repeated awakenings, interrupted sleep and consequent extreme sleepiness during the day. The prevalence of the disease will rise over the coming years with the increasing epidemic of obesity, the most significant risk factor for OSA, thereby posing an important public-health issue. [9]

The link between OSA and asthma, metabolic syndrome, diabetes, heart failure, coronary artery disease, arrhythmias, stroke, pulmonary hypertension, neurocognitive disorders, and mood disorders is currently acknowledged. Diagnosis is based on the combined clinical manifestation assessment and the findings of the objective sleep study. Snoring, sleepiness and severe occurrences of sleep apnea episodes are among the cardinal symptoms. In order to confirm the clinical suspicion of OSA syndrome, to assess its severity and to guide therapeutic decisions, polysomnography is the gold standard. Behavioral, medical and surgical options are available for the treatment. In most patients, Continuous Positive Airway Pressure (CPAP) constitutes the treatment of choice. Efficient in minimizing symptoms, coronary morbidity and mortality and neurocognitive sequelae, CPAP has been proven to be effective, but is also poorly accepted. Surgery and pharmacological therapy as first-line treatment are not validated by the results of clinical studies, although these treatments can be effective in chosen patients. A greater understanding of disease-based processes could enhance clinical strategies and decrease the social impact of OSA syndrome. [9]

2.3 Neurocognitive somatic symptoms of OSA

Impaired neurocognitive activity is associated with OSA. Like focus and concentrating, visuospatial and verbal memory, executive function, constructional skills and psychomotor functioning, all cognitive domains are affected [10].

Reduced gray matter associated with OSA severity was revealed by magnetic resonance imaging[11]. Beebe [12] found that vigilance was significantly compromised in a meta-analysis of 1092 patients with OSA; thus, patients with OSA often have difficulties focusing and retaining focus for long periods. The disease also greatly impairs the area of executive functioning, the capacity to establish and maintain a coordinated response to problem circumstances, and for fine-motor synchronization it is deleterious.

OSA can, primarily through intermittent hypoxia, promote cognitive impairment. An animal model of chronic episodic hypoxia develops performance-impaired neurodegenerative changes in the hippocampus and cortex during cognitive spatial task acquisition[13]. Some research in humans reported a substantial association between hypoxemia severity and neuropsychological impairment [14] [15] [16]. Findley and colleagues [17] found that people who have sleep apnea with concurrent hypoxemia have more serious cognitive impairment than those without hypoxemia.

Hypersomnolence can also play a role in the development of neurocognitive impairment due to sleep fragmentation [18] [19]. The association between OSA and depression is not fully clear. Peppard [20]

