UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN

STATISTICAL ANALYSIS OF THE CONNECTION
BETWEEN SLEEP APNEA AND SPEECH

TESIS DOCTORAL

ANA MONTERO BENAVIDES
LICENCIADA EN CIENCIAS Y TÉCNICAS ESTADÍSTICAS

2017

**DEPARTAMENTO DE SEÑALES, SISTEMAS Y
RADIOCOMUNICACIONES**

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

**STATISTICAL ANALYSIS OF THE CONNECTION
BETWEEN SLEEP APNEA AND SPEECH**

Autora: Ana Montero Benavides
Licenciada en Ciencias y Técnicas Estadísticas

Directores: Luis Alfonso Hernández Gómez
Doctor Ingeniero de Telecomunicación
José Luis Blanco Murillo
Doctor Ingeniero de Telecomunicación

Madrid, 2017

Department: Señales, Sistemas y Radiocomunicaciones
Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad Politécnica de Madrid (UPM)

PhD Thesis: **Statistical analysis of the connection
between sleep apnea and speech**

Author: **Ana Montero Benavides**
Licenciada en Ciencias y Técnicas Estadísticas

Advisors: **Dr. Luis A. Hernández Gómez**
Doctor Ingeniero de Telecomunicación
Dr. José Luis Blanco Murillo
Doctor Ingeniero de Telecomunicación

Year: 2017

Board named by the Rector of Universidad Politécnica de Madrid, on the 21st of June 2017.

Board: **Dr. Eduardo López Gonzalo**
Universidad Politécnica de Madrid (UPM)
Dr. Rubén San Segundo Hernández
Universidad Politécnica de Madrid (UPM)
Dra. Belén Ruiz Mezcu
Universidad Carlos III de Madrid (UC3M)
Dra. Seba Almedawar
Technical University of Dresden (TUD), Alemania
Dr. Doroteo Torre Toledano
Universidad Autónoma de Madrid (UAM)
Dra. Ascensión Gallardo Antolín
Universidad Carlos III de Madrid (UC3M)
Dr. Daniel Ramos Castro
Universidad Autónoma de Madrid (UAM)

After the defense of the PhD Thesis on the 26th of September 2017, at the E.T.S.I. de Telecomunicación, the board agrees to grant the following qualification:

CHAIR

SECRETARY

MEMBERS

This work was developed at *Grupo de Aplicaciones de Procesado de Señales (GAPS)* between 2011 and 2017, and funded by a *Formacion Personal Investigador (FPI)* Fellowship from the Spanish *Ministerio de Economía, Industria y Competitividad (MINECO)*.

RESUMEN

Esta tesis se ocupa de la Apnea Obstructiva del Sueño (AOS), que consiste en el cese involuntario de la respiración durante unos segundos al nivel de la faringe, que se repite a lo largo de la noche impidiendo un correcto descanso. Se trata de uno de los desórdenes del sueño más destacados dada su alta prevalencia y sus dramáticas consecuencias pero a pesar de ello su importancia no se ha reconocido hasta épocas recientes. Aunque hace ya dos siglos que se vienen observando anomalías en la respiración de algunos pacientes, tanto durante el sueño como en estado de vigilia, hasta las últimas décadas no se ha profundizado en el estudio de la AOS. Sin tratamiento, los pacientes que sufren este trastorno se enfrentan diariamente al cansancio permanente, que dificulta su vida cotidiana y aumenta su propensión a sufrir accidentes, involucrando así al resto de la población. Además, se han detectado otros efectos secundarios como hipertensión, diabetes y riesgo de infarto.

A pesar de que últimamente ya se ha reconocido la importancia a nivel mundial de este síndrome, su diagnóstico todavía es tedioso e intensivo en recursos. Comienza en la consulta del médico de familia, quien ante la sospecha de que su paciente puede tener apnea, lo deriva a la unidad del sueño, para que se le realice una polisomnografía. Este examen consiste en pasar la noche allí, mientras se graban múltiples variables que posteriormente se tendrán que analizar. Previamente se inscribe al paciente en una lista de espera y, en países como España, debe aguardar al menos un año para realizar la prueba, lo cual pone de manifiesto la necesidad de métodos alternativos de diagnóstico y cribado. El objetivo de esta tesis es ayudar a los médicos en la tarea de cribado de los casos más graves de AOS, para poder dar prioridad a estos enfermos reorganizando las listas de espera. De esta forma, los pacientes más afectados se beneficiarían de un diagnóstico más rápido y comenzarían antes con el tratamiento.

Dado que tanto la voz, como la AOS dependen de las vías aéreas, se ha supuesto y comprobado que existen rasgos en el habla que están relacionados con este síndrome. Investigaciones biomédicas previas han demostrado la existencia de patrones morfológicos vinculados con esta enfermedad que afectan al habla, junto con rasgos característicos en la dicción de estos pacientes. En el presente trabajo se profundiza en la relación entre estas variables teniendo en cuenta los estudios previos, que hemos conocido a través de la investigación, ya que en ellos se abordaban tres aspectos del habla de pacientes de AOS: resonancia, articulación y fonación. Nuestros experimentos están condicionados por la disponibilidad de datos. Trabajamos tanto con habla continua, como con vocales sostenidas y realizamos un cribado para clasificar a los pacientes según su estado, estableciendo así dos grupos: sanos y casos graves. Por otra parte, analizamos todo el espectro de la enfermedad, estratificando a la población en función del grado de severidad de la misma. En cada uno de estos estratos, analizamos las relaciones entre el habla y las variables clínicas asociadas a la AOS, para entender mejor las vinculaciones existentes entre ellas.

En primer lugar, añadimos información contextual al análisis estadístico; a continuación seleccionamos rasgos discriminativos para el cribado de AOS y analizamos las variables clínicas y el habla en un escenario clínico real. Aprovechando la existencia de una base de datos de nuestro grupo de investigación de tamaño moderado formada por 80 sujetos, incluyendo sanos y casos graves de apnea, con la peculiaridad de que se habían intentado controlar características físicas de estos pacientes (edad, peso, índice de masa corporal), nos centramos en analizar las diferencias de la señal de voz utilizando por primera vez en la detección de la AOS la información contextual. Para ello, construimos dos tipos de modelos ocultos de Markov, uno dependiente de frase y el otro dependiente del contexto de los fonemas. Con esta metodología logramos mejorar las tasas de clasificación con respecto al sistema de referencia independiente del texto. Con una perspectiva diferente, en lugar de basarnos únicamente en medidas de voz de la componente espectral, nos centramos en el análisis de medidas acústicas de distinta naturaleza: diferencias de formantes, propiedades fonéticas y prosódicas y una medida de nasalidad. A partir de un total de 16 medidas de voz, seleccionamos las que mejor diferencian entre las dos poblaciones, combinándolas en distintos tipos de discriminadores, resultando ser el mejor el de 8 medidas. Nuestro modelo basado en medidas del habla presenta un mejor rendimiento que el de referencia, basado en la edad y el índice de masa corporal, que es la técnica habitual de cribado usada en las clínicas.

Finalmente, tuvimos a nuestra disposición una base de datos nueva, diseñada y grabada por el mismo equipo investigador que la anterior. Esta nueva base de datos cumplía con dos propiedades importantes: (i) cubría el rango entero desde personas sin AOS hasta casos graves de AOS, incluyendo casos leves, y (ii) había sido grabada en un escenario clínico real de pacientes sospechosos de tener algún trastorno del sueño. El primer punto nos permite determinar correlaciones del índice de apnea/hipopnea, es decir, de la gravedad de la apnea, con otros parámetros clínicos (peso, altura, perímetro cervical, edad) y con características del habla (las frecuencias y anchos de banda de los primeros tres formantes). Debido al segundo punto, nuestro estudio es realista para el cribado que queremos diseñar. A la vez, el problema de clasificación se vuelve más difícil, porque la mayoría de los no-apnéicos tiene algún otro trastorno, ya que fueron referenciados para someterse a la polisomnografía en la unidad del sueño. Determinamos las frecuencias y los anchos de banda de los formantes de las vocales e investigamos las correlaciones con variables clínicas y con la AOS, tanto como las diferencias entre los grupos de los no-apnéicos, los casos leves y los casos graves de AOS. Al contrario de lo esperado, solo encontramos correlaciones escasas de la apnea con las frecuencias y anchos de bandas de los formantes.

En conclusión, demostramos que en el escenario controlado se puede mejorar la detección de AOS usando la información contextual (modelos ocultos de Markov) y combinando varias medidas acústicas seleccionadas. Sin embargo, en un escenario clínico real de una unidad de sueño, donde hay que distinguirla de otras enfermedades del sueño, el problema se vuelve más complejo. Se demuestra que algunas hipótesis aceptadas de otros autores (resonancias de formantes más bajas y más anchas), no se confirman. El problema de detectar la apnea todavía no se ha resuelto en un escenario clínico real. Sin embargo, en esta tesis se han puesto de manifiesto las interrelaciones entre dicha enfermedad, el habla y variables clínicas, lo que ayudará a guiar futuras investigaciones que traten la conexión entre el habla y la AOS.

ABSTRACT

This thesis deals with sleep disorders, namely with obstructive sleep apnea (OSA), one of the most important ones. Although breathing abnormalities that occur either during wakefulness or sleep have been reported since the 1800s, the high prevalence of disordered breathing that occurs only during sleep was not recognized until recently. In the last decades, deeper research has been conducted on sleep disorders, and in particular on OSA. This disease consists in episodes of involuntary cessation of breath during sleep by the collapse of the pharynx that may last a few seconds and reproduce throughout the night, preventing proper rest.

Left untreated, OSA leads to an increased risk of accidents and can lead to serious health risks, such as hypertension, diabetes, and stroke. Although recently OSA has been acknowledged as a worldwide problem, its diagnosis is still tedious and resource intensive. Currently, the diagnosis of OSA starts at the family doctor, who, after looking for cues of OSA, sends the patients to a sleep unit for a polysomnography test, which involves the recording of several variables overnight, as well as the posterior analysis of the results. In Spain, for instance, waiting lists for this test are longer than a year. This implies a strong need for alternative diagnosis and screening methods. With regard to the great improvement of life quality of these patients when they are properly diagnosed and treated, the scientific community should look for a solution to improve on the diagnosis of OSA.

Throughout this Thesis, we aim to help the clinicians on the screening of the most severe cases in order to prioritize those patients and reorganize the waiting lists. This way, the most severe OSA patients could benefit from a swift diagnosis and treatment.

Since both OSA and speech are related to the functioning of the upper airway, it has been expected and shown that there are traits in the speech that relate to the OSA disease. Previous biomedical research has shown the existence of morphological patterns related to OSA that have a direct influence on speech. Moreover, characteristic traits have been found in the speech of OSA patients. In this thesis, we deepen the understanding of the relationships between speech and OSA, taking into account previous works, which have shown that there are differences in resonance, articulation, and phonation.

Our experiments are conditioned on the available data. We work with continuous speech as well as with sustained vowels. We realize screening to detect those patients with severe OSA condition, i.e., discriminate between the groups *healthy* and *severe OSA*. On the other hand, we analyze the full spectrum, stratifying the population as a function of the severity of their condition, on data from a real-practice scenario recorded at a sleep unit. In each stratum, we analyze the relations of speech and clinical variables.

We advance in this research line and study not only the speech but also some clinical variables related to OSA and speech to get a deeper understanding of their relationship. We add text-dependent information to the statistical analysis; we select discriminative speech features to do screening; we analyze clinical variables and speech from a real clinical practice scenario.

Using an existing database with the speech samples from 80 subjects, split into *healthy* and *severe* OSA groups with balanced physical properties (age, weight, body-mass index), we applied, for the first time in apnea detection, models that make use of text-dependent information, namely hidden Markov models, which take into account the order in which the phonetic units are pronounced. We employed two different model architectures, one on the whole-sentence level and the other one on the level of phonemes. We achieved improvements in the classification rate with respect to the text-independent baseline system.

With a different perspective on the same data, instead of only considering the spectral envelope, we consider voice features of different nature that are expected to contain information on the speaker's OSA condition. The set of features included differences of formant frequencies, phonetic and prosodic properties, and a nasality measure. From the total number of 16 features, we select the most discriminative ones and construct combined discriminators, the best of which made use of 8 features. Our speech-based model performs better than a baseline model based on age and body-mass index, which is the screening method commonly used by clinicians.

Finally, a new, bigger database has become available, with two important properties: (i) it covers the whole range from non-apneic subjects over mild OSA cases to severe OSA cases, and (ii) was recorded in real practice from patients suspected to suffer from OSA at a sleep unit. The former property allows us to determine correlations of the apnea-hypopnea index, i.e., the severity of the OSA condition, with other clinical variables (weight, height, cervical perimeter, and age) and with voice features (namely the first three formants and their associated bandwidths). The latter property brings us very close to the real-practice scenario but makes the classification problem more difficult because most subjects without OSA have some other conditions whose symptoms brought them to the sleep unit in the first place. We determine the frequencies and bandwidths of the sustained-vowel formant resonances and investigate correlations with clinical variables and OSA, as well as differences between the groups non-OSA, mild OSA, and severe OSA. Contrary to the expected, in this real-practice scenario, we find only poor correlations between formant frequencies and bandwidths and the severity of OSA.

In conclusion, we show that the OSA detection can be improved by means of text-dependent information (hidden Markov models) and selected voice features, improving the classification rate in the controlled scenario. However, things become more complex in the real-practice scenario of a sleep unit, where OSA has to be discriminated from other diseases. We show that accepted conjectures from other authors (lower and wider formant resonances) are not confirmed.

The problem of robust OSA detection is not yet solved for the real-practice scenario, but this Thesis sheds light on the interdependences of OSA, speech and other clinical variables and will help to guide the future research on the connection between speech and OSA.

A mi familia y amigos

*"Lo vulgar es el ronquido, lo inverosímil, el sueño. La humanidad ronca,
pero el artista está en la obligación de hacerla soñar
o no es artista."*

Enrique Jardiel Poncela
(1901 – 1952)

AGRADECIMIENTOS

Me gustaría agradecer a todos los que de algún modo han colaborado a que esta Tesis se realizara, tanto a las personas que me han apoyado a nivel personal como a las del ámbito académico.

En primer lugar quiero agradecer especialmente a mi Director de Tesis, al Dr. Luis A. Hernández Gómez por todo lo que me ha enseñado y he podido aprender de él. Por haberme dedicado su tiempo y como ya dijo mi padre una vez, por haber sacado lo mejor de mí como investigadora. Por haber confiado en mí y haber apoyado mis ideas y cada uno de mis pasos a lo largo de estos años.

También quiero agradecer a mi compañero de investigación en esto de la apnea, al Dr. José Luis Blanco Murillo, quien después de un tiempo pasó a codirigir esta tesis. Porque siempre ha tenido tiempo para ayudarme, explicarme cosas y aconsejar tanto como compañero como director de Tesis.

Tanto a Luis como a José Luis quiero agradecerles el haberme acompañado, guiado y supervisado a lo largo de estos años en esta aventura de la investigación. Ha sido un honor poder contar con ellos durante esta etapa y poder aprender tanto de ellos, tanto por sus enseñanzas científicas como por sus lecciones de humildad. Por haberme dejado volar, al facilitarme el irme de estancia y seguir trabajando en la distancia para poder finalizar esta Tesis. Porque los dos son un ejemplo a seguir. Sin su dedicación esta Tesis no tendría sentido.

Me gustaría agradecer también a los revisores, por haber dedicado su tiempo y haberme aconsejado en la redacción de esta Tesis: Dra. Seba Almedawar, Dr. Rubén San Segundo y Dr. Eduardo Moreira.

A mis compañeros del G.A.PS.: A Rubén, Bea, David y los Álvaros por haberme presentado a Christopher. A los del nuevo despacho: J.L., Virginia y Nelson por compartir inquietudes como el Maus y hacer tan fácil el día a día en el despacho. A mis chinitas Qifei y Deng, por esos cafés y enseñanzas sobre su cultura.

Quiero aprovechar estas líneas para agradecer a toda mi familia, en especial a mis padres. Porque se han esforzado cada día de su vida en darme una esmerada educación y me han enseñado también grandes valores. Por todo su tiempo, dedicación, cariño y el apoyo económico para mandarme a estudiar a Madrid, a París... Todas esas clases del Conservatorio, Escuela de Idiomas, actividades extraescolares y sobre todo por esas vacaciones en Galicia.

A mi hermano de León, Jose, compañero en el exilio alemán quien siempre me enseña a ver el lado positivo de las cosas.

A mis tías las Doctoras Montero, que de alguna forma inspiraron el que siguiera sus pasos, yéndome de Erasmus, echándome novio guiri, haciendo un doctorado... En especial a mi madrina Josefa Montero que me ha apoyado siempre en cada una de mis ideas por muy descabelladas que fueran. También al tío Víctor, por su ayuda tecnológica.

También me gustaría dedicar unas líneas a mis abuelos que tanto me enseñaron de pequeña y que estarían muy orgullosos. A mi abuelita Pilar, al abuelito Julio y al abuelito Pedro. Y a mi abuelita Argentina que afortunadamente sigue cada uno de nuestros pasos muy de cerca.

Quiero agradecer también a todos los que me han acogido en su casa durante mis estancias en Madrid. En especial a mis tíos Jorge y Lola, por su apoyo logístico incondicional, que ya durante la carrera me adoptaron como a una más cada verano y me han aconsejado tantas veces. A mis primos Jorge, Pablo y Sofía, por compartir su casa y por su compañía. También quiero agradecerle a la tía Veva, a M.A. y a Lupe por acogerme en sus casas durante mis visitas a Madrid.

A mis compañeros de piso de La Latina: Raúl, Maria Chiara y Ximena. Por tantos buenos ratos juntos, incluso cuando tocaba limpiar. Por haber mantenido el contacto por lejos que estuviéramos.

A Pedro, mi compañero chileno de Máster. Porque sobrevivió al TECI e hizo más divertidas las noches de amigos con su guitarra y por cruzarse el charco para venir a vernos a Alemania.

A mis amigas de la universidad Gema y María, por todos estos años juntas: esas tardes en la biblioteca, las fiestas de la resi, haber venido a verme a París durante la Erasmus, las tardes de La Latina...

A mis amigas de León, en especial a Cristina y Marta por estar ahí siempre. A Alberto por tantas aventuras vividas.

A mi amiga Sofía, por compartir tantas cosas en Dresden.

Por último quiero agradecer de manera especial al Doctor Christopher Gaul, ese chico de ojos azules que apareció en mi vida en mi primer mes de doctoranda y que tantas alegrías me ha dado a lo largo de todos estos años. Sin su ayuda y su perseverancia esta Tesis tampoco habría tenido sentido. Por su paciencia, comprensión, dedicación y ayuda inestimable: Vielen Dank!

Y también quiero agradecerle a mi hijo Martín, que llegó a nuestras vidas este último año para revolucionarnos a todos y hacernos más felices.

Una vez más a mis padres y también a mis suegros Andrea y Stephan, por haber cuidado de Martín para que yo pudiera dedicar tiempo a terminar esta Tesis.

A todos los que habéis formado parte de esta aventura

¡Muchas gracias!

TABLE OF CONTENTS

Resumen	I
Abstract	III
Agradecimientos	VII
1. Introduction	1
1.1. Motivation	3
1.2. Goals	3
1.3. Structure of This Thesis	6
2. Obstructive Sleep Apnea and Speech, State of the Art	7
2.1. Introduction	7
2.2. History of OSA	8
2.2.1. OSA in Charles Dickens' Novels	9
2.2.2. Phenotypes of OSA Patients	10
2.3. OSA Symptoms	11
2.3.1. Definition of the Apnea-Hypopnea Index (AHI)	11
2.3.2. Snoring, the Most Typical Symptom	11
2.3.3. Severe Consequences of OSA	12
2.4. OSA Risk Factors and Prevalence	14
2.4.1. OSA Prevalence	14
2.4.2. OSA Risk Factors	15
2.5. Clinical Diagnosis	17
2.5.1. Screening Techniques for OSA	19
2.5.2. Treatment	20
2.6. Evolution of Speech and OSA	21
2.6.1. How Human Speech works	22
2.7. OSA Assessment via Speech Technologies	26
2.7.1. Perceptive Studies (Fox, Monoson et al.)	26
2.7.2. First Quantitative Studies via Formants (Fiz, Robb et al.)	30
2.7.3. OSA and Snoring	32

2.7.4. Further Developments	33
2.8. Summary	33
3. The Apnea Databases	35
3.1. Introduction	35
3.2. Apnea Database 1.0	36
3.2.1. Speech Corpus	36
3.2.2. Speech Data Acquisition	38
3.2.3. Collection Procedure for Apnea Database 1.0	39
3.2.4. Clinical Parameters of Apnea Database 1.0	40
3.3. Apnea Database 2.0	42
3.3.1. Resulting Database Size and Properties	44
3.3.2. Validation of Apnea Database 2.0	44
3.4. Summary	47
4. Hidden Markov Models, a Text-Dependent Approach	49
4.1. From Speech Signal to a Sequence of Feature Vectors (MFCCs)	50
4.1.1. Motivation	50
4.1.2. MFCC Features	50
4.2. On HMMs and the Algorithms Used	54
4.3. Designing HMM Experiments for Our Classification Problem	56
4.3.1. Whole-Sentence HMMs	57
4.3.2. Sentence-Dependent Phonemes	64
4.4. Results	66
4.5. Summary and Conclusions	68
5. Voice Features related to OSA, an Analysis on a Controlled OSA Database	69
5.1. Introduction	69
5.2. Selection and Definition of Voice Features Related to OSA	70
5.2.1. Formant Difference $(F_3 - F_2)_{/i/}$	70
5.2.2. Phonation Features (HNR, Jitter, Shimmer, SDRseg)	72
5.2.3. Nasality Measurements	75
5.2.4. Prosodic Features: Sentence Duration and Silence Features	76
5.2.5. Voice and Data Analysis Tools	76
5.3. Analysis and Results	76
5.3.1. Analysis of Individual Features	77
5.3.2. Analysis of Feature Combination	79
5.3.3. Classification and Diagnosis Performance	84
5.3.4. Analysis of Diagnosis Performance on a Separate Test Set	87
5.3.5. Comparison to Screening Using Only BMI and Age	88

5.4. Conclusions and Discussion	89
5.4.1. Summary	90
6. Formants and Clinical Variables in a Clinical Practice Scenario	93
6.1. Introduction and Motivation	94
6.1.1. Formants in Human Speech	94
6.1.2. Previous Studies About Formant Resonances and Apnea	95
6.1.3. Motivation for the Present Approach	96
6.2. Materials and Methods	97
6.2.1. Database	97
6.2.2. Acoustical Analysis	97
6.2.3. Statistical Analysis	98
6.3. Results	98
6.3.1. Formant Frequencies and Bandwidths	98
6.3.2. Cross Correlations	99
6.4. Discussion	103
6.4.1. Discussion of Results	104
6.5. Follow-up Discussion with Doctor Kemaloğlu and Doctor Mengü	108
6.5.1. Letter to the Editor	108
6.5.2. Reply: Additional Results	109
6.5.3. Discussion	112
6.6. Summary and Conclusions of the Chapter	113
6.6.1. Conclusions	113
7. Conclusions and Future work	115
7.1. Motivation of this Thesis Summarized	115
7.2. Summary of Contributions and Results of this Thesis	116
7.2.1. Results of Contribution 1: Hidden Markov Models	117
7.2.2. Results of Contribution 2: Voice Features	118
7.2.3. Results of Contribution 3: Formants and Bandwidths	119
7.2.4. Limitations	120
7.3. Conclusion	121
7.4. Future Work	121
A. Scientific Contributions, Collaborations, and Publications	125
A.1. Peer-Reviewed Scientific Articles	125
A.2. Awards	126
A.2.1. IberSPEECH 2012	126
A.2.2. ApneApp: Popularizing automatic severe OSA detection via speech analysis	126
A.3. Other Contributions	128
A.3.1. Research Stay at IMB Dresden	128

Table of Contents

A.3.2. Diagnosis of Multiple Sclerosis	128
B. Definitions and Abbreviations	129
Bibliography	131
Statement of Originality	145

LIST OF FIGURES

Introduction	1
1.1. Types of apnea with their relative prevalence and role in this Thesis.	1
1.2. Interdependencies studied in this Thesis.	2
1.3. Groups to be classified in the diagnosis OSA.	4
1.4. Groups of Apnea Database 1.0 (chapter 3), employed in chapters 4 and 5.	5
1.5. Groups of Apnea Database 2.0 (chapter 3), employed in chapter 6.	5
Obstructive Sleep Apnea and Speech, State of the Art	7
2.1. Apnea-Hypopnea episode development during sleep	8
2.2. Illustration of “Fat Joe” from Charles Dickens’ novel “The Posthumous Papers of the Pickwick Club”, the first description of the morphotype of an apnea patient	9
2.3. Illustration of normal breathing, snoring, and complete obstruction of the airway in an OSA patient	12
2.4. OSA primary sequence of events, physiological responses, and clinical features . . .	13
2.5. Illustration of polysomnography test data recording	19
2.6. Distribution of the nodes and antinodes A_x of the volume current in a uniform pipe and in the simplified model of a uniform vocal tract	23
2.7. Vowel chart for Spanish	24
2.8. Diagram of SVT_H and SVT_V	25
The Apnea Databases	35
3.1. Recording scheme for Apnea Database 1.0	39
3.2. Apnea Database 1.0.	40
3.3. Characteristics of Apnea Database 2.0 (as opposed to Apnea Database 1.0, Figure 3.2).	42
3.4. Recording scheme for Apnea Database 2.0	43
Hidden Markov Models, a Text-Dependent Approach	49
4.1. Block diagram: from speech signal to a sequence of feature vectors	51
4.2. Mel filter bank	52
4.3. MFCC feature vector layout	53
4.4. Representation of a (left-right) HMM	55

4.5. HMM operation scheme.	56
4.6. Leave-one-out scheme	57
4.7. Scheme for whole-sentence HMMs	58
4.8. Classification error for different HMM topologies	60
4.9. Error rates of different topologies with 1, 12, 40, and 100 states, with total number of Gaussians close to 500	61
4.10. The best topology of all tested whole-sentence models with 40 hidden states	63
4.11. Scheme of the approach of sentence-dependent phonemes	65
4.12. Topology of the single-phoneme HMMs, consisting of 3 states with 6 Gaussians each. .	66
4.13. DET curves corresponding to three severe-OSA detection systems GMM-based base- line system, sentence-dependent phoneme HMMs and whole-sentence HMMs	67
Voice Features related to OSA, an Analysis on a Controlled OSA Database	69
5.1. Differences between third and second formant for the vowel /i/ in the word “Suiza” for an OSA speaker and a control-group speaker	72
5.2. Examples for the calculation of jitter and shimmer	75
5.3. Correlation matrix for the 16 voice features	81
Formants and Clinical Variables in a Clinical Practice Scenario	93
6.1. F_1 vs. F_2 for the Andalusian group compared to the results of Rosique et al.	100
6.2. Map representation of clinical variables and their influence on formants frequencies and bandwidths	103
Scientific Contributions, Collaborations, and Publications	125
A.1. The ApneApp project.	126

LIST OF TABLES

Obstructive Sleep Apnea and Speech, State of the Art	7
2.1. Prevalence of OSA in different ethnic groups	15
2.2. Factors that contribute to the pathophysiology of OSA	16
2.3. Group means and standard deviations of judged speech disorder for three speech descriptors and treatment groups, from Fox & Monoson (1989)	27
The Apnea Databases	35
3.1. Sentences of Apnea Database 1.0 and Apnea Database 2.0, including phonetic transcription and English translation	36
3.2. Summary on the subjects included in the Apnea Database 1.0	41
3.3. Pairwise correlations for all variables included in Apnea Database 1.0 speakers' profiles	41
3.4. Comparison Apnea Database 1.0 vs. Apnea Database 2.0.	45
3.5. Descriptive Statistics on the 241 Andalusian Male Subjects.	45
3.6. Comparison of two apnea databases	46
Hidden Markov Models, a Text-Dependent Approach	49
4.1. Error rates per sentence on the different groups depending on the number of states.	64
4.2. EER values obtained for each classification scheme	66
Voice Features related to OSA, an Analysis on a Controlled OSA Database	69
5.1. OSA-related features under analysis.	71
5.2. Median, standard deviation and p -values for the proposed speech features computed on the training dataset	78
5.3. EER obtained with each one of the individual features on the training database. Features are sorted in increasing order of EER.	80
5.4. Distribution of age and BMI in healthy and apnea speakers in the databases.	82
5.5. Incremental feature combination using a MLR model and R_{adj}^2 selection criterion	84
5.6. Incremental feature combination using a LDA model and CER selection criterion	85
5.7. CER on the training dataset for LDA and MLR models for different numbers of selected features in the incremental feature analysis	85

5.8. Performance measures of the voice-based LDA model (8 best selected features) on the training database	87
5.9. Performance measures of our voice-based LDA model (8 best selected features) on the test database	88
5.10. Performance measures of our voice-based method vs. a model using only age and BMI	88
5.11. Correlation coefficient and p -values between speech features and Age and BMI . . .	89
Formants and Clinical Variables in a Clinical Practice Scenario	93
6.1. Statistical analysis of the first three formant frequencies and bandwidths of vowels	99
6.2. Spearman correlations between clinical variables and AHI.	100
6.3. Spearman correlations between clinical variables and formant frequencies	101
6.4. Spearman correlations of clinical variables and bandwidths	101
6.5. Contrast among OSA groups	107
6.6. Contrast among control and severe OSA groups as in Table 6.5, but on a subset without differences neither in age nor in height.	108
6.7. Contrast among control and severe OSA groups on a subset without differences neither in age nor in height	110
6.8. Spearman correlations of formants and clinical variables	111
6.9. Spearman correlations of formant differences and clinical variables	111
6.10. Sign of the change of the frequencies of formants F_1 to F_4 when the cross-sectional area of the tube is slightly decreased with regard to the neutral tract shape	112
6.11. Summary of experiment on OSA and formants	114

CHAPTER 1

INTRODUCTION

Sleep disorders are common health problems with serious consequences and lately have recently become a focus of study and growing interest (Puertas, Pin, María & Durán, 2005). Sleep-disordered breathing ranges from harmless snoring to repeated complete cessations of breathing during sleep. Progressively, not only clinicians but also the general public has become more conscious of their importance in daily life. There is an extremely high prevalence of sleep-disordered breathing, estimated to affect up to 25% of the world adult population (Peppard et al., 2013).

The sleep apnea syndrome is a sleep disorder in which the person repeatedly stops breathing during sleep. Although it is actually the most common sleep disorder, it is estimated that 90% of the patients are still undiagnosed (T. Young, Evans, Finn, Palta et al., 1997). The episodes of pauses in breathing are called apneas, which in Greek literally means “without breath”. An apnea is a period during which breathing either stops or is significantly reduced. Apnea can be classified into three types, namely central sleep apnea, obstructive sleep apnea and mixed sleep apnea (Figure 1.1). The most common type is the Obstructive Sleep Apnea (OSA) which represents 85% of the sleep apnea cases (Morgenthaler, Kagramanov, Hanak & Decker, 2006). While central sleep apnea arises from the brain’s respiratory control center, OSA is caused by weakness of the throat muscles; the soft tissue in the back of the throat collapses and closes, resulting in blocked airways. In this Thesis, we will focus on the study of OSA.

OSA is a highly prevalent disease, estimates ranging from 1.2% to 27% of the adult population (T. Young et al., 1993; Puertas et al., 2005; Jamie, Sharma & Bing, 2010). OSA is associated

Figure 1.1.: Types of apnea with their relative prevalence and role in this Thesis.

Figure 1.2.: Interdependencies studied in this Thesis.

with cardiovascular and neurocognitive impairments and has serious consequences if not treated (Coito & Sanches, 2011). It is difficult and very resource-intensive to diagnose because symptoms can remain unnoticed for years. The definitive diagnosis requires a full overnight study of patients held at a sleep unit, for the collection and posterior processing and evaluation of numerous signals collected during a minimum of 6 hours of sleep. The gold standard for diagnosis, the polysomnography (PSG), requires the patient to be monitored by specialized equipment during a whole night. It provides ca. 95% accuracy, but cannot be scaled to cope with the current demand. Other diagnosis techniques have also reached excellent performance rates in OSA detection in recent years. As an example, imaging diagnosis has proved that observable anatomical factors are strong predictors of this syndrome and can be used in its detection. Nevertheless, up to now all of these involve complex, time- and resources-consuming procedures, and accumulate waiting lists of over one year (Puertas et al., 2005). Stratification schemes have been proposed to reduce and reorganize waiting lists, providing clinicians with additional information for a better preliminary diagnosis; though no procedure has been found which matches clinicians' expectations. Therefore, faster and less costly screening and stratification techniques are needed for setting priorities to proceed to the PSG diagnosis.

Previous studies (Ayappa & Rapoport, 2003; Lan, Itoi, Takashima, Oda & Tomoda, 2006) have confirmed that patients with OSA commonly have narrower and more collapsible upper airways than patients without OSA, suggesting that OSA could be associated with anatomical and functional abnormalities of the upper airway. Evidences on the connection between OSA and speech have already been pointed out in the literature and analyzed in contributions published in scientific journals and presented at international conferences in medicine and engineering around the world. Unfortunately, only recently research has been carried out on the acoustic properties of speakers suffering from OSA. Nonetheless, abnormalities in phonation, articulation, and resonance have been found, although differences are somewhat unclear (Fox, Monoson & Morgan, 1989). Previous efforts devoted to the characterization of OSA patients' acoustic space to trace specific patterns connected to the apnea syndrome include (Fox et al., 1989; Robb, Yates & Morgan, 1997; J. L. Blanco-Murillo, Fernández-Pozo, Díaz-Pardo, Hernández et al., 2009; J. L. Blanco-Murillo, Fernández-Pozo, Díaz-Pardo, Alvaro Sigüenza et al., 2009; Fernández-Pozo et al., 2009a, 2009b). These contributions have pointed out to a relationship between speech and apnea as a first contribution. The goal of this Thesis is to work out the relationship between apnea, clinical variables, and speech from a statistical point of view (Figure 1.2).

1.1. Motivation

The work described in this Thesis is part of an ongoing project, developed at a research group that has been working on this subject for some years now. The link between apnea and speech has been one of the most outstanding research lines at Signal Processing Application Group (GAPS) from Universidad Politécnica de Madrid (UPM) since 2008. Pieces of evidence on the connection between OSA and speech based on the analysis of the spectral contour and the vocal tract have been found, which apparently match the pathogenesis of OSA. The work presented in this Thesis seeks to contribute to progress in previous research on the voice-apnea connection in general, and particularly in continuity with the previous contributions developed by GAPS.

In this context, this Thesis explores forms of characterization of speech from OSA patients that have not been considered before.

On the one hand there is the possibility to keep working on applying new innovative speech techniques on the characterization of the study of OSA that have not been applied before to reinforce our understanding of sleep apnea. As voice techniques move forward, there are new ways of studying the link between speech and apnea syndrome.

On the other hand, the contributions of this Thesis are somehow conditioned by the availability of data. Having a new database formed by a greater amount of patients in a real clinical scenario has a positive aspect, an opportunity to analyze the problem with more data. This will allow us to take into consideration additional phenomena that may be present in speech samples from sleep apnea patients. In turn, the fact that more data is available hampers the problem as more variability will appear. This Thesis focuses on speech analysis, but also on the study of clinical variables that are involved in this syndrome and that affect speech of the OSA patients.

There is a long way to go in research studying the link between speech, clinical variables and OSA (Figure 1.2). In this Thesis, we will delve into the study to clarify this topic.

1.2. Goals

The goals of this Thesis were conditioned by the availability of data, as it always happens when a data-driven study is carried out. In this Thesis, different groups attending to the severity of OSA disease are analyzed, see Figure 1.3.

Initially, we had a limited and controlled database according to certain, specific clinical variables including age and body mass index. This means that we had a rather small number of patients who had similar age and body mass index, which allowed us to focus on the analysis of the speech and the apnea disease. Hereafter, we will refer to this database as Apnea Database 1.0.

Then since 2012 a second database, which we will refer as Apnea Database 2.0, is being recorded in a clinical practice sleep unit. Its main characteristic is that it contains much more patients, additional data like clinical variables as well as speech. This means that this scenario has a great variability and is more challenging to be analyzed.

This fact has been taken into account and consequently has largely conditioned the goals and results of this Thesis, posing as goals to study different new forms of speech characterization over

Figure 1.3.: Groups to be classified in the diagnosis OSA.

the controlled database, Apnea Database 1.0. Namely, the contributions of this Thesis are:

Goal 1: enhance OSA-detection performance on Apnea Database 1.0 by adding context-dependent information by means of hidden Markov models (HMMs). Statistical acoustic models that represent the spectral characteristics of different linguistic units such as sentences and phonemes are used, instead of the context independent approach based on Gaussian mixture models (GMMs) that was previously used. GMMs do not take into account the order of phonemes; they model the sentence as a whole, while the HMMs take into account the order in the sequence of phonemes.

Goal 2: detection of OSA patients from the analysis of a set of acoustic variables, not only to focus on the spectral envelope but also acoustic variables of different nature that could be related to the speech of OSA patients on Apnea Database 1.0. Variables are defined and analyzed, based on previous results contributions in the literature and our own analysis. These features are analyzed first in isolation and then in combination to assess their discriminative power to classify voices as corresponding to OSA patients and healthy subjects. The set of features selected focus on parameters that were traditionally used in the diagnosis of pathological voices, peculiarities of OSA patient voices such as anomalies in articulation, phonation and resonance and also prosodic measures.

It is worth to point out another important difference between Apnea Database 1.0 and Apnea Database 2.0 that will condition our experiments and results. Apnea Database 1.0 let to screen the population of study as it is formed by healthy people and severe apnea patients, [Figure 1.4](#). This fact makes it a great database for screening with the purpose of detecting the most severe cases of OSA, those who really need to be diagnosed and treated as soon as possible, the ones that should be the first of being attended in a sleep unit in order to be diagnosed by a specialist.

On the other hand, another remarkable fact about Apnea Database 2.0 is that all the population included was recorded because they went to a sleep unit, which means that even if they were not detected as apnea patient they were not healthy people either, [Figure 1.5](#). This means that they may suffer from another condition and are not properly a healthy group. However, it is interesting because as in this database we also have apnea patients in different grades of the disease, mild OSA and severe OSA, we can stratify our population and classify them according to the severity

Approach 1: HMMs
and Approach 2: Voice Features

Approach 3: Formants

Goal 3: revision of the basis of the detection of apnea through speech studying formants and bandwidths of sustained vowels in a real clinical practice scenario. The basis of the detection of apnea disease through speech are reviewed by means of the study on this elementary speech feature that are formants and bandwidths focusing on the analysis of the smallest phonetic unit, the vowels. This will let us guide future research on a real diagnosis scenario. We focus not only on the speech but also on the analysis of clinical variables from each subject that somehow are related to the speech and the apnea syndrome on Apnea Database 2.0; and on their interdependencies. Given the complexity and sources of variability of the clinical scenario, in order to have

the speech more controlled, in this part of the Thesis we only focus on sustained speech. Due to the lack of enough data, this could not be done on Apnea Database 1.0. This Thesis analyze the impact of different clinical variables on different acoustic characteristics particularly on the study of formants and bandwidths, which have been said to be intimately affected by the presence of OSA traits.

The work presented in this Thesis will guide future research lines for this type of scenario of clinical practice.

1.3. Structure of This Thesis

In the remaining chapters of this Thesis, we will first summarize the State of the Art on Obstructive Sleep Apnea and Speech and discuss the available data in chapters 2 and 3, respectively. The three subsequent chapters 4–6 present the progress achieved on the three goals mentioned above; each of them corresponds to a published article and seeks to give the background information necessary for the thesis to be self-contained. Finally, we conclude with chapter 7, which provides a summary and perspectives for future work.

CHAPTER 2

OBSTRUCTIVE SLEEP APNEA AND SPEECH

STATE OF THE ART

In this chapter, the Obstructive Sleep Apnea (OSA) syndrome is introduced. Since this Thesis is about speech and OSA, it is important to know the backgrounds and to grasp the relevant ideas of how patients are affected by this syndrome. In this chapter, the main symptoms, risk factors, prevalence, treatment, and diagnosis, as well as its dangerous consequences are presented. We will see that a great number of factors is involved in this syndrome and that despite its high prevalence around the world, this syndrome is still not well understood. Further on, we will discuss the clinical variables involved in this syndrome and the relation to OSA and speech, which forms the basis for this Thesis. Finally, we outline previous efforts from the literature to relate OSA and speech. This puts into context the contributions of this Thesis in the following Chapters.

2.1. Introduction

The pathogenesis of OSA has been a focus of research in the last decades. According to the American Academy of Sleep Medicine (AASM), OSA is a sleep-related breathing disorder that involves a decrease or complete halt in airflow despite an ongoing effort to breathe ([The Report of an American Academy of Sleep Medicine Task Force, 1999](#)). The condition is usually associated with loud snoring and hypoxemia (abnormally low level of oxygen in the blood). Apneas are typically terminated by brief arousals, which result in sleep fragmentation and diminished amounts of slow-wave sleep and (REM) rapid-eye-movement sleep.

In [Figure 2.1](#), we can observe the process of an hypopnea/apnea episode. The process starts once the patient has fallen asleep. OSA occurs when the muscles in the back of the throat relax. These muscles support the soft palate, the triangular piece of tissue hanging from the soft palate (uvula), the tonsils, the side walls of the throat and the tongue. When these muscles relax, the upper airway resistance increases and the airway narrows or closes as the patient breathes in, and, as a consequence, there is a lack of adequate breath-in which causes a decrease in the level of oxygen in blood.

If the obstruction of the airway is only partial, it will cause an hypopnea, and if the obstruction

Figure 2.1.: Apnea-Hypopnea episode development during sleep. An episode starts once the patient has fallen asleep (bottom) and breathing turns disordered and finally ceases; the episode finishes after arousal as the patient returns to sleep. Afterwards, the process will start all over again, making it difficult for OSA patients to get proper rest.

is total we refer to this as an apnea episode. In any case, in order to overcome airflow cessation, to balance gas concentrations and to recover normal breath, the patient must wake up. That means, as the brain senses this inability to breathe freely, it sends a signal to rouse in order to reopen the airway. This central nervous alarm reaction, an arousal or micro awake is a transient awakening that lasts less than ten seconds. This widens the pharyngeal airway, alleviating the obstruction and leading to the resumption of breathing. Soon after the patient falls back to sleep, the tongue and soft tissues relax again, with consequent complete or partial obstruction and loud snoring, which is a characteristic of patients with OSA.

Most patients are not aware of these micro awakes, but although they do not remember these frequent arousals, they will suffer their consequences: They do not reach the desired deep, restful phases of sleep, and the most common consequence is that they will be sleepy during daily activities.