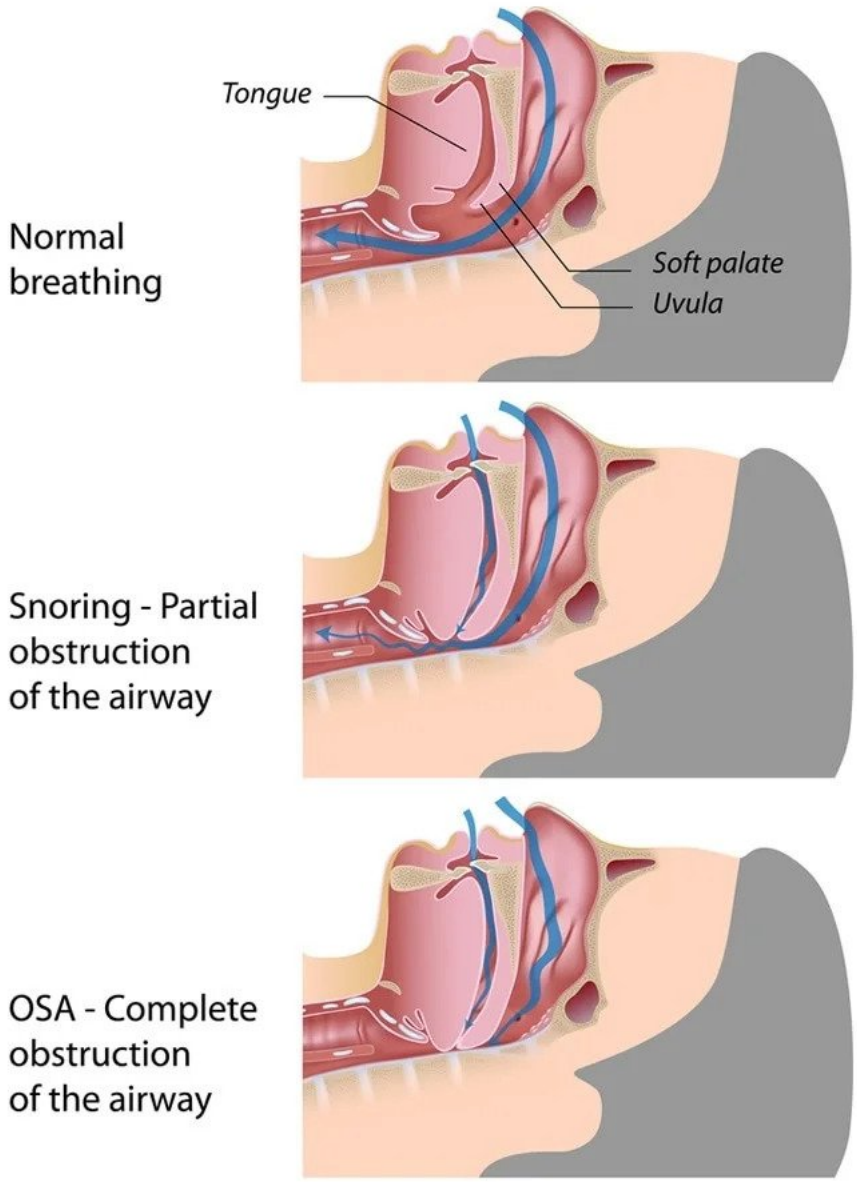

Figure 2.1: OSA Syndrome. `https://d1n5s2tett0dwr.cloudfront.net/0Vn0ajLL9U1M9mJspkjzkn_VLU0=/323x404/d3b3by4navws1f.cloudfront.net/shutterstock_119606590.jpg`

also found that patients with minor sleep breathing disorders have a 2-fold elevated risk of developing depression in a prospective cohort study of 1408 patients. No important association between OSA and depression was observed in other studies [21] [22]. McMahon and colleagues found that constant positive airway pressure (CPAP) had an important and positive effect on depression in a systematic literature review [23].

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. These sleep problems are transient or of minimal significance for most. Prolonged sleeplessness, however, is frequently associated with severe

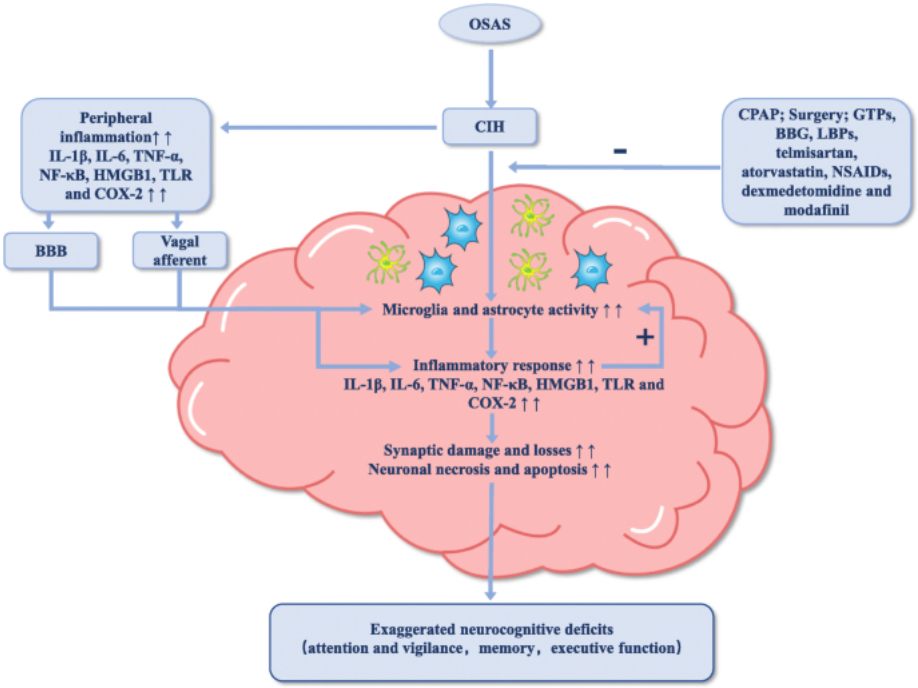

Figure 2.2: The relationship between inflammation and neurocognitive dysfunction in OSA. https://doi.org/10.1186/s12974-020-01905-2

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. [24]

Insomnia symptoms may also be an independent risk factor for suicide attempts and deaths from suicide, independent of depression. In dynamic cognitive functions, neuropsychological research reveals deficits, including working memory and attention splitting, which are not simply linked to diminished alertness. On the basis of the inferred original cause of the sleeplessness, older diagnostic systems sought to separate 'primary' from 'secondary' insomnia. However, since causal associations are frequently bidirectional between multiple medical and psychological conditions and insomnia, such assumptions are unreliable. [24] Some more symptoms of this disorder are represented in **Figure 2.3**.

In addition, the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders takes a strictly analytical approach, focusing on the occurrence and length of symptoms, attributable to the inadequate efficacy of insomnia subtyping based on phenotype or patho-physiology, making a diagnosis of insomnia disorder irrespective of any coexisting psychiatric disorders. Such symptoms can be: dissatisfaction with sleep quantity or quality, with one or more of the following symptoms: difficulty initiating sleep, difficulty maintaining sleep, characterized by frequent awakenings or trouble returning to sleep after awakenings, early-morning awakening with inability to return to sleep; the sleep disturbance causes clinically significant distress or impairment in daytime functioning, as evidenced by at least one of the following: fatigue or low energy; daytime sleepiness, Impaired attention, concentration, or memory; mood disturbance; behavioral difficulties; impaired occupational or academic function; impaired interpersonal or social function; negative effect on caregiver or family functioning; the sleep difficulty occurs at least 3 nights per week, is present for at least 3 months, and occurs despite adequate opportunity for sleep.