2.2. History of OSA

OSA as a diagnosis has grown in prominence over the last four decades. Until recently, OSA was regarded as a merely medical curiosity, and snoring was considered as a humor topic more than something to study. However, this way of thinking has dramatically changed, and finally, OSA has

Figure 2.2.: (a) Illustration of Fat Joe, the first description of the morphotype of an apnea patient, in a somnolence state on a daily activity extracted from Charles Dickens' novel "The Posthumous Papers of the Pickwick Club" (Dickens, 1837). (b) Another representation of Fat Joe, which is not from the original book, but from a cigarette collector's card *The Fat Boy, Pickwick Papers*, published by John Player, early 20th century [<http://www.lookandlearn.com/history-images/M115405/The-Fat-Boy-Pickwick-Papers>].

been acknowledged as a worldwide problem.

First references to OSA were found in the Hippocratic Corpus (V-IV century BC). The Roman author Pliny the Younger (79 AD) reported the death of a man with obesity, sleepiness, and snoring. Distinguished personalities that had OSA were Emperor Napoleon Bonaparte (1769 – 1821), Queen Victoria (1837 – 1901), and President Franklin D. Roosevelt (1882 – 1945), among others (Conti, Conti & Gensini, 2006).

2.2.1. OSA in Charles Dickens' Novels

For some time, OSA was known as "Pickwickian syndrome" after the symptoms of the character "Fat Joe" described by Charles Dickens in his novel "The Posthumous Papers of the Pickwick Club" published in 1836 – 1837 (Dickens, 1837). The detailed description of the character of Fat Joe throughout this novel and the descriptions about other diseases granted him a place in the annals of medicine, see also Figure 2.2. Besides describing the syndromes of diseases of his time, Dickens played an important role in medicine: he also promoted the treatment of children, helped establish medical institutions and brought us face to face with the humanity of the poor, the deformed, and the crippled (Kryger, 2012).

Coming back to the descriptions of Fat Joe and OSA, we quote some extracts from "The Posthu-

mous Papers of the Pickwick Club” to guide us and give us some clues about OSA:

Joe! – damn that boy, he’s gone to sleep again!

This expression or its variations appear often along the book. Dickens describes that Joe is very sleepy during the daytime at any moment. Joe also falls asleep at the wrong time and place:

The fat boy rose, opened his eyes, swallowed the huge piece of pie he had been in the act of masticating when he last fell asleep.

... the fat boy laid himself affectionately down by the side of the cod-fish, and, placing an oyster-barrel under his head for a pillow, fell asleep instantaneously.

The combination of sleepiness and snoring are recurrent in Dickens’ character Fat Joe through the novel.

‘Sleep!’ said the old gentleman, ‘he’s always asleep. Goes on errands fast asleep, and snores as he waits at table.’ ‘How very odd!’ said Mr. Pickwick.

It is even a surprise when Joe is awake:

Joe; why, damn the boy, he’s awake!’

The quoted description of Joe is the first written reference to sleep disorders enumerating almost every related symptom (Burwell, Robin, Whaley & Bickelmann, 1956). The description of Joe’s symptoms is enriched as the novel flows, suggesting that the young servant would have suffered from sleep apnea, most probably due to his obese condition. Nevertheless, the question whether Joe suffered from sleep apnea or obesity hypoventilation syndrome—which is related to the OSA condition (McNicholas & Phillipson, 2005)—is still under discussion.

2.2.2. Phenotypes of OSA Patients

If we take a look at Fat Joe, see [Figure 2.2](#), we appreciate some of the most common phenotypes of sleep apnea. We can observe that he is fat, he hardly has a neck because of the thick fat tissues in his cervical perimeter, he has a round face and he seems extremely tired. These characteristics, which can be observed just from a quick view of a person are the first clues doctors have to get an idea if a patient is prone to have apnea. In fact, some doctors have reported that just from a first impression, they get a first opinion if a person suffers from this disease. This revealing fact has been the focus of research of some contributions:

- [Cakirer et al. \(2001\)](#) characterized the association between anthropometric measures of cranial and facial form with the severity of the OSA syndrome, in a large sample of whites and African-Americans,
- [Johal and Conaghan \(2004\)](#) studied the maxillary morphological and found statistical significant differences between OSA and control subjects in the maxillary skeletal morphology and oropharyngeal dimensions.

- Moreover, [Johal and Conaghan \(2004\)](#) sought to identify craniofacial and pharyngeal anatomical factors directly related to OSA.
- [R. W. W. Lee, Chan, Grunstein and Cistulli \(2009\)](#) demonstrated that craniofacial phenotypic differences in OSA in Caucasian subjects, using a photographic analysis technique.
- [R. W. W. Lee et al. \(2010\)](#) concluded that there is a relationship between surface facial dimensions and upper airway structures in subjects with OSA, which supports the potential role of surface facial measurements in anatomic phenotyping for OSA.

In conclusion, it is generally accepted that there are some characteristics that make an individual more prone to be an apnea patient. Now it is time to focus on the symptoms of this disease.

2.3. OSA Symptoms

2.3.1. Definition of the Apnea-Hypopnea Index (AHI)

OSA patients have recurrent episodes of obstruction of the airway during sleep ([Figure 2.1](#)). Usually, they are not aware of the apneas or hypopneas and the related micro awakes. A widely accepted measure of the severity of the condition is the apnea-hypopnea index (AHI), which is defined as the average number of apnea and hypopnea events per hour of sleep. According to the AASM, an AHI of 15 marks the threshold from mild to moderate OSA. Details on the diagnosis of OSA and how the AHI is determined are given in [section 2.5](#).

2.3.2. Snoring, the Most Typical Symptom

Regarding the symptoms, chronic snoring is the hallmark symptom associated with OSA. Although loud disruptive snoring is observed in 85% of OSA cases, not everyone who snores has OSA ([Strohl, 2015](#)). Snoring is very common in general, but snoring prevalence is difficult to estimate and highly subjective as it depends on bed partner's perception and it varies from night to night. It is estimated that the overall share of snorers is about 57% in men and 40% in women, with increasing snoring prevalence with age ([Doghramji, 2014](#)).

According to the World Health Organization, OSA syndrome is a clinical chronic disorder marked by frequent pauses in breathing during sleep usually accompanied by loud snoring ([WHO, 2004](#)).

The habitual snoring is a sign of increased pharyngeal airflow resistance. Unlike central apnea, OSA is related to the anatomy of the respiratory system, which is intimately linked to the speech production system. The narrowing of the upper airways during sleep, which predisposes to OSA, inevitably results in snoring. In [Figure 2.3](#), we observe that there is no obstruction in the normal breathing subject, however, in the snorer, there is a partial obstruction of the airway and in the OSA patient, the obstruction of the airway is complete. Apart from loud snoring, the typical symptoms experienced while sleeping include choking or gasping for breath and pauses in breathing, abrupt awakenings accompanied by shortness of breath, nocturia and insomnia.

OSA patients are not the only ones that suffer from this disease: Because OSA patients usually are loud snorers, the sleep quality of their spouses or bed partners is affected leaving them

Figure 2.3.: Differences between a normal breathing, partial obstruction of the airway observed while snoring and complete obstruction of the airway observed in an OSA patient. Picture taken from ([Altuna & Saga Policlínica Gipuzkoa, Paseo Miramón 174 20014 Donostia San Sebastián, 2013](#)).

sleep-deprived as well, which may lead to separate sleeping rooms and disrupted relationships. Regarding how OSA indirectly affects bed partners, it has been referred to as “spousal arousal syndrome”.

Other direct symptoms. OSA symptoms can happen during sleep but also while patients are awake ([Lurie, 2011](#)). When the patient wakes up, it is also common that he has chest pain and dry mouth or a sore throat. While patients are awake they may experience hypersomnolence, chronic fatigue, lack of energy, morning headaches, unintentional sleep episodes, episodes of irregular heartbeat, chronic elevation in daytime high blood pressure, impaired concentration, mood swings, anxiety, and depression, as well as alterations in quality of life that can impact social and familial relationships and professional performance.

2.3.3. Severe Consequences of OSA

The effects of OSA include, but are not limited to, memory loss, weight gain, cardiovascular disease, and falling asleep during daytime activities. When left untreated, OSA often results in motor vehicle accidents, missed days of work, overnight hospital stays, and chronic fatigue.

If not treated, OSA is a serious health threat. As we have mentioned already in [section 2.1](#), the condition is a risk factor for hypertension and cardiovascular diseases. [Figure 2.4](#) picks up the apnea cycle [Figure 2.1](#) in the introduction and relates each factor to its physiological consequences (middle column) and the clinical features, i.e., health risks, in the right column. Everything is related. For example, the drop of oxygen causes systemic vasoconstriction which causes hypertension. Several epidemiological studies have implicated OSA as a risk factor for the development of systemic hypertension due to systemic vasoconstriction ([Golbin, Somers & Caples, 2008](#)).

Left untreated, OSA can lead to dreadful serious health consequences, including increased mortality and an increased incidence of hypertension, stroke, heart failure, coronary artery disease, cardiac rhythm problems, type 2 diabetes, gastroesophageal reflux disease, nocturnal angina, hypothyroidism, or neurocognitive difficulties ([Yaggi et al., 2005](#); [Marin et al., 2012](#)).

Figure 2.4.: OSA primary sequence of events, physiological responses and clinical features.
Adapted from (Goel et al., 2012).

2.3.3.1. Traffic Accidents and OSA

In addition to the medical issues mentioned, the lack of rest leads to sleepiness and lack of concentration, which increases the risk of accidents (Lloberes et al., 2000). George (2007) has estimated that individuals with untreated OSA are two to ten times more likely to have a traffic accident because of impaired driving performance. According to different studies, there has been established a strong relation between traffic accidents and OSA, it varies depending on every study.

Currently, traffic accidents are among the main causes of death. It is one of the main concerns of governments to reduce the mortality rates because of accidents through traffic campaigns and other means. There are countries such as Sweden that have forbidden to drive if you suffer from sleep apnea or other illnesses with sleep disturbance unless successfully treated (Valham, 2011; VVFS, 2008).

Sleep apnea, snoring and the risk of traffic accidents has been investigated in several studies. OSA, as well as snoring, is associated with an increased risk of traffic accidents (Barbé et al., 1998). Patients with sleep apnea perform worse in traffic simulators compared to control groups (Barbé et al., 1998; Risser, Ware & Freeman, 2000). Treatment of sleep apnea with a CPAP mask during sleep (see section 2.5.2.2 below) reduces the risk of traffic accidents and improves the simulator performance.

2.4. OSA Risk Factors and Prevalence

2.4.1. OSA Prevalence

Many people suffer from OSA: according to the World Health Organization, it has been estimated that 100 million individuals worldwide have some degree of OSA. The most frightening numbers suggest that 80% of moderate and severe OSA cases are still undiagnosed (Kapur et al., 1999).

In order to get an idea to quantify how many people are affected by this disease, we should focus on prevalence's studies. In Table 2.1 we find a comparison of OSA prevalences in different countries such as United States (T. Young et al., 1993; Bixler, Vgontzas, Ten Have, Tyson & A., 1998; Bixler et al., 2001), Spain (Durán, Esnaola, Rubio & Izutueta, 2001), Asia (Ip et al., 2001, 2004; Kim et al., 2004), India (Udwadia, Doshi, Lonkar & Singh, 2004; Sharma, Kumpawat, Banga & Goel, 2006) and Australia (Bearpark et al., 1995). In agreement with Jordan and McEvoy (2003), the table shows that OSA is two to three-times more common in men than in women. The likelihood of developing the condition increases with age. Among people under 35 years, OSA is more common in blacks. It is also more likely to develop in African-Americans, Hispanics, and Pacific Islanders than in Caucasians. OSA usually occurs in adults aged between 18 and 60, but it can occur at any age.

Although OSA has a higher prevalence in males, also females suffer from this disease. It has been shown that women have a higher risk of developing OSA during pregnancy. This happens because OSA is highly correlated with weight, and usually during pregnancy women gain some weight. Usually, the OSA condition disappears after pregnancy. Pregnant women are not the only female

Table 2.1.: Prevalence of OSA ($AHI \geq 5$) in different ethnic groups, assessed with standard polysomnography. From (Jamie et al., 2010).

Reference	Study population	Age (years)	Prevalence (%)	
			Men	Women
T. Young et al. (1993)	American men and women	30 – 60	4* – 25	2* – 19
Bixler et al. (1998)	American men	20 – 100	17	
Bixler et al. (2001)	American men and women	20 – 100	3.9*	1.2*
Durán et al. (2001)	Spanish men and women	30 – 70	14* – 26	7* – 28
Ip et al. (2001)	Chinese men	30 – 60	4.1* – 8.8	
Ip et al. (2004)	Chinese women	30 – 60		2.1* – 3.7
Kim et al. (2004)	Korean men and women	40 – 69	4.5* – 27	3.2* – 16
Udwadia et al. (2004)	Indian men	25 – 65	7.5* – 19.5	
Sharma et al. (2006)	Indian men and women	30 – 60	4.9* – 19.7	2.1* – 7.4

* OSA defined as $AHI \geq 5$ and excessive daytime sleepiness

OSA patients, rather, the prevalence increases with age, reaching its highest values in women after menopause. Thus, the typical female candidate is a middle age woman in menopause. In this case, the OSA symptoms can be confused with the menopause symptoms making it more difficult to diagnose the OSA syndrome in this stratum of the population. In postmenopausal women, the OSA prevalence increases, particularly in women without hormone replacement therapy, but it remains still lower than men of the same age stratum (Jordan & McEvoy, 2003).

2.4.2. OSA Risk Factors

As mentioned above [section 2.2.1, section 2.2.2], there are anatomical factors that predispose to OSA such as large tongue, large tonsils, short mandible, and obesity with an increased deposition of fat in tissues surrounding the neck and throat. In the above picture Figure 2.2 of Fat Joe, the famous character from Dickens' novel, we could observe some of these anatomical factors. Now, in Table 2.2 we can observe a collection of factors that contribute to suffering from OSA. Other predisposing factors are male sex, age greater than 65 years, sleeping in a supine position, and having a relative that suffers from this syndrome. According to (T. Young et al., 1993; Gibson, 2004), OSA is strongly associated with obesity, but is also increasingly identified in normal subjects, whose particular craniofacial structure is an important contributory factor. Regarding obesity, not everyone with OSA is overweight and, vice versa, also thin people can develop OSA.

Thus, OSA is associated with anatomic alterations due to adiposity around the pharynx. Furthermore, central obesity has been associated with the reduction of lung volume, which leads to a loss of caudal traction on the upper airway, and hence, an increase in pharyngeal collapsibility. In the following, we list a few epidemiological studies that have investigated the associations between sleep apnea and obesity:

Table 2.2.: Factors that contribute to the pathophysiology of OSA. Table extracted from (Deegan & McNicholas, 1995).

General factors	Anthropometric (male sex, age, obesity), Drugs (ethanol, hypnotics), Genetics
Reduced upper airway caliber	Specific anatomical lesions (enlarged tonsils, micrognathia), Neck flexion, Nasal obstruction
Mechanical factors	Supine posture, Increased UA resistance, Increased UA compliance
Upper airway muscle function	Abnormal UA dilator muscle activity, Impaired relationship of UA muscle and diaphragm contraction
Upper airway reflexes	Impaired response to negative pressure, Feedback from the lungs
Central factors	Reduced chemical drives, Increased periodicity of central drive, Inadequate response to breath loading
Arousal	Impaired arousal responses, Postapneic hyperventilation

UA: upper airway.

- [T. Young et al. \(1993\)](#) considered a community-based cohort of middle-aged Caucasian subjects and associated a 1-SD increase in body mass index with a 4-fold rise in the prevalence of sleep apnea.
- [Peppard, Young, Palta, Dempsey and Skatrud \(2000\)](#) demonstrated that a 10% body weight reduction was associated with a 26% decrease in the AHI, the measure for the severity of the disease defined in [section 2.3.1](#).

Thus, weight reduction is an important conservative treatment for sleep apnea, as we will see later in [section 2.5.2](#) on OSA treatment.

Another predisposal factor for OSA is the size of the neck, which is commonly referred to as the cervical perimeter or the neck circumference. A thick neck may narrow the airway and may be an indication of excess weight. According to some contributions, a large cervical perimeter is associated with an increased risk of OSA. Anything that blocks the upper airway will be a predisposing factor for OSA: narrowed airway, big tonsils, adenoids and nasal congestion. OSA is relatively common in people with hypertension and diabetes ([Punjabi, 2008](#)).

There are studies that show that if a relative has OSA you will have an increased risk. There is also some genetic tendency related to jaw morphology. The condition is also more common among people who have smaller airways in their noses, throats, or mouths. The small airway could be related to the actual size and shape of the airway, or to obstructions or other medical conditions that are causing obstructions ([T. Young, Skatrud & Peppard, 2004](#)).

Regarding habits that predispose to OSA people who smoke are more likely to have OSA and also people that consume alcohol are more prone to OSA. So quitting these habits will be another way to reduce the symptoms as we will see later in this chapter in [section 2.5.2](#).

2.5. Clinical Diagnosis

OSA must be considered as a multifactorial disease in which multiple genes, environmental influences, and development factors are involved. As OSA is a complex disease, it is not surprising that its diagnosis is also complex. One of the things that most surprised me when I started my research on this disease is the lack of a standard protocol. Even when I focused on one country, i.e. Spain, wondering what kind of protocol physicians used, I only found a consensus document that was written more than 10 years ago ([Puertas et al., 2005](#)) about this disease but still nowadays every physician or sleep unit has its own way of dealing with this disease.

The diagnosis of OSA is tedious and resource-intensive. It starts when a patient feels that something is not well and decides to see his family doctor. Due to the lack of information and awareness of sleep disorders in society, this first step may take years. Once in primary care, the patient or often his bed partner complains about not resting properly during sleep because of heavy snoring. The family doctors will check the medical history and a physical examination looking for OSA clues. The doctor should also ask about sleeping habits, degree of snoring and daytime sleepiness. If the family doctor suspects the patient might suffer from OSA, the patient will be referred to a specialist at a sleep unit in order to have more tests done. At the sleep

unit, the sleeping habits will be investigated more deeply. A full overnight sleep study, called polysomnography (PSG), is employed for the final diagnosis, which confirms or excludes the the OSA disorder.

The word polysomnography, derived from the Greek roots “poly”, meaning many, “somno”, meaning sleep, and “graphy” meaning to write, refers to multiple tests performed on patients while they sleep. According to the AASM, the PSG test is the gold standard test for quantifying and qualifying sleep disorders. PSG includes recordings of:

- electrographic variables (electroencephalogram, electrooculogram, and submental electromyogram) that permit the identification of sleep and its various stages
- ventilatory variables that permit the identification of apneas and their classification as central or obstructive
- arterial oxygen saturation by ear or finger oximetry
- heart rate

These measurements allow for assessment of sleep, wakefulness, arousals, breathing, heart function, and brain and nerve function.

In order to record all these signals, a series of painless electricity conducting pads are placed on the head, face, chest, and legs. These measure electric signals of the brain, respiration, heart, lungs, and muscles. This method, normally performed during a single night, is very reliable in estimating the sleep stages but is quite uncomfortable for the patient since he has to be all wired up and is restricted in his sleeping position. There is also a night-vision video camera to record actions and movements. After having all these records, data has to be analyzed by a technician to perform an appropriate sleep analysis.

A broadly used criterion for diagnosing OSA is having an AHI over 15. The AASM classification of OSA severity considers both the AHI (mild defined as an AHI of 5–15, moderate as 15–30, and severe as > 30) and the degree of daytime sleepiness (mild: unwanted sleepiness or involuntary sleep episodes occurring during activities that need little attention; moderate: during activities that need some attention, such as during meetings; severe: during activities that need more active attention such as during conversation or driving).

PSG is resource-intensive and expensive and requires patients to spend the night under observation in a foreign environment and to have at least 6 hours of sleep recordings. It is not uncommon that the patient is not able to sleep that night those 6 hours and another PSG is needed to make a diagnosis. Recently the demand for diagnostic studies for OSA has increased ([George, 2007](#)). In countries such as Spain, the waiting list to do the PSG is more than one year long, which makes the diagnosis extremely long ([Puertas et al., 2005](#)). Therefore, there is a strong need for alternative methods to PSG of screening apnea patients in order to detect patients with high risk of OSA and reduce these considerable delays.

Sleep disorders are a recent branch of modern medicine ([Haba Rubio & Krieger, 2012](#)). It was less than a century ago when the first human electroencephalogram (EEG) was recorded in 1924 by Hans Berger ([Berger, 1929](#)). It took some years, until 1935 when Albert Grass first

Figure 2.5.: Illustration of polysomnography test data recording (A) and profile of the curves collected including oxygen saturation and breathing events detection (B). [National Heart, Lung, and Blood Institute, “What To Expect During a Sleep Study”, <https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0062979/>]

manufactured a 3-channel EEG and founded his own company to supply EEGs (Collura, 1993). Also at that time sleep patterns were discovered. Later, in 1952 rapid-eye-movement (REM) sleep was discovered. In 1970 William Dement was the first in creating a sleep-disorder clinic. Until recently, sleep disorders were not part of medical studies in universities. Now, that the general public is more aware of them, sleep disorders are included in modern medical programs at faculties worldwide.

2.5.1. Screening Techniques for OSA

In the published literature, we find numerous efforts of complementary clinical methods of predicting OSA (Ramachandran & Josephs, 2009). These methods are broadly classified as questionnaires about sleepiness (Ahmadi, Chung, Gibbs & Shapiro, 2008) and clinical prediction models (Gurubhagavatula, Maislin, Nkwuo & Pack, 2004; Friedman et al., 2010) using anthropometric characteristics (i.e. BMI or morphology) and epidemiological and medical history parameters (i.e. snoring).

Currently, there are several questionnaires that can be used for screening OSA patients. One of the most useful is the Berlin questionnaire (Netzer, Stoohs, Netzer, Clark & Strohl, 1999), that contains questions addressing three symptom categories: snoring, sleepiness, and high blood pressure/weight. This questionnaire is able to identify patients at high or low risk of OSA.

Among sleep questionnaires for screening purposes, there is the sleep disorders questionnaire which is formed by 176 questions. Although is a valid tool to diagnose common sleep disorders, due to its length is not considered as a practical instrument in primary settings (Netzer et al.,

1999).

The Pittsburgh Sleep Quality Index is formed by 24 questions and can be useful for measuring changes in sleep quality over time in patients, alerting physicians of their need for further evaluation (Buysse, Reynolds, Monk, Berman & Kupfer, 1989). More recently, Roth et al. (2002) proposed the global sleep assessment questionnaire as a suitable screening tool for sleep disorders in general primary care population, although they declared some study-limitations. Furthermore, Casale et al. (2008) proposed the Rome questionnaire, a 7-items questionnaire. They proved its usefulness in identifying adult patients at risk of OSA syndrome. This questionnaire together with BMI seems to be a useful tool to make a selection of the patients at higher risk of moderate-severe OSA, who need a prompt PSG evaluation.

2.5.2. Treatment

The treatment for OSA depends on the specific cause of the obstruction, the individual's medical history and the severity of the disorder. Treatments are focused on controlling the risk factors, oral appliances, or airway surgery.

2.5.2.1. Controlling risk factors

Lifestyle changes that can help with OSA are positional therapy, weight loss, quitting smoking, alcohol, and drugs.

Regarding positional therapy, as OSA is more common in a supine position, it will be helpful to sleep at a 30 degree or higher elevation of the upper body in order to prevent the gravitational collapse of the airway. Sleeping in lateral positions as opposed to supine positions is also recommended (Neil, Angus, Sajkov & McEvoy, 1997), because the gravitational component is smaller in the lateral position. This approach can be used in combination with other treatments and may be particularly effective in obese.

As obese are at high risk for this disease, it is recommendable to lose weight. A weight loss may eliminate apneic episodes by reducing the adipose level of parapharyngeal tissues. Peppard et al. (2000) have demonstrated that a 10% body weight reduction was associated with a parallel 26% decrement in the AHI. Since smoking worsens apnea, it will be recommended to quit or reduce its consume. Also, it will be better to avoid alcohol within 4 hours before sleep and to avoid sedatives and sleeping medications.

2.5.2.2. Oral appliances

At present, the most effective and widespread treatment for OSA is nasal CPAP, which prevents apnea episodes by providing a pneumatic splint to the airway. A machine generates and pumps a controlled stream of constant pressurized air through a face mask worn over the nose, mouth, or both. The additional pressure holds the relaxed muscles, keeping the airway open. The right pressure is prescribed by the patient's technician, based on the results from the PSG test. There

is evidence supporting improvements in measures of OSA severity, daytime sleepiness, and neurocognitive function after using nasal CPAP. Although CPAP is very useful for treating OSA patients, adherence to therapy continues to be a major problem. There are even some patients finding CPAP intolerable. This is in general related to the facial mask and the pressure required to prevent airway collapse. Proper humidification of the inspired air, careful selection of the appropriate mask, and the addition of a pressure ramp all may improve compliance.

Another approach are oral appliances which are designed to alleviate OSA by repositioning of the anterior mandibular. These appliances are prepared by a dentist and act by pulling the lower jaw (and, thus, the tongue) forward, thereby enlarging the pharyngeal airway.

2.5.2.3. Airway surgery

Several types of surgeries exist that can help against OSA, which are all invasive treatments ([McDonald, 2003](#)).

Uvulopalatopharyngoplasty involves the removal of part of the soft palate, uvula and redundant peripharyngeal tissues, sometimes including the tonsils. Although this procedure is often effective in eliminating snoring, it is not necessarily curative for OSA, because areas of the airway other than the soft palate also collapse in most patients.

Jaw surgery is the most invasive surgical procedure used to treat OSA. Patients who have a small, retrognathic mandible and have failed more conservative treatments may be candidates for this approach. The surgical procedure may be limited to pulling the tongue forward. An extensive procedure may involve moving both the mandible and maxilla. It has to be taken into account that extensive jaw surgery has a higher complication rate and a longer recovery time.

Tracheostomy consists in opening a direct airway through an incision in the trachea. Because of its high associated morbidity, this procedure applied only in rare severe cases of OSA.

Gastric surgery is an indirect measure to alleviate OSA by weight reduction. It might be a consideration in some morbidly obese patients with OSA, who failed to reduce weight by natural means.

2.6. Evolution of Speech and OSA

The biological mechanisms leading to the collapse of the upper airway during sleep are related to a set of anatomic and physiological factors that constrict space for the soft tissues surrounding the pharynx ([Ryan & Bradley, 2005](#)). These anatomic factors of the pharynx necessarily affect the upper-airway structure, so it seems reasonable to look for characteristic signatures in OSA patients' voices.

In this section, we will review the acoustics of the human voice and discuss the related anatomic changes in the evolution of *Homo sapiens*.

The first signs of human verbal communication date back from 100 000 years ago ([Lieberman, Klatt & Wilson, 1969](#)). These must have been preceded by three main requisites for verbal communication, which are:

1. The vocal production mechanism needs to be able to produce a variety of clearly discriminable sounds (the formant patterns, see below), and thus a vocal stream of adequate intricacy to communicate complex linguistic concepts.
2. The ability to imitate novel sounds vocally is a prerequisite for the formation of large vocabularies that typify all human languages. This ability is often taken for granted, but in fact is very unusual among mammals and requires an evolutionary explanation.
3. Finally, the mental capability is needed to make use of the voice and to communicate in a sensible way.

Thus, the evolution of verbal communication required a specialized anatomy and a set of neural mechanisms that confer fully human speech, language, and cognitive ability.

Let us take a look at a few examples. Parrots are able to utter words, but they are limited to imitations and it is not possible to have a communication with them, i.e., they lack prerequisite 3. Chimpanzees, on the other hand, are very social beings, capable of sign language or other visual systems to develop large vocabularies and produce multi-symbol sentences. However, they are not able to use spoken vocabulary or even imitate speech sounds ([Hayes, 1951](#); [Savage-Rumbaugh et al., 1993](#)). They can be trained to imitate a few words using a limited range of sounds, which remains insufficient to make sentences, i.e., they lack prerequisites 1 and 2. Only *Homo sapiens* fulfills all three prerequisites.

Homo sapiens and chimpanzee are much more similar than *Homo sapiens* and parrots. Indeed, they share 98 percent of their genes. But which genes among the other 2 percent encode the anatomic differences that have such enormous consequences that one can speak and the other one cannot? ([Diamond & Shermer, 1993](#); [Davidson, 2003](#); [Davidson, Sedgh, Tran & Stepnowsky Jr, 2005](#))

2.6.1. How Human Speech works

The basic function of speech is the following. Air from the lungs flows through the trachea (wind-pipe) and passes the larynx, where it passes a narrow split between the vocal folds, known as the glottis. Depending on the air-pressure difference across the glottis, the tension of the vocal folds, and the opening of the glottis, the vocal folds are set in vibration. This process is the origin of the vocal sound, a process referred to as *phonation*. The primary sound from the vocal folds is then filtered in the supra-laryngeal vocal tract (SVT), whose *resonances* amplify certain frequencies, the so-called *formant* frequencies, and suppress others. The resonances of the SVT are tuned by the position of the back of the tongue and the velopharyngeal opening, which controls the coupling of the vocal tract with the nasal cavity. Finally, lips, teeth and the tip of the tongue control the foremost part of the vocal tract, namely the mouth opening. This fine-tuning of sound is referred to as *articulation*.

Figure 2.6.: Distribution of the nodes (N_x) and maximum points (antinodes, A_x) of the volume current in a uniform pipe (above) and in the simplified model of a uniform vocal tract (below). Adapted from (Chiba & Kajiyama, 1941).

Basic understanding of formants. In the simplest model, the vocal tract is a tube with one open end (mouth) and one closed end (larynx, the source of sound) (Chiba & Kajiyama, 1941; Behrman, 2007). The resonance condition is to have a standing sound wave with a node (of volume current) at the larynx and an antinode at the mouth opening, see Figure 2.6. It is fulfilled whenever the tube length l is an odd multiple of a quarter of the wavelength $\lambda = c/f$, i.e.,

$$f_n = (2n - 1) \frac{c}{4l}, \quad n \in \{1, 2, \dots\}. \quad (2.1)$$

With $c \approx 340 \text{ m/s}$ and $l \sim 20 \text{ cm}$, this yields rough estimates for the first formant frequencies $f_1 \approx 425 \text{ Hz}$ and $f_2 = 1.3 \text{ kHz}$. The true formant frequencies $F_1, F_2, \dots$ depend on the exact shape of vocal tract, which is, of course, not a straight tube of uniform diameter, but is bent, has varying diameter and is connected to the nasal cavity. Constrictions located at antinodes lower the formant frequencies, while constrictions at nodes will cause higher formant frequencies. Movements of the tongue are capable of producing great changes in the diameter of the SVT, which are needed to shift the formant frequencies. Already the first two formants are sufficient to distinguish the different vowels, see Fig. 2.7.

The human vocal tract. The evolution of human speech was closely tied to the mechanisms of sound production presented above. The most obvious speech-related difference between humans and other mammals concerns the structure of the human vocal tract (Fitch, 2000). The human SVT is formed by two sections, see Figure 2.8:

- SVT_H, a horizontal portion in the oral cavity, which includes the mouth and oropharynx
- SVT_V, a vertical portion in the throat called the pharynx, which is located behind the tongue and above the larynx. This vertical portion of the vocal tract extends from the palate down to the vocal cords.

Figure 2.7.: Vowel chart for Spanish. From (Quilis, 1981). On the vertical axis, first formant frequencies, on the horizontal axis, second formant frequencies

The human larynx rests much lower in the throat than in the apes. Due to the change in larynx position, the tongue descends into the pharynx, which allows it to move both vertically and horizontally within the vocal tract, which greatly expands our phonetic repertoire (Lieberman et al., 1969). Thus, in normal adults, SVT_H and SVT_V form an angle of 90 degrees and are approximately equal in length, in a 1:1 proportion. This anatomical modification includes a shortening of the maxillary, ethmoid, palatal, and mandibular bones; narrowing and elongation of the SVT, anterior foramen magnum, and oropharyngeal tongue; a descended larynx and shortened soft palate with loss of the epiglottic-soft palate lock-up; and an acute oral cavity-skull base angle (Johal, Patel & Battagel, 2007).

The recognition of the central importance of formants paved the way for breakthroughs in speech science, and insights into the evolution of speech. Already Charles Darwin noted in his “On the Origin of Species” that the human vocal tract differs from that of other living primates in a way that makes possible the human speech and the use of language to communicate (Darwin, 1871).

Additionally to the mentioned evolution of the vocal tract, human speech also requires a precise nervous control for the interplay of all articulators involved in the production of speech, such as tongue, lips, velum, jaw, etc.. The tongue body that shapes the formants must be closely synchronized with the lips and the palate and with the vibrations of the larynx. To quote an example from Fitch’s review (Fitch, 2000):

The distinction between ‘pat’ and ‘bat’ is primarily a question of when the larynx begins vibrating relative to vocal tract movements. A difference of tens of milliseconds is enough to differentiate these sounds perceptually.

Figure 2.8.: Diagram of SVT_H and SVT_V (extracted from [Goldshstein et al., 2011](#)).

The Meaning of Speech for Human Society. The structure of the larynx, tongue, and associated muscles give us fine control over spoken sounds. In contrast, the vocal tracts of other living primates are physiologically incapable of producing such vowels. Also, Neanderthals' necks were too short and their faces too long ($SVT_H/SVT_V > 1$) to have accommodated equally proportioned SVTs. Therefore Neanderthals were incapable of producing the full range of sounds made by humans today. The mentioned changes in anatomy resulted in the human capacity for speech, whose significance was pointed out by Jared Diamond in his book "The Rise And Fall Of The Third Chimpanzee" as follows ([Diamond, 1991](#)):

But if the missing ingredient did consist of changes in our vocal tract that permitted fine control of sounds, then the capacity for innovation that constitutes the Great Leap Forward would follow eventually. It was the spoken word that made us free.

Thus, the achievement of speech and communication opened the way for all the cultural and technological advances of humanity from 40.000 years ago until today.

Connection with OSA. But, all magic comes with a price: the low position of the larynx, as it is necessary for the development of speech, also brings inconveniences. It requires a complicated swallowing process because the ways of air and food cross in the pharynx. When swallowing, the larynx closes to prevent food from falling into the trachea (windpipe), a mechanism that bears the risk of choking. At birth, the human larynx has a high position, similar to that of other animals. As infants grow, the larynx begins a slow descent to its lower adult position, giving space for a wider range of formants. A second smaller descent occurs in human males at puberty ([Fitch & Giedd, 1999](#)). Infants during their first three months have their larynx still located high enough in the throat to allow for a direct connection of the trachea to the nasal passages. At the same time, the food can pass on both sides of the larynx to the esophagus. This way, simultaneous breathing and swallowing is possible, as it is for most animals ([Negus, 1949](#)).

Once the larynx has lowered, the flexible and extended pharynx is a region of the upper airway that is not supported by rigid structures as are the nasal passage supported by bones and the trachea supported by tracheal rings of cartilage. The vocal tract is supported only by soft tissues and muscles, which bear the risk of collapse and obstruction, especially during sleep. In such a way, the anatomic changes that enabled the development of voice, speech, and language, also have provided the structural basis for OSA (Davidson et al., 2005).

In turn, OSA should be reflected in the voice, opening a new way for the diagnosis of OSA.

2.7. OSA Assessment via Speech Technologies

After outlining the OSA syndrome and the evolution of speech, we now focus on the relationship between OSA and speech. OSA involves several physical alterations such as craniofacial abnormalities, dental occlusion, longer distance between the hyoid bone and the mandibular plane, relaxed soft tissues, large tongue base, etc., which, in general, cause a longer and more collapsible upper airway. Consequently, abnormal or particular speech features in OSA speakers may be expected from such an altered structure or function of their upper airways. This fact promoted the first perceptual studies from Fox, Monoson and colleges (Monoson & Fox, 1987; Fox et al., 1989), based on the previous bibliography of their times that already showed the link between OSA and speech.

2.7.1. Perceptive Studies (Fox, Monoson et al.)

2.7.1.1. Conceptual Study by Monoson and Fox (1987)

The contribution by Monoson and Fox (1987) was the preliminary study that laid the groundwork for the later study. It was a perceptual study in which skilled judges compared voices of apnea patients with those of a healthy group (control group). They judged speech samples as normal or abnormal regarding resonance, articulation, and phonation. Hereafter, we quote resonance disorder, articulation disorder and voice disorder as defined by Monoson and Fox (1987):

1. “Resonance disorder is characterized by abnormality in vocal quality, typically related to coupling of the vocal tract (pharynx and oral cavity) to the nasal cavity. Resonance disorder can also be related to an abnormality in vocal tract damping and compliance.”
2. “Articulation disorder is characterized by mispronunciation of the speech sounds related to faulty functional or organic control of the tongue, lips, hard, and soft palate.”
3. “Phonation disorder is defined as abnormal sound production related to laryngeal dysfunction.”

This first study (Monoson & Fox, 1987) was designed to confirm the conjecture that the presence of speech disorder occurred more commonly in sleep apnea patients than in matched control subjects. In this study, they had a population of 39 subjects, formed by 13 sleep apnea patients, 13 chronic obstructive pulmonary disease (COPD) patients, and 13 control subjects. Most of the

Table 2.3.: Group means and standard deviations (SD) of judged speech disorder for three speech descriptors and treatment groups, rated on an equally-appearing-interval scale ranging from 1 (normal) to 7 (severe deviation from normal). From (Fox et al., 1989).

Group	Articulation		Phonation		Resonance		Mean
	Mean	SD	Mean	SD	Mean	SD	
Sleep Apnea*	1.9	1.06	3.0	1.55	2.2	1.01	2.6
COPD	1.5	1.08	2.3	1.19	1.9	0.85	1.8
Normal	1.0	0.07	1.2	0.45	1.3	0.34	1.1
Mean	1.5	0.94	2.2	1.35	1.8	0.88	

* Mostly (85%) OSA

apnea subjects were non-smokers, with AHI greater than 14. 61.5% of them suffered from severe apnea (AHI > 30), and the other 38.5% from moderate apnea.

For the listening task, three certified speech pathologists served as judges. Their main objective was to check if there were speech disorders in the sleep apnea population. For that purpose they had to listen to a test sentence spoken by each subject in three repetitions.

From the apnea group, 62% of the speakers were judged as abnormal, whereas the COPD and control group had 23% and 8% respectively. From this 62% of the apnea group, 87.5% were unanimously judged abnormal, representing a high interobserver agreement between judges. The speech of the abnormal sleep apnea speakers was characterized by the judges as dysarthric with abnormal resonance. Dysarthria is the lack of muscular control over the speech mechanism due to damage of the central peripheral nervous system and covers motor disturbance of articulation, resonance, and voice.

Moreover, the design carried out in this preliminary study may have promoted interobserver disagreement. The judges may have tended to find a disorder when it was not present because they knew they would be listening to some disordered speakers.

2.7.1.2. Quantitative Study by Fox et al. (1989)

Two years later, Fox et al. (1989) published a follow-up study, whose aim was to perform a more precise and rigorous analysis by analyzing the same three descriptors resonance, articulation, and phonation, and quantifying the speech abnormalities. To this end, Fox et al. employed a larger population of 27 subjects in each group, and they also increased the number of judges to 10 graduate students of speech pathology. From the 27 apnea patients, 23 suffered from OSA, 4 from mixed-type apnea and none of them from central apnea.

Discriminative Power. Results of detected anomalies from (Fox et al., 1989) are shown in Table 2.3. Discriminant function analysis was used to differentiate the three groups using the weighted measures of articulation, phonation, and resonance. As results from this work, in 74%

of the sleep apnea subjects, they detected abnormal resonance, articulation or phonation, whereas in the COPD and control groups the percentage was 53% and 7%, respectively.

In a second setting, the COPD group was excluded, which increases the discrimination power for the sleep apnea/normal classification problem. Based on the three speech descriptors, they identified correctly 96.3% of normal subjects, whereas the percentage for the correct identification for the apnea group was only 63%.

The Observed Anomalies. Although acoustic cues for differences between the groups of speakers remain somewhat unclear, this study has pointed out several differences and has confirmed that the voices of apnea speakers had certain anomalies that might be due to an altered structure or function of their upper airway. Consequently, the occurrence of a speech disorder in OSA population should be expected including anomalies in articulation, phonation, and resonance. These anomalies can be described as follows:

Articulatory anomalies. A neuromotor dysfunction could be found in apnea population due to a “lack of regulated innervation to the breathing musculature or upper airway muscle hypotonus.” This dysfunction is normally related to speech disorders, especially dysarthria. There are several types of dysarthria, resulting in various different acoustic features, but all types affect the articulation of consonants and vowels causing the slurring of speech.

Phonation anomalies. One of the conclusions of Fox and Monoson’s work was that the most discriminative feature to distinguish apnea subjects from healthy subjects was related to phonation anomalies. This may be surprising considering that phonatory organs, and in particular the larynx, should have a very limited role in the physiopathology of the disease. This study found certain roughness alterations in the speech of apnea patients related to a lack of periodicity of the sounds. One of the hypotheses still to be confirmed is that these anomalies may be due to the characteristic pattern and heavy snoring of sleep apnea patients, which can cause inflammation in vocal cords (Boyd, Petrof, Hamid, Fraser & Kimoff, 2004) and also fatigue due to the extra respiratory effort in OSA patients (Payne et al., 2006).

Resonance anomalies. OSA voices can also present abnormal resonances that might be due to an altered structure or function of the velopharyngeal mechanism, related to the coupling of the vocal tract with the nasal cavity. As a result, an abnormal vocal quality characteristic can be expected in the speech of these patients, that may be revealed through two features:

- First, speakers with a defective velopharyngeal mechanism can produce speech with inappropriate nasal resonance. The term nasalization can refer to two different phenomena in the context of speech; hyponasality (no nasalization is produced when the sound should be nasal) and hypernasality (nasalization during the production of non-nasal sounds). In Spanish, vowels only acquire either a nasal or a non-nasal quality depending on the presence or absence of adjacent nasal consonants. The work by Fox

et al. (1989), concluded that these resonance abnormalities could not be perceived either as hyponasality or as hypernasality. Instead of this, OSA speakers exhibited smaller intra-speaker differences between non-nasal and nasal vowels due to this velopharyngeal dysfunction. The term applied to this speech disorder is “cul-de-sac” resonance that causes the sound to be perceived as if it were resonating in a blind chamber. This matches perfectly with the results of (J. L. Blanco-Murillo, Fernández-Pozo, Díaz-Pardo, Alvaro Sigüenza et al., 2009) where it is confirmed that apnea speakers showed smaller intra-speaker differences between non-nasal and nasal vowels than healthy speakers, measuring distances between statistical models.

- Secondly, due to this velopharyngeal dysfunction, different formant frequency values between apnea and healthy voices can be expected. For instance, according to Hidalgo and Quilis (2002) the position of the third formant is related to the size of the velopharyngeal opening, and as a consequence, OSA-characteristic lowering of the velum might produce higher third formant frequencies in apnea patient’s voices. This performance was also studied by Robb et al. (1997), in which vocal tract acoustic resonance was evaluated in a group of OSA males. Statistically significant differences were found in formant frequency and bandwidth values between apnea and healthy groups. In particular, first and second formant values among the OSA group were generally lower than those in the control group. The lower values were attributed to an OSA-specific altered structure of upper airway and a greater vocal tract length. This finding has been reported in cephalometric studies indicating that the distance from the hyoid bone to the mandibular plane is significantly longer in patients with OSA compared to non-OSA individuals (Maltais, Carrier, Cormier & Series, 1991).

Limitations. Concerning the results of these perceptual studies (Monoson & Fox, 1987; Fox et al., 1989), some weak points should be noted:

Although it seems that there is a high agreement between judges, looking at the results of the intrajudge agreement, it does not seem consistent, they don’t seem so sure of their scores. To measure the interobserver agreement, correlations were used. The correlation found among judges for articulation and phonation judgments was $r = 0.92$, whereas for resonance it was slightly less, $r = 0.83$. For the intrajudge reliability calculated from the repeated utterances, the correlations obtained were $r = 0.77$ for articulation, $r = 0.72$ for phonation and $r = 0.48$ for resonance. In both cases, results for resonance were lower, reflecting the fact that resonance is the descriptor that is hardest to detect, especially taking into account nasality phenomena.

From the judges’ assessments shown in Table 2.3 extracted from their paper, it can be appreciated that the severity of each factor is quite low. They use a range from 1 to 7, and the highest achieved mean is 3 in phonation disorder for the apnea group. The descriptor phonation was the most predictive of the three descriptors to discriminate between normal and OSA patients. However, taken by itself, it was not enough to discriminate; the discrimination was more accurate when the three descriptors were combined. This result of phonation being the most discriminative

descriptor is surprising since phonation is a function attributed to the larynx, and the role of the larynx in OSA was believed to be minor (Guilleminault, Tilkian & Dement, 1976).

2.7.2. First Quantitative Studies via Formants (Fiz, Robb et al.)

Despite their weaknesses, the studies by Fox and Monoson (see above) were the source of acoustic studies such as (Fiz et al., 1993; Robb et al., 1997) and the use of voice-processing techniques for assessment OSA.

Fiz et al. (1993). The work by Fiz et al. (1993) is a good reference on vowel sounds analysis considering direct inspection on the spectral representation of the collected data. They studied a small population formed by 28 male subjects, 18 of them were OSA patients and 10 formed the healthy control group. In a sitting position subjects pronounced three times the sustained vowels /a/, /e/, /i/, /o/ and /u/ after listening C major note from a xylophone to keep the pitch uniform. The main conclusion of the study was that the vocalization of vowel /i/ has discriminative power to distinguish between OSA patients and healthy subjects. To reach that conclusion they analyzed the anthropometric parameters and spirometric values, the sleep characteristics and the maximum vowel harmonic frequency, mean frequency of harmonics and number of harmonics of the sustained vowels between OSA patients and healthy subjects. They only found significant differences at the 0.01 level for BMI, FVC¹, maximum frequency of harmonics of /e/ and /i/ and the number of harmonics of /i/. The OSA patients were more overweight, had lower FVC, and lower maximum frequency of harmonics of /e/ and /i/ and the number of harmonics of /i/. The differences on the acoustic variables could be due to the differences in weight between the groups, so they divided the OSA group in two; OSA with normal BMI and OSA overweight patients and found not significant differences between both groups at the 0.01 level. It has to be pointed out that the significant level that they chose is very restrictive, a *p*-value lower than 0.01 instead of the broadly used *p*-value lower than 0.05.

The acoustic analysis carried out by Fiz et al. (1993) showed that voice emissions of OSA patients have a narrower range of frequencies with lower maximum of harmonic frequency for utterances /e/ and /i/ and lower number of harmonics for /i/ with respect to healthy subjects. A clear advantage of Fiz et al.'s study over Fox and Monoson is that their results are based on a simple and objective acoustic analysis instead of subjective assessment of different experts.

A noticeable explanation for these acoustic results is that fat deposit in the tongue and pharyngeal walls and changes in pharyngeal shape typical in obese OSA patients could influence the mechanical properties of the anatomic structures determining in part the characteristics of the voice of these patients. After their acoustic study for differences between OSA obese and OSA non obese group, they conclude that the explanation for the differences found in /e/ and /i/ vowels were not explained by obesity *per se* but by the presence of the OSA syndrome.

¹The forced vital capacity (FVC) is the volume of air breathed out after the deepest inhalation from a maximally forced expiratory effort

Robb et al. (1997). Later, in 1997, Robb et al. published an interesting paper evaluating vocal tract acoustic resonances in a small male population formed by 10 OSA patients and 10 non OSA subjects. In this study, they did an acoustic analysis of formant frequencies and formant bandwidths of the sustained vowels /i/, /u/ and /a/. They found lower formant frequencies in the OSA group, which they attributed to a greater vocal tract length. The formant bandwidths were wider in the OSA group, which they interpreted to reflect greater vocal-tract damping in the OSA subjects. This work, together with (Fiz et al., 1993) is the origin of the research on the use of formants for OSA detection purposes. In the following paragraphs, we will delve deeper into this topic.

The rationale of Robb et al.'s paper is to study the vocalization behavior to have a quick, non-invasive approach to evaluating associated laryngeal and vocal tract airway function which are intimately related to both the speech process and OSA syndrome. According to Fiz's findings, OSA could be reflected on the acoustic waveform of the voice and for this purpose the study of formants can be a very useful tool as they reflect the sound resonating properties of the vocal tract during vowel production.

Robb et al. believed that the examination of vocal tract resonance of OSA and non-OSA subjects could be useful for OSA diagnosis for two reasons:

Lower formant frequencies. From cephalometric studies, it was known that the distance from the hyoid bone to the mandibular plane was significantly longer in OSA patients (Maltais et al., 1991) and therefore they have longer vocal tracts than non-OSA individuals. Thus, according to Eq. (2.1) and acoustic theory in general (K. N. Stevens, 1955), OSA subjects would have lower formants frequency values when compared to non-OSA individuals. Thus a reasonable hypothesis would be that formant frequencies of OSA patients are lower compared to non-OSA subjects owing to differences in vocal tract length.

Wider bandwidths. According to research on the sound damping, OSA subjects increase in both velar and pharyngeal compliance, i.e., softening of the named tissues (Brown, Bradley, Phillipson, Zamel & Hoffstein, 1985; Smith, Wise, Gold, Schwartz & Permutt, 1988). Recent studies have demonstrated a reduction in pharyngeal cross-sectional area and in upper airway muscle tone in patients with OSA. These findings suggest that the pharynx in such patients may be more compliant in OSA subjects than in non-OSA subjects even in the awake state. These findings indicate that OSA patients have increased pharyngeal compliance. The increase in tissue compliance of these structures would be expected to alter vocal tract elasticity. During vowel production, excessive tissue compliance would tend to increase the sound damping within the vocal tract. Therefore, to evaluate the vocal tract compliance formant bandwidth could be useful as they provide an estimate of sound damping within the vocal tract and also the rate of sound energy absorption. As it is believed considerable vocal-tract compliance and sound damping translates to a wide formant bandwidth, Robb et al. expected to find wider formant bandwidths on the OSA population due to the differences in vocal tract damping between OSA patients and non-OSA subjects.

In order to test these two hypothesis, [Robb et al. \(1997\)](#) recorded sustained vowels /a/, /u/ and /i/ of 10 male OSA patients and 10 non-OSA male subjects. This study was limited by the small size of the population and by the quality of the recordings, which were made with an audio-cassette deck, which may not have had a good resolution. Despite these limitations, they carried out a pioneer work on the analysis of formants for OSA diagnosis.

From their speech records, they calculated the first and second formants and their associated first and second bandwidths and compared between their two groups: OSA and non-OSA subjects. In all cases analyzed, the first and second formant values of the OSA population were lower than in the non-OSA population. These differences were significant for the first formant of the /i/ and /a/ with a respective p -value of 0.03 and 0.05 and for the second formant of the /i/ and /u/ with a p -value of 0.05 and 0.01 respectively. These results suggest that the most robust results are found for the sustained vowel /i/ as found also in ([Fiz et al., 1993](#)).

Focusing on the results for the bandwidths, it was found that in all cases the bandwidths were wider in the OSA population. Once again the /i/ analysis was always significant for the first and second bandwidth with p -values of 0.05. The /u/ showed also significant differences in the first and second bandwidths also with p -values of 0.05. The differences between OSA and non-OSA were also significant for the second bandwidth of the /a/ with a p -value of 0.01. From their statistical analysis, we can see that they have used parametric methods for the analysis of the formants and bandwidths between the two populations. However, as they do not mention that their data follows a normal distribution and they have samples of only 10 subjects in every group it would have been more accurate to have applied non-parametric methods on their contribution to have a more reliable results.

As they did not submit cephalometric data on their subjects, they could not confirm if the values achieved of lower formants on F_1 and F_2 on the OSA population were due to the presumable differences of the length of vocal tract lengths. The same occurred for their second hypothesis of finding wider formant bandwidths on the OSA population due to a greater vocal tract compliance, vocal tract size or to the acoustic energy generated at the larynx.

2.7.3. OSA and Snoring

There is a strong relationship between OSA and snoring. Snoring is the audible sign of increased upper airway resistance and is also an important clinical hallmark of OSA. The analogy between speech and snoring lies in the fact that both are generated in the vocal tract. In speech, phonation consists in the vibration of the vocal folds as the result of air passing through the narrowed glottis, whereas snoring originates from the vibratory activity of pharyngeal structures, not by the vocal folds. There are physiological similarities between snoring and speech, which have promoted the use of speech analysis for snoring analysis in the last years. Acoustic analysis of snoring and vocal sounds demonstrates that fundamental frequencies, harmonics, and formant features can be observed in both processes ([Beck, Odeh, Oliven & Gavriely, 1995](#); [Ng et al., 2008](#)).

Among the literature consulted, there is one that stands out especially, the physiological review of the acoustics of snoring by [Pevernagie, Aarts and De Meyer \(2010\)](#). It may be interesting to

take a look at table 2 of that contribution, where they analyzed in detail different papers about the acoustic findings in snoring produced at different anatomical sites. Some of the conclusions are that depending on the snoring type (soft palate, tongue base, combined, larynx ...) the fundamental frequencies of the snores take also different values. There is also a significant difference in acoustic characteristics between snorers that were administered with sedatives and natural snorers.

2.7.4. Further Developments

The work of [Fernández-Pozo et al. \(2009a\)](#) reinforces the existence of differences between acoustic features of OSA and non-OSA speakers. In this piece of research, generative statistical modeling techniques were applied to describe the acoustic space of apnea voices in a way similar to that being used in speaker recognition systems and using continuous speech, unlike most research in pathological voices analysis. Other related and relevant reference addressing the use of both connected and sustained speech for OSA detection can be found in ([J. L. Blanco-Murillo, Fernández-Pozo, Torre-Toledano, Caminero & López, 2011](#)).

These early contributions have driven more recent proposals for using acoustic analysis of speech in OSA detection, where different classification techniques have been proposed for Hebrew ([Goldshtein et al., 2011](#)) and Spanish ([Fernández-Pozo et al., 2009a](#)) languages. Results have been reported for different types of speech (i.e. sustained and/or continuous speech), ([J. L. Blanco-Murillo, Fernández-Pozo, Díaz-Pardo, Alvaro Sigüenza et al., 2009](#); [Goldshtein et al., 2011](#); [Montero Benavides et al., 2012](#); [Montero Benavides et al., 2016a](#); [J. L. Blanco-Murillo, Hernández-Gómez, Fernández-Pozo & Ramos, 2012](#)), different speech features ([Goldshtein et al., 2011](#); [Fernández-Pozo et al., 2009a](#); [Montero Benavides et al., 2014](#)), and modeling different linguistic units ([Montero Benavides et al., 2012](#)) as well.

Despite the remarkable results obtained from these recent proposals, there is an important issue that must be addressed to guide future progress in this field. Reported results have been based on small databases recorded under unrealistic clinical practice scenarios so it will be desirable to study these measures in a more realistic database with a greater number of OSA patients. The design of a bigger database was inspired by the literature review on OSA effects on patients' voices, and is briefly described in the following [chapter 3](#).

Considering these previous contributions and the results obtained in them, and having into account the great impact on the welfare of these patients when they are properly diagnosed and treated, it seems reasonable to develop the research on this topic to analyze if the study of the speech and other clinical variables of these patients can be helpful in the diagnosis of OSA, which is the focus of this Thesis.

2.8. Summary

Through this chapter we have introduced the OSA syndrome, we have taken a look at the history of OSA and we have seen that even Emperor Napoleon suffered from this condition. We have met Fat

Joe, a character from Dickens who has been an example of a typical OSA patient through medical literature. We have reviewed that OSA has a high prevalence, particularly among men between 30 and 60 years old. We have reported the associated symptoms and risk factors that predispose to suffer from this condition. We have seen that overweight brings the risk of fat deposits around the upper airway that may obstruct breathing, making the subject more prone to develop OSA. We have also explained the different treatment options, in particular, the nasal CPAP, as well as some of the dangerous consequences if OSA is not treated properly such as traffic accidents.

Furthermore, we have explained the clinical diagnosis, the gold standard being the PSG, which is very tiring and time-consuming. Some screening techniques for OSA such as the Berlin questionnaire have been introduced. Due to a lack of cost-effective, scalable solutions for screening, stratification, and diagnosis of sleep apnea, waiting lists for diagnosis currently take over a year and no standard protocol for stratification is available.

These waiting lists are too long for severe OSA patients due to comorbidities with high health risks and co-occurring disorders, such as hypertension, resulting in reduced quality of life and impaired work performance. Thus, novel solutions are needed to cope with this situation and to ensure that patients get treated swiftly according to their need. Furthermore, these one-year-long waiting list result from 10%–15% of the patients being diagnosed. Thus, reaching the 100% of the target would require 8.5–9 times the time and resources being spent nowadays. This unscalable, worrying situation could be improved once scalable and cost-effective technological solutions are introduced into the market.

This motivates the research line of this Thesis, namely the detection of OSA via speech analysis, based on statistical analysis.

Therefore, we have reviewed the evolution of speech and OSA, where we have seen that the evolution of the upper airway configuration made human beings more prone to suffer from OSA. Then, we have made an extensive review of the existing literature on speech and OSA, starting from the first perceptual studies that already showed the link between OSA and speech and motivated the use of speech technologies for helping clinicians in the diagnosis of OSA which is the research line that motivated this Thesis and several contributions.

CHAPTER 3

THE APNEA DATABASES

Once the Obstructive Sleep Apnea (OSA) has been presented in detail along with the state of the art on the relation between speech and OSA in the previous [chapter 2](#), in this chapter the apnea databases used in this Thesis are described. For the experiments of this Thesis, two apnea databases are used, Apnea Database 1.0 from ([Fernández-Pozo et al., 2008](#)) that is available since 2008 and Apnea Database 2.0, which is being recorded since 2012.

3.1. Introduction

The statistical analysis of the relation of OSA and speech relies on the availability of specific data, i.e., on speech samples from apnea patients and a control group organized in a database. Although the first references addressing the connection between OSA and speech date from the late 1980s, when the studies of Fox and Monoson were presented ([Monoson & Fox, 1987](#); [Fox et al., 1989](#)), there are no references about the design and collection of databases including records from healthy and OSA patients until 2008.

From 2007 until 2008, a Spanish database was collected at Hospital Clínico de Málaga in Spain ([Fernández-Pozo et al., 2008](#)), and another apnea database was collected at that time in Israel ([Zigel, Tarasiuk & Goldshtein, 2008](#)). More recently, in 2012 it was published that another Spanish database was being recorded at Hospital Txagorritxu de Vitoria ([Durán-Cantolla, 2012](#)). Since then, no reports on other databases focusing on the OSA syndrome have been identified in the literature consulted.

For the experiments carried out through this Thesis, we have used two databases, referred to as Apnea Database 1.0 and Apnea Database 2.0. The latter is bigger, has a slightly modified speech corpus, a different recording protocol, and several modifications. The collection of both databases was supervised by researchers from Universidad Politécnica de Madrid and clinicians from the respective hospitals. The person in charge of the medical issues was Medical Doctor José Alcázar Ramírez.

Table 3.1.: Sentences of Apnea Database 1.0 and Apnea Database 2.0, including phonetic transcription [International Phonetic Alphabet (IPA)] and English translation, separated according to melodic groups. From (Montero Benavides et al., 2014).

1.	Francia, Suiza y Hungría ‘fraN θja ‘suj θa i uŋ ‘gri a France, Switzerland and Hungary	ya hicieron causa común. ya j ‘θje roŋ ‘kaw sa ko ‘mun already made common cause	
2.	Julián no vio la manga roja xu ‘ljan no ‘βjo la ‘maŋ ga ‘ro xa Julián did not see the red sleeve	que ellos buscan, ke ‘e ʎoz ‘βus kan they seek,	en ningún almacén. en niŋ ‘gun al ma ‘θen in any store
3.	Juan no puso la taza rota xwan no ‘pu so la ‘ta θa ‘ro ta Juan did not put the broken cup	que tanto le gusta ke ‘taN to le ‘γus ta that he likes so much	en el aljibe. en el al ‘xi βe in the tank
4.	Miguel y Manu llamarán entre ocho y nueve y media. mi ‘γel i ‘ma nu ʎa ma ‘ran ‘eN tre ‘o tʃo i ‘nwe βe i ‘me ðja Miguel and Manu will call between eight and half past nine		

3.2. Apnea Database 1.0

By the time I started my Ph.D., the Apnea Database 1.0 had already been recorded, which let me do the first experiments. The design and collection of this first database was part of a previous work (Fernández-Pozo et al., 2008), where the set of criteria for its design and collection had been established and the speech corpus was specified. It was designed taking into account the distinctive patterns found in the literature for automatic OSA detection on speech records.

To make this Thesis self-contained, we review the main properties of the Apnea Database 1.0, which has already been described in (Fernández-Pozo et al., 2008; Fernández-Pozo, 2011, Chap. 5; Montero Benavides et al., 2012, Sec. 2; Blanco Murillo, 2013, Sec. 3.2.1). In particular, we highlight features that are important for the development of the research in Chapters 4 – 6. An extensive analysis on the speech corpus is included, as well as the selection process of speakers, the clinical variables and the definition of Apnea and Control groups. These decisions will affect the results of our experiments, and thus need to be addressed.

The Apnea Database 1.0 was designed by researchers from GAPS-UPM and collected in the Respiratory Department at Hospital Clínico de Málaga (Spain). Most of the speakers are from the same area in southern Spain, Andalusia, and were recorded just before going into the PSG test.

3.2.1. Speech Corpus

The speech corpus was designed to include phonetic contexts in which the anomalies in the speech of apnea patients, see section 2.7.1, can manifest themselves. Instead of recording sustained vowels, as in conventional studies on voice pathologies [see, for example (Parsa & Jamieson, 2001)], it was decided to use continuous speech in the form of four sentences containing as many relevant phonetic contexts as possible. Continuous speech was expected to offer more possibilities in the detection of features related to sleep apnea.

The analysis of specific traces within speech signals that can be related to OSA requires a special

effort to design and collect a consistent database. The design of our corpus was based on covering all the relevant linguistic/phonetic contexts in which physiological OSA-related peculiarities derived from the acoustics as described in the work by Fox et al. (1989). These peculiarities include the articulation, phonation and resonance anomalies revealed in the previous research review in chapter 2, so sentences were designed that include instances of the following specific phonetic and linguistic contexts:

Phonation: Continuous voiced sounds were included to explore irregular phonation patterns (as the lack of periodicity) that could be related to fatigue and respiratory troubles in apnea patients (Payne et al., 2006).

Resonance: To look at resonance anomalies, sentences were designed that allow intra-speaker variation measurements; that is, assessing differential voice features for each speaker, for instance to compare the degree of vowel nasalization in nasal and non-nasal contexts. Also, we included specific allophones of the vowel phonemes where may occur resonance anomalies referred to expected different formant values in OSA patients (Robb et al., 1997).

Articulation: In relation to articulatory anomalies, we collected voiced sounds affected by certain preceding phonemes that have their primary locus of articulation near the back of the oral cavity (specifically, guttural sounds as Spanish voiced and voiceless velar plosives /g/ and /k/, respectively). This anatomical region has been shown to exhibit physical anomalies in speakers suffering from OSA, so it is reasonable to suspect that different coarticulatory effects may occur in these phonemes with apnea and healthy speakers.

The speech corpus contains readings of four sentences in Spanish repeated three times by each speaker. The sentences used are listed in Table 3.1.

The first sentence, Table 3.1, was taken from the ALBAYZIN database, a standard phonetically balanced speech database for Spanish (Moreno et al., 1993). It was chosen because it contains some interesting allophones of the vowel phoneme /i/ ([i] and [j] in IPA notation), where different formant values for OSA patients may be expected. It also contains examples of guttural sounds as voiced and voiceless velar plosives /g/ and /k/, respectively.

The second and third sentences are both negative statements and have a similar grammatical and intonational structure. They are potentially useful for contrastive studies of vowels in similar positions in the sentence but in different linguistic contexts. Some examples of these contrastive pairs arise from comparing a nasal context, e.g., first /a/ of “manga roja” [ˈmaŋ ga ˈro xa], with a neutral context, e.g., first /a/ of “taza rota” [ˈta θa ˈro ta]. These contrastive analyses could be helpful to confirm whether indeed the voices of speakers with apnea have an abnormal nasality and display smaller intra-speaker differences between non-nasal and nasal vowels due to OSA-related velopharyngeal dysfunction (Fox et al., 1989).

The fourth sentence, see Table 3.1, has a single and relatively long melodic group containing mainly voiced sounds. The rationale for this fourth sentence is that apnea speakers usually show fatigue in the upper airway muscles due to respiratory problems. Therefore, this sentence may be

helpful to discover phonation anomalies during the sustained generation of voiced sounds, or the presence of differences in speech dynamics, and in particular pause patterns.

The four sentences contain 42 phones on average, and their durations are less than 3 seconds. They exhibit a correct grammatical structure, but do not have any meaning. Although the meaning of the speech is not important for our experiments, a meaning in the sentences might have affected the speakers' pronunciation during data collection. The task of reading aloud in front of a microphone can be strange and unnatural to most of the people. Also, they may be nervous because of the expectations of the screening result after analyzing their records.

As the reading of the 4 sentences 3 times might be too long for some patients, clinicians expressed the need of a shorter data collection procedure, which is why Apnea Database 1.0 also contains a recording of the sustained vowel /a/ in four repetitions, which could be useful for a text-independent approach which is broadly used when analyzing pathological voices. These differences are also related to the choice of which experiments are developed on which database.

In [chapter 4](#) of this Thesis, we use Apnea Database 1.0, which is the same data as in ([Fernández-Pozo, 2011](#)) for a different approach, namely OSA detection through Hidden Markov Models. Apnea Database 1.0 is also used in the experiments carried out also in [chapter 5](#), where we analyze different speech features for OSA detection purposes.

3.2.2. Speech Data Acquisition

Speech was recorded using a sampling frequency of 16 kHz in an acoustically isolated room under the supervision of an expert. The recording equipment consisted of a standard laptop computer equipped with a SP500 Plantronics headset microphone with A/D conversion and digital data exchange through a USB-port.

Speakers were requested to read the sentences listed in [Table 3.1](#) in a comfortable and natural mode in three satisfactory repetitions of each of the four sentences, observing the specific rhythmic structure previously indicated and under the assistant's supervision. Apart from the four sentences, speakers were requested to sustain the vowel /a/ for at least one second, providing four repetitions. The assistant assured that speakers pronounced the correct words without making mistakes, made sure there were no background noises, artifacts between or after the uttered sentence, and checked that there was neither distortion nor saturation in the speech signal.

The speech material for the apnea group was recorded and collected in two different sessions: one just before being diagnosed (at the same time the control group was recorded) and the other after several months under CPAP treatment. This will allow for future studies of the progress of apnea voice characteristics for a particular patient before and after treatment. In this Thesis we have only used the first session of the apnea patients and the single session of the healthy subjects. The total length of the speech data is over 1.4 h for the train database and 45 min for the test database.

Figure 3.1.: Recording scheme for Apnea Database 1.0

3.2.3. Collection Procedure for Apnea Database 1.0

The first database was recorded in the Respiratory Department at Hospital Clínico de Málaga from 2007 to 2008 under the supervision of an assistant and using the equipment provided by a team of researchers from Universidad Politécnica de Madrid.

An important characteristic of Apnea Database 1.0 is that there are people without any symptoms which form part of the control group. It is important to note that they do not snore, which is one of the most common symptoms of OSA population. This fact will be one of the main differences between Apnea Database 1.0 and Apnea Database 2.0. The Apnea Database 1.0 was built as an effort to adjust and control the clinical variables of age and weight between the Control and the OSA group. Even after trying to build the database without any differences in these variables, when we later tested statistically, we found differences among the control subjects and the apnea patients.

On [Figure 3.1](#) we have summarized the recording protocol of Apnea Database 1.0. The OSA speakers were selected from the list of patients visiting the Pneumology Department of this Hospital. Patients were people who had experienced recurrent apnea episodes, often referred by their bed partners. When they visited their family doctor they were suspected to suffer from apnea disease so they were derived to the sleep unit to do a PSG test before having a final diagnosis and an adequate treatment. On the day of the arrival for this test, patients were selected for the database according to the clinician's experience and additional criteria (e.g. similar age, BMI, etc.), aiming at homogeneity in the apnea and control populations. The selected ones were taken to a separate room with the recording equipment and adequate acoustic conditions, where, after obtaining the patient's consent, his voice was recorded as explained above ([section 3.2.2](#)).

The Control group was formed by healthy population who did not suffer from any disorder. This peculiarity is very important when we analyze the results of the experiments carried out on this database.

After the speech material had been recorded, the patients were taken for the overnight PSG test

Figure 3.2.: Apnea Database 1.0.

for the final diagnosis. According to this diagnosis, the speaker's recordings were assigned to the OSA group if the AHI was over 30.

Patients with severe OSA were facilitated with a suitable CPAP treatment. After three months of treatment, patients took part in a second recording session, which was meant to provide additional information to evaluate any alternations introduced by the CPAP treatment. In this Thesis, we are only interested in the detection of OSA and do not use this information.

Regarding demographic characteristics, all speakers lived in the same area of Andalusia and exhibited a similar dialect, which makes Apnea Database 1.0 homogeneous in terms of dialect and helps to control the variability in the data.

3.2.4. Clinical Parameters of Apnea Database 1.0

The Apnea Database 1.0 contains the readings of 82 male subjects; 42 of them suffer from severe sleep apnea ($AHI > 30$). The rest of the speakers, which are asymptomatic population and can be considered as healthy population form the control group.

Thus, the database is designed for learning to discriminate severe cases of OSA from healthy subjects, which is exactly the need in medical practice, namely to identify subjects in major need for treatment [See also [Figure 3.2](#)]. In Apnea Database 1.0, the two classes OSA ($AHI > 30$) and healthy are labeled as a dichotomic variable 0 or 1, which means that, unfortunately, the continuous value of the AHI, which measures the severity of the disease, is not available.

As mentioned, subjects were selected in both groups aiming at similar physical characteristics in both groups. However, there are still significant statistical differences between both groups. [Table 3.2](#) shows the clinical variables for the two groups OSA and Control and also for all subjects together. In order to study the statistical differences of medians between the Apnea and Control group, a common Wilcoxon test has been carried out at the 95% confidence value. We have tested the null hypothesis of equal medians between groups and found that there are statistically significant differences for age and BMI in the first apnea database although the doctors tried to

Table 3.2.: Summary on the subjects included in the Apnea Database 1.0, [median $\pm$ standard deviation (SD) of the variables collected along with the speech records]. The last column shows the p-values estimated by the non-parametric Wilcoxon rank-sum test applied to each variable comparing the two groups Control and Severe OSA. From (Blanco Murillo, 2013).

	All		Control		OSA		p-value
	mean	$\pm$ SD	mean	$\pm$ SD	mean	$\pm$ SD	
age (y.o.)	46.0	10.5	42.2	8.8	49.9	10.7	0.001
weight (kg)	89.1	17.2	81.7	13.5	96.8	17.4	<0.001
height (cm)	173.6	8.1	176.3	6.3	170.9	8.8	0.002
BMI (kg/m ²)	29.6	5.8	26.2	3.9	33.2	5.5	<0.001
cervical perimeter (cm)	42.7	4.0	40.1	2.6	45.1	3.6	<0.001
group cardinality	80		40		40		
AHI	-		< 10		> 30		

Table 3.3.: Pairwise correlations for all variables included in Apnea Database 1.0 speakers' profiles: age, height, weight, body mass index, cervical perimeter, including the binary decision on the OSA condition based on the AHI. From (Blanco Murillo, 2013).

	OSA	age	weight	height	BMI	c. perim.
OSA	1.000	0.372	0.442	-0.338	0.597	0.635
age	0.372	1.000	0.124	-0.526	0.385	0.320
weight	0.442	0.124	1.000	0.249	0.874	0.799
height	-0.338	-0.526	0.249	1.000	-0.246	-0.073
BMI	0.597	0.385	0.874	-0.246	1.000	0.847
c. perim.	0.635	0.320	0.799	-0.073	0.847	1.000

select the subjects controlled for these variables. We also found statistically significant differences for the other clinical variables tested, weight, height and cervical perimeter.

In order to study the relationship between variables, we have done a correlation analysis between the clinical variables and the AHI-based binary grouping of the speakers "OSA" (coded as 0 or 1). We have calculated the pair-wise Pearson's correlation coefficient among them. These statistics should highlight those variables that are related one to the other, pointing out those which might be more informative when coming to the detection of the OSA syndrome. The correlations are collected in [Table 3.3](#).

The most relevant correlation identified is the one between weight and BMI, while the correlation between height and BMI is fairly low. This is not surprising since the body mass index is proportional to the individual's weight and inversely proportional to the square of the height. Far more interesting is the large correlation between BMI and cervical perimeter, which is most probably originated by the accumulation of fat deposits around the neck.

The same argument may be used to explain the correlation between weight and cervical perimeter, though the BMI is usually preferred over weight due to the normalization introduced by the squared height.

Finally, the correlation between OSA and the cervical perimeter, as well as with the BMI, are neither surprising according to the previous correlation coefficient values and the literature review in [chapter 2](#). Fat deposits may be found in the upper airway tissues, particularly in severe OSA cases, predisposing the collapse of the pharynx and the appearance of recurrent apnea episodes due to the induced narrowing in the pharynx cross-section.

3.3. Apnea Database 2.0

With the experience gained from the first experiments carried out on Apnea Database 1.0, the design of a new database was developed. This new database, called Apnea Database 2.0 in this Thesis, was recorded in a new campaign started in 2010 at a different hospital, the Hospital Quirón at Málaga (Spain) but under the supervision of the same medical doctor, Dr. Alcázar Ramírez and GAPS-UPM. In the following, we will go in detail through the differences with respect to Apnea Database 1.0, which are also collected in [Table 3.4](#).

Speech Corpus. It had turned out that for estimating certain parameters of the voice characteristics, sustained vowels are more suitable than continuous speech because the signal is more stable and independent of sources of variability like intonation and differences in dialect. Thus,

Figure 3.3.: Characteristics of Apnea Database 2.0 (as opposed to Apnea Database 1.0, [Figure 3.2](#)).

Figure 3.4.: Recording scheme for Apnea Database 2.0

for the recording of Apnea Database 2.0, in addition to the four sentences and the sustained /a/, the other four sustained vowels /e/, /i/, /o/ and /u/ were also recorded. This opens the way for studying the formant frequencies of sustained vowels in [chapter 6](#).

Moreover, the speakers were asked to read out their phrases and vowels in two different positions: sitting and lying on their back (supine position). This was meant to facilitate the analysis of differences related to the airway configuration, which should vary from one position to the other. According to the previous literature review, the lack of muscular tonus in the upper-airway muscles and a compliant pharynx would explain most of apnea episodes during the night, as well as symptoms such as characteristic snoring. Therefore, reproducing the same conditions during recording (i.e., having the patient in the same position) may help in detecting OSA characteristic patterns ([Durán-Cantolla, 2012](#)). However, recording in two different positions makes the recording process much longer and patients got tired. Due to this fact, there are only some recordings of patients in both positions.

Speech Data Acquisition. Speech was recorded at a sampling frequency of 50 kHz and encoded in 16 bits. Recording equipment was a standard laptop computer equipped with a SP500 Plantronics headset microphone that includes integrated A/D conversion and digital data exchange through USB-port. Afterwards, speech records were down-sampled to 16 kHz and converted to wav files for acoustical analysis.

Selection Procedure. On [Figure 3.4](#) we have summarized the recording protocol of Apnea Database 2.0. Speaker selection and recording took place with patients at the Sleep Disorders Unit of Hospital Quirón at Málaga (Spain). Subjects referred typical OSA symptoms during the preliminary interview with the family doctor, such as excessive daytime sleepiness, snoring, choking during sleep, or somnolent driving. The family doctor sent them to the sleep unit for a final diagnosis of a suspected sleep disorder. Once in the sleep unit all subjects underwent a detailed

interview concerning their clinical history and a thorough physical examination focusing on sleep-related symptoms, associated conditions, comorbidities and anthropometric measures. Subjects' weight and height were recorded while wearing light clothes. Body mass index (BMI) was calculated as the ratio of body weight (in kilograms) and the square of height (in meters), and included in the clinical history along with the cervical perimeter (in centimeters), measured at the level of cricothyroid membrane.

The recorded subjects included also females and more than one dialect. Patients were recorded as they were derived to the Sleep Unit, which means that none of them was really healthy because even if not diagnosed with OSA, they were likely to suffer from some other sleep disorder. This means that when using the Apnea Database 2.0 with an OSA group and such a non-OSA group, we have to take into account this information for our conclusions. The conclusions are different opposed to the situation of Apnea Database 1.0, where we have a healthy control group.

Clinical variables. Speaker's profiles include all variables present in Apnea Database 1.0, along with additional information from the clinical interview to gain a better insight into the upper airway condition. The information collected and stored in these profiles should prove useful in the posterior grouping of speakers. In This Thesis, we will only use the records of the male population. Though, it is expected that the female records will be used in the near future to analyze the differences between male and female subjects while dealing with the automatic detection of OSA syndrome based on speech cues.

3.3.1. Resulting Database Size and Properties

By the end of 2013, the speech of 403 subjects under evaluation and suspected to suffer from sleep disordered breathing has been recorded and included into Apnea Database 2.0. At this stage, the research of [chapter 6](#) of this Thesis was conducted.

3.3.1.1. Andalusian Male Sub-Database

For the experiments carried out in [chapter 6](#) of this Thesis, we will focus only on the male population, which are 281 subjects. Among these, we focused on 241 male subjects from the same region in Spain and with a homogeneous dialect, Andalusian, which allowed us to minimize the impact of cross-dialectal variability. Among these 241 subjects, the AHI ranges from 0 to 87, with a mean of 22.4 and a standard deviation of 19.3. A statistical summary of some more clinical variables is given in [Table 3.5](#).

3.3.2. Validation of Apnea Database 2.0

3.3.2.1. Reliability of clinical measurements collection

To validate our database on a larger scale, we compared the distribution of our data in the Apnea Database 2.0 with that of a similar but larger database collected in Greece, which did not contain

Table 3.4.: Comparison Apnea Database 1.0 vs. Apnea Database 2.0.

	Apnea Database 1.0	Apnea Database 2.0
Size	80 subjects (~ 50% OSA)	403 subjects (~ 50% OSA)
Speakers	males	males and females
Speech Corpus sustained vowels	4 sentences with 3 repetitions /a/	4 sentences with 3 repetitions /a/, /e/, /i/, /o/ and /u/
Clinical Variables	attend to have a prior controlled environment for the variables age and BMI. Age limited to middle-age subjects.	no prior control on any variable, patients are recorded as they arrive at the sleep unit. These are no healthy patients so we will refer to them as <i>non</i> -OSA patients. Most of them suffer from some other disorder than OSA.
AHI	coded as 0 for healthy subjects and 1 to severe OSA patients.	AHI value collected for every patient. For some experiments re-coded as 0 for non-OSA, 1 for mild OSA patients and 2 for severe OSA patients.
Position at recording	sitting	sitting and lying on back
		Additional information (when-ever available) related to snoring, smoking and the upper-airway condition.

Table 3.5.: Descriptive Statistics on the 241 Andalusian Male Subjects.

Clinical Variable	Mean ($n = 241$)	SD	Range
AHI	22.4	19.3	0.0–87.0
Weight (kg)	92.8	17.4	61.0–162.0
Height (cm)	175.5	7.2	159.0–197.0
BMI (kg/m ²)	30.1	5.1	21.4–52.3
Age (years)	47.6	11.6	21.0–78.0
Cervical perimeter (cm)	42.5	3.2	34.0–53.0

Abbreviation: SD, standard deviation.

Table 3.6.: Comparison of two apnea databases

	Greek database (Bouloukaki et al., 2011)		Spanish database (Apnea Database 2.0)	
Population	n	percentage	n	percentage
Males	2061	76.62%	281	71.14%
Females	629	23.38%	114	28.86%
Apnea patients	2130	79.18%	189	47.85%
Healthy patients	560	20,82%	206	52,15%
Total	2690	100%	395	100%
Clinical variables	Mean $\pm$ SD		Mean $\pm$ SD	
Age (years)	50.7 $\pm$ 12.2		49 $\pm$ 11.9	
BMI (kg/m ²)	33.3 $\pm$ 6.7		30.1 $\pm$ 5.7	
Cervical Perimeter (cm)	41.9 $\pm$ 3.6		40.9 $\pm$ 4	
AHI	34.4 $\pm$ 23.8		20.2 $\pm$ 18.8	

Abbreviation SD: standard deviation

any speech records ([Bouloukaki et al., 2011](#)). The population of our database, which was collected in a sleep unit in the south of Spain, and the Greek population collected in a sleep unit in Greece, shall share remarkable factors as both are Mediterranean countries and the phenotype and the diet are similar. Nonetheless, in ([Bouloukaki et al., 2011](#)) authors stress their concern that the database they collected may not be regarded as representative of the general population. Nonetheless, it suits the comparison with our database as both were collected in sleep units under similar conditions; and we merely focus on their demographic analysis.

In order to have a fair comparison, we used the same definition they used for the OSA group (AHI over 15) and Healthy (otherwise). This is a commonly recommended threshold value according to [The Report of an American Academy of Sleep Medicine Task Force \(1999\)](#), though it varies among clinicians.

The percentage of male population for the Greek database was 76,62%, while for the Spanish database it was 71,14%. Apnea prevalence also varied from one database to another. For the Greek it was 79,18%, while for the Spanish database this was much lower, down to 47,85%. The total population for the Greek database was 2690, however the Spanish database by the time this Thesis was written was formed by 403 subjects.

Table 3.6 provides some numbers on the two databases (Greek and Spanish). Both populations exhibit similar ratios in terms of male versus female; The most remarkable difference between both databases are in terms of prevalence, being much higher on the Greek database. Such high prevalence was already pointed out by [Bouloukaki et al. \(2011\)](#), and clearly exceeds the prevalence estimated for OSA among an arbitrary group of people. Anyhow, based on these numbers it is clear that among people that decide to visit a physician, OSA prevalence is very high.

In [Table 3.6](#), we also present some statistics on clinical measures extracted from both databases, including age, BMI, cervical perimeter, and AHI. Clinical parameters are very similar in two databases, except for AHI, which is related to the apnea prevalence as already explained. If we take a look at the BMI we see that patients in the Spanish database have lower average BMI, 30.1, whereas in the Greek database the average BMI is 33.3, though still, 1-sigma intervals around the means do overlap.

After having compared these databases, we can conclude that our Apnea Database 2.0 is very similar in clinical parameters to other bigger databases found in the literature. Later, in [chapter 6](#) we will also compare the speech features, formants and bandwidths to another Spanish speech database to be sure of the reliability of our Apnea Database 2.0.

3.4. Summary

In this Chapter, we have detailed the databases that will be used for the experiments developed in this Thesis. We have explained the importance of the speech corpus recorded and also the main differences between Apnea Database 1.0 and Apnea Database 2.0, which are summarized in [Table 3.4](#).

These differences are also related to the choice of which experiments are developed on which database. In [chapter 4](#) of this Thesis, we use Apnea Database 1.0, which is the same data as in ([Fernández-Pozo, 2011](#)) for a different approach, namely OSA detection through Hidden Markov Models. This way we can compare the new results obtained with those from the baseline system in order to discuss the improvements or losses observed, working on the same database. Apnea Database 1.0 is also used in the experiments carried out also in [chapter 5](#), where we analyze different speech features for OSA detection purposes. In [chapter 6](#), in contrast, Apnea Database 2.0 is used, which became available later in 2013 and which made available sustained vowels for the study of formants and bandwidths.

CHAPTER 4

HIDDEN MARKOV MODELS

A TEXT-DEPENDENT APPROACH

OSA classification using speech signals has been previously assessed through text-independent approaches (J. L. Blanco-Murillo, Fernández-Pozo, López-Gonzalo & Hernández-Gómez, 2008; J. L. Blanco-Murillo, Fernández-Pozo, Díaz-Pardo, Hernández et al., 2009; J. L. Blanco-Murillo, Fernández-Pozo, Díaz-Pardo, Alvaro Sigüenza et al., 2009; Fernández-Pozo et al., 2009a, 2009b; Fernández-Pozo et al., 2010). In this Chapter, we will go a step forward and explore text-dependent acoustic models, in particular, hidden Markov models (HMMs), to characterize and discriminate continuous speech from severe OSA and healthy speakers.

The databases used in this Thesis contain four pre-established sentences that are read by all the speakers (chapter 3). This scenario represents a similar environment as text-dependent speaker recognition (Hébert, 2008). HMMs are widely used in both text-dependent speaker recognition and speech recognition. The key idea of a HMM is to model a given observation sequence with a model that runs through an (unobserved) sequence of states with certain transition probabilities from one state to another. The sequence of states of a HMM models the sequence of linguistic units, i.e., phonemes, syllables, words, or whole sentences. Thus, a HMM represents a more precise model of the sequence than a Gaussian mixture model (GMM), which could be considered a single-state HMM.