The clinician should monitor whether treatment of such coexisting disorders normalizes sleep, and if not, treat the insomnia disorder independently. [24]

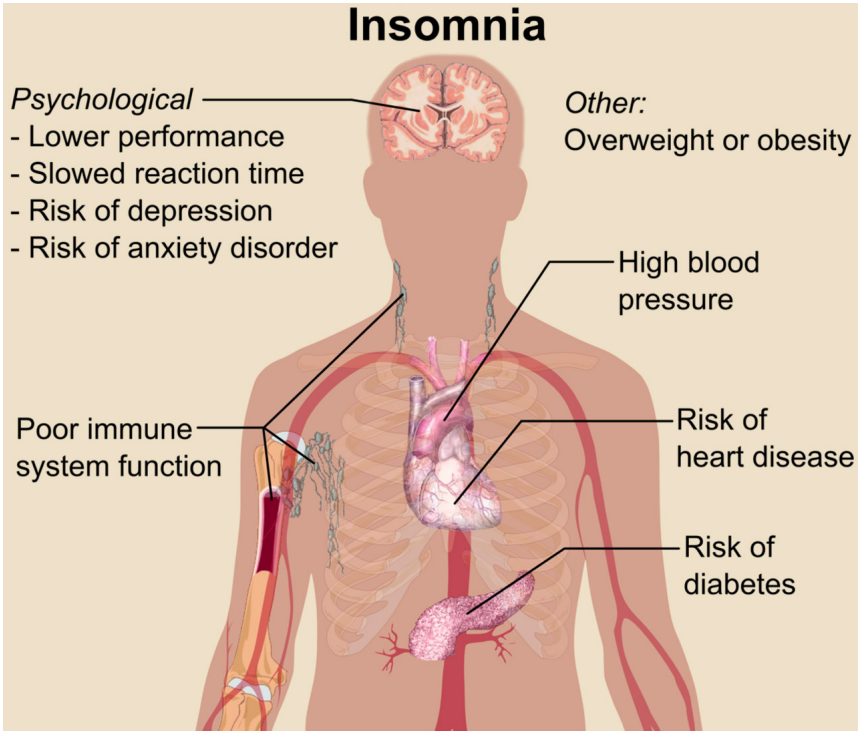

Figure 2.3: Insomnia symptoms. `http://www.sci-news.com/medicine/gene-regions-insomnia-06977.html`

Chapter 3

Related Work

3.1 Speech Signals

Some useful papers that can help to understand the mechanics and the already attempted processes in identifying Obstructive Sleep Apnea using the speech biomarker are ”Diagnosis of Obstructive Sleep Apnea Using Speech Signals From Awake Subjects” [25] and ”Prediction of Sleepiness Ratings from Voice by Man and Machine” [26].

The first one talks about how the most widely used objective measure of OSA severity is the number of obstructive events per hour of sleep, known as the apnea-hypopnea index (AHI). The AHI is often used to assess the severity of the syndrome. This index reflects the average number of obstructive apnea and hypopnea events per hour of sleep, and is calculated by dividing the total number of events by the number of hours of sleep. In the literature, two AHI thresholds of either 10 or 15 events per hour have been suggested to classify OSA into severity groups. In this study, the authors refer to a subject with AHI greater than or equal to 15 as an OSA patient and a subject with AHI less than 15 as a non-OSA patient.

This study reports an innovative system to identify OSA subjects while they are awake, not asleep, using speech signal processing techniques. The assumption is that OSA affects the acoustic parameters of speech because it is associated with anatomical and functional abnormalities of the upper airway. The primary goal of this study was to develop an OSA monitoring tool that estimates OSA severity from a short speech utterance recorded prior to bedtime. This tool may increase accessibility and decrease the percentage of undiagnosed cases. The system associates three different sub-systems based on features extracted from breathing segments within continuous speech signals, information acquired from sustained vowels using a convolutional neural network, and inherent information in continuous speech signals using a recurrent neural network.

The three main extractions from the speech data base used in this work were:

- **Breathing Segments**

  This study only analyzed the tracheal breath sounds that were recorded while in the supine and upright sitting positions, using both nose and mouth breathing at a medium flow rate (without any

talking). In the current study OSA was estimated from spontaneous breathing sounds within a continuous speech signal.

- **Sustained Vowels**

  In this algorithm, the performance of a CNN in estimating AHI from speech signals containing sustained vowel sounds was analyzed. Sustained vowels have numerous advantages such as being produced with less effort than continuous speech signals and the fact that they are not affected by speech rate, vocal pauses or stress.

- **Continuous Speech Signal**

  In RNNs all the inputs are related to each other, so the RNN to some extent keeps the information that was fed into it [27], [28]. Speech is a typical example of an inherently dynamic process with complex correlations. Hence, the RNN can exploit this information to estimate OSA severity. Here the authors used the long-short-term memory (LSTM) RNN architecture.