Certain parameters of a HMM, such as the number of states or the parametrization of the probability distributions, are fixed in advance. The other parameters, like the transition probabilities between states and the probability distributions of the states, are “learned” from training data.

In text-dependent speaker recognition, each speaker is represented by his or her own HMM. In our case, there will be one HMM representing the severe OSA group of speakers and another one representing the control group. Once these HMMs have been trained, they can be applied to an unknown signal, whereupon each HMM yields a probability score, the highest of which identifies the class of the signal (apnea/healthy).

In this Chapter, we design two different architectures of HMMs for the discrimination of speech samples from severe OSA patients and healthy subjects. Both outperform the baseline system (Fernández-Pozo et al., 2009a), which relies on text-independent GMMs.

4.1. From Speech Signal to a Sequence of Feature Vectors (MFCCs)

In the outline of this chapter, we have already given a rough idea on how classifying speech signals with HMMs should work. HMMs work on a sequence of observations that is discrete in time, with all observations preferably containing meaningful information on what is said and how it is said. Markov chains only work when the states are discrete, and text satisfies this property since there are a finite number of characters in a sequence.

4.1.1. Motivation

For the HMMs to work on audio recordings, which represent a continuous signal in time, namely the microphone voltage sampled at 16 kHz, we need a suitable speech parametrization or feature extraction. Human perception works on a time scale of some 10 milliseconds, but it does distinguish precisely the different frequencies that are present in the signal. To get access to the frequencies, a Fourier transform would be the straightforward way, but by performing a standard Fourier transform on the whole signal, one would lose all notion of temporal resolution.

Thus, the observation sequence for our HMMs should be something between the raw speech signal and a full Fourier transform. Meaningful feature vectors should vary on the timescale of human perception (i.e., 10 ms – 50 ms) and contain the information of shorter timescales in some transformed form. It should capture pitch, tone, and, in the case of speech, information on the phoneme pronounced. All this information is contained in the short-time spectrum, i.e., the Fourier transform of a sliding time window of, say, 50 ms length. In fact, the short-time power spectrum reflects the shape of the vocal tract during speech [see [section 2.6](#)]. Thus, the basic idea is to cut the signal into time windows and to perform a frequency analysis within each of these windows. These spectra may be further processed and compressed for the specific task to be solved, which yields the desired sequence of feature vectors for the HMM to work on.

4.1.2. MFCC Features

More specifically, the features that were selected for our experiments are the mel-frequency cepstral coefficients (MFCCs). The MFCC parametrization is designed relying on human perception and provides relatively independent coefficients and high discrimination between sounds, which is why it is widely used for automatic speech recognition. In this Thesis, we use the HTK implementation ([S. Young et al., 2009](#)) of a mel-filter bank of 20 filters and 12 MFCCs without velocity or acceleration coefficients. In the following, we describe in detail what this means, how it is done, and why it was done this way.

The feature-extraction process is summarized in the block diagram in [Figure 4.1](#), from the original time series of the recorded signal $s[n]$, to the final sequence of feature vectors $c_i[m]$. Good references on the MFCC feature extraction are ([Bimbot et al., 2004](#), Section 2) and ([S. Young et al., 2009](#)).

Figure 4.1.: Block diagram of the workflow from speech signal to the sequence of feature vectors described in the text. Control parameters (arrows from above) are the length of the Hamming window, the overlap of the Hamming windows, the number N_{chan} of mel filters, and the number of cepstrum coefficients. Figure adapted from (Blanco Murillo, 2013).

Pre-Processing. The input signal $s[n]$ is sampled at 16 kHz, i.e., the sound amplitude sampled at time steps of 62.5 ms labeled with the index n . The pre-processing step applies a low-cut filter and normalizes the amplitude. This removes low-frequency noise, such as direct current (DC) components or electric hum (50 Hz) from the pre-processed signal $x[n]$.

Windowing. The windowing step cuts the signal into overlapping pieces, the so-called frames. Mathematically is a multiplication of the signal with a function that is only non-zero within the given time window. The Fourier transform in the next step is then the convolution of the spectrum of the signal with the Fourier transform of the window function. Thus, the shape of the window will affect the quality of the spectrogram. We choose the Hamming window,

$$w(n) = \begin{cases} \alpha - \beta \cos\left(\frac{2\pi n}{N-1}\right), & \alpha = \frac{25}{46}, \beta = 1 - \alpha, \text{ for } 0 \leq n < N \\ 0, & \text{otherwise,} \end{cases} \quad (4.1)$$

which has been proven a good compromise of dynamic range and resolution. The length of the window, T_{win} , and the overlap of windows, T_{rate} , are parameters to be chosen. Typical values are $T_{\text{win}} = 20$ ms and $T_{\text{rate}} = 10$ ms. Longer T_{win} increase frequency resolution at the cost of temporal resolution.

Fourier Transform. Each time window is labeled with an index i , while within the time window the time steps are indexed by n . Each of these windows was Fourier transformed, where we extended the time window to the closest power of two in order to make use of the fast Fourier transform (FFT) algorithm. The fast Fourier transform (FFT) of each time window transforms from time steps n to frequency indices k . The modulus squared of the complex amplitudes is taken, which is also referred to as the energy of the signal.

Mel Filter Bank. While the Fourier transform is a bijective transformation, taking the modulus (squared) maps complex to real numbers and reduces the dimension of the feature vector by a factor 2. Still, we have feature vectors of dimension $\frac{1}{2}T_{\text{win}}f_{\text{samp}} = 160$ and need a further reduction of dimension and compression of information. As is common practice in speech and speaker recognition, we apply a filter bank to the power spectrum from the FFT step, which yields

Figure 4.2.: Mel filter bank. Panel (a) shows the relation of frequency f and perceived frequency, “mel scale” of Eq. (4.2). A mel filter bank is constructed from $N_{\text{chan}} = 20$ triangular filters distributed uniformly on the mel scale [panel (b)], which, according to the mel mapping results in the non-uniform filter bank in frequency space [panel(c)].

one number per filter, namely the dot product of the filter and the spectrum. *A priori*, many different types of filter banks are conceivable. For example, the frequency range may be divided into N intervals of equal length with one filter corresponding to each interval, which would be a uniform coarse-graining of the spectrum.

In our work, we chose the mel-filter bank with $N_{\text{chan}} = 20$ filters. The mel scale maps the physical frequency f to the perceptual scale of pitches perceived by human hearing,

$$m(f) = A \ln(1 + f/B), \quad A = 1127 \text{ s}^{-1}, \quad B = 700 \text{ s}^{-1}, \quad (4.2)$$

see (S. S. Stevens, Volkman & Newman, 1937) and Figure 4.2(a). The decreasing slope of the curve means that, with increasing frequency, human hearing distinguishes frequencies with less sensitivity. The mel filter bank is uniform in mel space, mimicking the human perception of frequencies [Figure 4.2(b)]. Mapped back to frequency space Figure 4.2(c), the resulting warped

Figure 4.3.: MFCC feature vector layout. The c_i are the MFCCs, ranging from 1 to $N = N_{\text{cep}} = 12$ or 10 in our computations; E is the logarithm of the energy of the frame (i.e., determined right after the windowing step), c_0 is the 0-th MFCC (the sum of logarithms of the energies in each mel channel). The operator d indicates the difference $dc_i[m] = c_i[m+1] - c_i[m-1]$ (velocity, or Δ), and the operator D indicates the difference of differences, i.e., $Dc_i[m] = dc_i[m+1] - dc_i[m-1]$ (acceleration, or $\Delta\text{-}\Delta$).

filter bank is coarser at high frequencies than at low frequencies. The output of each mel filter represents the loudness of the frame at the frequency sampled by the filter.

Note that it is not clear a priori, that for the OSA detection problem it is adequate to mimic the perception of human hearing. However, since in (Fox et al., 1989) skilled judges were able to identify several abnormal patterns on OSA speech records, we may assume that such a configuration is at least a sensible one.

Extracting the Cepstral Coefficients. The last steps are taking the logarithm of each of the outputs of the filters, which, again, reflects the human perception of loudness. The logarithms of the outputs from all mel filters from a given frame are kept in a vector, which is transformed according to a discrete cosine transform (DCT), which yields the final *mel-frequency cepstral coefficients*¹ (MFCCs) c_i , $0 \leq i < N_{\text{chan}}$. The first DCT coefficient is the sum of the logarithms of the energies of all filters, while the highest DCT coefficients contain differences of neighboring mel filters. This last step helps to de-correlate the features from the overlapping mel filters.

It has proven helpful to discard the highest MFCCs, which do not contain relevant information and to keep only the c_i with $1 \leq i \leq N_{\text{cep}}$, typical values being $N_{\text{chan}} = 26$ and $N_{\text{cep}} = 12$ (S. Young et al., 2009). We will use $N_{\text{chan}} = 20$ and $N_{\text{cep}} = 12$, as it was shown by Blanco Murillo (2013) that this is a very good choice for OSA detection in with the GMM baseline system.

Apart from these direct MFCCs, in speech and speaker recognition usually dynamic information is included in the feature vector, i.e., differences of MFCCs from one frame to another referred to as Δ or velocity. Analogously, $\Delta\text{-}\Delta$, or acceleration, coefficients may be included, see Figure 4.3 for some common choices.

We did some experiments trying the following configurations:

- **MFCC_0_D_A** with $N_{\text{cep}} = 10$, which includes the 0-th cepstral coefficient with delta and

¹The term ‘cepstrum’ refers to the DCT. The word stems from ‘spectrum’, i.e. (forward) Fourier transform, with the first letters spelled backward, i.e. an inverse Fourier transform.

acceleration. This configuration was previously used as default configuration in our research group for speech recognition. Feature-vector dimension $M = 33$.

- **MFCC_E_D_A** with $N_{\text{cep}} = 12$, which includes the energy E instead of c_0 , again with delta and acceleration. Feature-vector dimension $M = 39$
- **MFCC** with $N_{\text{cep}} = 12$, i.e., the plain mel-frequency cepstral coefficients without any energy measure or dynamic information. Feature-vector dimension $M = 12$.

Surprisingly the best results were achieved when we used only the MFCC without velocity or acceleration. The same behavior has been found previously in (Blanco Murillo, 2013).

In conclusion, we have transformed the signal $s[n] \in \mathbb{R}$, n sampled at steps of $62.5 \mu\text{s}$ ($\sim 16 \text{ kHz}$) to $c_i \in \mathbb{R}^M$ (M being the size of the feature vector) sampled at $T_{\text{rate}} \approx 20 \text{ ms}$.

4.2. On HMMs and the Algorithms Used

In this section, we focus on the use of HMMs. The basic theory of HMMs was first introduced by Baum and colleagues in a series of articles between 1966 and 1972 (Baum & Petrie, 1966; Baum, 1972). HMMs have a solid statistical basis, efficient learning algorithms and a great adaptability to the changing conditions of the voice or the transmission channel. HMMs are widely used in classification problems, In particular, for speech recognition, as well as in speaker recognition tasks (Rabiner & Gold, 1975; Rabiner, 1989).

A HMM is a doubly stochastic process consisting of an unobservable hidden stochastic process (transitions between states) and an observable stochastic process whose output is the sequence of feature vectors [see Figure 4.4]. The hidden process is a finite state machine, which changes state at discrete time steps. In each time interval, an observation is drawn from the state's probability distribution. In our case, the observations are the vectors of $M = 12$ MFCCs described in section 4.1 above, and the probability distributions of the states are modeled by GMMs. These probability distributions, as well as the transition probabilities from one state to another have to be “learned” from observation data, while the sequence of underlying states of the system remains unknown.

The three basic problems for HMMs can be stated as follows (Rabiner, 1989):

1. Given a HMM and an observation sequence, determine the probability of this sequence to be generated from the model.
2. Given a HMM and an observation sequence, determine the underlying state sequence that “explains” how the observation sequence was generated.
3. Adjust the parameter of a HMM to given observation data to maximize the probability of the observation data to be generated from the HMM (training).

Typically, several HMM are set up and trained (problem 3) to characterize all recognition units of interest (classes): phonemes, words, phrases. At deploy time (testing), the HMMs are applied to an unknown signal, and the HMM with the highest probability score (problem 1, typically

Figure 4.4.: Representation of a (left-right) HMM as typically used in speech recognition. The hidden states s_i are connected with transition probabilities a_{ij} to other states s_j (including the remain probability s_{ii}). Each state is has a probability distribution (sketched in the middle row) to generate observations y_k .

making use of problem 2) leads to a decision, or labeling of the input signal, which could be a phoneme, a word, or the classification ‘severe OSA patient’ or ‘healthy subject’ in our case. The whole operation scheme of the HMM approach is sketched in [Figure 4.5](#).

Training. For each class, the topology of the HMM has to be chosen. This includes the number of states, directions of transitions, and the parametrization of the probability distributions [for an example, see [Figure 4.4](#)]. For a single phoneme or word, the number of states would be much smaller than for a whole sentence. The probability distribution of each state lives in the ($M = 12$)-dimensional space of MFCCs and is represented by a GMM. The parameters of the HMM are trained taking several repetitions of the speech sample. A lot of training data is needed to achieve a good estimate of the model parameters. The training process starts with an initial alignment of parameters (transition probabilities and density functions at each state) and a re-estimation using the Baum-Welch algorithm. For each possible class, a HMM is adapted to the respective training data.

Classification of an unknown signal. To assess the probability that the signal could have been generated by a given HMM, the Viterbi algorithm firstly finds the best state sequence for a given observation sequence [problem 2, explained in ([Rabiner, 1989](#), at Eq. (30))]. That means, the maximum likelihood path is used to determine the probability of that sequence, ([Rabiner, 1989](#), at Eq. (114))].

If there are only two classes to distinguish, it is straightforward to apply a threshold to the logarithm of the likelihood ratio, which allows shifting the working point towards low false-alarm

Figure 4.5.: HMM operation scheme.

rates (also known as “false positive” or type-I error) or low miss rates (also known as “false negative” or type-II error), according to the requirements of the classification task at hand. For a more detailed description see, for example (Blanco Murillo, 2013, Sec. 3.3.6).

4.3. Designing HMM Experiments for Our Classification Problem

The baseline system for our experiments is the text-independent GMM-based system described in (Fernández-Pozo et al., 2009a), which was trained using the BECARs open-source toolkit (Blouet et al., 2004). For the training and evaluation of HMMs in the present study were used the Hidden Markov Model Toolkit (HTK) (S. Young et al., 2009).

Since the Apnea Database 1.0, the one available at the time these experiments were carried out, is limited to only very few speakers [see section 3.2], we make use of the leave-one-out scheme to get the best estimate of the performance and generalizability of our system. “Leave one out” means that the HMMs for both classes (severe OSA and healthy) are trained with the whole database except one speaker. Then, the model is tested with this one speaker, and the probability ratio $p(\text{OSA})/p(\text{healthy})$ is recorded see Figure 4.6. The leave-one-out procedure is applied to all healthy subjects, which gives the false-alarm rate as a function of the threshold applied to the probability ratio. In the same manner, the procedure is applied to all OSA subjects, which gives the miss rate as a function of the threshold. Combining the data of both groups, we can draw the so-called detection error tradeoff (DET) graph (A. Martin, Doddington, Kamm, Ordowski & Przybocki, 1997), which is the miss probability plotted against the false-alarm probability, parametrized by the threshold.

We stress that is important to leave one speaker out, not just one recording of a speaker. This way, we ensure that the system is never tested with a speaker that was in the training data.

Figure 4.6.: Leave-one-out scheme

Compared to the GMM approach from the baseline system (Fernández-Pozo et al., 2009a, 2009b), the HMM approach may reflect what is said in its hidden state sequence. This way, it can model how the short-time spectrum (i.e., the vector of MFCCs) changes during the pronunciation of the sentence. Thus, whole-sentence HMMs are expected to be more powerful than a simple GMM that produces MFCCs from the same probability distribution, no matter what part of the sentence is pronounced at that moment. In other words, HMMs add text information to the classifier system.

Context-dependent (CD) phonetic modeling is widely used in large-vocabulary automatic speech recognition (ASR) systems. The most frequent units for CD modeling are triphones, which represent a phone in a particular left and right context. However, as the specificity of the model increases, the number of parameters to train also increases, and as the amount of triphone repetitions in our database is limited (e.g. the first triphone “/f/-/r/-/a/” of the first sentence only appears once) we decided to test the following alternative approaches.

- In [section 4.3.1](#) we consider whole-sentence HMMs for different numbers of states and numbers of Gaussian mixtures in the probability distributions of each state. We come to the conclusion that models with about 40 states and 13 Gaussian mixtures perform best.
- In [section 4.3.2](#) we discuss an alternative to the whole-sentence HMMs, namely sentence-dependent phonemes.

The performance results are presented and discussed with respect to the baseline system in the following [section 4.4](#).

4.3.1. Whole-Sentence HMMs

The database [see [section 3.2](#)] contains four different sentences spoken by all speakers. The first approach is to model each sentence as a whole with one HMM, as sketched in [Figure 4.7](#). For

Figure 4.7.: Scheme for whole-sentence HMMs. Each of the four sentences is trained independently to represent the specific linguistic structure of each sentence through the sequence of states.

these kinds of models, the number of states and the number of Gaussian mixtures in the GMM of each state needs to be chosen beforehand.

It is expected that a model with few states and few mixtures is too simple to grasp the difference between the severe OSA group and the control group. On the other hand, a model with too many states and mixtures tends to produce overfitting on the training data and will generalize badly to unknown data.

In the baseline system (Fernández-Pozo et al., 2009a), 256 mixtures were used to adapt both the OSA and control models from a universal background model by means of maximum a posteriori (MAP) adaptation. The use of a higher number of mixtures up to 2048 for the GMM was tested without obtaining a noticeable improvement in the classification accuracy.

Our experiments were developed in two different stages. The first one served the purpose to get familiar with the HTK and HMMs. In this initial stage, we will study the errors per sentence for all the four sentences that were analyzed and also the errors calculating the mean score of all sentences together. The same scheme will also be used for the second stage.

4.3.1.1. 1st Stage

On this first stage, different models with very different numbers of Gaussians were tried:

- 1 state with 2 mixtures would be equivalent to a model with 2 Gaussians

- 12 states with 2 mixtures would be equivalent to a model with 24 Gaussians
- 3 states with 16 mixtures would be equivalent to a model with 48 Gaussians
- 12 states with 4 mixtures would be equivalent to a model with 48 Gaussians
- 26 states with 2 mixtures would be equivalent to a model with 52 Gaussians
- 40 states with 2 mixtures would be equivalent to a model with 80 Gaussians
- 54 states with 2 mixtures would be equivalent to a model with 108 Gaussians
- 100 states with 2 mixtures would be equivalent to a model with 200 Gaussians

In [Figure 4.8](#), we can see the results we obtained with these models for every sentence.

In [Figure 4.8\(a\)](#) (sentence 1), we observe that for the severe OSA patients the error decreases as the number of Gaussians increases. From 40, 54 and 100 states the error remains very similar. However for the control group, the healthy subjects, the best result is achieved with the 12 states and 2 mixtures. In terms of total subjects, the lowest error rate is achieved with the models of 40, 54 and 100 states with 2 mixtures respectively.

The figure on sentence 2, [Figure 4.8\(b\)](#), shows how the error decreases for the severe OSA group as the number of states increases getting the best results for 40 and 100 states with 2 Gaussians. The same occurs for the total subjects. The results obtained for sentence 3, [Figure 4.8\(c\)](#), show very low error rates for the control group, being lower than 22.5%. The results for the severe OSA group improve from 54 and 100 states with 2 mixtures respectively. For the fourth sentence, [Figure 4.8\(d\)](#), we observe the same that occurred for the third sentence, the control group has a very low rate and the severe OSA group improve their results as the number of states increases.

Finally, we have also calculated the errors of the weighted average of the four sentences in [Figure 4.8\(e\)](#). For the calculation of this figure, we have given the same weight to every sentence and calculated the mean error for every topology studied. From this figure, it can be appreciated that we obtained best results in terms of lower error rates for the healthy group and also that the modelization of the severe OSA group improves as the number of states increases.

4.3.1.2. 2nd stage: HMM Topologies at fixed number of Gaussian.

Once we have got familiar with the HTK we will deepen into the use of HMMs for modeling this four sentences. We have observed that the best results are usually achieved by 40 states so we will keep this in mind for the selection of topologies studied in the second stage.

On a second stage, we choose for the HMMs models a similar order of magnitude as in the baseline GMM system. To establish the proper topology for the whole-sentence HMMs, a maximum number of Gaussian mixtures close to 500 was considered, as proved to be the most appropriate considering the amount of training data. Different topologies for different numbers of states are used, but it is intended so that the complexity of each model in terms of total number of Gaussian mixtures is comparable. The models tested are:

- 1 state with 512 mixtures: 512 Gaussians

Figure 4.8.: Classification error for different HMM topologies with different numbers of states and Gaussian mixtures in the probability distribution of each state, resulting in the total number of Gaussians ranging from 2 to 200.

4.3. Designing HMM Experiments for Our Classification Problem

Figure 4.9.: Error rates of different topologies with 1, 12, 40, and 100 states, with total number of Gaussians close to 500 as listed in the main text. The data shown is also collected in [Table 4.1](#).

- 12 states with 43 mixtures: 516 Gaussians
- 40 states with 13 mixtures: 520 Gaussians
- 100 states with 5 mixtures: 500 Gaussians

The first model, 1 state with 512 Gaussian mixtures is considered a pure GMM to have a single state. In Figure 4.9, the results are presented phrase by phrase.

Sentence 1, “Francia, Suiza y Hungría ya hicieron causa común”. As we have seen in more detail in chapter 3 regarding the databases used in this Thesis, this sentence 1 is taken from the ALBAYZIN database, a reference database on speech recognition in Castilian. It was selected because it includes /i/s marked where anomalies related to the resonance of the voice of patients with OSA phenomena may occur.

As we can appreciate from Figure 4.9(a)], the total error diminishes as the number of states increases. The best result on sentence 1 is achieved by the model of 100 states and 5 mixtures which has a complexity of 500 Gaussians.

Sentence 2, “Julián no vio la manga roja, que ellos buscan, en ningún almacén”, and Sentence 3, “Juan no puso la taza rota, que tanto le gusta, en el aljibe”. Both sentences were designed to study the nasalization phenomena that was observed in OSA patients by Fox and Monoson in their preliminary studies on 1987 and 1989 (Monoson & Fox, 1987; Fox et al., 1989). They can be compared to different parameters and acoustic measurements calculated for the first /a/ in a nasal context "manga roja" in the second sentence with those obtained in other /a/, with the same intonation in the melodic group, but in a no nasal context as "taza rota".

From Figure 4.9(b), we can observe that the best results in sentence 2 are achieved by the model of 40 states in terms of total population and also for the severe OSA group. For this model, we achieved a total error of 20%, being 15% the error rate in the severe OSA group and 25% in the control group. There is a tendency of increasing the error rate of the control group as the number of states increases.

For sentence 3, the results for severe OSA patients stands at 30% for all topologies except for 40 states when it drops to 22.50%. The error rates in the healthy group are much more moderate and gradually increase with increasing number of states ranging between 12.5% and 27.5%. We can also conclude that the best model would be that of the 40 states with 13 mixtures in which the total error rate stands at 20%.

Comparing the results of sentence 2 and sentence 3: Looking only at the total error rate of models of 40 and 100 states, it can be appreciated that values are the same for sentence 2 and sentence 3 (20% and 28.75%, respectively). However, it seems that there is a difference between groups for sentence 2 and sentence 3 in these two models.

In a context of nasalization like sentence 2, Figure 4.9(b), which is characteristic of patients with severe OSA, for the model of 40 states the OSA group has an error of 15% versus 25% error obtained in CTRL. This is a difference of 10% from one group to another. In the model of 100 states

Figure 4.10.: The best topology of all tested whole-sentence models was the one with 40 hidden states. Each state has a probability distribution (in the 12-dimensional MFCC feature space) represented by a mixture of 13 Gaussians (here shown schematically in two dimensions).

the difference is even greater, a 17.50% obtained by subtracting the error of the control group from the severe OSA group (37.50% – 20%). However in sentence 3, Figure 4.9(c), in a context of no nasalization, this difference is much smaller, yielding a 5% difference (22.50% – 17.50%) and 2.5% (30% – 27.50%).

A possible explanation could be that in the context of nasalization, characteristic of patients with OSA, they pronounce differently than healthy people and because of this it will be easier to detect that a patient is from the OSA group and not from the healthy group.

Sentence 4, “Miguel y Manu llamarán entre ocho y nueve y media”. This sentence was designed to detect patterns of irregular phonation because patients with OSA usually show a consequence of excessive fatigue nocturnal snoring (Boyd et al., 2004; Payne et al., 2006). Figure 4.9(d) shows that this phrase yields particularly good results. CTRL rates remain fairly stable between 20% and 22.5%. For severe OSA subjects, the error rates remain between 17.5% and 22.5% for all models except for the GMM which has significantly higher errors, 27.5%.

Averaging the error rates over all four sentences, Figure 4.9(e), we see that the best results among all whole-sentence HMMs were obtained when using the same number of states equal to 40 for each sentence and 13 mixtures of Gaussians for state, which results in a total of $4 \times 40 \times 13 = 2080$ Gaussian components). The architecture of this HMM is shown in Figure 4.10. The number of states 40 coincides roughly with the number of phonemes in each sentence.

4.3.1.3. Percentage of Error per Sentence and Group Depending on the Number of States

The best topology according to the minimal error is that of 40 states for every sentence but sentence 1. For this sentence, we achieve also good results with the models of 100 and 12 states.

Sentence 4 is the one with the lowest error rates. We can conclude that this sentence models well irregular patterns that are characteristics of the OSA population. In respect to the second

Table 4.1.: Error rates per sentence on the different groups depending on the number of states.

		1 state	12 states	40 states	100 states
Sentence 1	OSA	27.5%	25.0%	22.5%	30.0%
	Control	27.5%	25.0%	27.5%	12.5%
	Total	27.5%	25.0%	25.0%	21.25%
Sentence 2	OSA	35.0%	40.0%	15.0%	20.0%
	Control	15.0%	15.0%	25.0%	37.5%
	Total	25.0%	27.5%	20.0%	28.75%
Sentence 3	OSA	30.0%	30.0%	22.5%	30.0%
	Control	12.5%	15.0%	17.5%	27.5%
	Total	21.25%	22.5%	20.0%	28.75%
Sentence 4	OSA	27.5%	17.5%	17.5%	22.5%
	Control	20.0%	22.5%	22.5%	22.5%
	Total	23.75%	20.0%	20.0%	22.5%

and third sentence, we would choose the models of 40 and 100 states. It can be appreciated that on a nasalization context such as sentence 2 there are much more differences between the severe OSA and the Control group, which makes easier modeling the OSA population.

The selection of these 4 sentences for our corpus is appropriate since we achieve good results with all of them. Every sentence has an important role. Sentence 1 was chosen from ALBAYZIN database, a referenced database on Spanish speech recognition. The importance of the second sentence lies in its nasal context that makes it a good candidate when modeling the disease. On the other hand, sentence 3 is a good candidate to model the non nasal context which makes it adequate to model the healthy group. Sentence 4 is relevant for being able to characterize the irregular phonation patterns typical from nocturnal snoring.

4.3.2. Sentence-Dependent Phonemes

After having studied how to model every sentence with a different topology of HMMs, we focus our analysis now on sentence-dependent phonemes.

The approach of sentence-dependent phonemes, was designed to use context-independent (CI) phonemes but training a different CI phoneme depending on each one of the four sentences in our corpus. So, e.g. the CI HMM model for phoneme /a/ in sentence 1 is different from CI HMMs for phoneme /a/ in sentences 2, 3 and 4 respectively, see [Figure 4.11](#). In that way, HMMs of sentence-dependent phonemes should be able to model possible severe OSA/control voice differences related to the specific different phonetic contexts defined in the design of each one of the sentences (as previously described in [section 4.2](#)).

For the sentence-dependent phonemes, we used a total number of 80 HMMs to represent all

(a) Sentence-dependent phonemes for sentence 1

(b) Sentence-dependent phonemes for sentence 2

Figure 4.11.: Scheme of the approach of sentence-dependent phonemes. All occurrences of a given phoneme within a sentence are modeled by a single phoneme model, for example, all occurrences of /a/ in sentence 1 are modeled by a_1 , while all occurrences of /a/ in sentence 2 are modeled by a_2 . Each phoneme is modeled by a HMM with 3 states and 6 Gaussians per state, as sketched in [Figure 4.12](#).

Figure 4.12.: Topology of the single-phoneme HMMs, consisting of 3 states with 6 Gaussians each.

the sentence-dependent phonemes in the four sentences (17, 24, 20, 19 different phonemes in sentences 1, 2, 3, 4, respectively [see [chapter 3](#)]). Each HMM phoneme model used a standard left-to-right 3-states topology, and the best results were obtained using 6 mixtures per state (resulting in 1440 Gaussians). HMM models for sentence-dependent phonemes were obtained using standard MLLR + MAP for adapting a set of CI phoneme models trained from the Spanish phonetically balanced corpus of ALBAYZIN ([Moreno et al., 1993](#)), to the severe OSA and control groups (only records from male subjects were used as our Apnea Database 1.0 includes only male speakers).

4.4. Results

Results using the leave-one-out procedure on the three automatic severe OSA detection systems [GMM baseline ([Fernández-Pozo et al., 2009a](#)), whole-sentence HMM ([section 4.3.1](#)), sentence-dependent phonemes ([section 4.3.2](#))] are shown in [Figure 4.13](#). LLRs (i.e. Log-Likelihood Ratio (severe OSA/control)) scores, corresponding to each speaker (40 control + 40 severe OSA speakers) while uttering each phrase (4 sentences, 3 repetitions each one) were used to estimate the plotted DET curves in the Figure (a total of 960 tests), considering all possible operating points.

As we can see in [Figure 4.13](#), the initial classification results obtained with the text-independent GMMs, which mostly disregard the linguistic context information, were enhanced as context information modeled by HMMs was added. So, when compared to the GMM baseline system, a relative moderate improvement is achieved using sentence-dependent phoneme-based HMM models. However, better results are achieved by whole sentence HMM models with 40 states, producing

Table 4.2.: EER values obtained for each classification scheme

Model	EER on one sentence	EER combining all sentences
GMM	21.46%	20.00%
sentence-dependent phonemes HMMs	21.25%	17.50%
whole sentence HMMs	18.96%	15.00%

Figure 4.13.: DET curves corresponding to three severe-OSA detection systems: GMM-based baseline system (red), sentence-dependent phoneme HMMs (blue), and whole-sentence HMMs (black). Figure published in (Montero Benavides et al., 2012).

an overall 11.6% relative reduction in Equal Error Rate (EER) (see second column in Table 4.2). In order to have a global figure-of-merit using all the available audio data for each speaker in our database, LLR scores obtained from the three repetitions of the four sentences were fused (i.e. averaged) for every speaker and used, again following a leave-one-out protocol, for severe OSA/control classification. ERRs for each classification scheme are presented in the third column of Table 4.2. These results show that the best classification performance is again achieved when using the whole sentence HMM classifier, providing a 25% relative reduction in the EER value compared to the baseline GMM system.

It has to be taken into account that as these experiments were carried out on Apnea Database 1.0, i.e., the classification is done between severe OSA subjects and healthy people. In this setting, it is expected to find more differences between the extreme groups ‘severe OSA’ and ‘healthy’ than in the case of classifying different groups of OSA patients according to the severity of their disease.

4.5. Summary and Conclusions

In this chapter, we have studied two different kinds of HMMs to take advantage of the context information of our corpus that was pre-established beforehand. First, we have analyzed an approach that uses HMMs to characterize every sentence of our corpus. For this purpose, we have tried different topologies, from the most simple model which is a GMM, an HMM of just one state until the more complex, 40 states with 13 mixtures of Gaussians per state. We have “played” with the number of states and the number of mixtures of Gaussians per state as well. These experiments were divided into two stages: In the first one, we freely tried different models with different complexity that should not be compared between them. In the second stage, we focus on different models with similar complexity, around 500 and 520 mixtures of Gaussians. This number of Gaussians was chosen according to the best results achieved on a previous work with GMMs to model the speech from severe OSA subjects.

After having applied HMMs per sentence we have developed another interesting approach that models every phoneme of every sentence through HMM. This analysis is done by sentence, which means that all the /a/ from sentence 1 are modeled together and all the /a/ for sentence 2 also, as well as those /a/ from sentence 3 and those from sentence 4. This way every phoneme /a/ is modeled depending on the /a/s of the respective sentence. If for instance, there was only one /a/ in a sentence, it was modeled alone. For the individual phoneme models, we used standard left-to-right HMMs with 3-states topology (Figure 4.12). The best results were achieved using 6 mixtures per state (resulting in a total of 1440 Gaussians).

The work described through this Chapter is published in (Montero Benavides et al., 2012). Moreover, as a result of this contribution, the authors of the paper were recognized by the Microsoft Best Student Paper Award at the IberSPEECH 2012 conference.

CHAPTER 5

VOICE FEATURES RELATED TO OSA

AN ANALYSIS ON A CONTROLLED OSA DATABASE

After having seen the advantages of modeling the speech of apnea patients with HMMs focusing on the spectral envelope of continuous speech, in this chapter, we will analyze a set of acoustic variables of different nature that could be related to the speech of OSA patients. In contrast, for example, to whole-sentence HMMs, which work essentially as a black box, the present approach works on features that have a clear meaning and interpretation and may be related to certain physical properties of OSA patients.

We will work on the same controlled Apnea Database 1.0, which means working with the extreme groups *severe OSA patients* versus *Healthy population*. The speech variables are defined based on previous research from other authors and also from our research group. We start with an analysis of the discriminative power of each variable in isolation. After identifying the most promising candidates, we incrementally select features to build linear discriminant models (LDAs) and multiple linear regression models (MLRs). We will see that a model with just 8 variables is the best one for discrimination purposes.

5.1. Introduction

OSA diagnosis is time-consuming and resource-intensive, such that alternative diagnosis and screening techniques are highly needed ([section 2.5](#)). This chapter tackles the problem of OSA screening via speech technologies by investigating the acoustical characteristics of the voice in patients with OSA for the purpose of learning whether OSA may be detected via speech analysis. In this case, we are going to study the analysis of some acoustic features related to previous findings on OSA speech.

In order to have a controlled experimental framework to study apnea voice characterization we have used Apnea Database 1.0, the speech database designed following linguistic and phonetic criteria derived from previous research in the field ([section 3.2](#)). Our work here is focused on continuous speech rather than on sustained vowels, the latter being the standard approach in pathological voice analysis. This is justified because we are interested in the acoustic analysis of

the characteristics of the speech signal in specific linguistic and phonetic contexts, and also in the analysis of other features not used previously in the characterization of apnea voices such as the dynamics of speech, and in particular the patterns of pauses, which we hypothesized as valuable discriminative features.

Together with speech-dynamics and pause features, OSA-related voice characteristics associated with articulation, phonation, and nasalization will be evaluated over the voices of severe-OSA patients and healthy subjects to have a contrastive study on the acoustic discrimination attainable in our database using either individual features or a combination of some or all of the proposed features. In addition, and due to the crucial dependency of OSA disease with BMI and age factors, we finished with a study about relations between OSA-related voice features and these two important factors.

5.2. Selection and Definition of Voice Features Related to OSA

The topic of physiological features of OSA patients and the impact on their voices has been introduced in the previous [section 2.7](#). Since the obstruction occurs in the upper airway, it is logical to assume the existence of specific characteristics of the speech signals of patients with OSA.

To continue, we defined a set of voice features that will be analyzed on the speech corpus of Apnea Database 1.0 in order to determine its utility in sleep apnea screening, both in isolation and combined. The definition of these features is based on:

- Parameters traditionally used in the diagnosis of pathologic voices.
- Previous studies about the characteristics of the voice of apnea patients, in particular ([Monoson & Fox, 1987](#); [Fox et al., 1989](#)), see also [section 2.7.1](#).
- A preliminary perceptual informal analysis of our corpus.

All acoustic variables to be analyzed are collected in [Table 5.1](#). In the following, we motivate and explain all these features, grouped in OSA-related, phonation-related, nasality-related and prosodic features. The set of features is also directly related to the types of peculiarities that were detected in previous works for apnea voices ([Fox et al., 1989](#)), and henceforth we can organize each of the proposed features as features related to anomalies in articulation, phonation, and resonance. Most features are clearly related to one of these abnormalities, although one of them, $(F_3 - F_2)_{/i/}$ may be related to more than one of these types of anomalies.

5.2.1. Formant Difference $(F_3 - F_2)_{/i/}$

According to previous research, there are differences found when comparing the snoring signal of apnea patients and healthy subjects. Snoring is generated by the vibrating soft tissue in the upper airway, and it is a common symptom of OSA. Thus, snore signals may provide an excellent framework for non-invasive diagnosis of OSA. Among the bibliography consulted we found that formant features of snore signals carry useful information for OSA diagnosis, and demonstrate

Table 5.1.: OSA-related features under analysis.

Feature	Feature Name	Description	Related to abnormalities in
1	$(F_3 - F_2)_{/i/}$	Difference between F_3 and F_2 in the phone /i/ of the first sentence of the corpus (Hz)	Articulation and resonance
2	HNR	Harmonic to Noise Ratio (dB) measured on the fourth sentence	Phonation
3	Jitter	Jitter (s) measured on the fourth sentence	Phonation
4	SDRseg	Segmental Signal to Dysperiodicity Ratio SDR (dB) measured on the fourth sentence	Phonation
5	Shimmer	Shimmer (dB) measured on the fourth sentence	Phonation
6	$a1h1max800_{\text{nasal}}$	Nasality measure for vowels in nasal contexts (second sentence of corpus)	Resonance
7	$a1h1max800_{\text{non-nasal}}$	Nasality measure for vowels in non-nasal contexts (third sentence of corpus)	Resonance
8	$a1h1max800_{\text{diff}}$	$a1h1max800_{\text{nasal}} - a1h1max800_{\text{non-nasal}}$	Resonance
9	Duration1	Average duration of sentence 1 (s)	Articulation
10	Duration2	Average duration of sentence 2 (s)	Articulation
11	Duration3	Average duration of sentence 3 (s)	Articulation
12	Duration4	Average duration of sentence 4 (s)	Articulation
13	PercSil1	Average Percentage of Silences in sentence 1 (%)	Articulation
14	PercSil2	Average Percentage of Silences in sentence 2 (%)	Articulation
15	PercSil3	Average Percentage of Silences in sentence 3 (%)	Articulation
16	PercSil4	Average Percentage of Silences in sentence 4 (%)	Articulation

that the use of snore signals can be a convenient, inexpensive, and reliable diagnostic approach for mass screening of OSA:

- [Sola-Soler, Jane, Fiz and Morera \(2003\)](#) analyzed the spectral envelope of the snoring signal and found out that the standard deviation of some snoring formant frequencies is significantly lower in simple snorers than in OSA patients, even when nonpostapneic snores are considered in this group. They found out significant differences in formant frequencies variability between simple snorers and OSA patients.
- Later, [Sola-Soler, Jane, Fiz and Morera \(2008\)](#) analyzed the formants of inspiration and exhalation sounds in snoring subjects with and without OSA and found out that formant F_1 is significantly lower in inspiration episodes of OSA patients and that this formant has a significantly higher variability in pathological subjects, for both inspiration and exhalation episodes. A higher variability of formant frequencies seems to be an indicator of the presence of OSA.
- [Ng et al. \(2008\)](#) found that apneic snores exhibit higher formant frequencies than benign snores, especially F_1 , which can be related to the pathology of OSA.
- [Ng, Koh and Puvanendran \(2007\)](#) presented the use of formant features of snore signals for OSA diagnosis. Acoustical changes due to the collapsing upper airways can be reflected on the

Figure 5.1.: Differences between third and second formant for the vowel /i/ in the word “Suiza” (‘suj θa), for an OSA speaker (left) and a control-group speaker (right). Figure published in (Montero Benavides et al., 2014).

formant features in the frequency spectrum, which were extracted to discriminate between benign and apneic snores.

Nevertheless, since no records on snoring are available in our database we have decided to focus on the analysis of formants on continuous speech, namely on the word “Suiza”.

An interesting conclusion from our initial perceptual contrastive study was that, when comparing the distance between the third (F_3) and second formant (F_2) for the vowel /i/, differences between the apnea and healthy subjects were found. For apnea speakers, this value was greater, and this was especially clear in diphthongs with /i/ as the stressed vowel, as in the Spanish word “Suiza” (‘suj θa) of the first sentence (see Figure 5.1). This distance $F_3 - F_2$ is obtained as the average difference of the values for these two formants. We measured absolute distances in spite of the fact that the location of the formants is speaker-dependent. Nevertheless, we considered that normalization was not necessary because our database contains only male subjects with similar relevant physical characteristics, and the formants should lie roughly in the same regions for all of our speakers. We measured this feature in utterances of the first sentence in our corpus listed above, which contains good examples of stressed allophones for the vowel /i/. For the calculations of the formant values, we used the Praat software.

This finding is in agreement with the conclusion of Robb et al. (1997) that the F_2 formant value in the vowels produced by apnea subjects is lower than healthy ones due to greater length of the vocal tract of OSA patients (see section 2.7.2); but also, and perhaps more importantly, with the OSA-characteristically abnormal velopharyngeal opening and lowering of the velum which may cause higher third formant frequencies (Hidalgo & Quilis, 2002). Therefore, this is a voice feature related to articulatory and resonance OSA-related anomalies.

5.2.2. Phonation Features (HNR, Jitter, Shimmer, SDRseg)

Trying to characterize phonation abnormalities related to OSA patients, we will study several acoustic features traditionally used in the detection of pathologic voices: harmonic-to-noise ratio (HNR), signal-to-dysperiodicity ratio (SDRseg) and features that measure the perturbation of fun-

damental frequency (Jitter) and energy (Shimmer). This type of features are generally measured on sustained vowels, although their use has been successfully extended to continuous speech (Parsa & Jamieson, 2001). However, in this work, we have calculated this measures on the fourth sentence of our corpus due to its acoustic characteristics. The rationale behind is that continuous speech can take into account important attributes of speech such as rapid voice onset and termination, changes in voice frequency and amplitude, and sudden discontinuities in speech.

A brief description of each of these features follows:

Harmonic-to-noise ratio (HNR) is a phonation feature used to estimate the level of noise in human voice signals. The HNR can be used to measure the degree of periodicity in a voice waveform. It is a measurement of voice pureness where ‘healthy’ voices will tend to have a higher HNR than a ‘pathological’ voice. Studies have reported large differences in HNR between normal and disordered talkers, as well as strong correlations with subjective ratings of hoarseness severity (Yumoto, Gould & Baer, 1982; Yumoto, Sasaki & Okamura, 1984). In our case, we employ a time-domain approach (Boersma, 1993) that will be computed on voiced segments in connected speech. The parameters based on the relation between the harmonic energy and the noise energy have an extensive application in the voice analysis in medical applications (Yumoto et al., 1982).

Due to the random frequency distribution of noise, the HNR measurement tend to be very sensitive to the recording systems used, since they can introduce noise levels that are different from those of the voice itself, altering the results (D. Martin, Fitch & Wolfe, 1995).

The calculation of the HNR for our database was performed using the Praat software (Boersma & Weenink, 2013), which uses an algorithm based on the analysis in the time domain of the voice signal and in the use of a measure of autocorrelation of the signal for the calculation of the fundamental frequency. All this makes this method guarantee good results and is especially insensitive to possible disturbances in pitch values (Boersma, 1993). So, in short, we see that the autocorrelation of a signal is given by:

$$r_x(\tau) \triangleq \int x(t)x(t+\tau)dt \quad (5.1)$$

The fundamental period, T_0 , is defined as the value of τ which corresponds to the maximum value (excluding zero) of the autocorrelation function (hereafter $r_x(\tau)$). The energy of the speech signal within a time window is the value of the autocorrelation function with its zero index:

$$r_x(0) = r_p(0) + r_{ap}(0) \quad (5.2)$$

Where $r_p(0)$ and $r_{ap}(0)$ are the energies of the periodic and aperiodic component, respectively. The normalized autocorrelation is defined by:

$$r'_x(\tau) = \frac{r_x(\tau)}{r_x(0)} \quad (5.3)$$

Given the periodicity of the periodic component of the autocorrelation function and assuming uncorrelated additive white noise, the energy of the normalized periodic component is given as

$$r'_p(0) = r'_p(T_0) = r'_x(T_0), \quad (5.4)$$

and the energy of the aperiodic component is:

$$r'_{ap}(0) = 1 - r'_p(0) = 1 - r'_x(T_0). \quad (5.5)$$

In this way, the harmonic-to-noise ratio (HNR) is defined as

$$\text{HNR} \triangleq \frac{r'_p(0)}{r'_{ap}(0)}. \quad (5.6)$$

Segmental signal-to-dysperiodicity ratio (SDRseg). Dysperiodicity is a common symptom of voice disorders that refers to anomalies in the glottal excitation signal generated by the vibrating vocal folds and the glottal airflow. SDRseg, measured in dB, is a parameter used to summarize these vocal dysperiodicities in connected speech. Many variants of the signal-to-dysperiodicity ratio are known to be good estimators of perceived speech quality, among which we implemented a variant of the algorithm “bidirectional distal linear predictive analysis” presented in the work of [Bettens, Grenez and Schoentgen \(2005\)](#). SDRSeg is used to predict the perceived degree of hoarseness of a voice.

Jitter measures perturbations in voice production, especially in vowel phonation, due to small fluctuations in the opening and closing times of the vocal chords. Jitter can be used to discriminate between healthy and dysphonic speakers. A high degree of jitter occurs in a voice with roughness. In our case, we computed jitter on vocal segments of continuous speech (measured in seconds). For the calculation of the jitter we used the software Praat and the following formula:

$$\text{Jitter}_{\text{absolute}} = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|, \quad (5.7)$$

where T_i are the length of the periods of the fundamental frequency and N is the number of periods, see also Figure 5.2(a).

Shimmer measures perturbations in glottal cycle lengths related to the fluctuations of amplitude (measured in dB). It is again calculated on vocal segments of continuous speech. Shimmer was also calculated with Praat.

$$\text{Shimmer}_{\text{absolute}} = \frac{1}{N-1} \sum_{i=1}^{N-1} |20 \log(A_{i+1}/A_i)|, \quad (5.8)$$

where A_i are the peak-to-peak amplitudes and N is the number of periods of fundamental frequency extracted, as shown in the Figure 5.2(b).

Specifically, these four features (HNR, SDRseg, Jitter, and Shimmer) were computed only for the voice sounds of the fourth sentence of the corpus. In theory, a lower HNR and SDRseg should be related to a more pathological voice, while a lower Jitter and Shimmer should correspond to a less pathological voice.

(a) Periods of the fundamental frequency for the calculation of jitter

(b) Amplitudes of the fundamental frequency for the calculation of shimmer

 Figure 5.2.: Examples for the calculation of jitter and shimmer according to equations (5.7) and (5.8) for $N = 4$ periods of fundamental frequency (Farrús & Hernando, 2009).

5.2.3. Nasality Measurements

In the work of Fox et al. (1989), the existence of certain irregular resonance characteristics and some abnormalities in nasalization on apnea voices was hypothesized for the first time. The hypothesis to be confirmed is an unusual variation of nasalization between nasal and non-nasal contexts in apnea subjects, exhibiting smaller intra-speaker differences between these contexts. In order to try to characterize these anomalies, we used one of the nine best parameters for the measurement of nasal character of vowels proposed in the Ph.D. Thesis of Pruthi (2007). This feature is based on the difference between the amplitude of the first formant (A_1) and the first harmonic (H_1). The first formant is taken as the maximum in the 0–800 Hz frequency band (this explains the name of the feature a1h1max800) and the first harmonic is the maximum closest to 0 Hz with amplitude over 10 dB and wider than 80 Hz in bandwidth. These thresholds were defined in the latter reference. Its calculation is given by:

$$\text{a1h1max800} = A_1 - H_1 \quad (5.9)$$

This measure is indicative of the vowel nasal character, being lower for nasal vocal segments as compared with non-nasal ones (i.e. it can be interpreted more as a measure of hyponasality than hypernasality since it is higher for less nasality).

In our speech corpus, this parameter was measured separately on vowels in nasal contexts ($\text{a1h1max800}_{\text{nasal}}$) and vowels in non-nasal contexts ($\text{a1h1max800}_{\text{non-nasal}}$), and was also computed the difference between both contexts ($\text{a1h1max800}_{\text{diff}}$). In particular, in nasal contexts it

was measured on the highlighted /a/ phoneme of the words “manga” and “buscan” (in the second sentence of the corpus); in non-nasal contexts it was measured on the highlighted /a/ phoneme of the words “taza” and “gusta” (in the third sentence of the corpus).

5.2.4. Prosodic Features: Sentence Duration and Silence Features

Besides spectral parameters, there are other speech features that are believed to contribute to the recognition and understanding of speech by humans. One such category of non-spectral speech features is prosody as it is manifested on both the segmental and the supra-segmental level.

The last features are those related to the duration of the pronunciation of the sentences and the percentage of silences in the sentence (considering only silences in the middle of the sentence, not before or after). These measurements, which can be categorized as suprasegmental features, have already been successfully applied in speaker recognition domain (Reynolds et al., 2003) and in the characterization of pathological voices (Zhang & Jiang, 2008). Our motivation to include these features is based on the conclusions of the work of (Fox et al., 1989), regarding the phonation anomalies of apnea subjects. We already argued that these abnormalities in glottal activity could produce fatigue in the muscles of the larynx and respiratory troubles, which could result in difficulty of apnea subjects to pronounce sentences without pauses (such as the fourth sentence in our corpus). Therefore we would expect more sentence duration and more silences in the apnea patients than in the control subjects. In order to assess the importance of the length and absence of pauses in the sentence we have measured these features for each of the sentences separately, and therefore we have four sentence durations and four percentages of silences per subject.

In Table 5.1 we can see all the speech features analyzed among with their description and their related abnormalities: articulation, phonation, and resonance.

5.2.5. Voice and Data Analysis Tools

Regarding voice analysis tools used in this study, we applied Praat software (Boersma & Weenink, 2013) to compute formant values, harmonic-to-noise ratio (HNR), Jitter and Shimmer. The signal-to-dysperiodicity ratio (SDRseg) was implemented with Matlab, which was also used for feature selection and classification algorithms (described in the following section). Finally, the automatic phonetic segmentation of the recordings was obtained using standard speech recognition technology based on the HTK Toolkit (S. Young et al., 2009) and a set of acoustic models trained on the phonetic corpus of the ALBAYZIN database (Moreno et al., 1993).

5.3. Analysis and Results

The main goals of our analysis are:

- Analyze each one of the selected features in order to determine their utility to discriminate voices of severe OSA patients from voices of healthy subjects.

- Propose ways to combine those features in order to obtain a classifier that could be helpful to assist in OSA screening between the extreme groups, severe OSA versus healthy population.
- Measure the performance of such classifiers both in terms of classification error and in the characteristic metrics used in clinical practice.

After measuring the proposed voice features on our speech databases and organizing properly these measurements, we have structured our analysis and results section into three parts:

Feature analysis and classification model training using the training database (section 5.3.3). Performance evaluation of the best system on the test database (section 5.3.4). Comparison to screening based only on age and BMI factors (section 5.3.5).

5.3.1. Analysis of Individual Features

Our first study is focused on the analysis of the individual features. This analysis has been performed by means of two different techniques: statistical and discriminant analysis.

5.3.1.1. Statistical analysis of individual features using Mann–Whitney U test

First, using the training database, we want to determine whether the distributions of each selected feature present statistically significant differences between severe OSA and Healthy subjects (and henceforth could potentially be used for OSA screening purposes). As our features are not normally distributed, we used the Mann–Whitney U test (UMW), as it is commonly used as an alternative non-parametric to a standard t-test (Mann & Whitney, 1947). This test is based on the median of the difference on ranked data and it shows us that some of the individual features present statistically significant differences at the 95% confidence value between both groups.

For each particular feature and group, Table 5.2 shows its median (as this is the statistical output of the UMW test), its standard deviation and the correspondent p -value provided by this test. The features with the lowest p -values are considered the most promising features in order to be used for discrimination. According to the results of UMW test the best features for discrimination are the following: SDRseg, $(F_3 - F_2)_{/i/}$, HNR, PercSil4, Duration4, a1h1max800_{diff}, Jitter and Duration1 while the features with the higher p -values: Shimmer, a1h1max800_{non-nasal}, PercSil1-3, Duration2-3, a1h1max800_{nasal} would be considered the less promising features for which we do not find statistically significant differences between both populations.

Other interesting conclusions can be extracted from these results. It is remarkable that sentence duration and percentage of silence can be very discriminative features but only if the sentence pronounced is designed especially for this purpose (as fourth sentence of our speech corpus was). Otherwise, these suprasegmental features are not very discriminative. Regarding the nasal features, we defined these features based on the findings of Fox et al. (1989), in which most experts perceived some abnormalities in nasalization but they were not capable of describing them consistently (neither as hyponasality nor as hypernasality). Our experiments show that differences in nasalization in nasal (a1h1max800_{nasal}) and non-nasal (a1h1max800_{non-nasal}) contexts are not statistically significant at the 95% confidence value, which could explain why experts found diffi-

Table 5.2.: Median, standard deviation and p -values for the proposed speech features computed on the training dataset. p -values were obtained comparing the groups severe OSA/Control median values using the UMW test. Results are sorted in decreasing order of p -value.

Feature #	Feature name	Median (Std) Apnea	Median (Std) Control	p -value
4	SDRseg	30.9 (4)	34 (3)	< 0.001*
1	$(F_3 - F_2)_{i/}$	614 (80)	586.5 (66)	< 0.001*
2	HNR	10.5 (1.8)	11.3 (1.6)	0.003*
16	PercSil4	0.08 (0.08)	0.06 (0.05)	0.004*
12	Duration4	3.6 (0.9)	3.2 (0.7)	0.006*
8	a1h1max800 _{diff}	-1.0 (1.8)	-2.5 (2.5)	0.011*
3	Jitter	0.0177 (0.003)	0.0170 (0.004)	0.011*
9	Duration1	4.0 (0.8)	3.9 (0.7)	0.044*
11	Duration3	4.7 (1.2)	4.6 (1.1)	0.096
6	a1h1max800 _{nasal}	5.1 (2.2)	3.9 (2.4)	0.096
13	PercSil1	0.16 (0.07)	0.15 (0.07)	0.119
10	Duration2	5.0 (1.0)	4.8 (1.1)	0.205
15	PercSil3	0.20 (0.08)	0.17 (0.1)	0.264
14	PercSil2	0.18 (0.07)	0.19 (0.09)	0.315
7	a1h1max800 _{non-nasal}	6.2 (2.6)	6.8 (3.0)	0.328
5	Shimmer	0.11 (0.02)	0.11 (0.02)	0.458

* Statistically significant differences found at the 95% confidence level.

cult to find what the abnormality in nasalization was. We have found, however, that the difference in nasalization between nasal and non-nasal contexts (a1h1max800_{diff}) is statistically significant at the 95% confidence value. Looking at the medians of the nasality measures it seems that apnea patients have a smaller dynamic range of nasalization (i.e. they tend to pronounce with a more similar nasality level in nasal and non-nasal contexts) than the healthy subjects that have more difference in nasalization between both contexts.

This finding allows us to propose the new feature a1h1max800_{diff} as a good candidate to be used in OSA detection, and is also consistent with both the hypothesis suggested by [Fox et al. \(1989\)](#), and the results of our previous work ([J. L. Blanco-Murillo, Fernández-Pozo, Díaz-Pardo, Alvaro Sigüenza et al., 2009](#)), which shows us that OSA speakers exhibited smaller intra-speaker differences between non-nasal and nasal vowels.

5.3.1.2. Discriminative Analysis of Individual Features: EER

The analysis on the statistical differences of the distributions for each individual parameter presented before describes their underlying potential for OSA discrimination. In this section, their classification capability, when used in isolation, is measured through performance metrics relevant to their application for OSA diagnosis. For each feature, we have built a very simple classifier

by means of setting a decision threshold in order to separate apnea patients from healthy subjects using the training database. Next, we have computed the error rates obtained for both groups which correspond to the two types of errors that can be made in the OSA diagnosis: errors in the Apnea group because apnea subjects are incorrectly diagnosed as healthy, and errors in the Control group because healthy subjects are incorrectly detected as patients with OSA. The point at which those two error rate curves cross defines the equal error rate (EER), which is very interesting because it measures the discriminative power of the feature with a single number, hence facilitating comparison of features.

As we can see in [Table 5.3](#), when analyzing EER per feature we see that the best individual feature is the SDRseg with an EER of 34%, followed by $\text{a1h1max800}_{\text{diff}}$ and PercSil4 with EERs slightly below 38%. With lower discriminative power, HNR and Duration4 have EERs a little below 40%, and the rest of the features provide EERs over 40%. Shimmer, the worst feature, still provides some discrimination since the EER is still below 48%. Although similarities with the conclusions from the previous statistical analysis are evident, there are also notable differences. For instance, the feature $(F_3 - F_2)_{/i/}$ seemed to be one of the two best features according to the statistical analysis, but it does not appear among the best features according to EER (the opposite applies to $\text{a1h1max800}_{\text{diff}}$ feature).

5.3.2. Analysis of Feature Combination

In the previous section, we have analyzed the behavior of the individual features proposed to discriminate voices of OSA patients from voices of healthy subjects. We have made this analysis from different points of view (i.e. statistical analysis and discriminative power) and have reached conclusions not totally equivalent (for instance regarding features $(F_3 - F_2)_{/i/}$ and $\text{a1h1max800}_{\text{diff}}$). This makes it difficult to choose the best features to be used in our purpose of detecting OSA with speech signals. This selection is more difficult if we take into account that the features could be highly correlated between them, what could lead to a problem of multicollinearity that could worsen the detection model. In other words, once a feature is included in the set of best features adding another feature that individually seemed to provide a great deal of information does not necessarily add any new information. For this reason, in this section, we first analyze the correlation between different features, and then we present the outcomes of an incremental feature selection study for identifying the best feature subsets from 1 to 16 features.

5.3.2.1. Correlation Analysis

Aiming to study the possible linear dependencies between the selected features, we have computed a correlation matrix that is shown in [Figure 5.3](#). We can see that, in general, there is slight correlation between the features proposed: almost 60% of the correlation coefficients are below 0.2 in absolute value and almost 80% are below 0.5. There are only a few exceptions. Features 2 (HNR) and 4 (SDRseg) are highly correlated, which is reasonable because both measurements characterize disorders related to phonation anomalies. Features 6, 7 and 8 ($\text{a1h1max800}_{\text{nasal}}$, $\text{a1h1max800}_{\text{non-nasal}}$ and $\text{a1h1max800}_{\text{diff}}$) are also clearly correlated, what is expected since

Table 5.3.: EER obtained with each one of the individual features on the training database. Features are sorted in increasing order of EER.

Feature #	Feature name	EER (%)
4	SDRseg	34.0
8	a1h1max800 _{diff}	37.6
16	PercSil4	37.8
2	HNR	39.7
12	Duration4	39.7
1	$(F_3 - F_2)_{/i/}$	40.2
11	Duration3	40.3
10	Duration2	40.7
7	a1h1max800 _{non-nasal}	41.6
15	PercSil3	41.6
3	Jitter	41.7
9	Duration1	42.0
14	PercSil2	42.8
6	a1h1max800 _{nasal}	43.3
13	PercSil1	43.7
5	Shimmer	48.0

Figure 5.3.: Correlation matrix showing absolute values of correlation coefficients computed on the training database for the 16 voice features (Table 5.1) proposed to characterize apnea voices. Darker means stronger correlation.

$a1h1max800_{diff}$ is defined as the difference of the other two features. Finally, all features related to sentence durations and percentages of silence are highly correlated from sentence to sentence of the corpus. This fact is quite logical since a speaker that tends to read slower will do it on all sentences. The same happens with the silences. It is interesting to see that Duration4 and PercSil4 are less correlated with the equivalent measures on the other sentences. This may be due to the different nature of fourth sentence: one single melodic group that is asked to be pronounced without pauses.

5.3.2.2. Analysis of Incremental Subsets of Features

In order to determine the best possible combination of features, we have performed an incremental study in which we built models that start with one feature and are augmented with the next best feature remaining according to some feature comparison criteria.

We consider two different models for the combination of features: the multiple linear regression (MLR) model (Draper & Smith, 1998) and the linear discriminant analysis (LDA) model (Duda, Hart & Stork, 2001). These models need to be trained on a specific set and might probably produce overfitting to that set. To avoid overfitting, we have divided the training database (42 apnea and 40 healthy subjects) into two similar-size subsets containing the same number of apnea (20) and healthy (20) subjects. One of these subsets will be used for training the model and selecting the best next feature to add (train train subset) and the other subset only to evaluate each model selected on data not seen during training (train dev subset). In fact, we evaluate the models on both subsets, so that we can study in more detail the problem of overfitting, but conclusions are drawn on results obtained on the train dev subset since they are more representative of the situation of a real system that has to classify patterns not seen during training.

Table 5.4.: Distribution of age and BMI in healthy and apnea speakers in the databases.

Database	Group	Number of subjects	Age			BMI		
			Mean	Std. dev.	<i>p</i> -value	Mean	Std. dev.	<i>p</i> -value
TRAIN	Control	40	42.2	8.8	0.002*	26.2	3.9	0.000*
	Apnea	42	49.5	10.8		32.8	5.4	
TEST	Control	20	46.5	7.8	0.936	29.1	3.29	0.086
	Apnea	20	46.3	7.7		30.7	2.79	

*Statistically significant differences found at the 95% confidence level.

5.3.2.3. Summary of Selected Features to Characterize Apnea Speech

Table 5.4 summarizes the 16 OSA-related features considered in this study that have been described in detail in [section 5.2 \(Table 5.1\)](#). In this table, we have computed the UMW test to look for differences in Age and BMI between the groups of Control and Apnea. For the train database we have found statistically significant differences in age and BMI between Control and Apnea patients. However, for the test database, there are no significant differences between Control and severe OSA regarding the clinical variables age and BMI, which makes it a good subset to classify only according to speech.

To evaluate the models we used two different criteria:

- Classification error rate (CER) attained by the model. It is the ratio between the number of incorrectly classified subjects and the total number of subjects. It would be desirable to have as lower CER as possible (its minimum is 0% by definition). Measuring CER implies setting a threshold to take decisions. For LDA classification, this threshold has been adjusted based on the train train subset, while for the MLR model it has been set to 0.5: the average between the 0 output value (i.e. dependent variable value) assigned to healthy subjects and the 1 output value assigned to apnea subjects.
- The adjusted R^2 is defined as

$$R_{\text{adj}}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \quad (5.10)$$

or equivalently

$$1 - R_{\text{adj}}^2 = (1 - R^2) \frac{n - 1}{n - p - 1} = (1 - R^2) \left[1 + \frac{p}{n - p - 1} \right], \quad (5.11)$$

where p is the total number of explanatory variables in the model (not including the constant term), and n is the sample size, and R^2 statistic ([Vittinghoff, Glidden, Shiboski & McCulloch, 2004](#)) is the amount of variance that the model explains. R^2 statistic is also the square of the Pearson's correlation coefficient between the values predicted by the model and the actual values. It would be desirable to have as high R^2 as possible (its maximum is 1 by definition). But the R^2 has some disadvantages: Every time a predictor is added to a model, the R^2 increases,

even if due to chance alone. It never decreases. Consequently, a model with more terms may appear to have a better fit simply because it has more terms. If a model has too many predictors and higher-order polynomials, it begins to model the random noise in the data, what is known as overfitting the model and it produces misleadingly high R^2 values and a lessened ability to make predictions. Due to these disadvantages, we have computed instead a more robust measure which is the adjusted R^2 .

After establishing the classification models and performance criteria, the incremental study starts by considering all the individual features in isolation and choosing the best individual feature computed on the train train subset according to the evaluation criteria. It would be natural to choose R^2_{adj} for the MLR model (R^2 is closely related to the regression residual, the parameter to optimize in MLR) and CER for the LDA model (CER is the parameter that LDA optimizes). The reason for this decision is that, after having split the training corpus into two subsets, the amount of subjects in each subset is only 41. This makes CER a very granular measurement because it only changes when at least one decision change. In our experiments, this happened only a few times, and therefore the criterion had not enough resolution to choose among different features (hence the selection was many times random). In contrast, R^2_{adj} takes into account the predicted value and the actual value, and therefore is computed before a decision is taken, which allows finding differences in R^2_{adj} even when the decisions are exactly the same. Once the best individual feature is chosen according to R^2_{adj} computed on the train train subset, all the remaining features are tested in combination with this best feature, and the next best feature is chosen, again according to R^2_{adj} of the predictions made by the model combining the two features on the train train subset. By repeating this procedure we continue adding the best next feature until all features are contained in the models.

The outcomes of this procedure are a list of the features sorted in decreasing order of importance for each of the models, the ‘best’ models considering 1–16 features, and also the evaluation metrics (R^2_{adj} and CER) for the MLR and CER for the LDA on the train train and train dev subsets. All this information, and in particular results on the train dev subset, will allow us to determine which features really provide important, discriminative and subset-independent (with good generalization capability) information, not previously provided by other (more important) features, and which ones provide little discriminative information or poor generalization capabilities. Therefore this experiment will allow us to rank the analyzed features and define the subsets of the most important ones for OSA assessment.

Tables 5.5 and 5.6 show the incremental feature selection outcomes for the MLR and LDA model respectively. It can be seen that the order in which features are added is the same, irrespective of the model employed in the experiment. This consistency in the results of feature selection using two different models makes these results more reliable. Comparing the columns corresponding to results evaluated in the train train subset and the train dev subset, it can be observed that outcomes obtained on the train train subset tend to improve as more and more features are included in the model. However, results for the train dev subset improve clearly only with the first 6–9 features and from that point on results are clearly degraded. This is indicating an overfitting situation when

Table 5.5.: Incremental feature combination using a MLR model and R^2_{adj} selection criterion. Best outcomes for each column are highlighted.

Feature#	Feature name	train train subset		train dev subset	
		R^2_{adj}	CER	R^2_{adj}	CER
4	SDRseg	0.113	39.0%	0.163	26.8%
+8	+a1h1max800 _{diff}	0.217	26.8%	0.187	36.6%
+2	+HNR	0.293	24.4%	0.200	29.3%
+3	+Jitter	0.411	22.0%	0.297	17.1%
+1	+ $(F_3 - F_2)_{/i/}$	0.426	19.5%	0.343	22.0%
+12	+Duration4	0.443	19.5%	0.422	14.6%
+13	+PercSil1	0.470	24.4%	0.417	17.1%
+9	+Duration1	0.476	17.1%	0.422	17.1%
+14	+PercSil2	0.493	19.5%	0.385	12.2%
+10	+Duration2	0.532	14.6%	0.336	19.5%
+16	+PercSil4	0.549	9.8%	0.198	26.8%
+15	+PercSil3	0.545	9.8%	0.146	26.8%
+11	+Duration3	0.552	9.8%	0.161	24.4%
+5	+Shimmer	0.548	9.8%	0.164	24.4%
+6	+a1h1max800 _{nasal}	0.544	7.3%	0.156	24.4%
+7	+a1h1max800 _{non-nasal}	0.551	9.8%	0.125	26.8%

more than 6–9 features are added to the models. This could be due to the nature of the features (the features themselves do not provide additional valuable information) or to the limitations of the size of the corpus (we have 41 subjects in the train dev subset and the models require estimating one parameter for each of the features considered, plus a bias term, so having 7 features requires estimating 8 parameters with 41 subjects). In the future it would be very interesting to expand these experiments to determine the reason for this overfitting.

In any case, what we can conclude with certainty is that, with this amount of material and with these features, the best set of features is the one containing the first 6–9 features, which include, in order: SDRseg, a1h1max800_{diff}, HNR, Jitter, $(F_3 - F_2)_{/i/}$, Duration4, PercSil1, Duration1 and PercSil2. This set of features includes: classical measurements used in the characterization of pathological voices (such as SDRseg, HNR, and Jitter); features already proposed for the specific problem of apnea detection (Fernández-Pozo et al., 2009a), $(F_3 - F_2)_{/i/}$; and other features not used previously in the characterization of apnea voices such as Duration1, Duration4, PercSil1, PercSil2 and a1h1max800_{diff}, being this last measure based on previous perceptual findings (Fox et al., 1989).

5.3.3. Classification and Diagnosis Performance

As a result of the incremental feature combination analysis in the previous section, we have obtained two classifiers (MLR and LDA) combining all the features or a subset of best 6–9 features.

Table 5.6.: Incremental feature combination using a LDA model and CER selection criterion. Best outcomes for each column are highlighted.

Feature#	Feature name	CER on train train subset	CER on train dev subset
4	SDRseg	36.6%	29.3%
+8	+a1h1max800 _{diff}	31.7%	36.6%
+2	+HNR	26.8%	31.7%
+3	+Jitter	22.0%	22.0%
+1	+ $(F_3 - F_2)_{/i/}$	17.1%	24.4%
+12	+Duration4	22.0%	14.6%
+13	+PercSil1	22.0%	14.6%
+9	+Duration1	17.1%	14.6%
+14	+PercSil2	22.0%	22.0%
+10	+Duration2	17.1%	24.4%
+16	+PercSil4	12.2%	29.3%
+15	+PercSil3	12.2%	29.3%
+11	+Duration3	12.2%	24.4%
+5	+Shimmer	12.2%	26.8%
+6	+a1h1max800 _{nasal}	12.2%	26.8%
+7	+a1h1max800 _{non-nasal}	12.2%	26.8%

Table 5.7.: CER on the training dataset using a leave-one-out procedure for LDA and MLR models using all features and the 6–9-best features found in the incremental feature analysis. Best outcomes are highlighted for LDA and MLR models.

	LDA model	MLR model
6-best features	18.3%	19.5%
7-best features	20.7%	22.0%
8-best features	17.1%	18.3%
9-best features	18.3%	18.3%
All features	23.2%	22.0%

In this section, we firstly analyze the performance when using the two classification models (MLR and LDA) with all the features or the best 6–9 features on the training database in terms of classification performance, using CER. See [Table 5.7](#).

Then, in a second and probably most important analysis, diagnosis performance, we evaluate figures of merit traditionally employed in medical diagnosis such as sensitivity, specificity, positive predictive value, negative predictive value, and F1-measure.

Given that the division of the training database into two similar-size subsets limits very much the amount of training examples (only 41 subjects), which could possibly jeopardize the accuracy achieved, we have performed these experiments using a leave-one-out procedure, that is repeatedly using all the subjects in the training database except one to train the models and perform a test on the excluded subject.

It should also be noted that as the training database was used in the feature selection process, results in this section can be considered as a resubstitution upper bound that will be later compared with those obtained on a separate test database (Sections [5.3.4](#) and [5.3.5](#)). Another limitation for the training database is the significant differences in age and BMI between OSA and Healthy subjects, this issue will be also discussed in [section 5.3.5](#).

5.3.3.1. Classification Performance

In practice, if our system is going to be used in diagnosis support it should make a decision and classify the voice as corresponding to a severe OSA patient or a healthy subject. For this reason, we have performed this decision based on a threshold trained in the case of LDA (the threshold is trained on the 81 remaining subjects of corpus in the leave-one out testing protocol) and a fixed threshold in the case of MLR (0.5, corresponding to the middle point between the 0 for healthy subjects and 1 for apnea patients).

Classification error rates (CER) are presented in [Table 5.7](#). Results show that by combining all features with the LDA model and performing a leave-one-out experiment on the training database we attain a reasonably good classification performance, with a CER of 23.2%. This result is slightly better with the MLR model (22.0%). If we limit our models to the set of 6–9 best features found in [section 5.3.2.2](#), models attain better classification performance: LDA reaches a CER of 17.1%, while MLR achieves a CER of 18.3%. These best results are achieved with the 8 best features (8 or 9 in the case of the MLR model). This gives an idea of the importance of feature selection in this context. In general, we can conclude that using all features produces worse results than using only the best 6–9 features and that the LDA model produces slightly better results than the MLR model in this case. Best performance is achieved using 8 features.

5.3.3.2. Diagnosis Performance

Previous results have been presented in the typical form for classification and verification problems. However, the goal of the selected features and proposed methods is to help in the OSA screening to identify those patients that have severe apnea and who will most benefit from a swiftly diagnosis and treatment. Therefore in this section, we evaluate the best of the proposed

Table 5.8.: Performance measures (sensitivity, Specificity, Positive and Negative Predictive Value and F1-measure) of the voice-based LDA model (8 best selected features) on the training database. Results obtained using a leave-one-out approach

Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1-measure
78.6% (33/42)	87.5% (35/40)	86.8% (33/38)	79.6% (35/44)	82.5%

methods (the combination of the 8-best features with the LDA model) using the criteria commonly used in clinical diagnosis (Knottnerus & Buttinx, 2008):

- Sensitivity: ratio of correctly classified OSA-suffering speakers (true positives) to total number of speakers actually diagnosed with OSA.
- Specificity: ratio of correctly classified healthy subjects (true negatives) to total number of speakers diagnosed as not suffering from OSA.
- Positive Predictive Value (PPV): ratio of true positives to total number of patients classified (correctly or incorrectly) as having an OSA voice.
- Negative Predictive Value (NPV): ratio of true negatives to total number of patients classified (correctly or incorrectly) as not having an OSA voice.
- F1-measure: a single measure of performance obtained as the harmonic mean of positive predictive value (precision) and sensitivity (recall). $F1\text{-measure} = 2 \times (PPV \times \text{sensitivity}) / (PPV + \text{sensitivity})$.

Results in Table 5.8 are very encouraging and seem to indicate that our approach could be useful for performing fast and convenient screening for OSA. Comparing these results with previous work (Fernández-Pozo et al., 2009a) where we reported a sensitivity value of 77% and specificity value of 85%, the obtained values here are slightly better using a completely different approach, which leads us to believe we could achieve better results by combining the two different approaches.

5.3.4. Analysis of Diagnosis Performance on a Separate Test Set

Up to this point, we have used the training database to analyze and select features and to train and test models. We will now take the best of the proposed methods (8 best features selected and LDA model) and we will train a model on the whole training database in order to evaluate it on a different test dataset (test database) that has not been used so far. Diagnosis performance results obtained on this independent test set are shown in Table 5.9.

Results obtained on the test database are only slightly worse in terms of performance than those obtained on the training database (Table 5.8). Features were selected using results of the training database; therefore, we could expect some degradation when moving to a different database. It is also noteworthy to highlight that the trade-off between sensitivity and specificity has changed considerably between Tables 5.8 and 5.9. This means that the calibration (or threshold setting) of the method proposed for apnea screening would probably require further research.

Table 5.9.: Performance measures (sensitivity, specificity, positive and negative predictive value and F1-measure of our voice-based LDA model (8 best selected features) on the test database. Results obtained using a leave-one-out approach.

Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	F1-measure
85.0% (17/20)	75.0% (15/20)	77.3% (17/22)	83.3% (15/18)	81.0%

Table 5.10.: Performance measures (sensitivity, specificity, positive and negative predictive value and F1-measure) of our voice-based method vs. a model using only age and BMI on training and test databases.

Database	Method	Sensitivity	Specificity	Positive PV	Negative PV	F1-measure
TRAIN	Voice-based	78.6%	87.5%	86.8%	79.6%	82.5%
	BMI + age	78.9%	77.5%	76.9%	79.5%	77.9%
TEST	Voice-based	85.0%	75.0%	77.3%	83.3%	81.0%
	BMI + age	65.0%	55.0%	59.1%	61.1%	61.9%

PV: predictive value

5.3.5. Comparison to Screening Using Only BMI and Age

Our voice-based screening method seems to be comparable in accuracy to other OSA screening strategies that employ clinical prediction models using anthropometric characteristics (like BMI and age) and epidemiological parameters (Gurubhagavatula et al., 2004; Friedman et al., 2010). However, although these approaches report similar sensitivity values (around 80%), they generally present appreciably worse specificity values (around 70%).

To study our particular case, we compared the results in Tables 5.8 and 5.9 (using the selected LDA model and 8 features) with those obtained with a model that does not contain information from speech but only uses age and BMI data. On the training database accuracy, results shown in Table 5.10 using only the variables mentioned, are lower than when using our voice-based approach. The only exceptions are the results for the Sensitivity and Negative Predictive Value, which are similar. In terms of the F1-measure the difference between our proposed approach and BMI and age alone is 4.6% in absolute terms.

On the test database, both methods suffer some degradation. However, our voice-based features method only suffers an absolute reduction of 1.5% in terms of F1-measure while the absolute reduction reaches a 16% for the method based on age and BMI. The reason for the important degradation in performance when using only age and BMI is that the test database was carefully designed to avoid statistically significant differences in terms of both factors between the two groups. Thus this low performance on this database was expected. The main conclusion from Table 5.10 is that the method based on voice features continues working almost as well as on the training dataset even when there are not statistically significant differences between the two

Table 5.11.: Correlation coefficient and p -values between speech features and Age and BMI. Statistically significant correlations are highlighted.

FEATURE	TRAIN database				TEST database			
	Correlation coefficient		p -Value		Correlation coefficient		p -Value	
	AGE	BMI	AGE	BMI	AGE	BMI	AGE	BMI
SDRseg	-0.21	-0.26	0.06	0.02*	-0.08	-0.23	0.60	0.14
a1h1max800 _{diff}	-0.02	0.04	0.87	0.72	-0.11	-0.06	0.47	0.70
HNR	-0.16	-0.02	0.16	0.89	-0.09	-0.02	0.57	0.90
Jitter	0.16	0.15	0.16	0.20	-0.11	0.10	0.48	0.51
$(F_3 - F_2)_{i/}$	0.22	0.28	0.06	0.01*	-0.05	0.11	0.72	0.48
Duration4	0.35	0.28	0.002*	0.01*	0.13	-0.1	0.40	0.54
PercSil1	0.23	0.07	0.04*	0.56	0.04	-0.06	0.77	0.71
Duration1	0.34	0.10	0.003*	0.37	0.076	-0.14	0.64	0.37

* Statistically significant differences found at the 95% confidence level.

groups of speakers in terms of age and BMI. This clearly indicates that our method provides information related to apnea not only through their relation to age and BMI.

We have also delved into the study of the correlations between the 8 selected voice features and the age and BMI variables. Table 5.11 shows the correlation coefficients and their associated p -values for both training and test databases. The first noticeable difference between the two databases is that in the training database there is a number of statistically significant correlations, while in the testing database no correlation is close to be statistically significant at the 95% confidence value. The main difference between the two databases is that in the training database there are statistically significant differences in terms of BMI and age between the Apnea and Control group, while the test database was designed to make those differences not statistically significant. The absence of statistically significant correlations in this second test seems to indicate that the correlation found for some speech features in the training database could be due to their correlation with the apnea condition and the correlation of this condition with BMI and age.

5.4. Conclusions and Discussion

Through this chapter, a set of 16 speech features that could be related to obstructive sleep apnea (OSA) has been studied. Some of these features are traditionally used in the detection of voice pathologies (such as HNR, Jitter, etc.), while others have been taken from previous research on voice characteristics of OSA patients or have been proposed from the analysis of our speech apnea databases collected for this work. From our results on discriminative power of individual OSA-related voice features, we can highlight that we have found among the most discriminative features some already proposed by authors in previous works (i.e. the difference between the

third and second formants for the /i/ phoneme) and some first time implemented, although already hypothesized in related works (i.e. the difference in nasality measured between vowels in nasal and non-nasal contexts).

OSA detection based on a selected subset of 8-best speech features and a LDA model has been assessed. Classification results reached an overall performance of over 80%, similar to other OSA screening strategies based on clinical data such as BMI and age ([Gurubhagavatula et al., 2004](#)). Given that speech features could be related to BMI and age, which are known to be related to apnea, classification results using the selected speech features have been compared to those obtained using only BMI and age. Results using the voice-based approach on a test database with no statistical differences in BMI and age distributions, and showing no significant correlations between speech features and both factors, have provided a F1-measure value 20% higher in absolute terms compared to using only BMI and age factors. This result can lead us to conclude that the proposed voice measures have the potential to improve OSA detection on segments of population where age and BMI have no discriminative power.

At this point, it must be noted that, in addition to BMI and age, there are many other clinical data and symptoms in OSA patients, such as snoring, daytime sleepiness, smoking history, etc. that can influence their voice characteristics. The study of these dependencies is beyond the scope of this work, as we actually do not have complete clinical data for the patients in this database, but it will be a line for future research. Nevertheless, in some cases, these studies could be faced through the articulation of complementary research lines under the speech analysis field; for example, it will be worth considering how our speech features are correlated to the phonetic features already related to sleepiness in the recent research presented by [Krajewski, Schnieder, Sommer, Batliner and Schuller \(2012\)](#).

Despite these limitations, we consider that our results are encouraging and clearly show that the distinctive apnea traits identified by our incremental feature selection method can be combined to provide reasonable diagnosis performance for screening purposes. Furthermore, such promising outcomes were obtained with just 8 speech features and with very simple and basic classification models. In previous works on the same database ([Fernández-Pozo et al., 2009a](#)) we reported results slightly worse than those presented here, but following a completely different approach. We expect to achieve even better results by combining these models with those previously proposed models.

Finally, further experimentation with a larger population of control and OSA subjects would be also desirable both to deepen our understanding of the analyzed and new speech features.

5.4.1. Summary

In this chapter, we have proposed the use of speech analysis as an alternative or complement to existing screening methods. A set of voice features that could be related to apnea was defined, based on previous results from other authors and our own analysis. These features were analyzed first in isolation and then in combination to assess their discriminative power to classify voices as corresponding to apnea patients or healthy subjects. This analysis was performed in Apnea

Database 1.0 containing three repetitions of four carefully designed sentences read by 40 healthy subjects and 40 subjects suffering from severe apnea. As a result of the analysis, a linear discriminant model (LDA) was defined including a subset of eight features (signal-to-dysperiodicity ratio, a nasality measure, harmonic-to-noise ratio, jitter, difference between third and second formants on a specific vowel, duration of two of the sentences and the percentage of silence in one of the sentences). This model was also tested on a separate database containing 20 healthy and 20 apnea subjects yielding a sensitivity of 85% and a specificity of 75%, with a F1-measure of 81%. These results indicated that the proposed method, only requiring a few minutes to record and analyze the patient's voice during the visit to the specialist, could help in the development of a non-intrusive, fast and convenient PSG-complementary screening technique for OSA.

CHAPTER 6

FORMANTS AND CLINICAL VARIABLES IN A CLINICAL PRACTICE SCENARIO

After having employed complex analysis techniques such as Hidden Markov Models and multi-variate analysis on different speech features from different nature on Apnea Database 1.0 in the previous Chapters 4 and 5, we now turn to Apnea Database 2.0, which became available in 2013 and has been described in detail in [section 3.3](#). We recall that it was recorded from patients of the sleep unit of Hospital Quirón in Málaga and that it contains sustained vowels apart from the four sentences of Apnea Database 1.0. Since all subjects of this database were patients suspected to suffer from sleep disorders, even the non-OSA group cannot be expected to be healthy (see [Figures 1.5](#)). and [3.4](#). This makes the classification problem more similar to the use case of clinical pre-diagnosis, but also more difficult than in our previous research on Apnea Database 1.0. Therefore, we have decided to take a step back in the analysis and to study some more elementary speech features, namely the formant frequencies and bandwidths of the sustained vowels.

In this chapter, we review the basis of the detection of apnea disease through speech by means of formant frequencies and bandwidths focusing on the analysis of the smallest phonetic unit, the vowels (confer also [sections 2.6.1](#) and [2.7.2](#)). We focus not only on speech features but also on the analysis of clinical variables related to speech and the apnea syndrome and on their interdependencies. Our analysis is focused hereafter on a real diagnosis scenario. Given the complexity and sources of variability of the clinical scenario, in order to have the speech more controlled, in this part of the Thesis, we only focus on sustained vowels. Due to the lack of data, this could not be done on Apnea Database 1.0, which has far less speakers and only the vowel /a/ recorded. This Chapter analyzes, in particular, the formant frequencies and bandwidths on all sustained vowels /a/, /e/, /i/, /o/ and /u/, which are believed to be intimately affected by the presence of OSA traits, and brings them in relation to different clinical variables.

6.1. Introduction and Motivation

6.1.1. Formants in Human Speech

We have already discussed the basics of human speech and the meaning of formants in sections 2.6.1 and 2.7.2, respectively. We remind that simply speaking, air flows from the lungs and air tube through the larynx and causes the vocal cords to vibrate, which is where the sound is created. The airflow is then filtered along the vocal tract (laryngeal cavity, the pharynx, the oral cavity, and the nasal cavity), where certain frequencies are suppressed and others amplified by the so-called formant resonances. Human speech makes extraordinarily heavy use of these formants, which are the most important acoustic parameter. By adjusting tongue, lips, etc., the vocal tract changes its resonances characteristically, determining whether we pronounce an /a/, /i/, or some different vowel (confer Figure 2.7).

In the power spectrum, a formant is usually characterized by its central frequency and its bandwidth, defined as the width where the power has decayed by 3 dB. Formants and bandwidths differ from person to person and depend on the vowel that is pronounced. More concretely, they depend on the upper airway properties including its compliance, shape, and dimensions. Hence, these may contain information from specific physiological characteristics in OSA patients.