Each of these sub-systems provided an AHI estimation and were combined with age and body mass index (BMI) to produce a composite system that estimates AHI using a linear regression. As expected, the composite AHI estimation yielded the superior results, with a Pearson correlation coefficient of 0.61 between the estimated and diagnosed AHI. To distinguish between OSA and non-OSA subjects, a classification decision was made using an AHI threshold of 15 events per hour. The system achieved an average accuracy of 77.14%, a sensitivity of 75%, and a specificity of 79%.

The second above mentioned paper [26] approaches the prediction of sleepiness using the speech biomarker. This paper 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 of this document 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” [29], has some useful information regarding speaker recognition systems using speech signals.

Data augmentation was used in this paper to boost the efficiency of deep neural network (DNN) embedding for speaker recognition. The DNN, which is trained to differentiate across speakers, maps utterances of variable length to fixed-dimensional embeddings called x-vectors.

A very active research field currently is the use of deep neural networks (DNN) to obtain speaker characteristics. Representations called x-vectors are extracted from a DNN in the authors' approach and used like i-vectors. This paper builds on their recent architecture for DNN embedding systems [30]. Increasing training data with noise and reverberation artificially is shown to be a highly effective strategy for improving performance in DNN embedding systems. Speaker recognition systems can be based on i-vectors[31]. The standard approach consists of a universal background model (UBM) and a large **T** projection matrix that are learned to maximize the probability of data in an unsupervised manner. In a low-dimensional representation, known as an i-vector, the projection maps high-dimensional statistics from the UBM. To compare i-vectors, a probabilistic linear discriminant analysis (PLDA)[32] classifier is used to allow same-or-different speaker decisions[33] [34] [35].

There were four recognition systems developed for this study, which consist of two i-vector baselines and the DNN x-vector system. All systems were built using the Kaldi speech recognition toolkit [36]:

- **Acoustic i-vector**

  As an acoustic-feature basis system, a traditional i-vector system based on the GMM-UBM recipe described in [37] has been used. The features are 20 MFCCs that are mean-normalized over a sliding window of up to three seconds with a frame length of 25ms. To construct 60 dimensional

function vectors, Delta and acceleration are appended. Features corresponding to speech frames are chosen by an energy-based speech activity recognition (SAD) system. The UBM is a GMM full-covariance 2048 component. For scoring, the system uses a 600 dimensional i-vector extractor and PLDA.

- **Phonetic bottleneck i-vector**

This i-vector system from an ASR DNN acoustic model integrates phonetic bottleneck features (BNF) and is similar to [38]. With p-norm non-linearities, the DNN is a time-delay acoustic model. On the Fisher English corpus, the ASR DNN is trained and uses the same recipe and design as the system[37], except that a 60 dimensional linear bottleneck layer replaces the penultimate layer. Excluding the softmax outputlayer, which is not needed to compute BNFs, the DNN has 9.2 million parameters. The BNFs are concatenated with the same 20 dimensional MFCCs described in the Acoustic i-vector Section plus deltas to create 100 dimensional features. The remaining system components (feature processing, UBM, i-vector extractor, and PLDA classifier) are similar to those of the Acoustic i-vector Section.

- **The x-vector system**

This section describes the x-vector system. It is based on the DNN embeddings in [30] and also described in greater detail. The software framework has been made available in the Kaldi toolkit.

The features are 24 25ms frame-length dimensional filter banks, mean-normalized over a sliding window of up to 3 seconds. Non-speech frames are filtered out the same energy SAD as used in the baseline structures. Suppose there are T frames for an input segment. The first five layers work on speech frames, with the current frame t focusing on a small temporal context. The input to layer frame 3, for instance, is the spliced output of frame 2, at frames t3, t and t+3. This builds on the earlier layers' temporal context, so that frame 3 sees a cumulative context of 15 frames. Embeddings are extracted from the affine component of layer segment 6 after training. There are a total of 4.2 million parameters, excluding the softmax output layer and segment 7 (because they are not needed after training).

- **PLDA classifier**

The same type of PLDA [32] classifier is used for the x-vector and i-vector systems. The repre-

sentations (x-vectors or i-vectors) are centered, and projected using LDA. The LDA dimension was tuned on the SITW development set to 200 for i-vectors and 150 for x-vectors. The representations are length-normalized and modeled by PLDA after dimensionality reduction. Using adaptive s-norm [39], the scores are normalized.

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. The x-vectors achieved much lower error rates than the best baseline on Speakers in the Wild, after including a large amount of augmented microphone speech. In the best i-vector system, bottleneck features from an ASR DNN are used, and so transcribed data is required during training. On the other hand, only speaker labels are required to train the x-vector DNN, making it possibly ideal for domains with little transcribed speech. More broadly, it seems that x-vectors are now a dominant candidate for speaker recognition representations of the next generation.

3.3 Simulated snoring

In this next paper, Michael Herzog et al.[40] 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. There was no change in the AHI caused by the patient's position during simulated snoring. Patients with much dorsal mobility were more likely to have a greater AHI.
- An increase in pharyngeal collapse at the tongue's base level was associated with a rise in the AHI. The position of the patient during the simulated snoring had no bearing on the results of the test. AHI was expected to be higher in patients with a high degree of pharyngeal collapse at tongue-base level.
- During simulated snoring, pharyngeal collapse at the velum level was not associated with a high overnight AHI. It is impossible to predict the AHI based on the velar pharyngeal collapse.
- 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 in this paper 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 anatomic features of the upper

airway during simulated snoring and the AHI, in contrast to ”static” exams such as dorsalization of the tongue base, tonsil size, Mallampati index, and Mueller maneuver that do not correlate at all.