For speech analysis, the first three formants are the most informative in the production of vowels. The variations of the formant frequencies that are needed to pronounce different vowels can be explained in terms of the oral cavity width and shape, i.e., the vocal-tract length and constrictions at the location of nodes and antinodes of certain resonances. The effective vocal-tract length changes with the mouth opening and rounding of the lips. All formant frequencies decrease with the vocal-tract length, as already indicated in equation (2.1). The effect of vocal-tract constrictions is formant-selective because different formants have their nodes and antinodes located at different locations in the vocal tract, as sketched in Figure 2.6. Constrictions at nodes of a formant's standing wave increase the frequency of that formant, while constrictions at antinodes lower the frequency. Rounding the lips, i.e., constricting the antinode of all formants at the mouth opening, lowers all formant frequencies. In the following, we collect some formant-specific details.

F_1 is the first resonance and has only a node at the larynx and an antinode at the mouth opening and no further nodes or antinodes in-between. Thus, F_1 is determined by the vocal-tract length and width, i.e., the pharynx and the degree of mouth opening at the lips. As the tongue moves from a high to low position, the pharyngeal cavity decreases in volume. “High” vowels, like /i/, have more pharyngeal space which resonates at lower frequencies. As the pharyngeal space increases, the value of the first formant gets lower. Also, a small mouth opening produces low values of the first formant, as it is the case for all formants.

F_2 has a node in the oral cavity and an antinode in the pharynx and is correlated primarily with tongue retraction. A retracted tongue constricts the antinode in the larynx and gives room in the oral cavity, where the node is located. Both effects produce a lower F_2 . Front vowels, like /i/, have higher F_2 while back vowels, like /a/ have lower F_2 . Also, the shorter the

length in the front cavity the higher the F_2 .

Again, lip rounding lowers also F_2 by elongating the oral cavity and constricting the antinode at the mouth opening. (/u/ has the lowest F_2 of all vowels because the tongue is retracted and lips are rounded).

F_3 is related with the nasal cavity; understanding it goes beyond the simple tube-and-constriction model. A constriction in the back produces a high F_3 while a constriction in the front produces a low F_3 .

6.1.2. Previous Studies About Formant Resonances and Apnea

Before developing the approach of this chapter, we recapitulate some works related to formants, bandwidths and OSA already exposed at the end of [section 2.7.2](#), namely the first acoustic analysis of speech for OSA detection by [Fiz et al. \(1993\)](#), and the study by [Robb et al.](#) who picked up this approach in 1997.

[Fiz et al. \(1993\)](#) analyzed five sustained vowels pronounced in a sitting position and found that the vocalization of vowel /i/ showed differences between OSA patients and healthy subjects. OSA patients had a narrower range of frequencies with lower maximum of harmonic frequency for utterances /e/ and /i/ and lower number of harmonics for /i/ with respect to healthy subjects, which may point at differences in laryngeal behavior between OSA and non-OSA speakers. A noticeable explanation for this acoustic results is that fat deposit in the tongue and pharyngeal walls and changes in pharyngeal shape typical in obese OSA patients could influence the mechanical properties of the anatomic structures determining in part the characteristics of the voice of these patients. However, they conclude that the explanation for the differences found in /e/ and /i/ vowels were not explained by obesity *per se*, but rather by the presence of the OSA syndrome.

Later on, [Robb et al. \(1997\)](#) presented an acoustic analysis on a small male population of just 10 OSA patients and 10 non-OSA subjects, focusing on the supra-laryngeal level by means of vocal-tract formant frequencies and bandwidths, where OSA-related alterations should have a larger impact according to the pathogenesis of the disorder. From the analysis of sustained phonations of vowels, they found significant differences in the first and second formants (F_1 , F_2) and the corresponding bandwidths (BW_1 , BW_2). In particular, the formant frequencies and corresponding bandwidths of the sustained vowels /i/, /u/ and /a/ were considered. [Robb et al.](#) found lower formant frequencies in the OSA group, which they attributed to a longer vocal tract. The wider formant bandwidths in the OSA group were interpreted to reflect greater vocal-tract damping in the OSA subjects.

According to [Robb et al. \(1997\)](#), two main reasons may support the hypothesis of a relationship between formants and OSA. First, cephalometric research has indicated that the distance from the hyoid bone and the mandibular plane is significantly longer in patients with OSA compared to non-OSA individuals; this suggests that longer vocal tracts in OSA patients can lead to lower formant frequencies compared to non-OSA subjects. A second reason relates to an increase in both velar and pharyngeal compliance commonly found among OSA patients. This increased

collapsibility of the pharyngeal airway has been related to an excessive fat deposition (Shelton, Woodson, Gay & Suratt, 1993; Isono, 2012) that would increase the sound damping inside the vocal tract (Robb et al., 1997). Consequently, formant bandwidths characterizing OSA voices may be expected to be wider than those for non-OSA subjects.

6.1.3. Motivation for the Present Approach

Our aim in this chapter is to review the previous findings of Robb et al. (1997) on our larger Apnea Database 2.0 and to examine possible indirect or mediated effects of clinical variables, which may shed more light on the connection between speech and OSA.

We investigate whether or not the significant lower formant frequencies and wider bandwidths, as found in the small group of 10 male OSA patients compared to 10 non-OSA subjects in Robb's work, are reproduced over our larger population in Apnea Database 2.0, which is representative for the clinical practice scenario (see section 3.3 for more details on our Apnea Database 2.0). As pointed out by Robb et al. (1997), the review of their initial research results on larger populations is essential in order to fully realize the applicability of acoustic analysis in the assessment of OSA.

Another important aim of our research was to investigate to what extent these possible connections between formants and OSA may be affected by the specific clinical characteristics in OSA populations. The objective is to evaluate whether or not clinical variables such as height, age, cervical perimeter or BMI can present a stronger impact on vocal-tract resonances than specific OSA-related alterations. This is an important issue that has already raised a debate when evaluating recent speech-based detection methods (see, for example, Goldshtein et al., 2011, or Montero Benavides et al., 2014): due to human voice changes with age, height or BMI detection methods may be predicting these demographic and anthropometric variables rather than OSA. To contribute to the research in this important topic, for each individual in our database not only speech records but also the clinical features height, weight, BMI, cervical perimeter, gender and age were made available.

From a methodological point of view, we expanded the analysis in (Robb et al., 1997) to include the first three formant central frequencies (i.e. F_1 , F_2 , and F_3) and their associated bandwidths (i.e. BW_1 , BW_2 , and BW_3), measured over sustained phonations of the vowels /a/, /e/, /i/, /o/, and /u/, as they are contained in Apnea Database 2.0. We decided to extend our analysis to all five Spanish vowels and to follow a class-dependent approach as it has been suggested that acoustic cues for OSA detection can be phonetic-dependent (J. L. Blanco-Murillo et al., 2008). Although connected speech may provide further insights on the connections between speech and OSA, in this study, we have followed the same methodological approach designed in (Robb et al., 1997), where sustained phonations were used to avoid spurious deviations due to phonetic variability (Parsa & Jamieson, 2001).

In summary, in this study we test whether or not the observed lower formant frequencies values and wider formant bandwidths found among OSA subjects in (Robb et al., 1997) are recovered in a large database of subjects suspected to suffer from OSA submitted to PSG test in a Sleep Unit in Spain. This result will elucidate if plausible explanations of longer vocal tracts in OSA

patients leading to lower formant frequencies and higher sound damping due to an increase in tissue compliance may be maintained. We also evaluate whether clinical variables such as age, height, cervical perimeter or BMI may have stronger impact on vocal-tract resonances than specific OSA-related alterations.

6.2. Materials and Methods

6.2.1. Database

In order to keep under control the dialect, we have employed a subset of the Apnea Database 2.0 containing only speakers with an Andalusian accent, as described in [section 3.3.1.1](#). The resulting population of study includes 241 male speakers of Andalusian, which is characterized by the elimination of final consonants, namely final /s/ is aspirated or elided and is also characterized by ceceo/seseo, i.e., the neutralization of /s/ and /θ/ to [θ].

6.2.2. Acoustical Analysis

Spoken language contains frequencies in the range from a few hertz to a few kilohertz. In the case of sustained vowels, the spectrogram has an essentially stationary envelope. Now, the formants are the (local) maxima of the envelope function, i.e., frequencies where the acoustical signal is particularly strong: the resonances of the vocal tract. Formants are counted from low frequencies to high frequencies, and each formant can be associated with a width, the so-called bandwidth.

Measuring formant frequencies can be difficult as it is highly influenced by multiple factors, including the method of analysis that is chosen and the analysis settings. Moreover, higher resonances are much more difficult to determine than lower ones due to a natural decrease of energy with frequency. Our evaluation of acoustical measurements has shown that for the fourth formant and above no reliable information could be extracted, and therefore we restrict our analysis to the first three formants which we will refer as F_1 for the first formant, F_2 for the second formant and F_3 for the third formant.

Tools. To extract a consistent set of measures on formants' central frequencies and bandwidths we have followed a specific protocol. First, we compute the values for the three first formant central frequencies and bandwidths using two different tools, namely *Praat* ([Boersma, 1993](#)) and *Snack* (<http://www.speech.kth.se/snack/>), which are widely used open-source tools ([Woehrlich, de Mareüil & Boula, 2007](#)). Formant frequencies and bandwidths were estimated every 5 ms using 25 ms-long analysis windows. Their values were finally obtained by averaging along the most stable regions of the sustained phonations of each vowel, selecting a steady-state segment of 800 ms where the standard deviation of formant contours was the lowest, excluding initial and ending silences in each utterance.

Error correction. To guarantee a reliable estimation, we have measured the absolute differences between estimated values obtained from Praat and Snack for each formant F_1 to F_3 . We then manually reviewed those cases for which differences exceeded 70 Hz for F_1 and F_2 , and 150 Hz for F_3 . These thresholds match the level of accuracy in (Robb et al., 1997) and seem consistent with values used in studies that compare Praat and Snack (Woehrli et al., 2007). In most cases, for which deviations exceeded the pre-specified thresholds, one of the two values (either the one from Snack or the one from Praat) was found to be obviously incorrect. Most often, one of the formant was skipped. In those cases, the erroneously estimated value was removed and the value provided by the other software was kept. In some other cases both Snack and Praat failed in providing precise results. In those cases values for formant central frequencies and bandwidths had to be manually selected using spectrograms and linear predictive coding analysis. The decision on the number of poles for an optimal fitting of the linear predictive coding envelope was based on the general knowledge about the formant structure of each vowel. Values for formants' central frequencies were obtained as maximum values of the linear predictive coding spectral slope. The associated bandwidths were determined as the frequency region around formant's central frequency in which the spectral envelope amplitude remains within 3 dB below the maxima values.

Finally, after the error correction described above, outliers were checked to detect cases where both Praat and Snack may have failed to deliver sensible results but matched one another. In doing so we identified 6 files which contained audio recording errors (distorted recordings or recordings with wrong content). Those 6 speech files were discarded.

6.2.3. Statistical Analysis

To start with, normality tests were carried out for every variable. Among the clinical variables, we found only height and age to follow a normal distribution. The same was done on speech variables and once again most of them were not normally distributed. Hence non-parametric methods had to be used, for example, Spearman's correlations rather than Pearson's.

Thus, Spearman correlation coefficients ρ_s were used to compute correlations between the AHI and certain pathophysiologically related variables. A p -value of $p < 0.05$ was considered significant in this analysis, as it is widely used in clinical studies. P -values larger than that were regarded as nonsignificant.

To assess the significance of differences between the different populations (non-OSA, mild OSA, severe OSA), the Wilcoxon rank-sum test was used. We conducted our statistical analysis using IBM SPSS program, Version 21.

6.3. Results

6.3.1. Formant Frequencies and Bandwidths

Table 6.1 summarizes the formant central frequencies and bandwidths we have obtained using the procedure described in section 6.2.2 on the Andalusian sub-database. In the table, one may

Table 6.1.: Statistical analysis of the first three formant frequencies and bandwidths of vowels. All frequencies in the table are given in Hertz (Hz).

Vowel	Formant	Mean	SD	Range	Bandwidth	Mean	SD	Range
/a/	F_1	707	65	556–1012	BW_1	137	70	34–588
	F_2	1257	90	990–1513	BW_2	149	89	49–994
	F_3	2458	193	1894–3426	BW_3	196	147	26–1650
/e/	F_1	445	36	341–547	BW_1	76	37	15–263
	F_2	1879	119	1561–2332	BW_2	109	101	40–1485
	F_3	2474	137	2126–2857	BW_3	193	128	56–1508
/i/	F_1	303	33	203–402	BW_1	53	26	5–258
	F_2	2183	152	1762–2677	BW_2	99	97	19–1363
	F_3	2690	170	2232–3334	BW_3	225	160	61–1570
/o/	F_1	482	42	330–600	BW_1	93	44	20–338
	F_2	858	74	668–1032	BW_2	111	61	20–460
	F_3	2499	205	1820–3005	BW_3	144	94	31–600
/u/	F_1	336	39	235–460	BW_1	62	28	11–197
	F_2	738	75	558–981	BW_2	116	76	12–504
	F_3	2422	187	1922–2951	BW_3	155	94	22–682

Abbreviation: SD, standard deviation.

observe the mean, standard deviation and range of the first three formant frequencies and bandwidths for the vowels /a/, /e/, /i/, /o/ and /u/. [Figure 6.1](#) shows a graphical representation of the data, where each vowel is represented by an ellipse in the two-dimensional formant space span by the first two formants. Locations and diameters of the ellipses are given by the mean values and standard deviation from [Table 6.1](#), respectively.

Reliability of acoustic measurements To verify our estimation of formants, [Figure 6.1](#) shows data from [Rosique et al. \(2003\)](#) for comparison. We observe that the mean values for F_1 and F_2 reported by [Rosique et al.](#) are consistent with our estimated values for the Andalusian subset. According to this, we have found that our speech recordings are consistent with those reported in Rosique’s contribution, and consequently, we may think that these could match those from an arbitrary group of Spanish speakers.

6.3.2. Cross Correlations

6.3.2.1. Correlations of clinical parameters with AHI

[Table 6.2](#) summarizes the correlation coefficients between all clinical variables that are considered in this paper and AHI values obtained from the PSG test. Only the correlation coefficient relating height and AHI was found to be non-significant.

For the rest of this Chapter, the strength of the correlation coefficient will be described as in ([Fowler, Cohen & Jarvis, 2013](#)), i.e., values between 0 and 0.19 will be described as very weak,

Figure 6.1.: F_1 vs. F_2 for the Andalusian group compared to the results of Rosique et al. (2003).

The elliptical areas in color were built from results on our database so that the center is located at the mean values for (F_2, F_1) and the x -radius and y -radius correspond to the standard deviations obtained for F_2 and F_1 , respectively. Dots report the location of the mean values reported by Rosique et al..

Table 6.2.: Spearman correlations between clinical variables and AHI.

Correlation with AHI	Weight	Height	BMI	Age	Cervical Perimeter
Spearman correlation coefficient ρ_s	0.32†	0.00	0.37†	0.15*	0.42†
P value	0.00	0.97	0.00	0.02	0.00

* Correlation is significant at the 0.05 level (two-tailed).

† Correlation is significant at the 0.01 level (two-tailed).

between 0.2 and 0.39 as weak and in the range 0.40 to 0.59 as modest correlation. Stronger correlations do not occur.

The strongest correlation found was a modest correlation between cervical perimeter (neck circumference) and AHI ($\rho_s = 0.42$). Correlations between weight, BMI and AHI are also significant but weak ($\rho_s = 0.32$ and 0.37 , respectively). The three of them, cervical perimeter, weight and BMI, are strongly related to obesity, which is known to be a risk factor of OSA (Jamie et al., 2010). Surprisingly, the correlation between AHI and age is not as strong as expected being age one of the most important risk factors of OSA disease. From these results, we can conclude that the risk factors related to overweight such as a high BMI, high weight, and big cervical perimeter are more relevant than the age.

6.3.2.2. Correlations between formants and clinical variables

Tables 6.3 and 6.4 collect all significant correlations between formants and clinical variables. No significant correlation was found between formant frequencies and AHI, thus, AHI does not appear in Table 6.3.

Among all clinical variables shown in Table 6.3, most of the significant relationships with for-

Table 6.3.: Spearman correlations ρ_s between clinical variables and formant frequencies. The numbers in parentheses are the corresponding p -values.

Vowel	Formant	Weight	Height	Age	Cervical Perimeter
/a/	F_1			−0.14* (0.03)	
	F_2		−0.13 (0.05)	−0.13 (0.05)	
/e/	F_1		−0.12 (0.05)	−0.12 (0.05)	−0.17 [†] (0.01)
	F_2	−0.15* (0.02)	−0.20 [†] (0.00)		−0.16* (0.02)
	F_3		−0.21 [†] (0.00)		
/i/	F_1		−0.15* (0.02)	0.16* (0.01)	
	F_2		−0.21 [†] (0.00)		
	F_3				−0.20 [†] (0.00)
/o/	F_1				−0.13* (0.05)
	F_2			−0.27 [†] (0.00)	
	F_3		−0.17 [†] (0.01)		
/u/	F_1		−0.14* (0.03)		
	F_2			−0.24 [†] (0.00)	
	F_3	−0.14* (0.04)	−0.20 [†] (0.00)		

Notes: For clarity, nonsignificant correlation values are omitted.

* Correlation is significant at the 0.05 level (two-tailed).

[†] Correlation is significant at the 0.01 level (two-tailed).

Table 6.4.: Spearman correlations ρ_s (p -values in parentheses) of clinical variables and bandwidths.

Vowel	Band width	AHI	Weight	Height	BMI	Age	Cervical Perimeter
/a/	BW ₁					−0.21 [†] (0.00)	
	BW ₂	0.13* (0.05)					0.14* (0.03)
/e/	BW ₁					−0.17 [†] (0.01)	
	BW ₂						
	BW ₃	−0.17 [†] (0.01)	−0.14* (0.04)		−0.13* (0.05)		
/i/	BW ₁					−0.13* (0.05)	
	BW ₃		−0.15* (0.02)	−0.14* (0.03)			
/o/	BW ₁						
	BW ₂		−0.16* (0.02)		−0.14* (0.03)	0.15* (0.02)	

Notes: For clarity, nonsignificant correlation values are omitted.

* Correlation is significant at the 0.05 level (two-tailed).

[†] Correlation is significant at the 0.01 level (two-tailed).

formants were for age and height. F_1 and F_2 of /a/, F_1 of /e/ and F_2 for /o/ and /u/ were negatively (or inversely) correlated with age ($\rho_s = -0.14, -0.13, -0.12, -0.27, -0.24$, respectively). Only a positive, very weak correlation was detected between age and F_1 of /i/ ($\rho_s = 0.16$) and two significant, weak correlations with age ($\rho_s = -0.27$ and -0.24 for F_2 in /o/ and /u/). Height was also found negatively correlated with up to nine formants. F_2 of /a/, F_1, F_2, F_3 of /e/, F_1 and F_2 for /i/, F_3 of /o/ and F_1 and F_2 of /u/ were negatively (or inversely) correlated with height ($\rho_s = -0.13, -0.12, -0.2, -0.21, -0.15, -0.21, -0.17, -0.14$ and -0.2 , respectively). Fewer correlations were found between formants and weight and cervical perimeter; weight correlated negatively with F_2 of /e/ ($\rho_s = -0.15$) and F_3 of /u/ ($\rho_s = -0.14$), and cervical perimeter with F_1 and F_2 of /e/, F_3 of /i/ and F_1 of /o/ ($\rho_s = -0.17, -0.16, -0.20$ and -0.13 , respectively). No significant correlations were observed between BMI and formants.

If we focus in the number of significant correlations found per vowel with formants we found:

- 7 significant formant correlations with /e/,
- 4 significant formant correlations with /i/ and /u/,
- 3 significant formant correlations with /a/ and /o/.

As shown in Table 6.4, there is a positive correlation between BW_2 from vowel /a/ and AHI ($\rho_s = 0.13$); and a negative one between BW_3 from /e/ and AHI ($\rho_s = -0.17$). Three significant correlations were found with weight: BW_3 from /e/ and /i/ and BW_2 from /o/ ($\rho_s = -0.14, -0.15$ and -0.16 , respectively). There is only one negative significant correlation between formant bandwidth and height, $\rho_s = -0.14$ with the third bandwidth of /i/, and two negative correlations with BMI for BW_3 from /e/ and the BW_2 from /o/ ($\rho_s = -0.13$ and -0.14). Also three significant, negative correlations were detected between age and BW_1 for /a/, /e/ and /i/ ($\rho_s = -0.21, -0.17$ and -0.13) and a positive correlation between BW_2 for /o/ and age ($\rho_s = 0.15$). Finally, there is a correlation with the second bandwidth of /a/ and cervical perimeter ($\rho_s = 0.14$). If we focus in the number of significant correlations found per vowel with bandwidths we found

- 4 significant bandwidth correlations with /e/,
- 3 significant bandwidth correlations with /a/, /i/, and /o/,
- 0 significant bandwidth correlations with /u/.

In Figure 6.2, we represent graphically, which clinical variables and formant features are correlated. The vowel /e/ is the one that has most significant relations both in formants and bandwidths correlated with clinical variables. With respect to (Robb et al., 1997) this vowel was not studied.

Focusing on the total number of significant correlations for every clinical variable, we find:

- AHI correlates with 2 bandwidths and 0 formants, which makes a total of 2 significant correlations with speech variables.
- Weight correlates with 3 bandwidths and 2 formants, which makes a total of 5 significant correlations with speech variables.
- Height correlates with 1 bandwidth and 9 formants, which makes a total of 10 significant correlations with speech variables

Figure 6.2.: Map representation of clinical variables and their influence on formants frequencies and bandwidths. Each arrow represents a significant correlation from Table 6.3 or Table 6.4. Legend: thin: very weak correlation; thick: weak correlation; solid: p -value < 0.05; red wiggled: $\rho_s > 0$; black straight: $\rho_s < 0$, dotted: p -value = 0.05 (p -values larger than 0.05 are not drawn).

- BMI correlates with 2 bandwidths and 0 formants, which makes a total of 2 significant correlations with speech variables.
- Age correlates with 4 bandwidths and 6 formants, which makes a total of 10 significant correlations with speech variables.
- Cervical perimeter correlates with 1 bandwidth and 4 formants, which makes a total of 5 significant correlations with speech variables.

The clinical variables that have more significant correlations with speech are height and weight, with a total of 10 relationships.

6.4. Discussion

In this chapter, we have analyzed formants and bandwidths of sustained vowels on a male OSA population with a homogeneous accent, Andalusian. We have focused the analysis on groups according to the severity of the disease (non OSA, mild OSA and severe OSA) and check for statistical differences between the non OSA and mild OSA and between non OSA and severe OSA, taking into account the whole range of the severity of the disease.

6.4.1. Discussion of Results

The results of this investigation clearly indicate that only poor correlation exists between formants frequencies and bandwidths and clinical variables in the OSA population under study. Although all the detected correlations can be described as very weak, weak or modest [graded according to Fowler et al. (2013)], there are more statistically significant correlations between formants and clinical variables (21 correlations) than between bandwidths and clinical variables (13 significant correlations).

6.4.1.1. Correlation between formants/bandwidths and clinical variables

If we focus on formants, Table 6.3, the highest, though weak, correlations are found with age, height, and cervical perimeter. Surprisingly only two formants are correlated with weight and no one is correlated with BMI or AHI. Considering bandwidths, Table 6.4, again only low correlations are found: four with age, three with weight and two with AHI and BMI. For both the cervical perimeter and height variables, only one significant correlation with a bandwidth was detected.

6.4.1.2. Correlation with age

The highest number of correlations between speech measurements and clinical variables was found between formants and age (the highest being that with F_2 in /o/, a Spearman's coefficient of $\rho_s = -0.27$). The presence of several correlations between formants and age is in line with existing studies on age-related acoustic characteristics; where different speech measurements, such as fundamental frequency, formants or speech velocity, have been reported to correlate with age and have been linked to changes in anatomy and physiology of the speech mechanism (Torre, P III & Barlow, 2009). Some specific studies have reported age-related changes to formants, particularly in the production of vowels. According to these studies, a negative correlation among formants and age, as is also found in our study, can be expected (Torre, P III & Barlow, 2009). This lowering of vowel formants with age can presumably be a by-product of the lowering of the vocal folds over the lifespan, which results in a longer vocal cavity (Benjamin, 1997; Xue & Hao, 2003), as well as with a trend to vowel centralization in older speakers (Liss, Weismer & Rosenbek, 1990; Rastatter, McGuire R & Kalinowski J, 1997). In some cases, these changes have been found to occur only on particular vowels (Rastatter et al., 1997; Xue & Hao, 2003).

In records from our OSA population, a decrease in formant values with age was observed for six formants (in decreasing order of the Spearman correlation coefficient): F_2 in /o/ and /u/, F_1 and F_2 in /a/, F_1 in /e/ and F_1 in /i/. This latter result exhibits a similar Spearman coefficient as the one reported in (Torre, P III & Barlow, 2009). Even though these results are consistent with those previously reported for general populations, two considerations must be made when analyzing the impact of age on OSA voices. First, most of the studies on voice aging generally compare young to elderly speakers, while OSA is most frequently associated to an age-range better described as middle-aged to elderly; in our case, from 21 to 78 years (average 47.6 years, SD 11.6). Second, research on craniofacial risk factors for OSA using 3D-MRI (J. W. Young & McDonald, 2004) has

reported an inferior position of the hyoid bone in apnea patients that correlates with enlarged tongue volumes and longer vocal tracts. Therefore research on the acoustic effect of vowel formant lowering associated to longer vocal tracts (for example [Lammert & Shrikanth, 2013](#)) when applied to OSA voices should consider not only OSA-specific anatomic features but also the particular age range in OSA populations.

6.4.1.3. Correlation with height, weight, and BMI

Besides correlations with age, ten speech variables also showed significant correlations with height, but only two did so when correlated with weight. Similar results on weak relationships between body size and speech formants have been already reported by González for a general population ([Gonzalez, 2004](#)). However, in González's research stronger correlations were reported for male speakers, mainly between F_2 /e/ and height ($\rho_s = -0.57$) and F_4 /o/ and weight ($\rho_s = -0.48$); whereas in our OSA population the higher correlation coefficients' values were obtained between F_2 /i/ and height ($\rho_s = -0.21$), and F_2 /e/ and weight ($\rho_s = -0.15$). Thus, despite differences between our results and those from González, both point towards a similar direction: formants seem to be weak predictors of body size. Just as in the previous discussion regarding age, it is possible to hypothesize that these significant though weak correlations with height or weight may interfere with specific acoustic characteristics related to OSA.

The weight-height relationship measured through the BMI is known to be a relevant indicator for OSA ([Jamie et al., 2010](#)). From [Table 3.5](#), our OSA population is clearly overweight. Nevertheless, no significant correlation was detected between formants and BMI. This could be due to a nonlinear relationship which breaks the correlation between BMI and OSA.

One may expect formant bandwidths to be larger for OSA speakers as an increase in both velar and pharyngeal compliance could result in increased sound damping within the vocal tract ([Robb et al., 1997](#)). However, only two significant correlations between bandwidths and BMI were obtained, and both were negative: BMI correlation with BW_3 /e/ ($\rho_s = -0.13$) and BW_2 /o/ ($\rho_s = -0.14$).

6.4.1.4. Correlation with cervical perimeter

The cervical perimeter is another feature that is commonly used in discriminating between healthy subjects and OSA patients. More specific than BMI, cervical perimeter can describe how excessive weight may increase tissue bulk in the neck, which will also increase the dynamic loading of the airway thus contributing to the pathogenesis of OSA ([Katz, Stradling, Slutsky, Zamel & Hoffstein, 1990](#)). In the OSA population under study, and similarly to what we have found for the other body-size measurements, only a few significant and also weak correlations appeared between cervical perimeter and speech: with F_3 /i/ ($\rho_s = 0.2$) and BW_2 /a/ ($\rho_s = 0.14$). Several previous studies have similarly failed to find obvious relationships between voice acoustics and body-size effects measured through BMI ([B. J. Lee, Ku, Jang & Kim, 2013](#)), body mass composition ([Hamdan et al., 2013](#)) or weight and cervical perimeter ([Rendall, S. Kollias & Lloyd, 2005](#)). Nevertheless, in our study, five significant correlations were found between voice features and weight (2 with formant

frequencies and 3 with formant bandwidths), also five with cervical perimeter (4 frequencies, 1 bandwidth) and two with BMI (0 frequencies, 2 bandwidths). These results suggest a slight increase in voice to body-size relations in OSA populations due to the characteristic presence of extreme cases of over-weighted subjects.

6.4.1.5. Correlation with AHI

Considering now the correlation between speech variables and AHI, no significant correlations were found with formant frequencies and only two very weak correlations were detected with bandwidths: BW_2 /a/ ($\rho_s = 0.13$) and BW_3 /e/ ($\rho_s = -0.17$). These results may be contrasted with those in the main previous reference to our study (Robb et al., 1997); though in that work, Robb et al. only could evaluate first and second formants frequencies and bandwidths from vowels /u/, /i/ and /a/, measured on a rather limited population (10 male patients with diagnosed OSA and 10 non-OSA male subjects). In Robb's research, a discriminant analysis of formant frequencies between groups of healthy and OSA subjects discovered significant differences in F_1 values for vowels /i/ and /a/, and F_2 values for /i/ and /u/. Also in (Robb et al., 1997) formant bandwidths BW_1 /i/ and /u/, and BW_2 /i/, /u/ and /a/ were found to be significantly different for both groups.

6.4.1.6. Contrast among OSA groups without controlling any variable

As the Apnea Database 2.0 contains the AHI value for every OSA patient, we can now stratify our apnea population to study every stratum of the disease. To compare our results to those by Robb, Table 6.5 provides statistical information from the U Mann Whitney tests on three OSA groups extracted from our database: one group for $AHI \leq 10$, another including subjects with mild AHI between 10 and 30, and a third group for severe $AHI > 30$.

Looking at results in Table 6.5, most of the discriminative voice features reported in Robb's research are not detected in our study. Only the significant difference in BW_2 /a/ is present, while a few novel differences arise for F_3 /i/, BW_1 and BW_3 /o/ and BW_3 /e/. The discrepancy between these results can be attributed to several factors, such as the small sample size in Robb's analysis (10 OSA and 10 no-OSA subjects) or the broad AHI range (from 9.3 to 109.5) for OSA patients that Robb et al. compared to a control healthy group; while in our case the control group includes OSA subjects with AHI below 10.

Results in Table 6.5 also suggest that the relationship between speech and the apnea syndrome can be very different for each of the stratified groups; as most of the discriminative speech variables are not common to both OSA groups (mild and severe). This partially supports recent research presented in (Solán-Casals et al., 2014) where the capabilities of voice features to detect severe cases ($AHI > 30$) vanish when trying to detect mild OSA speakers (AHI between 5 and 30).

Table 6.5.: Contrast among OSA groups. p -values are from the Wilcoxon rank-sum test.

	Non OSA AHI ≤ 10 ($n = 81$)	Mild OSA $10 \leq \text{AHI} \leq 30$ ($n = 87$)		Severe OSA $30 < \text{AHI}$ ($n = 73$)	
Measure	Median	Median	p -value	Median	p -value
Weight	86	90	NS	99	0.00*
BMI	27.31	28.93	0.01*	31.46	0.00*
Age	42	51	0*	48	0.02*
Cervical perimeter	41	42	0.00*	44	0.00*
F_3 /i/	2707	2682	0.05*	2642	0.03*
BW ₁ /o/	94	79	0.00*	85	NS
BW ₃ /o/	98	136	0.01*	107	NS
BW ₂ /a/	118	125	NS	148	0.01*
BW ₃ /e/	170	201	NS	140	0.03*

Abbreviation: NS, nonsignificant.

* There are significant differences between OSA groups at the 0.05 level.

6.4.1.7. Contrast among the most extreme groups controlling for age and height

Furthermore, based on the weak relationships between voice features and OSA that we report in this work, it may be suggested that future studies on voice features and OSA should evaluate the possible underlying impact of clinical variables such as height and age that, as it has been shown in this work, can influence speech measurements.

To illustrate this, Table 6.6 shows the same statistical contrasts as in Table 6.5, but now for two subgroups of subjects with $\text{AHI} \leq 10$ (i.e. control) and $\text{AHI} > 30$ (i.e. severe) selected to present no significant statistical differences regarding age or height; thus reducing the dependency between age and height and speech.

As it can be seen in Table 6.6, once the possible effect of age and height in speech is minimized, only significant differences on BW₂ /a/ and BW₃ /e/ remain, while the difference for F_3 /i/ reported in Table 6.5 disappears.

Although we did not find a significant correlation between AHI and F_3 /i/, correlation between AHI and cervical perimeter is $\rho_s = 0.42$ (Table 6.2) and between cervical perimeter and F_3 /i/ $\rho_s = -0.2$ (Table 6.3). Hence, an underlying connection between AHI and F_3 /i/ through the cervical perimeter may actually exist.

Consequently, and to reiterate, major results in our study indicate that correlation coefficients between AHI and clinical variables (shown in Table 6.2) are stronger than those with speech variables, and that some clinical variables present higher correlations with voice features than with AHI. If this is so, contributions linking voice and OSA shall carefully explore and report on possible indirect influences of speech and AHI mediated through these clinical variables. This

Table 6.6.: Contrast among control and severe OSA groups as in Table 6.5, but on a subset without differences neither in age nor in height.

Measure	Non OSA, AHI ≤ 10 , ($n = 73$)	Severe OSA, $30 < \text{AHI}$, ($n = 65$)	p -value
	Median $\pm$ SD	Median $\pm$ SD	
Weight	85 $\pm$ 11.1	99 $\pm$ 22.65	0.00*
Height	174 $\pm$ 7.26	176 $\pm$ 6.61	NS
BMI	27 $\pm$ 3.79	32 $\pm$ 6.63	0*
Age	43 $\pm$ 9.51	47 $\pm$ 8.55	NS
Cervical perimeter	41 $\pm$ 2.46	44 $\pm$ 3.7	0.00*
BW ₂ /a/	116 $\pm$ 40.31	148 $\pm$ 129.4	0.00*
BW ₃ /e/	161 $\pm$ 98.23	137 $\pm$ 94.6	0.03*

Abbreviation: NS, nonsignificant.

* There are significant differences between OSA groups at the 0.05 level.

same consideration should be paid to any other factor that could affect speech such as speakers' dialect, gender, mood state etc. As an example, in (Solán-Casals et al., 2014) OSA and non-OSA groups contain very different percentages of male and female speakers, and therefore, considering the notable differences between male and female voices, this is likely to bias the results obtained.

6.5. Follow-up Discussion with Doctor Kemaloglu and Doctor Mengü

6.5.1. Letter to the Editor

The results shown and discussed so far in this chapter were published in the Journal of Voice (Montero Benavides et al., 2016a) and have already received some citations.¹ In particular, our research got the attention of Professor Doctor Yusuf Kemaloglu and Doctor Mengü from Gazi University in Ankara, Turkey. They raised the question whether lower formant frequencies of the vowels could only be expected in subjects with Obstructive Sleep Apnea. This interesting discussion was promoted in a Letter to the Editor of the Journal of Voice (Kemaloglu & Mengü, 2016).

Basically, Doctor Kemaloglu and Doctor Mengü wanted to discuss if our previous results of “no significant and meaningful association between AHI and formant values” could be explained by using the background of speech science. Hence, in the view of the data presented in this Chapter and in our contribution (Montero Benavides et al., 2016a), they considered the question whether a longer vocal tract could be the sole explanation for our data in the subjects with OSA.

Kemaloglu and Mengü's hypothesis. Similar to our own reasoning, the authors of the Letter to the Editor argue that both speech and OSA are complex processes of the upper airway. There,

¹<https://scholar.google.de/scholar?cluster=2065746356089770200>

the soft-tissue structure, obesity and muscular tension factors which are assumed to cause OSA affect the cross section of the upper airway and cause constrictions at the tongue base, the tonsillar region, and the soft palate. The first constriction place within the supraglottal airway is the region just above the larynx, occupied with the tongue base (Davidson et al., 2005; Maltais et al., 1991; Sakakibara et al., 1999). Both the static shape of the vocal tract and dynamic processes during speech play a role.

Formant frequencies in OSA subjects are shifted by two competing effects, namely

- OSA patients are believed to have longer vocal tracts [section 2.7.2, Robb et al. (1997)], which lowers the formant frequencies [section 2.6.1 and Eq. (2.1)].
- The constrictions mentioned above are primarily located at the tongue base, i.e., close to a volume-current node, at least for the low formants [Figure 2.6]. Such constrictions at nodes increase the formant frequencies.

Both effects annihilate each other, at least partially, which explains why it is so difficult to detect correlations of formant frequencies with AHI (no AHI correlation in Table 6.3).

Thus, Kemaloğlu and Mengü suggest that the differences between the formants appear to be more important because

- (i) as it is known, longer vocal-tract length decreases the gap between the formants (Behrman, 2007) and
- (ii) the location of the constrictions (near or far to the nodes and antinodes) and the inability of the tongue/tongue base during phonation due to its size or intrinsic muscular factors, etc. can change the formant differences, as reported in some of the previous studies (Fernández-Pozo et al., 2009a; Goldshtein et al., 2011; Montero Benavides et al., 2014).

Kemaloğlu and Mengü's suggestions for further investigations are twofold.

1. First, they propose to use a different threshold between 'healthy' and 'mild OSA' $AHI = 5$ (instead of $AHI = 10$) which, according to them is a more conventional value. They suggest searching for correlations only within the OSA group ($AHI > 5$) instead of the whole population (no AHI restriction).
2. Instead of investigating the formant frequencies directly, consider differences of adjacent formant frequencies, e.g., $F_2 - F_1$.

6.5.2. Reply: Additional Results

After the interesting discussion promoted by Professor Doctor Kemaloğlu and Doctor Mengü, we wrote a reply (Montero Benavides et al., 2016b) to the Letter to the Editor (Kemaloğlu & Mengü, 2016) regarding our article "Formant Frequencies and Bandwidths in Relation to Clinical Variables in an Obstructive Sleep Apnea Population" (Montero Benavides et al., 2016a), following their suggestions. Hereafter we show some additional tables with the suggested results.

They had proposed exploring correlations within only an OSA group defined for subjects with

Table 6.7.: Contrast among control and severe OSA groups on a subset without differences neither in age nor in height. p -values are from the Wilcoxon rank-sum test. With respect to Table 6.6, the threshold in the first group is changed, and five subjects with incomplete information have been removed, see section 6.5.2 on page 110.

	AHI < 5 ($n = 44$)	AHI > 30 ($n = 64$)	
Measure	Median $\pm$ SD	Median $\pm$ SD	P value
AHI	2.45 $\pm$ 1.57	43.05 $\pm$ 14.85	0.00**
Weight	86 $\pm$ 14.3	103.5 $\pm$ 22.34	0.00**
Height	176 $\pm$ 8.12	176.5 $\pm$ 6.47	0.87
BMI	27.5 $\pm$ 4.21	31.77 $\pm$ 6.63	0.00**
Age	41.5 $\pm$ 11.41	46 $\pm$ 9.94	0.12
PerCerv	41 $\pm$ 2.99	44 $\pm$ 3.62	0.01*
F_3 /i/	2758.66 $\pm$ 193.48	2651.69 $\pm$ 151.24	0.01*
F_3-F_1 /i/	2433.78 $\pm$ 205.94	2341.83 $\pm$ 157.85	0.01*

* Significant differences at the 0.05 level.

** Significant differences at the 0.01 level.

Abbreviations: BMI, body mass index; SD, standard deviation.

AHI over 5. However, we ran such analysis without obtaining any relevant additional information. We believe that, as concluded in (Solán-Casals et al., 2014), for intermediate AHI cases, alterations in the upper airway may not be so pronounced to have an impact on speech characteristics. Therefore, we find it more informative to present results, as suggested also by them, exploring group differences (using Mann-Whitney U-tests) between formants, formant differences and clinical variables for non-OSA (now using AHI < 5 instead the cut-off of 10 used before) and OSA (AHI > 30) groups. These results are presented in Table 6.7. In this Table, trying to avoid the indirect effect of age and height in speech, non-OSA and OSA groups were selected without significant differences regarding age and height (as it was already done for results in Table 6.6 above). Again for the sake of clarity, Table 6.7 only shows those formants or formants differences delivering statistical significance at the 0.05 level between these most extreme groups (i.e., AHI < 5 and AHI > 30), and also we have included the p -values of height and age (0.87 and 0.12 respectively, i.e., non-significant).

Five subjects in our dataset (section 3.3.1.1) lack some clinical information, and we have decided to exclude them from the results we report here. That is why the severe-OSA group (AHI > 30) in Table 6.7 has one subject less than the same group in Table 6.6.

We have re-computed the correlations from Table 6.3 on this slightly reduced set in Table 6.8, which is identical to our previous finding, apart from the fact that four correlations that used to be just at the p -value threshold are now slightly over the threshold and have been dropped from the table. As suggested by Kemaloğlu and Mengü, Table 6.9 shows Spearman's correlations of formant *differences* with clinical variables, extending our previous analysis.

Table 6.8.: Spearman correlations ρ_s of formants and clinical variables (p -values in parentheses). With respect to Table 6.3, five subjects with incomplete information have been removed, see section 6.5.2 on page 110. (From our reply [Montero Benavides et al., 2016b](#))

Vowel	Formant	Weight	Height	Age	Cervical Perimeter
/a/	F_1			-0.14* (0.03)	
/e/	F_1				-0.14* (0.03)
	F_2	-0.15* (0.02)	-0.20** (0.00)		-0.16* (0.02)
	F_3		-0.21** (0.00)		
/i/	F_1		-0.15* (0.02)	0.16* (0.01)	
	F_2		-0.21** (0.00)		
	F_3				-0.20** (0.00)
/o/	F_1				-0.13* (0.05)
	F_2			-0.27** (0.00)	
	F_3		-0.17** (0.01)		
/u/	F_1		-0.14* (0.03)		
	F_2			-0.24** (0.00)	
	F_3	-0.14* (0.04)	-0.20** (0.00)		

Notes: For clarity, nonsignificant correlation values are omitted.

* Correlation is significant at the 0.05 level (two-tailed).