Even though some ear, nose, and throat professionals practice simulated snoring in awake individuals, the correlation with polysomnographic data has yet to be examined. The newly designed documentation and grading system allows for a better link between anatomic alterations and polysomnographic outcomes during simulated snoring.

3.4 Acoustic analysis of cough

William Thorpe et al.[41] conducted a study on the acoustic analysis of voluntary cough, which showed that it can be useful in the diagnosis of respiratory diseases.

They were able to describe voluntary cough sounds using numerous temporal and spectral descriptors in this investigation. These measurements were consistent across subjects and changed in response to airway narrowing caused by methacholine (parasympathomimetic drug administered in the form of its crystalline chloride especially to diagnose hypersensitivity of the bronchial air passages) inhalation. The data was simple to obtain because they recorded voluntary cough sounds, which eliminated the need for continuous sound monitoring from the patient when capturing spontaneous cough, which has the problem of inherent variability in the number of coughs and different lung volumes at which spontaneous coughs can occur.

The findings revealed that changes in cough sounds following a methacholine challenge were linked to changes in spirometric lung function. Both the durations of the cough phases and the entire spectral distribution of acoustic energy showed alterations. Furthermore, the cough sounds were substantially more consistent following methacholine, implying that the cough generation mechanism was confined to a smaller range due to the constricted airways.

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.

Chapter 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

Correia and colleagues [42] compiled a corpus from publicly available video blogs (vlogs) on YouTube as an attempt to overcome the problems associated with generating a large enough speech corpus for health applications.The in-the-wild speech medical corpus (WSM) originally included subsets from patients with either Parkinson's Disease and Depression, and a much smaller one with Cold. The WSM corpus was expanded during Catarina Botelho's MSc Thesis with a small subset including patients with OSA [1].

In the current thesis, the original in-the-wild OSA (WOSA) subset was expanded to more than twice the number of videos. The vlogs included in this subset are those in which people admit to having obstructive sleep apnea at some point. Other types of videos in which other individuals (e.g., doctors or students) talk about the disease, or talk about someone else who has the disease, or state they have had the disease in the past, must be filtered out. There is an added problem in the case of obstructive sleep apnea because there are two varieties of sleep apnea, obstructive and central, and people frequently claim to have "sleep apnea" without defining the type on such vlogs. On total, the WOSA corpus now includes speech recordings from 40 English-speaking subjects, 22 OSA patients, and 18 healthy controls (gender information on speakers that compose the corpus is presented in Table 4.1).

Table 4.1: 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. Some of them also reported that they were receiving different types of treatment, either continuous airway positive pressure, or oral appliances to keep the airway unobstructed. Others subjects reported they were not receiving any treatment or did not mention it at all, which is also another variable to consider. Other variable to consider is the age range, since we do not have detailed information about the subjects, we cannot know exactly their age. Furthermore, these speakers could be suffering from additional conditions that are not addressed but could influence their ability to talk. The control subjects' speech recordings were taken from a random selection of vlogs with unrelated topics. As a result, the corpus may be noisy, and precaution should be exercised when interpreting the results.

4.1.2 Feature Extraction

We extracted the audio of these vlogs and each audio file was processed by a Voice Activity Detector (VAD). A VAD is able to detect the audio segments where there is speech activity.With that information, we were able to filter out from those files most non-speech activity. 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. [43]

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.[44]

i-vectors and x-vectors

In order to extract i-vectors and x-vectors[29] from the audio files, we used pretrained models in kaldi toolkit [36].

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 [1], 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 [45]. 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 4.2 shows the best hyperparameters found by the grid search for each feature set. The best results are in Table 4.3.

Table 4.2: 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 4.3: 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 and comparing them to the previous study.

4.2.1 Corpus

The corpus for this experiment was compiled using data obtained from a JotForm [46] survey which was disclosed to the public and to the patients of the CENC [47] (Centro de Electroencefalografia e Neurofisiologia Clínica), which was important to gather medically diagnosed OSA patients.

Every subject that answered the survey was able to record the necessary audio files using their personal gadget's microphone. Even though they were not recording in a controlled environment, the audio files did not have any background noise nor music playing.

Four types of acoustic signals were requested in the survey:

Table 4.4: 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

- 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 (detailed information on speakers that compose the corpus is presented in Table 4.13).

4.2.2 Feature Extraction

Just like the previous speech study, all the audio files were split into 4 second segments with an overlap of 0.5 seconds for feature extraction. The same three feature sets were extracted from the audio files from the cough, snore, sustained vowel and speech signals: eGeMAPS, i-vectors and x-vectors.

4.2.3 Classifiers and hyperparameter choice

Just like in the previous speech study, we performed a grid search with leave-one-speaker-out cross validation in order to select the best SVM hyperparameter using the scikit-learn toolkit from python [45]. The exact same parameters were tested in the following experiments.

4.2.4 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 4.14 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 4.15.