** Correlation is significant at the 0.01 level (two-tailed).

Table 6.9.: Spearman correlations ρ_s (p -values in parentheses) of formant differences and clinical variables. (From our reply [Montero Benavides et al., 2016b](#))

Vowel	Formant	Weight	Height	Age	Cervical Perimeter
/a/	$F_3 - F_2$				-0.14* (0.03)
	$F_2 - F_1$				0.13* (0.05)
/e/	$F_3 - F_1$		-0.18** (0)		
	$F_2 - F_1$	-0.14* (0.03)	-0.16* (0.01)		
/i/	$F_3 - F_1$				-0.17** (0.01)
	$F_2 - F_1$		-0.17** (0.01)		
/o/	$F_3 - F_2$		-0.17** (0.01)	0.14* (0.03)	
	$F_3 - F_1$		-0.16* (0.02)		
	$F_2 - F_1$			-0.29** (0)	
/u/	$F_3 - F_2$	-0.13* (0.04)	-0.18** (0)	0.14* (0.02)	
	$F_3 - F_1$	-0.13* (0.04)	-0.17** (0.01)		
	$F_2 - F_1$			-0.27** (0)	

Notes: For clarity, nonsignificant correlation values are omitted.

* Correlation is significant at the 0.05 level (two-tailed).

** Correlation is significant at the 0.01 level (two-tailed).

Table 6.10.: Sign of the change of the frequencies of formants F_1 to F_4 when the cross-sectional area of the tube is slightly decreased with regard to the neutral tract shape. The lip and glottal regions are to the left and right respectively. From (J. Blanco-Murillo & Schoentgen, 2013, Table 1).

Upper Airway Regions	bilabial labiodental		dental / alveolar	post-alveolar palatal		velar		oropharyngeal		hypopharyngeal		epiglottopharyngeal	epilaryngeal	glottal
Formants	A	B	C	D	E	F	G	$\overline{G}$	$\overline{F}$	$\overline{E}$	$\overline{D}$	$\overline{C}$	$\overline{B}$	$\overline{A}$
F1	—	—	—	—	—	—	—	+	+	+	+	+	+	+
F2	—	—	—	+	+	+	+	—	—	—	—	+	+	+
F3	—	—	+	+	+	—	—	+	+	—	—	—	+	+
F4	—	+	+	+	—	—	+	—	+	+	—	—	—	+

6.5.3. Discussion

The analysis of the presented results does not show any variation on the general conclusion of the results from the first part of this Chapter where no significant correlations were detected between AHI and formant frequencies or bandwidths. Results in Table 6.8 evidence the same correlations between formants and clinical variables as in Table 6.3, except for F_2 /a/ and F_1 /e/ of height and age, which slightly exceed the threshold $p = 0.05$ and dropped out. The same number of 17 (very weak and weak) significant correlations now appears between the corresponding formant differences and clinical variables and Table 6.9. Our results do not show any significant correlation between AHI and formant differences.

Also using a cut-off of 5 in AHI (instead of 10 that was used before) we only found significant but weak differences for F_3 from vowel /i/ and the difference $F_3 - F_1$ for /i/ with p -values 0.013 and 0.014 respectively. Moreover, we did not find any significant correlation between BMI and formants or differences between formants as in our previous results.

In analyzing these results, we agree with the comments of Doctor Kemaloglu and Doctor Mengü on that the speech of OSA subjects could be affected not only by a longer vocal tract but also by a narrower or constricted airway. In fact, there is a previous contribution addressing this topic on a more restrictive dataset in (J. Blanco-Murillo & Schoentgen, 2013). As described there, we believe that rather than considering an obstruction at the tongue-base level causing deviations on all formant frequencies of vowels towards higher frequencies, shifts in formants frequencies from speech OSA patients could be better explained by a characteristic pharyngeal narrowing. This is reported on the pathophysiology of OSA patients and was consistent with the results obtained from continuous speech analysis via long-term average spectrum. This can be concluded from (J. Blanco-Murillo & Schoentgen, 2013, Table 1), which was built following the study of speech production from (Mrayati, Carré & Guerin, 1988), and which we quote in Table 6.10. Table 6.10

explains the OSA-related decrease of $F_4 - F_3$ observed in (J. Blanco-Murillo & Schoentgen, 2013) with a constriction at the epilarynx, the oropharynx or at the palate (gray columns), where the oropharynx is the region where constrictions are expected in OSA patients.

Nevertheless, we agree with the observation that the hypothesis of a peculiar structure of formants in OSA speakers should be studied considering more complex effects such as the interaction between the (localized) airway constriction and a longer vocal tract. However, we feel that, considering, for example, some recent studies using medical images in (Hatano et al., 2012), even well-known effects as those of the vocal-tract length on formants may be weaker than predicted by acoustic models.

6.6. Summary and Conclusions of the Chapter

We have carried out a retrospective study on a 241-male-subjects subset suspected to suffer from OSA population. We made use of the clinical variables usually reported as predictors of OSA: BMI, cervical perimeter, height, weight, and age. The AHI obtained from overnight PSG is available for every subject as ground truth. Voice samples of sustained phonations of the vowels /a/, /e/, /i/, /o/ and /u/ were analyzed. The approach via sustained vowels mitigates the impact of phonetic variability due to the speakers' dialect, which would largely affect continuous speech records. Formant frequencies F_1 , F_2 , F_3 and bandwidths BW_1 , BW_2 , BW_3 of the 5 sustained vowels were determined using spectrographic analysis. Correlations between AHI, clinical parameters, and formants, differences of adjacent formants and bandwidths were determined by nonparametric methods.

Correlations between AHI and clinical variables were stronger than between AHI and voice features. AHI only correlated poorly with BW_2 of /a/ and BW_3 of /e/. A number of further weak but significant correlations has been detected between voice and clinical variables. There were only two significant correlations with BMI, BW_3 of /e/ and BW_2 of /o/. Also using a cut-off of 5 in AHI instead of 10 that was used before we only found significant but weak differences for F_3 from vowel /i/ and the difference $F_3 - F_1$ for /i/.

The main conclusion of our analysis is that in our Apnea Database 2.0, no significant correlations were detected between AHI and formant frequencies and bandwidths. Correlations between voice and other clinical factors characterizing OSA were weak, but highlight the importance of considering indirect or mediated effects of such clinical variables in any future research on speech and OSA.

6.6.1. Conclusions

There is a poor correlation between formant frequencies and bandwidths and AHI, the feature on which we rely on measure the severity of apnea patients' condition. Interestingly, we have found some significant relationships between voice and clinical variables (age, height, weight, BMI and cervical perimeter). Consequently, future research should address indirect or mediated effects of clinical variables when studying the connections between voice and AHI, for example, through

Table 6.11.: Summary of experiment on OSA and formants

	negatives	input type	finding for OSA group
Robb et al. (1997)	healthy	sustained vowels	lower F_1 , F_2 , higher BW_1 , BW_2
J. Blanco-Murillo and Schoentgen (2013)	possibly other conditions	whole sentence	higher F_3 , lower F_4
Our Chapter 6	possibly other conditions	sustained vowels	poor correlations, higher BW_2 of /a/, lower BW_3 of /e/

the analysis of matched populations, the use of speaker normalization techniques or by defining specific statistical models able to isolate the different sources of variability.

Our results show that other physical characteristics of speakers (age, height, weight, state of mood, tiredness, etc.) also impact in the acoustic features of speech, as it was also found in a number of studies such as the work of [Gonzalez \(2004\)](#); [Poorjam, Bahari, Vasilakakis and Van hamme \(2015\)](#); [Bahari, McLaren, Van hamme and van Leeuwen \(2014\)](#). These sources of variability can be extended to other factors that range from socio-dialect varieties of a language to the acoustic effects during the speech production and recording process (e.g. loudness of variability during phonation, vowel prolongations or the influence elongation and rounding of the lips namely /u/ and /i/). Considering this complex scenario, we can conclude that the success of the acoustic-analysis assessment in speech pathology detection relies on the strong acoustic effects of the underlying pathologies (dysphonia, laryngitis, vocal-cord paralysis, ...) above other sources of variability. Hence, despite the possible impact of a peculiar upper airway in OSA voices, this is extremely difficult to isolate. The same conclusion was reached in ([Espinoza-Cuadros et al., 2016](#)) after reviewing previous studies on OSA assessment via continuous speech and state-of-the-art speech processing and machine learning techniques. For all these reasons we believe that is very interesting to go deeper into this complex topic to develop and improve the current research on this field.

In the real-practice scenario of Apnea Database 2.0, we have *neither* confirmed the finding of lower formant frequencies (F_1 , F_2) and wider bandwidths in OSA patients ([Robb et al., 1997](#)), *nor* have we recovered the increase of F_3 and the decrease of F_4 (the latter was not determined here). Possible explanations are differences in the data, in particular: (i) the control group, which was true healthy in these references and possibly included other breathing-related conditions in our study and (ii) spectrum of a whole spoken sentence ([J. Blanco-Murillo & Schoentgen, 2013](#)) vs. sustained vowels, as summarized in [Table 6.11](#).

In this chapter, we have learned that in real-practice scenario is much more complicated than the binary decision OSA or healthy. In this context, simple rules as found in ([Robb et al., 1997](#); [J. Blanco-Murillo & Schoentgen, 2013](#)) simply do not apply. Still, there is information in the bandwidths, which may contribute to an OSA decider, but which is by itself not sufficient for the decision.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

In this last chapter, we will summarize the work done throughout the whole Thesis. We will briefly recapitulate the motivation and purpose for this work, as well as the specific objectives and the three different approaches taken. We will summarize the major findings obtained from our experiments, draw conclusions and finally we explain recommendations for possible future research. We conclude with a description of the publications and awards achieved with this scientific research line.

7.1. Motivation of this Thesis Summarized

Sleep has been a fascinating subject for artists since antiquity but it was not until the last decades that it became a focus of research in the medical field. OSA is one of the most typical forms of sleep disorders that nowadays affect the population all around the world. As we have seen throughout this Thesis, obstructive sleep apnea (OSA) is a very complex disease with many factors involved, which make it difficult to diagnose. Left untreated, OSA leads to serious health consequences and risks of accidents, which make the need for alternative diagnosis and screening techniques so urgent.

The first cue about this disease was already presented by Charles Dickens in his famous novel *The Pickwick Papers* through the detailed description of the character Fat Joe (see [section 2.2](#)). Thus, Dickens introduced without knowing the typical apnea patient, someone fat who is so tired during the day that he falls asleep at any moment. Although this disease was introduced a long time ago, understanding it in depth is still an open question nowadays.

From the review of the pathogenesis of OSA, we know that OSA patients have some craniofacial signs, such as big cervical perimeter, short neck, fat tissue around the belly, among others, which predispose them to suffer from this disease. These patients also have a peculiar upper airway configuration (skeletal structure and surrounding tissues) which explain the recurrent partial (hypopnea) or total (apnea) obstruction episodes at the level of the pharynx, which is more prone to collapse than that of a healthy subject. To overcome the collapse and reopen the upper airway, patients awake for a few seconds, just enough to restore normal breathing. As these micro awakes

keep repeating during the whole night, it is impossible for them to rest properly. The syndrome affects the patients' quality of life and health condition, as cardiovascular disorders become more prevalent in the OSA population, and reported symptoms such as daytime sleepiness cause work impairment and traffic accidents.

From the time, the patient decides to consult a doctor for the first time until he gets his turn for the polysomnography test, which is considered the gold standard for its detection, it is very common that he has to wait for one year in developed countries such as Spain (Puertas et al., 2005). The polysomnography test is resource intensive, it requires a full overnight study of patients, held at the hospital, and the collection and posterior processing of numerous signals during at least 6 hours of sleep (see section 2.5). Once the patient is diagnosed with OSA, losing weight and the use of a continuous positive airway pressure (CPAP) mask will be recommended to avoid the apnea episodes during sleep (see section 2.5.2).

According to T. Young et al. (1997), there are more than 100 million of people worldwide estimated to have OSA of which the vast majority is undiagnosed. Moreover, OSA is a common disease with a high prevalence, affecting over 4% of male and 2% female population between ages 30 to 60, and related to domestic, occupational and traffic accidents and working impairment, its diagnosis currently is extremely expensive and time-consuming. Because of all these reasons, OSA has lately become an important concern for public health that raises several challenges regarding its diagnosis, assessment, and treatment. In order to overcome the long waiting times, alternative screening methods are needed to prioritize waiting lists. This is where the approach of this Thesis comes into play, namely the detection of OSA via statistical analysis of the patient's speech. Such techniques would be fast, cheap and non-invasive and would contribute to a fast and efficient diagnosis of OSA.

The connection of OSA and speech was first observed in the late 1980s and has been under study ever since. However, even nowadays it is still not well understood. Nevertheless, recently it has been shown that speech processing techniques may exploit the differences that have been observed between healthy subjects and severe OSA patients, and provide useful information for clinicians. Moreover, it has been shown that this information can be combined with clinical predictors to enhance the performance of this non-intrusive, simpler and reliable detection scheme (see section 2.7).

In this Thesis, we have advanced in this research line and studied not only the speech, but also some clinical variables related to OSA and speech to get a deeper understanding of their relationship. We have added text-dependent information to the statistical analysis, we have selected discriminative speech features for the screening problem, we have analyzed clinical variables and speech from a real clinical practice scenario.

7.2. Summary of Contributions and Results of this Thesis

The motivation of this Thesis consists in contributing to future screening techniques (or stratification techniques) for OSA based on speech samples of the patients. In order to get a deeper

understanding of the connection between OSA and speech, three different approaches were considered in this Thesis:

1. The first approach consisted in adding text-dependent information to the speech analysis of OSA patients using hidden Markov Models (HMMs). To the best of our knowledge, it was the first time that such text information was used in OSA detection via speech. For this purpose, we employed two different kinds of models on every sentence: whole-sentence HMMs and sentence-dependent phonemes. Both outperformed the text-independent baseline system.
2. The second approach was to select and evaluate 16 different acoustic speech measures related to articulation, phonation, and resonance. Some of them had already been proved to be involved in the OSA speech patients. We employed LDA and MLR models to discriminate with a subset of these features for classification as OSA or healthy subjects.
3. Finally, the last approach consisted in analyzing basic speech measures, such as formants and bandwidths, on a more complex database recorded at a sleep unit on a real clinical practice scenario, the Apnea Database 2.0. It contained the speech of OSA patients and other patients that were evaluated in the sleep unit as candidates for OSA. This means that even if they were not diagnosed with OSA, they could not be considered as a healthy control group either, since they were likely to suffer from other diseases, which made our analysis more difficult. This new scenario contained more patients and also more clinical data recorded per patient.

The contributions resulting from the three approaches are very different and have been tested in two different environments; the first two contributions were tested on Apnea Database 1.0 and the third one on Apnea Database 2.0. In general, our research was conditioned by the limited availability of data on apnea speakers. For this purpose, we have focused on the analysis of the OSA syndrome in our two different databases, Apnea Database 1.0 and Apnea Database 2.0 and exploit their nature for our experiments and study, considering these two different environments. More details about these databases can be found in [chapter 3](#).

7.2.1. Results of Contribution 1: Hidden Markov Models

The key idea of the first approach was adding context-dependent information to the existing approaches based on Gaussian mixture models (GMMs), which mostly disregard the linguistic context information. From the MFCC representation, experimental results were obtained using HMMs of two different architectures. The architecture “sentence-dependent phonemes” achieved a moderate relative improvement compared to the GMM baseline system. However, better results were achieved for whole-sentence HMMs with 40 states, producing an overall 11.6% relative reduction in equal error rate (EER). These 40 states configuration could be related to the number of different phonemes in each of the four sentences.

In order to have a global figure-of-merit using all the available audio data for each speaker in our database, LLR scores obtained from the three repetitions of the four sentences were fused (i.e. averaged) for every speaker and used, following a leave-one-out protocol, for severe OSA/control classification. These results showed that the best classification performance is again achieved

when using the whole-sentence HMM classifier, providing a 25% relative reduction in the EER value compared to the baseline GMM system.

These sentence-dependent models proved to be able to combine the potential of a flexible representation for specific linguistic contexts in each sentence together with an efficient training procedure with a limited amount of training data.

In conclusion, this first contribution offers an innovative perspective on how HMMs can be used to model context-dependent phonetic information for detecting voices of speakers suffering from severe OSA syndrome. We have achieved better classification rates than with the baseline system based only on GMMs, as explained in detail in [chapter 4](#).

7.2.2. Results of Contribution 2: Voice Features

We also got satisfactory results with the second of our approaches analyzing different acoustic variables and building an LDA model with a subset of 8 features instead of the 16 variables available. Our results on the discriminative power of individual OSA-related voice features lead to the selected features were SDRseg, a1h1max800_{diff}, HNR, Jitter, $(F_3 - F_2)_{/i/}$, Duration4, PercSil1, and Duration1. We highlight that this set contains

- classical measurements used in the characterization of pathological voices (such as SDRseg, HNR and Jitter),
- a feature already proposed for the specific problem of apnea detection, namely the difference between the third and second formants for the /i/ phoneme, $(F_3 - F_2)_{/i/}$ proposed by [Fernández-Pozo et al. \(2009a\)](#),
- and some new features (although already hypothesized in related works), namely the difference in nasality measured between vowels in nasal and non-nasal contexts, a1h1max800_{diff} and the prosodic features Duration and PercSil.

OSA detection based on a selected subset of 8-best speech features and a LDA model has been assessed. Our results outperformed the classification rates that are commonly achieved when using only BMI and Age as screening variables for the classification of OSA patient and healthy. Classification results reached an overall performance of over 80%, similar to other OSA screening strategies based on clinical data such as BMI and age ([Gurubhagavatula et al., 2004](#)). Given that speech features could be related to BMI and age, which are known to be related to apnea, classification results using the selected speech features have been compared to those obtained using only BMI and age. Results using the voice-based approach on a test database with no statistical differences in BMI and age distributions, and showing no significant correlations between speech features and both factors, have provided a F1-measure value 20% higher in absolute terms compared to using only BMI and age factors. To probe the validity of our model we tested it also on a separate subgroup containing 20 healthy and 20 apnea subjects yielding a sensitivity of 85% and a specificity of 75%, with a F1-measure of 81% thus demonstrating the effectiveness of the methodology. This result can lead us to conclude that the proposed voice measures have the potential to improve OSA detection on segments of population where age and BMI have no discriminative

power.

We consider that results in this work are encouraging and clearly show that the distinctive apnea traits identified by our incremental feature selection method can be combined to provide reasonable diagnosis performance for screening purposes. Furthermore, such promising outcomes were obtained with just 8 speech features and with very simple and basic classification models. In previous works on the same database ([Fernández-Pozo et al., 2009a](#)), we reported results slightly worse than those presented here, but following a completely different approach.

7.2.3. Results of Contribution 3: Formants and Bandwidths

Finally, in the third contribution, we have analyzed formants, differences of adjacent formant frequencies and bandwidths on the Apnea Database 2.0. In this contribution, we have evaluated on the full AHI range. Our conclusions showed that through the simple analysis of basic speech measures extracted from sustained vowels, we could also achieve very interesting conclusions. Surprisingly, we found only poor correlations between the formant frequencies and bandwidths with the AHI, the feature on which we rely to measure the severity of apnea patients' condition. No significant correlation was found between formants and BMI, which is one of the most common measures used as screening tool for OSA patients. Interestingly, we found some significant relationships between voice and other clinical variables (age, height, weight, BMI and cervical perimeter). We found 21 significant correlations between formants and clinical variables and 13 significant correlations between bandwidths and clinical variables, as summarized in [Figure 6.2](#).

The clinical variables that got most significant correlations with speech measurements were height and age, with a total of 10 correlations. For each of the clinical variables weight and cervical perimeter, a total of 5 significant correlations with speech measures were obtained.

We have also focused our analysis on every stratum of our population regarding the severity of the disease, and we have compared the three groups (non OSA, mild OSA, and severe OSA). From these experiments, we have observed that the disease has a different behavior in every stratum. The discriminant variables between non OSA and mild OSA, and between non OSA and severe OSA are not always the same. The clinical variables BMI, weight, age and cervical perimeter showed significant p -values. The only speech measure that remains significant in the contrasts was F_3 /i/. Between non OSA and mild OSA the first and third bandwidth of /o/ resulted significant. Between mild OSA and severe OSA the second bandwidth of /a/ and the third bandwidth of /e/ was significant.

After having seen the influence on speech of the variables age and height, we have done a new contrast between non OSA and severe OSA groups. Once their impact is minimized, the only significant differences that remain between these extreme groups are the second bandwidth of /a/ and the third bandwidth of /e/. Although no significant correlation was found between AHI and F_3 /i/, there was a correlation of 0.42 between AHI and cervical perimeter and a correlation of -0.2 between cervical perimeter and F_3 /i/, which points at a possible underlying connection between AHI and F_3 /i/ through the cervical perimeter. Our main conclusion for this study is the important role of the influence of the clinical variables and other factors on speech measurements.

Before going deeper into the analysis of OSA and speech, it is essential to control all the possible effects of variables that may have an impact on speech.

7.2.4. Limitations

All results presented in this Thesis are limited to the speech from speakers of a single Spanish dialect, namely the Andalusian from the Málaga region; therefore these should be contrasted and compared to those obtained for other languages and dialects.

Finally, this study has only considered male speakers; its extension on female speakers could be especially interesting as apnea disease and its interdependencies with speech are still unclear in women.

Work on Apnea Database 1.0. The first experiments explained in chapters 4 and 5 of this Thesis were made on Apnea Database 1.0, a special database with some limitations such as the size, only 40 subjects in each class (apnea/healthy). Then it has to be taken into account that when trying models on a very limited database the conclusions based on the results may not be extrapolated on larger databases. If also the new database such as Apnea Database 2.0 is of different nature having more variability, this fact will totally change the results and the conclusions will be very different. Moreover, Apnea Database 2.0 contains not only more people but also more variables of every subject which add a great source of variability and makes its analysis and interpretation more complex.

Work on Apnea Database 2.0. Regarding the experiments carried out on our bigger database Apnea Database 2.0 in chapter 6 there are also some remarkable limitations:

First, techniques for automatic estimation of formants frequencies and bandwidths always exhibit limited precision. Being aware of that, we have tried our best to be rigorous based on both current state-of-the-art techniques and a careful review of the estimations. Other more robust spectral representations as, for example, linear predictive coding or MEL Cepstrum coefficients, could also be used.

Regarding the statistical analysis, a matched study on the clinical variables instead of selecting the subset without differences neither in age or height would have been desirable. Unfortunately, we could not do that due to the lack of enough data.

The lack of vocal-tract length measurements for each speaker in our database is another major limitation. This information could have been extracted from medical images and would have added great value to our analysis. Then, the conjecture by Robb et al. (1997), that the differences observed in formant frequencies could be explained by longer vocal tracts in patients with OSA, could have been tested.

Another limitation of this work is that it focuses only on sustained speech. Complementary information extracted from continuous speech, such as formant trajectories or unvoiced sounds, could add new insights on the relationship between voice and OSA.

Moreover, beyond formant frequencies and bandwidths, other speech measurements related to phonation and nasalization could have been explored, as they have been studied in the context of OSA in [chapter 5](#) and previously in the literature ([Fox et al., 1989](#); [Goldshtein et al., 2011](#); [Fernández-Pozo et al., 2009a](#); [Montero Benavides et al., 2014](#)).

7.3. Conclusion

In conclusion, we have shown that the detection of OSA via speech in the controlled scenario *Healthy vs. severe OSA* improves by including text-dependent information (using hidden Markov models) and enhancing the selection of voice features (chapters [4](#) and [5](#), respectively). Using a LDA model based on just 8 speech features, better results have been achieved than using a typical screening method commonly used in clinical practice based on BMI and age.

In the real-practice scenario of a sleep unit, however, it has turned out that the classification problem is more difficult because OSA has to be discriminated from other diseases and not simply against healthy speakers. In this scenario, we have found only poor correlations of the formants (and their bandwidths) with the OSA condition (AHI). We have shown that accepted conjectures from other authors (lower and wider formant resonances) are not confirmed.

The problem of robust OSA detection is not yet solved for the real-practice scenario, but this Thesis has shed light on the interdependencies of OSA, speech and other clinical variables and will help guiding the future research on the connection between speech and OSA.

7.4. Future Work

This Thesis is part of an ongoing effort in which clinicians and researchers from UPM have collaborated for almost a decade. As we have summarized above, the experiments conducted in this Thesis and the conclusions drawn have advanced our understanding of the complex relationship between OSA and speech. However, the story does not end here. In the following, we will outline how the research on this topic may go on.

Getting more out of Apnea Database 2.0

The Apnea Database 2.0 (see [section 3.3](#)) is very rich, and we have only scratched the surface by studying the formant frequencies, differences in adjacent formant frequencies and bandwidths in [chapter 6](#). It would be recommendable to do a perceptual evaluation of the speech material of Apnea Database 2.0. It would be interesting that phonetic experts review this database reproducing the tests described in ([Fox et al., 1989](#)). This phonetic study shall help in validating the database that is being used.

Regarding the large variability of the Apnea Database 2.0 due to a larger number of speakers and more clinical variables collected, our previous contributions [Chapters [4](#) and [5](#), and also the baseline system of [Fernández-Pozo et al. \(2009a\)](#)] will have to be reworked on the newly available

data to decide whether the conclusions presented in this Thesis and previous contributions were accurate or resulted from an inadequate interpretation of the previous results.

Further analysis should be carried out to confirm the results of [chapter 4](#) on a larger database, both in terms of the number of speakers and the speech corpus recorded from each speaker. Besides higher statistical significance, this would yield the possibility to explore the use of other alternative HMM modeling schemes (such as triphones) and to explicitly exploit the different discriminative power of different linguistic contexts.

Need for better data

As said, it is desirable for all methods to work on larger datasets, like our Apnea Database 2.0. However, the amount of the data is only one thing. Good control of the data, in particular, labeling of the clinical conditions of the subjects, is indispensable. Since we assume that our Apnea Database 2.0 contains many subjects with other sleep and breathing disorders, we should consider to improve the data before applying more techniques to Apnea Database 2.0. A possible Apnea Database 2.1 would consist of the OSA cases from Apnea Database 2.0 combined with a group of really *healthy* subjects. Ideally, the possible other health conditions of the subjects of Apnea Database 2.0 would be worked out in order to label the data accordingly. Depending on the outcome, several groups would be formed, leading to a multi-class classification problem.

Indirect dependencies of speech and OSA via other clinical variables

Our results from the analysis of clinical variables, formants, and bandwidths ([chapter 5](#)), have shown that correlations between AHI and clinical variables are stronger than those between AHI and speech variables. Further, some clinical variables (height, age) presented higher correlations with voice features (Tables [6.3](#) and [6.4](#)) than with AHI ([Table 6.2](#)). Thus, the dependencies of speech and OSA seem to be indirect, mediated by other clinical variables. Future research on the connection of voice and OSA shall carefully explore such possible indirect influences of speech and AHI mediated through these clinical variables. The same consideration should be paid to any other factor that could affect speech, such as speakers' dialect, gender, state of mood, etc. All these variables should be taken into account, for example analyzing data which contains very different percentages of male and female speakers, and therefore, considering the notable differences between male and female voices, this is likely to bias the results obtained ([Solán-Casals et al., 2014](#)).

In order to improve the parametrization of OSA-related patterns, it would be desirable to enhance speaker normalization. In particular, adding data on the vocal-tract length for each speaker to our databases would be of great value for our analysis. Differences in formants frequencies were reported already by [Robb et al. \(1997\)](#) and could be well explained by an increased length of the vocal tract in patients with OSA and could be quantitatively tested. The vocal-tract length can be extracted from medical images. To gather the data we would need to collect 3D images of the patients' vocal tracts, which would be a long and expensive process as ideally magnetic reso-

nance images (MRI) of all patients should be provided. Then we could also use this new clinical information combined with the variables that we already have to improve our results.

Future research should address indirect or mediated effects of clinical variables when studying the connections between voice and AHI, for example, through the analysis of matched populations, the use of speaker normalization techniques or by defining specific statistical models able to isolate the different sources of variability.

At this point, it must be noted that, in addition to BMI and age, there are many other clinical data and symptoms in OSA patients, such as snoring, daytime sleepiness, smoking history, drinking habits, mood, etc. that can influence their voice characteristics. Currently, we do not have complete clinical data for the patients in this database, but it will be a line for future research, so this is another example of the need of better data.

Other Languages

It would also be recommendable to evaluate our models on foreign databases. This task is not so easy to be accomplish due to the limited number of speech databases and their restricted access. According to the literature consulted there are only three Spanish speech databases, the two databases that we have analyzed in this Thesis and one more from ([Durán-Cantolla, 2012](#)). We have also found the existence of a Hebrew Apnea Database ([Zigel et al., 2008](#)). However, access to this material is highly restricted and the large differences between this two languages have prevented us from considering such analysis.

We believe that there are other languages more similar to Spanish whose analysis will be more interesting to test our models. Due to the nasalization typical in languages such as French, we believe that it would be more interesting to reproduce the experiments described in this Thesis on French records, and check whether differences in speakers' native languages do actually affect the results. Though, this will require a new corpus to be designed and the participation of speech technology experts and clinicians who may contribute to this effort and be willing to get involved in the collection of new material.

APPENDIX A

SCIENTIFIC CONTRIBUTIONS, COLLABORATIONS, AND PUBLICATIONS

A.1. Peer-Reviewed Scientific Articles

As a result from the work described through this Thesis, several contributions have been published, referring to the connection between OSA and speech.

Montero Benavides et al. (2012) Montero Benavides, A., Blanco Murillo, J. L., Fernández, A., Fernández Pozo, R., Torre Toledano, D. & Hernández Gómez, L. A. (2012). *Using HMM to detect speakers with severe obstructive sleep apnea syndrome*. In D. T. Toledano et al. (Eds.), *IberSPEECH* (Vol. 328, p. 121-128). Springer.

This paper contains the research of [chapter 4](#).

Montero Benavides et al. (2014) Montero Benavides, A., Fernández Pozo, R., Torre Toledano, D., Blanco Murillo, J. L., López Gonzalo, E. & Hernández Gómez, L. A. (2014). *Analysis of voice features related to obstructive sleep apnea and their application in diagnosis support*. *Computer Speech & Language*, 28 (2), 434-452.

This paper contains the research of [chapter 5](#).

Montero Benavides et al. (2016a) Montero Benavides, A., Blanco Murillo, J. L., Fernández Pozo, R., Espinoza Cuadros, F., Torre Toledano, D., Alcázar-Ramírez, J. D. & Hernández Gómez, L. A. (2016). *Formant frequencies and bandwidths in relation to clinical variables in an obstructive sleep apnea population*. *Journal of Voice*, 30 (1), 21-29.

Montero Benavides et al. (2016b) Montero Benavides, A., Blanco Murillo, J. L., Fernández Pozo, R., Espinoza Cuadros, F., Torre Toledano, D., Alcázar-Ramírez, J. D. & Hernández Gómez, L. A. (2016). Reply to the Letter to the Editor by Kemaloglu and Mengü. *Journal of Voice*.

The paper ([Montero Benavides et al., 2016a](#)) and the ensuing exchange of the letter to the editor ([Kemaloglu & Mengü, 2016](#)) and the reply ([Montero Benavides et al., 2016b](#)) were the basis for [chapter 6](#).

Figure A.1.: The ApneApp project.

Espinoza-Cuadros et al. (2014) Espinoza-Cuadros, F., Garcia-Zamora, M. A., Torres-Boza, D., Ferrer-Riesgo, C. A., Montero Benavides, A., Gonzalez-Moreira, E. & Hernández Gómez, L. A. (2014). *A spoken language database for research on moderate cognitive impairment: Design and preliminary analysis*. In IberSPEECH (p. 219–228).

The author of this Thesis has also contributed to this article. The database presented in this article is designed for the early recognition of moderate cognitive impairment (MCI), which affects elderly people and is often a precursor of Alzheimer. The database serves as a starting point to detect language impairment at a very early stage.

A.2. Awards

A.2.1. IberSPEECH 2012

This paper ([Montero Benavides et al., 2012](#)) received the *Best Student Paper* Award of the IberSPEECH conference 2012.

A.2.2. ApneApp: Popularizing automatic severe OSA detection via speech analysis

In 2013, the author of this Thesis and José Luis Blanco Murillo, who later became co-advisor of this Thesis, participated in the competition *IDEA*² announced by the MIT-Madrid *m+visión* consortium. They presented the project “ApneApp: Popularizing automatic severe OSA detection via speech analysis” and were awarded the first prize ([MIT-Madrid M+Visión Consortium, 2013](#)). This participation emerged voluntarily as own initiative in addition to our research work.

The Madrid-MIT M+Visión Consortium is a partnership of leaders in science, medicine, engineering, business, and the public sector dedicated to catalyzing change in Madrid’s healthcare innovation ecosystem by accelerating translational research and encouraging entrepreneurship.

It was founded by Comunidad de Madrid, through Fundación para el Conocimiento madri+d, in 2010, in partnership with the Massachusetts Institute of Technology (MIT).

The competition. First our team had to submit a short pre-proposal for a research or business idea. Following, our proposal was reviewed by a diverse panel and we were selected as IDEA² Madrid finalists. Later we were assigned three “Proposal Catalysts” (mentors) who, over the subsequent 4 months provided advice towards the development of a full proposal. Finally we presented our full proposal orally to a diverse review panel. Eventually our proposal ApneApp was selected as IDEA² Madrid awardees.¹

Our proposal ApneApp: “Popularizing automatic severe OSA detection via speech analysis” ApneApp was born as **an idea** of developing an *app* (mobile-device software application) that could help clinicians to detect the more severe cases of Apnea, namely OSA in order to prioritize diagnosis. The idea was to turn our current solution, designed and evaluated offline on speech and clinical data, into a prototype that could be used during clinical interviews, **Figure A.1**. The prototype shall finally provide clinicians with an index which correlates with the degree of severity.

Current solutions already had achieved over 90% classification rate on severe cases, though these results have been obtained on collected speech samples (recorded under controlled conditions). The developed systems had not yet been merged into a commercial solution, neither is there a prototype to be evaluated by clinicians in real conditions. Additionally, it was thought to expand our studies to other languages, extending the use of speech technologies for clinical disorders detection.

A solution such as the one proposed here shall have a direct impact on the diagnosis of OSA syndrome to reduce and reorganize waiting lists according to patients priority of need. Nonetheless, it could also be used for other pathologies.

Benefits from the competition. The most interesting thing about this Project was to make new contacts and let people know about our research line. In the following months, our Project got the attention and appeared in several press articles and scientific magazines.²

Among the contacts that we made there were what IDEA² Madrid called catalyzers, experts that help us with their expertise on the field to improve our proposal. As part of the program we could interview either in person or via Skype with:

¹<http://mvisionconsortium.org/mvision-programs/idea2madrid/>

²UPM: http://www.upm.es/Investigacion?fmt=detail&prefmt=articulo&id=8830174e84354410VgnVCM10000009c7648a_

Europa Press: <http://www.europapress.es/madrid/noticia-expertos-upm-reciben-premio-idea2-madrid-detectar-apnea-sueno-traves-voz-20140225192255.html>

Con Salud: <http://consalud.es/tecnologia/una-app-para-detectar-la-apnea-del-sueno-gana-el-idea2-madrid-8021>

Redacción Médica: <https://www.redaccionmedica.com/secciones/gestion/idea2-madrid-selecciona-9-finalistas-para-su-programa-de-apoyo-al-desarrollo-de-proyectos-biomedicos-innovadores-5254>

madri+d: <http://www.madrimasd.org/cienciaysociedad/entrevistas/revista-madrimasd/detalleEmpresa.asp?id=528>

- Doctor Med. Nicolás González-Mangado, Fundación Jiménez Díaz, Madrid
- Doctor Med. Arántzazu Álvarez de Arcaya Vicente, Hospital Clínico San Carlos, Madrid
- Doctor Med. Peter Farrell, CEO at ResMed Inc., San Diego, California

After talking to them, we could improve our proposal thanks to the point of view of two doctors that deal with the disease in real practice hospitals and with the CEO of ResMed, one of the most important suppliers of CPAP.

ApneApp was drawn up after the promising results for the controlled OSA-detection scenario of chapters 4 and 5. Because of the difficulties found later in the real-practice scenario (section 3.3, chapter 6), the idea of ApneApp has not been realized yet. Nevertheless, the prize showcases the potential of the idea and leaves ApneApp as a future branch of the research line of this Thesis.

A.3. Other Contributions

Apart from the research line “Statistical analysis of the connection between sleep apnea and speech”, the author of this Thesis, Ana Montero Benavides, has been involved in other applications of statistics and biometrics in medical diagnosis and health care.

A.3.1. Research Stay at IMB Dresden

During the academic year 2013/14 the author of this Thesis received a scholarship to a stay abroad from the Ministerio de Economía, Industria y Competitividad (EEBB-I-14-07917). The stay took place at the *Intitut für Medizinische Informatik und Biometrie* (IMB) in Dresden, Germany. The IMB is a cross-disciplinary institute with diverse research activities at the intersection of medicine, biology, mathematics, statistics, and bioinformatics. This institution belongs to the Medical Faculty of the Technical University (TU) of Dresden. The author of this Thesis made new contacts from the medical field, participated in the seminar series of the institute, and presented a poster titled “Statistical Methods for the Detection of Obstructive Sleep Apnea Using Speech Samples” at the *Workshop on Computational Models in Biology and Medicine* held at TU Dresden in October 2013.

A.3.2. Diagnosis of Multiple Sclerosis

Later, the author joined another research group at TU Dresden working as biostatistician: Starting in June 2015 she carried out research activities in the group of Dr. Hagen Kitzler at the *Institut und Poliklinik für Neuroradiologie* of *Uniklinikum Carl Gustav Carus*, Dresden.

This new research line is related with another complex disease, multiple sclerosis. It is a different problem from the field of data-driven medical diagnostics, namely the analysis of magnetic resonance (MRI) 3D brain images for the diagnosis of multiple sclerosis. In this project, image pattern recognition methods are applied to model every stratum of the disease and to look for differences between Early and Progressive groups.

APPENDIX B

DEFINITIONS AND ABBREVIATIONS

AHI apnea-hypopnea index, measures the number of apnea and hypopnea per hour of sleep

AASM American Academy of Sleep Medicine

Compliance “softness”, i.e., the inverse of stiffness of an organ etc.

CER classification error rate

COPD chronic obstructive pulmonary disease, [section 2.7.1](#)

CPAP continuous positive airway pressure, see [section 2.5.2.2](#)

EER equal error rate

GMM Gaussian mixture model

HMM hidden Markov model [chapter 4](#)

HNR Harmonic-to-noise ratio, [section 5.2.2](#)

LDA Linear discriminant analysis

LLR log likelihood ratio

MLR Multiple linear regression model

OSA obstructive sleep apnea

PSG polysomnography test, conducted in sleep units at hospitals

SD standard deviation

BIBLIOGRAPHY

- Ahmadi, N., Chung, S. A., Gibbs, A. & Shapiro, C. M. (2008). The Berlin questionnaire for sleep apnea in a sleep clinic population: relationship to polysomnographic measurement of respiratory disturbance. *Sleep and Breathing*, 12(1), 39–45. doi:[10.1007/s11325-007-0125-y](https://doi.org/10.1007/s11325-007-0125-y) 19
- Altuna & Saga Policlínica Gipuzkoa, Paseo Miramón 174 20014 Donostia San Sebastián. (2013). Retrieved January 24, 2017 from <http://www.altunasaga.com/patologia-y-tratamiento/garganta/ronquidos-y-sindrome-de-apnea-obstruktiva-del-sueno>. 12
- Ayappa, I. & Rapoport, D. (2003). The upper airway in sleep: physiology of the pharynx. *Sleep Med Rev*, 7(1), 9-33. doi:[10.1053/smr.2002.0238](https://doi.org/10.1053/smr.2002.0238) 2
- Bahari, M. H., McLaren, M., Van hamme, H. & van Leeuwen, D. A. (2014). Speaker age estimation using i-vectors. *Engineering Applications of Artificial Intelligence*, 34, 99–108. doi:[10.1016/j.engappai.2014.05.003](https://doi.org/10.1016/j.engappai.2014.05.003) 114
- Barbé, F., Pericás, J., Muñoz, A., Findley, L., Antó, J. M., Agustí, A. G. & de Lluc Joan, M. (1998). Automobile accidents in patients with sleep apnea syndrome: an epidemiological and mechanistic study. *American journal of respiratory and critical care medicine*, 158(1), 18–22. doi:[10.1164/ajrccm.158.1.9709135](https://doi.org/10.1164/ajrccm.158.1.9709135) 14
- Baum, L. E. (1972). An equality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, 3, 1–8. 54
- Baum, L. E. & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6), 1554–1563. Retrieved from <http://www.jstor.org/stable/2238772> 54
- Bearpark, H., Elliott, L., Grunstein, R., Cullen, S., Schneider, H., Althaus, W. & C., S. (1995). Snoring and sleep apnea: a population study in Australian men. *Am J Respir Crit Care Med*, 151, 1459-1465. doi:[10.1164/ajrccm.151.5.7735600](https://doi.org/10.1164/ajrccm.151.5.7735600) 14
- Beck, R., Odeh, M., Oliven, A. & Gavriely, N. (1995). The acoustic properties of snores. *European Respiratory Journal*, 8(12), 2120–2128. doi:[10.1183/09031936.95.08122120](https://doi.org/10.1183/09031936.95.08122120) 32
- Behrman, A. (2007). Resonance and articulation. In *Speech and voice science* (p. 215-351). Plural Publishing, San Diego. 23, 109
- Benjamin, B. J. (1997). Speech production of normally aging adults. *Seminars in Speech and Language*, 18(2), 135-141. doi:[10.1055/s-2008-1064068](https://doi.org/10.1055/s-2008-1064068) 104

- Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1), 527–570. doi:[10.1007/bf01797193](https://doi.org/10.1007/bf01797193) 18
- Bettens, F., Grenez, F. & Schoentgen, J. (2005). Estimation of vocal dysperiodicities in disordered connected speech by means of distant-sample bidirectional linear predictive analysis. *Journal of the Acoustic Society of America*, 117(1), 328-37. doi:[10.1121/1.1835511](https://doi.org/10.1121/1.1835511) 74
- Bimbot, F., Bonastre, J.-F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., ... Reynolds, D. A. (2004). A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, 2004(4), 430–451. doi:[10.1155/s1110865704310024](https://doi.org/10.1155/s1110865704310024) 50
- Bixler, E., Vgontzas, A., Lin, H., Ten Have, T., Rein, J., Vela-Bueno, A. & Kales, A. (2001). Prevalence of sleep-disordered breathing in women: effects of gender. *Am J Respir Crit Care Med*, 163, 608-613. doi:[10.1164/ajrccm.163.3.9911064](https://doi.org/10.1164/ajrccm.163.3.9911064) 14, 15
- Bixler, E., Vgontzas, A., Ten Have, T., Tyson, K. & A., K. (1998). Effects of age on sleep apnea in men: I. prevalence and severity. *American Journal of Respiratory and Critical Care Medicine*, 157, 144-148. doi:[10.1164/ajrccm.157.1.9706079](https://doi.org/10.1164/ajrccm.157.1.9706079) 14, 15
- Blanco Murillo, J. L. (2013). *Contributions and impact assessment of speech technologies on the automatic detection of severe obstructive sleep apnea syndrome* (Doctoral dissertation, Universidad Politécnica de Madrid). Retrieved from <http://oa.upm.es/22234/> 36, 41, 51, 53, 54, 56
- Blanco-Murillo, J. & Schoentgen, J. (2013). Vocal tract settings in speakers suffering from obstructive sleep apnea syndrome. In C. Manfredi (Ed.), *Proceedings of the 8th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)* (p. 211-214). Firenze University Press. Retrieved from http://oa.upm.es/28943/1/INVE_MEM_2013_166500.pdf 112, 113, 114
- Blanco-Murillo, J. L., Fernández-Pozo, R., Díaz-Pardo, D., Alvaro Sigüenza, Hernández-Gómez, L. A. & Alcázar-Ramírez, J. (2009). Analyzing GMMs to characterize resonance anomalies in speakers suffering from apnoea. In *Proceedings of the International Speech Communication Association Conference* (p. 1459-1462). 2, 29, 33, 49, 78
- Blanco-Murillo, J. L., Fernández-Pozo, R., Díaz-Pardo, D., Hernández, L., López, E. & Torre, D. (2009). Apnoea voice characterization through vowel sounds analysis using generative Gaussian mixture models. In *Proceedings of 3rd Advanced Voice Function Assessment International Workshop* (Vol. 1). 2, 49
- Blanco-Murillo, J. L., Fernández-Pozo, R., López-Gonzalo, E. & Hernández-Gómez, L. A. (2008). Exploring differences between phonetic classes in sleep apnoea syndrome patients using automatic speech processing techniques. *The Phonetician. Journal of the International Society of Phonetic Science*, 97, 36-55. 49, 96
- Blanco-Murillo, J. L., Fernández-Pozo, R., Torre-Toledano, D., Caminero, J. & López, E. (2011). Analyzing training dependencies and posterior fusion in discriminant classification of apnea patients based on sustained and connected speech. In *Proceedings of the International Speech Communication Association Conference* (p. 3033-3036). 33
- Blanco-Murillo, J. L., Hernández-Gómez, L. A., Fernández-Pozo, R. & Ramos, D. (2012). Improv-

- ing automatic detection of obstructive sleep apnea through nonlinear analysis of sustained speech. *Cognitive Computation*, 5(4), 458–472. doi:[10.1007/s12559-012-9168-x](https://doi.org/10.1007/s12559-012-9168-x) 33
- Blouet, R., Mokbel, C., Mokbel, H., Sanchez-Soto, E., Chollet, G. & Greige, H. (2004). BECARs: a free software for speaker verification. In *Proceedings of the Speaker and Language Recognition Workshop, ODYSSEY* (p. 145-148). 56
- Boersma, P. (1993). Accurate Short-Term analysis of the fundamental frequency and the Harmonics-to-Noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam*, 17, 97–110. Retrieved from http://www.fon.hum.uva.nl/paul/papers/Proceedings_1993.pdf 73, 97
- Boersma, P. & Weenink, D. (2013). *Praat: doing phonetics by computer [computer program]*. Version 4.5.01, retrieved 2 April 2013 from <http://www.praat.org/>. 73, 76
- Bouloukaki, I., Kapsimalis, F., Mermigkis, C., Kryger, M., Tzanakis, N., Panagou, P., ... Schiza, S. (2011). Prediction of obstructive sleep apnea syndrome in a large greek population. *Sleep and Breathing*, 15(4), 657-664. doi:[10.1007/s11325-010-0416-6](https://doi.org/10.1007/s11325-010-0416-6) 46
- Boyd, J. H., Petrof, B. J., Hamid, Q., Fraser, R. & Kimoff, R. J. (2004). Upper airway muscle inflammation and denervation changes in obstructive sleep apnea. *Am J Respir Crit Care Med*, 170(5), 541–546. doi:[10.1164/rccm.200308-1100OC](https://doi.org/10.1164/rccm.200308-1100OC) 28, 63
- Brown, I., Bradley, T., Phillipson, E., Zamel, N. & Hoffstein, V. (1985). Pharyngeal compliance in snoring subjects with and without obstructive sleep apnea 1–3. *American Review of Respiratory Disease*, 132(2), 211–215. 31
- Burwell, C. S., Robin, E. D., Whaley, R. D. & Bickelmann, A. G. (1956). Extreme obesity associated with alveolar hypoventilation— a Pickwickian syndrome. *American Journal of Medicine*, 811-818. doi:[10.1016/0002-9343\(56\)90094-8](https://doi.org/10.1016/0002-9343(56)90094-8) 10
- Buyse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R. & Kupfer, D. J. (1989). The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2), 193–213. doi:[10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4) 20
- Cakirer, B., Hans, M., Graham, G., Aylor, J., Tishler, P. & Redline, S. (2001). The relationship between craniofacial morphology and obstructive sleep apnoea in Whites and in African-Americans. *American Journal of Respiratory and Critical Care Medicine*, 163(4), 947-950. doi:[10.1164/ajrccm.163.4.2005136](https://doi.org/10.1164/ajrccm.163.4.2005136) 10
- Casale, M., Rinaldi, V., Bressi, F., Di Peco, V., Baptista, P., Sadun, B., ... Salvinelli, F. (2008). A suitable test for identifying high risk adult patients of moderate-severe obstructive sleep apnea syndrome. *Eur Rev Med Pharmacol Sci*, 12(4), 275–280. Retrieved from <http://www.europeanreview.org/article/510> 20
- Chiba, T. & Kajiyama, M. (1941). *The vowel: Its nature and structure*. Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo. 23
- Coito, A.-S. & Sanches, J. (2011). Assessment of obstructive sleep apnea syndrome by spectral analysis of physiological parameters. In *1st Portuguese meeting in bioengineering (ENBENG), 2011*. (p. 1-6). doi:[10.1109/ENBENG.2011.6026088](https://doi.org/10.1109/ENBENG.2011.6026088) 2
- Collura, T. F. (1993). History and evolution of electroencephalographic instruments and techniques. *Journal of clinical neurophysiology*, 10(4), 476–504. 19

- Conti, A. A., Conti, A. & Gensini, G. F. (2006). Fat snorers and sleepy-heads: Were many distinguished characters of the past affected by the obstructive sleep apnea syndrome? *Medical hypotheses*, 67(4), 975–979. doi:10.1016/j.mehy.2006.01.067 9
- Darwin, C. (1871). *On the origin of species*. D. Appleton and Co., New York. (<http://www.biodiversitylibrary.org/bibliography/28875>) Retrieved from <http://www.biodiversitylibrary.org/item/71804> 24
- Davidson, T. (2003). The great leap forward: the anatomic basis for the acquisition of speech and obstructive sleep apnea. *Sleep Med.*, 4(3), 185-194. doi:10.1016/S1389-9457(02)00237-X 22
- Davidson, T., Sedgh, J., Tran, D. & Stepnowsky Jr, C. (2005). The anatomic basis for the acquisition of speech and obstructive sleep apnea: evidence from cephalometric analysis supports the great leap forward hypothesis. *Sleep Med*, 6(6), 497-505. doi:10.1016/j.sleep.2005.03.007 22, 26, 109
- Deegan, P. & McNicholas, W. (1995). Pathophysiology of obstructive sleep apnoea. *European Respiratory Journal*, 8(7), 1161-1178. Retrieved from <http://erj.ersjournals.com/content/8/7/1161.abstract> 16
- Diamond, J. M. (1991). The great leap forward. In *The rise and fall of the third chimpanzee* (p. 27). Hutchinson Radius. 25
- Diamond, J. M. & Shermer, M. (1993). *The great leap forward*. Skeptics Society. 22
- Dickens, C. (1837). *The posthumous papers of the Pickwick Club*. London: Chapman & Hall. Retrieved from https://openlibrary.org/books/OL23001437M/The_posthumous_papers_of_the_Pickwick_Club 9
- Doghramji, K. (2014). *Snoring* (Tech. Rep.). Merck Manuals. Retrieved from <http://www.merckmanuals.com/professional/neurologic-disorders/sleep-and-wakefulness-disorders/snoring> 11
- Draper, N. R. & Smith, H. (1998). *Applied regression analysis*. Wiley. 81
- Duda, R., Hart, P. & Stork, D. (2001). *Pattern classification*. Wiley. 81
- Durán, J., Esnaola, S., Rubio, R. & Iztueta, A. (2001). Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr. *American Journal of Respiratory and Critical Care Medicine*, 163, 685-689. doi:10.1164/ajrccm.163.3.2005065 14, 15
- Durán-Cantolla, J. (2012). Necessity to treat also affected people with mild or moderate apnoea. In *Proceedings of the 3rd International Congress on Sleep Apnoea*. 35, 43, 123
- Espinoza-Cuadros, F., Fernández-Pozo, R., Toledano, D. T., Alcázar-Ramírez, J. D., López-Gonzalo, E. & Hernández-Gómez, L. A. (2016). Reviewing the connection between speech and obstructive sleep apnea. *BioMedical Engineering OnLine*, 15(1). doi:10.1186/s12938-016-0138-5 114
- Espinoza-Cuadros, F., Garcia-Zamora, M. A., Torres-Boza, D., Ferrer-Riesgo, C. A., Montero Benavides, A., Gonzalez-Moreira, E. & Hernández Gómez, L. A. (2014). A spoken language database for research on moderate cognitive impairment: Design and preliminary analysis. In *IberSPEECH* (p. 219-228). doi:10.1007/978-3-319-13623-3_23 126

- Farrús, M. & Hernando, J. (2009). Using jitter and shimmer in speaker verification. *IET Signal Processing, Special Issue on Biometric Recognition*, 3(4), 247–257. doi:[10.1049/iet-spr.2008.0147](https://doi.org/10.1049/iet-spr.2008.0147) 75
- Fernández-Pozo, R. (2011). *Técnicas de caracterización de voces como apoyo a la detección automática de la apnea del sueño* (Doctoral dissertation, Universidad Politécnica de Madrid). Retrieved from <https://serviciosgate.upm.es/tesis/tesis/6713> 36, 38, 47
- Fernández-Pozo, R., Blanco-Murillo, J., Díaz, D., Hernández, L., López, E. & Alcázar, J. (2010). Early detection of severe apnoea through voice analysis and automatic speaker recognition techniques. In A. Fred, J. Filipe & H. Gamboa (Eds.), *Biomedical engineering systems and technologies* (Vol. 52, p. 245-257). Springer Berlin Heidelberg. doi:[10.1007/978-3-642-11721-3_19](https://doi.org/10.1007/978-3-642-11721-3_19) 49
- Fernández-Pozo, R., Blanco-Murillo, J. L., Hernández-Gómez, L. A., López, E., Alcázar, J. & Torre-Toledano, D. (2009a). Assessment of severe apnoea through voice analysis, automatic speech, and speaker recognition techniques. *EURASIP J. Adv. Signal Process*, 6:1–6:11. doi:[10.1155/2009/982531](https://doi.org/10.1155/2009/982531) 2, 33, 49, 56, 57, 58, 66, 84, 87, 90, 109, 118, 119, 121
- Fernández-Pozo, R., Blanco-Murillo, J. L., Hernández-Gómez, L. A., López, E., Alcázar, J. & Torre-Toledano, D. (2009b). Severe apnoea detection using speaker recognition techniques. In P. Encarnação & A. Veloso (Eds.), *Proceedings of the BIOSIGNALS Conference* (p. 124-130). INSTICC Press. doi:[10.5220/0001546601240130](https://doi.org/10.5220/0001546601240130) 2, 49, 57
- Fernández-Pozo, R., Hernández-Gómez, L. A., López-Gonzalo, E., Alcázar-Ramírez, J., Portillo, G. & Torre Toledano, D. (2008). Design of a multimodal database for research on automatic detection of severe apnoea cases. In N. Calzolari (Conference Chair) et al. (Eds.), *Proceedings of the sixth international conference on language resources and evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA). Retrieved from <http://www.lrec-conf.org/proceedings/lrec2008/> 35, 36
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in cognitive sciences*, 4(7), 258–267. doi:[10.1016/S1364-6613\(00\)01494-7](https://doi.org/10.1016/S1364-6613(00)01494-7) 23, 24
- Fitch, W. T. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522. 25
- Fiz, J. A., Morera, J., Abad, J., Belsunces, A., Haro, M., Fiz, J. I., ... Rodenstein, D. (1993). Acoustic analysis of vowel emission in obstructive sleep apnea. *Chest*, 104(4), 1093-6. doi:[10.1378/chest.104.4.1093](https://doi.org/10.1378/chest.104.4.1093) 30, 31, 32, 95
- Fowler, J., Cohen, L. & Jarvis, P. (2013). *Practical statistics for field biology*. John Wiley & Sons. 99, 104
- Fox, A. W., Monoson, P. K. & Morgan, C. D. (1989). Speech dysfunction of obstructive sleep apnea. a discriminant analysis of its descriptors. *Chest*, 96(3), 589-95. doi:[10.1378/chest.96.3.589](https://doi.org/10.1378/chest.96.3.589) 2, 26, 27, 28, 29, 35, 37, 53, 62, 70, 75, 76, 77, 78, 84, 121
- Friedman, M., Wilson, M. N., Pulver, T., Pandya, H., Joseph, N. J., Lin, H.-C. & Chang, H.-W. (2010). Screening for obstructive sleep apnea/hypopnea syndrome: Subjective and objective factors. *Otolaryngology – Head and Neck Surgery*, 142(4), 531-535.

- doi:[10.1016/j.otohns.2009.12.038](https://doi.org/10.1016/j.otohns.2009.12.038) 19, 88
- George, C. F. (2007). Sleep apnea, alertness, and motor vehicle crashes. *American Journal of Respiratory and Critical Care Medicine*, 176(10), 954–956. doi:[10.1164/rccm.200605-629PP](https://doi.org/10.1164/rccm.200605-629PP) 14, 18
- Gibson, G. (2004). Obstructive sleep apnoea syndrome: underestimated and undertreated. *British medical bulletin*, 72(1), 49–64. doi:[10.1093/bmb/ldh044](https://doi.org/10.1093/bmb/ldh044) 15
- Goel, A., Loomba, A., Goel, P., Goel, P., Mathur, R. & Tandon, R. (2012). Obstructive sleep apnea: An overview. *Asian Journal of Oral Health & Allied Sciences*, 2(1), 26–30. Retrieved from http://www.sdclucknow.com/doc_file_uploads/5.pdf 13
- Golbin, J. M., Somers, V. K. & Caples, S. M. (2008). Obstructive sleep apnea, cardiovascular disease, and pulmonary hypertension. *Proceedings of the American Thoracic Society*, 5(2), 200–206. doi:[10.1513/pats.200708-143mg](https://doi.org/10.1513/pats.200708-143mg) 12
- Goldshtein, E., Tarasiuk, A. & Zigel, Y. (2011). Automatic detection of obstructive sleep apnea using speech signals. *IEEE Transactions on Biomedical Engineering*, 58(5), 1373–1382. doi:[10.1109/TBME.2010.2100096](https://doi.org/10.1109/TBME.2010.2100096) 25, 33, 96, 109, 121
- Gonzalez, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics*, 32, 277–287. 105, 114
- Guilleminault, C., Tilkian, A. & Dement, W. C. (1976). The sleep apnea syndromes. *Annu. Rev. Med.*, 27(1), 465–484. doi:[10.1146/annurev.me.27.020176.002341](https://doi.org/10.1146/annurev.me.27.020176.002341) 30
- Gurubhagavatula, I., Maislin, G., Nkwuo, J. E. & Pack, A. I. (2004). Occupational screening for obstructive sleep apnea in commercial drivers. *American journal of respiratory and critical care medicine*, 170(4), 371–376. doi:[10.1164/rccm.200307-968OC](https://doi.org/10.1164/rccm.200307-968OC) 19, 88, 90, 118
- Haba Rubio, J. & Krieger, J. (2012). Introduction to modern sleep technology. In R. P.-Y. Chiang & S.-C. Kang (Eds.), (pp. 19–31). Springer Science+Business Media, Dordrecht. doi:[10.1007/978-94-007-5470-6](https://doi.org/10.1007/978-94-007-5470-6) 18
- Hamdan, A., Barazi, R. A., Khneizer, G., Turfe, Z., Sinno, S., Ashkar, J. & Tabri, D. (2013). Formant frequency in relation to body mass composition. *J Voice.*, 27(5), 567–71. 105
- Hatano, H., Kitamura, T., Takemoto, H., Mokhtari, P., Honda, K. & Masaki, S. (2012). Correlation between vocal tract length, body height, formant frequencies, and pitch frequency for the five Japanese vowels uttered by fifteen male speakers. In *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association, Portland, Oregon, USA, September 9-13, 2012* (pp. 402–405). Retrieved from http://www.isca-speech.org/archive/interspeech_2012/i12_0402.html 113
- Hayes, C. (1951). *The ape in our house*. Harper, New York. 22
- Hébert, M. (2008). Text-dependent speaker recognition. In *Springer handbook of speech processing* (pp. 743–762). Springer. 49
- Hidalgo, A. & Quilis, M. (2002). *Fonética y fonología españolas* (T. L. Blanch, Ed.). 29, 72
- Ip, M., Lam, B., Lauder, I., Tsang, K., Chung, K., Mok, Y. & Lam, W. (2001). A community study of sleep-disordered breathing in middle-aged Chinese men in Hong Kong. *Chest*, 119, 62–69. doi:[10.1378/chest.119.1.62](https://doi.org/10.1378/chest.119.1.62) 14, 15
- Ip, M., Lam, B., Tang, L., Lauder, I., Ip, T. & Lam, W. (2004). A community study of sleep-

- disordered breathing in middle-aged Chinese women in Hong Kong: prevalence and gender differences. *Chest*, 125, 127-134. doi:[10.1378/chest.125.1.127](https://doi.org/10.1378/chest.125.1.127) 14, 15
- Isono, S. (2012). Obesity and obstructive sleep apnoea: mechanisms for increased collapsibility of the passive pharyngeal airway. *Respirology*, 17(1), 32–42. 96
- Jamie, C. L., Sharma, S. & Bing, L. (2010). Obstructive sleep apnoea: Definitions, epidemiology & natural history. *The Indian journal of medical research*, 131, 165-70. 1, 15, 100, 105
- Johal, A. & Conaghan, C. (2004). Maxillary morphology in obstructive sleep apnea: a cephalometric and model study. *Angle Orthod*, 74(5), 648-56. Retrieved from <http://www.angle.org/doi/abs/10.1043/0003-3219%282004%29074%3C0648%3AMMIOA%3E2.0.CO%3B2> 10, 11
- Johal, A., Patel, S. I. & Battagel, J. M. (2007). The relationship between craniofacial anatomy and obstructive sleep apnoea: a case-controlled study. *Journal of Sleep Research*, 16(3), 319–326. doi:[10.1111/j.1365-2869.2007.00599.x](https://doi.org/10.1111/j.1365-2869.2007.00599.x) 24
- Jordan, A. & McEvoy, R. D. (2003). Gender differences in sleep apnea: epidemiology, clinical presentation and pathogenic mechanisms. *Sleep Medicine Reviews*, 7(5), 377–389. doi:[10.1053/smr.2002.0260](https://doi.org/10.1053/smr.2002.0260) 14, 15
- Kapur, V., Blough, D. K., Sandblom, R. E., Hert, R., de Maine, J. B., Sullivan, S. D. & Psaty, B. M. (1999). The medical cost of undiagnosed sleep apnea. *Sleep*, 22, 750-755. Retrieved from <http://www.journalsleep.org/ViewAbstract.aspx?pid=24161> 14
- Katz, I., Stradling, J., Slutsky, A., Zamel, N. & Hoffstein, V. (1990). Do patients with obstructive sleep apnea have thick necks? *The American review of respiratory disease*, 141(5 Pt 1), 1228-31. 105
- Kemaloğlu, Y. K. & Mengü, G. (2016). Speech and OSA: Could lower formant frequencies of the vowels only be expected in subjects with obstructive sleep apnea? *Journal of Voice*. doi:[10.1016/j.jvoice.2016.05.001](https://doi.org/10.1016/j.jvoice.2016.05.001) 108, 109, 110, 125
- Kim, J., In, K., Kim, J., You, S., Kang, K., Shim, J., ... Shin, C. (2004). Prevalence of sleep-disordered breathing in middle-aged Korean men and women. *Am J Respir Crit Care Med*, 170, 1108-1113. 14, 15
- Knottnerus, J. & Buttinx, F. (Eds.). (2008). *The evidence base of clinical diagnosis: Theory and methods of diagnostic research*. Wiley, London. 87
- Krajewski, J., Schnieder, S., Sommer, D., Batliner, A. & Schuller, B. (2012). Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech. *Neurocomputing*, 84, 65 - 75. (From Neuron to Behaviour: Evidence from Behavioral Measurements) doi:[10.1016/j.neucom.2011.12.021](https://doi.org/10.1016/j.neucom.2011.12.021) 90
- Kryger, M. (2012). Charles Dickens: Impact on medicine and society. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 8(3), 333. doi:[10.5664/jcsm.1930](https://doi.org/10.5664/jcsm.1930) 9
- Lammert, A. & Shrikanth, N. (2013). On instantaneous vocal tract length estimation from formant frequencies. In *Proceedings of meetings on acoustics* (Vol. 19, p. 060027). Acoustical Society of America. doi:[10.1121/1.4799393](https://doi.org/10.1121/1.4799393) 105
- Lan, Z., Itoi, A., Takashima, M., Oda, M. & Tomoda, K. (2006). Difference of pharyngeal mor-

- phology and mechanical property between OSAHS patients and normal subjects. *Auris Nasus Larynx*, 33(4), 433–439. doi:[10.1016/j.anl.2006.03.009](https://doi.org/10.1016/j.anl.2006.03.009) 2
- Lee, B. J., Ku, B., Jang, J.-S. & Kim, J. Y. (2013). A novel method for classifying body mass index on the basis of speech signals for future clinical applications: A pilot study. *Evidence-Based Complementary and Alternative Medicine*, 13. doi:[10.1155/2013/150265](https://doi.org/10.1155/2013/150265) 105
- Lee, R. W. W., Chan, A. S. L., Grunstein, R. R. & Cistulli, P. A. (2009). Craniofacial phenotyping in obstructive sleep apnea—a novel quantitative photographic approach. *Sleep*, 32(1), 37–45. doi:[10.5665/sleep/32.1.37](https://doi.org/10.5665/sleep/32.1.37) 11
- Lee, R. W. W., Sutherland, K., Chan, A. S. L., Zeng, B., Grunstein, R. R., Darendeliler, M. A., ... Cistulli, P. A. (2010). Relationship between surface facial dimensions and upper airway structures in obstructive sleep apnea. *Sleep*, 33(9), 1249–1254. doi:[10.1093/sleep/33.9.1249](https://doi.org/10.1093/sleep/33.9.1249) 11
- Lieberman, P., Klatt, D. H. & Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science*, 164(3884), 1185–1187. doi:[10.1126/science.164.3884.1185](https://doi.org/10.1126/science.164.3884.1185) 22, 24
- Liss, J. M., Weismer, G. & Rosenbek, J. C. (1990). Selected acoustic characteristics of speech production in very old males. *Journal of Gerontology*, 45(2), 35–45. doi:[10.1093/geronj/45.2.P35](https://doi.org/10.1093/geronj/45.2.P35) 104
- Lloberes, P., Levy, G., Descals, C., Sampol, G., Roca, A., Sagales, T. & de la Calzada, M. (2000). Self-reported sleepiness while driving as a risk factor for traffic accidents in patients with obstructive sleep apnoea syndrome and in non-apnoeic snorers. *Respir Med*, 94(10), 971–6. doi:[10.1053/rmed.2000.0869](https://doi.org/10.1053/rmed.2000.0869) 14
- Lurie, A. (2011). Obstructive sleep apnea in adults. *Adv Cardiol Basel Karger*, 46, 1–42. doi:[10.1159/000327660](https://doi.org/10.1159/000327660) 12
- Maltais, F., Carrier, G., Cormier, Y. & Series, F. (1991). Cephalometric measurements in snorers, non-snorers, and patients with sleep apnoea. *Thorax*, 46(6), 419–423. doi:[10.1136/thx.46.6.419](https://doi.org/10.1136/thx.46.6.419) 29, 31, 109
- Mann, H. B. & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, 18(1), 50–60. Retrieved from <https://projecteuclid.org/euclid.aoms/1177730491> 77
- Marin, J. M., Agusti, A., Villar, I., Forner, M., Nieto, D., Carrizo, S. J., ... Jelic, S. (2012). Association between treated and untreated obstructive sleep apnea and risk of hypertension. *JAMA*, 307(20), 2169–2176. doi:[10.1001/jama.2012.3418](https://doi.org/10.1001/jama.2012.3418) 12
- Martin, A., Doddington, G., Kamm, T., Ordowski, M. & Przybocki, M. (1997). The det curve in assessment of detection task performance. In *Proceedings of the EUROSPEECH Conference* (pp. 1895–1898). Retrieved from http://www.isca-speech.org/archive/eurospeech_1997/e97_1895.html 56
- Martin, D., Fitch, J. & Wolfe, V. (1995). Pathologic voice type and the acoustic prediction of severity. *J Speech Hear Res*, 38(4), 765–71. doi:[10.1044/jshr.3804.765](https://doi.org/10.1044/jshr.3804.765) 73
- McDonald, J. (2003). A review of surgical treatment for obstructive sleep apnoea/hypopnoea syndrome. *The Surgeon*, 1(5), 259–264. 21

- McNicholas, W. T. & Phillipson, E. A. (2005). *Breathing disorders in sleep*. Saunders Ltd. 10
- MIT-Madrid M+Visión Consortium. (2013). 2013 IDEA² Madrid finalists announced. Retrieved January 16, 2017 from <http://mvisionconsortium.org/2013-idea2-madrid-finalists-announced/>. 126
- Monoson, P. & Fox, A. (1987). Preliminary observation of speech disorder in obstructive and mixed sleep apnea. *Chest*, 92, 670-75. doi:10.1378/chest.92.4.670 26, 29, 35, 62, 70
- Montero Benavides, A., Blanco Murillo, J. L., Fernández, A., Fernández Pozo, R., Torre Toledano, D. & Hernández Gómez, L. A. (2012). Using HMM to detect speakers with severe obstructive sleep apnoea syndrome. In D. T. Toledano et al. (Eds.), *IberSPEECH* (Vol. 328, p. 121-128). Springer. doi:10.1007/978-3-642-35292-8_13 33, 36, 67, 68, 125, 126
- Montero Benavides, A., Fernández Pozo, R., Torre Toledano, D., Blanco Murillo, J. L., López Gonzalo, E. & Hernández Gómez, L. A. (2014). Analysis of voice features related to obstructive sleep apnoea and their application in diagnosis support. *Computer Speech & Language*, 28(2), 434-452. doi:10.1016/j.csl.2013.08.002 33, 36, 72, 96, 109, 121, 125
- Montero Benavides, A., Blanco Murillo, J. L., Fernández Pozo, R., Espinoza Cuadros, F., Torre Toledano, D., Alcázar-Ramírez, J. D. & Hernández Gómez, L. A. (2016a). Formant frequencies and bandwidths in relation to clinical variables in an obstructive sleep apnea population. *Journal of Voice*, 30(1), 21-29. doi:10.1016/j.jvoice.2015.01.006 33, 108, 109, 125
- Montero Benavides, A., Blanco Murillo, J. L., Fernández Pozo, R., Espinoza Cuadros, F., Torre Toledano, D., Alcázar-Ramírez, J. D. & Hernández Gómez, L. A. (2016b). Reply to the Letter to the Editor by Kemaloglu and Mengü. *Journal of Voice*. doi:10.1016/j.jvoice.2016.05.008 109, 111, 125
- Moreno, A., Poch, D., Bonafonte, A., Lleida, E., Llisterri, J., Mariño, J. B. & Nadeu, C. (1993). Al-bayzin speech database: design of the phonetic corpus. In *Eurospeech '93* (p. 175-178). ISCA. Retrieved from http://www.isca-speech.org/archive/eurospeech_1993/e93_0175.html 37, 66, 76
- Morgenthaler, T., Kagramanov, V., Hanak, V. & Decker, P. (2006). Complex sleep apnea syndrome: is it a unique clinical syndrome? *Sleep*, 29(9), 1203-1209. doi:10.1093/sleep/29.9.1203 1
- Mrayati, M., Carré, R. & Guerin, B. (1988). Distinctive regions and modes: a new theory of speech production. *Speech Communication*, 7(3), 257-286. doi:10.1016/0167-6393(90)90058-H 112
- Negus, V. E. (1949). *The comparative anatomy and physiology of the larynx*. Hafner Pub. Company; New York. 25
- Neil, A. M., Angus, S. M., Sajkov, D. & McEvoy, R. D. (1997). Effects of sleep posture on upper airway stability in patients with obstructive sleep apnea. *American journal of respiratory and critical care medicine*, 155(1), 199-204. doi:10.1164/ajrccm.155.1.9001312 20
- Netzer, N. C., Stoohs, R. A., Netzer, C. M., Clark, K. & Strohl, K. P. (1999). Using the berlin questionnaire to identify patients at risk for the sleep apnea syndrome. *Annals of internal medicine*, 131(7), 485-491. doi:10.7326/0003-4819-131-7-199910050-00002 19, 20
- Ng, A. K., Koh, E., T. S. and Baey & Puvanendran, K. (2007). Diagnosis of obstructive sleep

- apnea using formant features of snore signals. In R. Magjarevic & J. H. Nagel (Eds.), *World Congress on Medical Physics and Biomedical Engineering 2006: August 27 – September 1, 2006 COEX Seoul, Korea “Imaging the Future Medicine”* (pp. 967–970). Springer Berlin Heidelberg. doi:[10.1007/978-3-540-36841-0_230](https://doi.org/10.1007/978-3-540-36841-0_230) 71
- Ng, A. K., San Koh, T., Baey, E., Lee, T. H., Abeyratne, U. R. & Puvanendran, K. (2008). Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Medicine*, 9(8), 894–898. doi:[10.1016/j.sleep.2007.07.010](https://doi.org/10.1016/j.sleep.2007.07.010) 32, 71
- Parsa, V. & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res*, 44(2), 327–39. doi:[10.1044/1092-4388\(2001/027\)](https://doi.org/10.1044/1092-4388(2001/027)) 36, 73, 96
- Payne, R. J., Kost, K. M., Frenkiel, S., Zeitouni, A. G., Sejean, G., Sweet, R. C., ... Kimoff, R. J. (2006). Laryngeal inflammation assessed using the reflux finding score in obstructive sleep apnea. *Otolaryngology – Head and Neck Surgery*, 134(5), 836–842. doi:[10.1016/j.otohns.2006.01.012](https://doi.org/10.1016/j.otohns.2006.01.012) 28, 37, 63
- Peppard, P. E., Young, T., Barnet, J. H., Palta, M., Hagen, E. W. & Hla, K. M. (2013). Increased prevalence of sleep-disordered breathing in adults. *American Journal of Epidemiology*, 177, 1006–1014. doi:[10.1093/aje/kws342](https://doi.org/10.1093/aje/kws342) 1
- Peppard, P. E., Young, T., Palta, M., Dempsey, J. & Skatrud, J. (2000). Longitudinal study of moderate weight change and sleep-disordered breathing. *Jama*, 284(23), 3015–3021. doi:[10.1001/jama.284.23.3015](https://doi.org/10.1001/jama.284.23.3015) 17, 20
- Pevernagie, D., Aarts, R. M. & De Meyer, M. (2010). The acoustics of snoring. *Sleep Medicine Reviews*, 14(2), 131–144. doi:[10.1016/j.smrv.2009.06.002](https://doi.org/10.1016/j.smrv.2009.06.002) 32
- Poorjam, A. H., Bahari, M. H., Vasilakakis, V. & Van hamme, H. (2015). Height estimation from speech signals using i-vectors and least-squares support vector regression. *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*. doi:[10.1109/tsp.2015.7296469](https://doi.org/10.1109/tsp.2015.7296469) 114
- Pruthi, T. (2007). *Analysis, vocal-tract modeling and automatic detection of vowel nasalization* (Doctoral dissertation, University of Maryland). Retrieved from <http://hdl.handle.net/1903/4273> 75
- Puertas, F. J., Pin, G., María, J. M. & Durán, J. (2005). Documento de consenso nacional sobre el síndrome de apneas-hipopneas del sueño. *Grupo Español De Sueño*, 164. Retrieved from <http://www.aeped.es/documentos/consenso-nacional-sobre-sindrome-apneas-hipopneas-sueno-sahs> 1, 2, 17, 18, 116
- Punjabi, N. M. (2008). The epidemiology of adult obstructive sleep apnea. *Proc Am Thorac Soc*, 5(2), 136–143. 17
- Quilis, A. (1981). *Fonética acústica de la lengua española*. Biblioteca Románica Hispánica, Editorial Gredos. 24
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286. doi:[10.1109/5.18626](https://doi.org/10.1109/5.18626) 54, 55
- Rabiner, L. R. & Gold, B. (1975). *Theory and application of digital signal processing*. Prentice-Hall Inc., Englewood Cliffs, NJ. 54

- Ramachandran, S. K. & Josephs, L. A. (2009). A meta-analysis of clinical screening tests for obstructive sleep apnea. *Anesthesiology*, 110(4), 928–939. doi:10.1097/ALN.0b013e31819c47b6 19
- Rastatter, P. M., McGuire R, A. & Kalinowski J, S. A. (1997). Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatr Logop*, 49, 1-8. doi:10.1159/000266431 104
- Rendall, D., S. Kollias, C. N. & Lloyd, P. (2005). Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *Journal of the Acoustical Society of America*, 117(2), 944–955. doi:10.1121/1.1848011 105
- Reynolds, D., Andrews, W., Campbell, J., Navratil, J., Peskin, B., Adami, A., ... Xiang, B. (2003). The supersid project: exploiting high-level information for high-accuracy speaker recognition. *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)..* doi:10.1109/icassp.2003.1202760 76
- Risser, M. R., Ware, J. C. & Freeman, F. G. (2000). Driving simulation with EEG monitoring in normal and obstructive sleep apnea patients. *Sleep: Journal of Sleep Research & Sleep Medicine*. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3400226/> 14
- Robb, M. P., Yates, J. & Morgan, E. J. (1997). Vocal tract resonance characteristics of adults with obstructive sleep apnea. *Acta Otolaryngol*, 117(5), 760-3. doi:10.3109/00016489709113474 2, 29, 30, 31, 32, 37, 72, 95, 96, 98, 102, 105, 106, 109, 114, 120, 122
- Rosique, M., Ramón, J., Canteras, M. & Rosique, L. (2003). Análisis discriminante aplicado a los formante de las vocales castellanas en la fonación con prótesis y erigmo fonía tras laringuectomía total. *Acta Otorrinolaringologica Espanola*, 54(5), 361-366. 99, 100
- Roth, T., Zammit, G., Kushida, C., Doghramji, K., Mathias, S. D., Wong, J. M. & Buysse, D. J. (2002). A new questionnaire to detect sleep disorders. *Sleep Medicine*, 3(2), 99–108. doi:10.1016/S1389-9457(01)00131-9 20
- Ryan, C. M. & Bradley, T. D. (2005). Pathogenesis of obstructive sleep apnea. *Journal of Applied Physiology*, 41(6), 323-330. 21
- Sakakibara, H., Tong, M., Matsushita, K., Hirata, M., Konishi, Y. & Suetsugu, S. (1999). Cephalometric abnormalities in non-obese and obese patients with obstructive sleep apnoea. *European Respiratory Journal*, 13(2), 403–410. Retrieved from <http://erj.ersjournals.com/content/13/2/403> 109
- Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M. & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the society for research in child development*, i–252. 22
- Sharma, S. K., Kumpawat, S., Banga, A. & Goel, A. (2006). Prevalence and risk factors of obstructive sleep apnea syndrome in a population of Delhi, India. *CHEST Journal*, 130(1), 149–156. doi:10.1378/chest.130.1.149 14, 15
- Shelton, K. E., Woodson, H., Gay, S. & Suratt, P. M. (1993). Pharyngeal fat in obstructive sleep apnea. *American Review of Respiratory Disease*, 148(2), 462–466. doi:10.1164/ajrccm/148.2.462 96
- Smith, P. L., Wise, R. A., Gold, A. R., Schwartz, A. R. & Permutt, S. (1988). Upper airway pressure-

- flow relationships in obstructive sleep apnea. *Journal of Applied Physiology*, 64(2), 789–795. 31
- Solán-Casals, J., Munteanu, C., Martín, O. C., Barbé, F., Queipo, C., Amilibia, J. & Durán-Cantolla, J. (2014). Detection of severe obstructive sleep apnea through voice analysis. *Applied Soft Computing*, 23(0), 346 - 354. doi:<http://dx.doi.org/10.1016/j.asoc.2014.06.017> 106, 108, 110, 122
- Sola-Soler, J., Jane, R., Fiz, J. & Morera, J. (2003). Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea. In *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE* (Vol. 3, pp. 2527–2530). doi:[10.1109/IEMBS.2003.1280430](http://dx.doi.org/10.1109/IEMBS.2003.1280430) 71
- Sola-Soler, J., Jane, R., Fiz, J. A. & Morera, J. (2008). Formant frequencies of normal breath sounds of snorers may indicate the risk of obstructive sleep apnea syndrome. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3500–3503). doi:[10.1109/IEMBS.2008.4649960](http://dx.doi.org/10.1109/IEMBS.2008.4649960) 71
- Stevens, K. N. (1955). Development of a quantitative description of vowel articulation. *J. Acoust. Soc. Am.*, 27(3), 484. doi:[10.1121/1.1907943](http://dx.doi.org/10.1121/1.1907943) 31
- Stevens, S. S., Volkman, J. & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3), 185-190. doi:<http://dx.doi.org/10.1121/1.1915893> 52
- Strohl, K. P. (2015). *Obstructive sleep apnea* (Tech. Rep.). Merck Manuals. Retrieved from <http://www.merckmanuals.com/professional/pulmonary-disorders/sleep-apnea/obstructive-sleep-apnea> 11
- The Report of an American Academy of Sleep Medicine Task Force. (1999). Review: Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep*, 22(5), 667-89. 7, 46
- Torre, P. III & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, 42(5), 324–333. 104
- Udwadia, Z., Doshi, A., Lonkar, S. & Singh, C. (2004). Prevalence of sleep-disordered breathing and sleep apnea in middle-aged urban indian men. *Am J Respir Crit Care Med*, 169, 168-173. 14, 15
- Valham, F. (2011). *Obstructive sleep apnea— the relationship to cardiovascular disease, diabetes mellitus, motor vehicle driving and ambient temperatur* (Doctoral dissertation, Umeå University, Sweden). Retrieved from <http://umu.diva-portal.org/smash/get/diva2:451994/FULLTEXT02.pdf> 14
- Vittinghoff, E., Glidden, D., Shiboski, S. & McCulloch, C. (2004). *Regression methods in biostatistics. Linear, logistic, survival, and repeated measures models*. Springer Science+Business Media, New York. (pages 43–44, ISBN 0-387-20275-7) 82
- VVFS. (2008). *Swedish traffic regulations: Vägverkets föreskrifter om ändring i föreskrifterna (VVFS 2008:158) om medicinska krav för innehav av körkort m.m.: Vägverkets Författningssamling VVFS 2008:166*. 14
- WHO. (2004). *WHO technical meeting on sleep and health report* (Tech. Rep.). WHO European

- Centre for Environment and Health. 11
- Woehrling, C., de Mareüil & Boula, P. (2007). Comparing praat and snack formant measurements on two large corpora of northern and southern french. In *Interspeech* (p. 1006-1009). ISCA. 97, 98
- Xue, S. A. & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production a pilot study. *Journal of Speech, Language, and Hearing Research*, 46, 689-701. doi:10.1044/1092-4388(2003/054) 104
- Yaggi, H. K., Concato, J., Kernan, W. N., Lichtman, J. H., Brass, L. M. & Mohsenin, V. (2005). Obstructive sleep apnea as a risk factor for stroke and death. *New England Journal of Medicine*, 353(19), 2034–2041. 12
- Young, J. W. & McDonald, J. P. (2004). An investigation into the relationship between the severity of obstructive sleep apnoea/hypopnoea syndrome and the vertical position of the hyoid bone. *Surgeon*, 2(3), 145-51. doi:10.1016/S1479-666X(04)80075-1 104
- Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Lui, A., ... Woodland, P. (2009). *The HTK Book (for HTK version 3.4)*. Cambridge University Engineering Department. Retrieved from <http://htk.eng.cam.ac.uk/docs/docs.shtml> 50, 53, 56, 76
- Young, T., Evans, L., Finn, L., Palta, M. et al. (1997). Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep*, 20(9), 705–706. 1, 116
- Young, T., Palta, M., Dempsey, J., Skatrud, J., Weber, S. & Badr, S. (1993). The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med*, 328, 1230-1235. doi:10.1056/NEJM199304293281704 1, 14, 15, 17
- Young, T., Skatrud, J. & Peppard, P. E. (2004). Risk factors for obstructive sleep apnea in adults. *JAMA*, 291(16), 2013–2016. doi:10.1001/jama.291.16.2013 17
- Yumoto, E., Gould, W. J. & Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am*, 71(6), 1544-9. 73
- Yumoto, E., Sasaki, Y. & Okamura, H. (1984). Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness. *Journal of Speech Language and Hearing Research*, 27(1), 2. doi:10.1044/jshr.2701.02 73
- Zhang, Y. & Jiang, J. J. (2008). Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *J Voice*, 22(1), 1-9. 76
- Zigel, Y., Tarasiuk, A. & Goldshtein, E. (2008). Analysis of speech signals among obstructive sleep apnea patients. In *Proceedings of the IEEE 25th Convention of Electrical and Electronic Engineers in Israel*. 35, 123

STATEMENT OF ORIGINALITY

DECLARACIÓN DE ORIGINALIDAD

I hereby declare that this Thesis is my own, original work and that to the best of my knowledge it contains no materials previously published or written by another person, except where due acknowledgement is made in the Thesis.

All the assistance received in preparing this Thesis has been acknowledged and information derived from the published and unpublished work of others has been acknowledged in the text and referenced.

Por la presente declaro que esta Tesis es mi trabajo original, y que según mi leal saber y entender no contiene material previamente publicado o escrito por otra persona, salvo cuando debidamente se reconociera en la Tesis.

Toda la ayuda recibida en la elaboración de esta Tesis ha sido señalada, y la información derivada de los trabajos publicados y no publicados de otros ha sido reconocida en el texto y debidamente referenciada.

Madrid, a de 2017.

Ana Montero Benavides
Licenciada en Ciencias y Técnicas Estadísticas