Table 4.5: 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 4.6: 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

4.2.5 Snoring Biomarker experiment

The same procedure as before was applied to the recordings of the simulated snore of the subjects.

Since it was a reasonable way to do so, we decided to analyse it in the same context as in these experiments since it can be a useful biomarker.

Table 4.16 shows the best hyperparameters found by the SVM grid search and the best results are in Table 4.17.

Table 4.7: 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 4.8: 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

4.2.6 Sustained Vowel experiment

The same experiment was done on the sustained vowel audios. Although it is not a different biomarker from the speech signal, it is useful in order to compare results between all the previous experiments. The Table 4.18 shows the best hyperparameters found by the SVM grid search and the best results are in Table 4.19.

Table 4.9: 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 4.10: 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

4.2.7 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 4.20 shows the best hyperparameters found by the SVM grid search and the best results are in Table 4.21.

Table 4.11: 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 4.12: 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 contain-

Table 4.13: 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

ing only 4 female subjects, using only the male subjects might be useful to see if the models behave differently or if the results maintain.

4.3.2 Experimental results

The feature extraction, the classifiers and hyperparameter choice were the same as the previous experiment, the only difference in this case is the corpus. The follwing tables refer to the best hyperparameters for the 4 different subsets (cough, snore, sustained vowel and speech) and their best results.

Table 4.14: 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 4.15: 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 4.16: 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 4.17: 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 4.18: 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 4.19: 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 4.20: 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

Table 4.21: 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

4.4 Result discussion

For the first experiment, the best result is 74.19% of accuracy with majority voting and it was obtained with an SVM with the radial basis function kernel, with C of 100 and the γ with the value of 1e-04. 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. It might also be due to over fitting of the model or distinction between male and female features of the voice clips instead of healthy and OSA features.

Chapter 5

Conclusions

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

For the in-the-wild corpus created from public vlogs to test the speech biomarker, we found that the best classification results were obtained by extracting the i-vectors embedding from the audio files. The best result for this experiment is 74.19% of accuracy with majority voting and it was obtained with an SVM with the radial basis function kernel, with C of 100 and the γ with the value of 1e-04.

The second experiment was based on a different corpus and focused on four biomarkers: cough, snore, sustained vowel and speech. This corpus was created using audio files from a JotForm survey. The best classification results for the cough, snore, sustained vowel and speech modalities were, respectively, 91.67% of accuracy with majority voting using the eGeMAPS features classified by an SVM with the sigmoid function kernel, with C of 10 and the γ with the value of 1e-02; 60% of accuracy with majority voting using the x-vectors embedding classified by a SVM with the linear function kernel, with C of 100 and the γ with the value of 1e-03; 72% of accuracy with majority voting using the eGeMAPS features classified by a SVM with the sigmoid function kernel, with C of 10000 and the γ with the value of 1e-03; and finally 84% of accuracy with majority voting using the x-vector embeddings classified by a SVM with the sigmoid function kernel, with C of 10000 and the γ with the value of 1e-03.

The third experiment was based on the previous experiment's corpus but without the female subjects (male only corpus) and focused on the same four biomarkers: cough, snore, sustained vowel and speech. The best classification results for the cough, snore, sustained vowel and speech modalities were, respectively, 72.72% of accuracy with majority voting using the i-vector embeddings classified by a SVM with the sigmoid function kernel, with C of 10 and the γ with the value of 1; 72.72% of accuracy with majority voting using the x-vectors embedding classified by a SVM with the radial basis function kernel, with C of 100 and the γ with the value of 1e-04; 81.82% of accuracy with majority voting using the i-vector embeddings classified by an SVM with the sigmoid function kernel, with C of 1 and the γ with the value of 1; and finally 81.81% of accuracy with majority voting using the i-vector embeddings classified by a SVM with the sigmoid function kernel, with C of 0.1 and the γ with the value of 1.

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.

Regarding the automatic detection of obstructive sleep apnea for the in-the-wild corpus, even though the subject number was relatively even gender wise (21 male subjects and 19 female subjects) the amount of recorded voice time was limited by data pre-processing and removal of any clips with background noise and music, and the results might have been affected by the small amount of data.

Regarding the automatic detection of obstructive sleep apnea for the JotForm corpus, the amount of data was limited and the subject number was unbalanced (22 male subject and 4 female subjects). This is due to the fact that the disease has significantly higher incidence in male 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, as well as the possibility that our models will classify patients based on factors other than obstructive sleep apnea.

Regarding the automatic detection of obstructive sleep apnea for the male-only JotForm corpus, the amount of data was reduced compared to the original JotForm corpus which can affect the results.

Other impediment was the current global pandemic situation that prevented us from collecting more information about different biomarkers in order to evaluate them as possible indicators of obstructive sleep apnea.

5.2 Future Work

We suggest, for future work a comparison between our extracted feature sets and different ones to see how the results would differ and what would also be a practicable method of diagnosing obstructive sleep apnea through different biomarkers. It is also important to increase all corpora size in order to have more reliable results in all experiments. It would also be interesting to see other experiments with different biomarkers other than the ones that were studied in this work. Finally, the next step would be using different models in order to evaluate the features extracted from the biomarkers and comparing the results with the ones obtained in this work.

Bibliography

[1] M. C. Botelho, I. Trancoso, A. Abad, and 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)*, pages 5851–5855, 2019. doi: 10.1109/ICASSP.2019.8682431.

[2] A. W. Fox, P. K. Monoson, and C. D. Morgan. Speech dysfunction of obstructive sleep apnea. a discriminant analysis of its descriptors. *Chest*, 96(3):589–595, Sept. 1989.

[3] J. Solé-Casals, C. Munteanu, O. C. Martín, F. Barbé, C. Queipo, J. Amilibia, and J. Durán-Cantolla. Detection of severe obstructive sleep apnea through voice analysis. *Applied Soft Computing*, 23: 346–354, 2014.

[4] 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.

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

[6] J. S. Durmer and D. F. Dinges. Neurocognitive consequences of sleep deprivation. In *Seminars in neurology*, volume 25, pages 117–129. Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New . . . , 2005.

[7] D. Hillman, S. Mitchell, J. Streatfeild, C. Burns, D. Bruck, and L. Pezzullo. The economic cost of inadequate sleep. *Sleep*, 41(8):zsy083, 2018.

[8] J. E. Ferrie, M. Kumari, P. Salo, A. Singh-Manoux, and M. Kivimäki. Sleep epidemiology—a rapidly growing field. *International Journal of Epidemiology*, 40(6):1431–1437, 12 2011. ISSN 0300-5771. doi: 10.1093/ije/dyr203. URL `https://doi.org/10.1093/ije/dyr203`.

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

[10] M. S. Aloia, J. T. Arnedt, J. D. Davis, R. L. Riggs, and D. Byrd. Neuropsychological sequelae of obstructive sleep apnea-hypopnea syndrome: a critical review. *Journal of the International Neuropsychological Society*, 10(5):772–785, 2004.

[11] P. M. Macey, L. A. Henderson, K. E. Macey, J. R. Alger, R. C. Frysinger, M. A. Woo, R. K. Harper, F. L. Yan-Go, and R. M. Harper. Brain morphology associated with obstructive sleep apnea. *American journal of respiratory and critical care medicine*, 166(10):1382–1387, 2002.

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

[13] D. Gozal, J. M. Daniel, and G. P. Dohanich. Behavioral and anatomical correlates of chronic episodic hypoxia during sleep in the rat. *J. Neurosci.*, 21(7):2442–2450, Apr. 2001.

[14] M.-A. Bédard, J. Montplaisir, F. Richer, I. Rouleau, and J. Malo. Obstructive Sleep Apnea Syndrome: Pathogenesis of Neuropsychological Deficits. *Journal of Clinical and Experimental Neuropsychology*, 13(6):950–964, Nov. 1991. ISSN 0168-8634. doi: 10.1080/01688639108405110. URL https://www.tandfonline.com/doi/full/10.1080/01688639108405110.

[15] G. D. Greenberg, R. K. Watson, and D. Deptula. Neuropsychological Dysfunction in Sleep Apnea. *Sleep*, 10(3):254–262, May 1987. ISSN 0161-8105, 1550-9109. doi: 10.1093/sleep/10.3.254. URL https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/10.3.254.

[16] B. Naëgelé, V. Thouvard, J.-L. Pépin, P. Lévy, C. Bonnet, J. E. Perret, J. Pellat, and C. Feuerstein. Deficits of cognitive executive functions in patients with sleep apnea syndrome. *Sleep*, 18(1):43–52, 1995.

[17] L. J. Findley, J. T. Barth, D. C. Powers, S. C. Wilhoit, D. G. Boyd, and P. M. Suratt. Cognitive Impairment in Patients with Obstructive Sleep Apnea and Associated Hypoxemia. *Chest*, 90(5): 686–690, Nov. 1986. ISSN 00123692. doi: 10.1378/chest.90.5.686. URL https://linkinghub.elsevier.com/retrieve/pii/S0012369215437821.

[18] E. Verstraeten. Neurocognitive effects of obstructive sleep apnea syndrome. *Current Neurology and Neuroscience Reports*, 7(2):161–166, Mar. 2007. ISSN 1528-4042, 1534-6293. doi: 10.1007/s11910-007-0012-8. URL http://link.springer.com/10.1007/s11910-007-0012-8.

[19] E. Verstraeten, R. Cluydts, J. Verbraecken, and J. D. Roeck. Neuropsychological functioning and determinants of morning alertness in patients with obstructive sleep apnea syndrome. *Journal of the International Neuropsychological Society*, 2(4):306–314, July 1996. ISSN 1355-6177, 1469-7661. doi: 10.1017/S1355617700001326. URL https://www.cambridge.org/core/product/identifier/S1355617700001326/type/journal_article.

[20] P. E. Peppard. Longitudinal Association of Sleep-Related Breathing Disorder and Depression. *Archives of Internal Medicine*, 166(16):1709, Sept. 2006. ISSN 0003-9926. doi: 10.1001/archinte.166.16.1709. URL http://archinte.jamanetwork.com/article.aspx?doi=10.1001/archinte.166.16.1709.

[21] G. Pillar and P. Lavie. Psychiatric Symptoms in Sleep Apnea Syndrome. *Chest*, 114(3):697–703, Sept. 1998. ISSN 00123692. doi: 10.1378/chest.114.3.697. URL https://linkinghub.elsevier.com/retrieve/pii/S0012369216325673.

[22] D. F. Kripke, S. Ancoli-Israel, M. R. Klauber, D. L. Wingard, W. J. Mason, and D. J. Mullaney. Prevalence of Sleep-Disordered Breathing in Ages 40–64 Years: A Population-Based Survey. *Sleep*, 20(1):65–76, Jan. 1997. ISSN 1550-9109, 0161-8105. doi: 10.1093/sleep/20.1.65. URL https://academic.oup.com/sleep/article-lookup/doi/10.1093/sleep/20.1.65.

[23] C. Varounis, V. Katsi, I. E. Kallikazaros, D. Tousoulis, C. Stefanadis, J. Parissis, J. Lekakis, C. Siristatidis, A. J. Manolis, and T. Makris. Effect of CPAP on blood pressure in patients with obstructive sleep apnea and resistant hypertension: A systematic review and meta-analysis. *International Journal of Cardiology*, 175(1):195–198, July 2014. ISSN 01675273. doi: 10.1016/j.ijcard.2014.04.240. URL https://linkinghub.elsevier.com/retrieve/pii/S016752731400936X.

[24] J. W. Winkelman. Insomnia Disorder. *New England Journal of Medicine*, 373(15):1437–1444, Oct. 2015. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMcp1412740. URL http://www.nejm.org/doi/10.1056/NEJMcp1412740.

[25] R. M. Simply, E. Dafna, and Y. Zigel. Diagnosis of obstructive sleep apnea using speech signals from awake subjects. *IEEE Journal of Selected Topics in Signal Processing*, 14(2):251–260, 2020. doi: 10.1109/JSTSP.2019.2955019.

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

[27] H. Sak, A. W. Senior, and F. Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. 2014.

[28] A. Graves, A. Mohamed, and G. Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, 2013. doi: 10.1109/ICASSP.2013.6638947.

[29] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur. X-vectors: Robust dnn embeddings for speaker recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5329–5333, 2018. doi: 10.1109/ICASSP.2018.8461375.

[30] D. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur. Deep neural network embeddings for text-independent speaker verification. In *Interspeech 2017*. ISCA, Aug. 2017. doi: 10.21437/interspeech.2017-620. URL https://doi.org/10.21437/interspeech.2017-620.

[31] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouelslet. Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4):788–798, 2011. doi: 10.1109/TASL.2010.2064307.

[32] S. Ioffe. Probabilistic linear discriminant analysis. In A. Leonardis, H. Bischof, and A. Pinz, editors, *Computer Vision – ECCV 2006*, pages 531–542, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-33839-0.

[33] P. Kenny. Bayesian Speaker Verification with Heavy-Tailed Priors. In *Proc. The Speaker and Language Recognition Workshop (Odyssey 2010)*, page paper 14, 2010.

[34] N. Brummer and E. Villiers. The speaker partitioning problem. *Proc. of Odyssey 2010*, 01 2010.

[35] D. Garcia-Romero and C. Espy-Wilson. Analysis of i-vector length normalization in speaker recognition systems. pages 249–252, 01 2011.

[36] 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*, number CONF. IEEE Signal Processing Society, 2011.

[37] D. Snyder, D. Garcia-Romero, and D. Povey. Time delay deep neural network-based universal background models for speaker recognition. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pages 92–97, 2015. doi: 10.1109/ASRU.2015.7404779.

[38] M. McLaren, Y. Lei, and L. Ferrer. Advances in deep neural network approaches to speaker recognition. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4814–4818, 2015. doi: 10.1109/ICASSP.2015.7178885.

[39] D. E. Sturim and D. A. Reynolds. Speaker adaptive cohort selection for tnorm in text-independent speaker verification. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 1, pages I/741–I/744 Vol. 1, 2005. doi: 10.1109/ICASSP.2005.1415220.

[40] M. Herzog, T. Metz, A. Schmidt, T. Bremert, B. Venohr, W. Hosemann, and 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):1456–1462, 11 2006. ISSN 0161-8105. doi: 10.1093/sleep/29.11.1456. URL `https://doi.org/10.1093/sleep/29.11.1456`.

[41] W. Thorpe, M. Kurver, G. King, and C. Salome. Acoustic analysis of cough. In *The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001*, pages 391–394, 2001. doi: 10.1109/ANZIIS.2001.974110.

[42] J. Correia, B. Raj, I. Trancoso, and F. Teixeira. Mining multimodal repositories for speech affecting diseases. In *Proc. Interspeech 2018*, pages 2963–2967, 2018. doi: 10.21437/Interspeech.2018-1806. URL `http://dx.doi.org/10.21437/Interspeech.2018-1806`.

[43] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong. The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2):190–202, 2016. doi: 10.1109/TAFFC.2015.2457417.

[44] Audeering. Opensmile. https://www.audeering.com/opensmile/, July 2021.

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

[46] JotForm. Jotform surveys. https://www.jotform.com/, 2021.

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