Early Detection of Severe Apnoea through Voice Analysis and Automatic Speaker Recognition Techniques

Ruben Fernández[1], Jose Luis Blanco[1], David Díaz[1],
Luis A. Hernández[1], Eduardo López[1], and José Alcázar[2],

[1] Signal, Systems & RadioComm. Department (GAPS). Universidad Politécnica de Madrid,
Avda. Complutense 30, 28040 Madrid, Spain.
{ruben, jlblanco, dpardo, luis, eduardo}@gaps.ssr.upm.es
[2] Respiratory Department. Hospital Torrecardenas, 04009, Almeria, Spain.
{jose44}@separ.es

Abstract. This study is part of an on-going collaborative effort between the medical and the signal processing communities to promote research on applying voice analysis and Automatic Speaker Recognition techniques (ASR) for the automatic diagnosis of patients with severe obstructive sleep apnoea (OSA). Early detection of severe apnoea cases is important so that patients can receive early treatment. Effective ASR-based diagnosis could dramatically cut medical testing time. Working with a carefully designed speech database of healthy and apnoea subjects, we present and discuss the possibilities of using generative Gaussian Mixture Models (GMMs), generally used in ASR systems, to model distinctive apnoea voice characteristics (i.e. abnormal nasalization). Finally, we present experimental findings regarding the discriminative power of speaker recognition techniques applied to severe apnoea detection. We have achieved an 81.25 % correct classification rate, which is very promising and underpins the interest in this line of inquiry.

Keywords: Apnoea, Automatic Speaker Recognition, GMM, nasalization

1 Introduction

Obstructive sleep apnoea (OSA) is a highly prevalent disease [1], affecting an estimated 2-4% of the male population between the ages of 30 and 60. It is characterized by recurring episodes of sleep-related collapse of the upper airway at the level of the pharynx (AHI > 15, *Apnoea Hypopnoea Index*, which represents the number of apnoeas and hypoapnoeas per hour of sleep) and it is usually associated with loud snoring and increased daytime sleepiness. OSA is a serious threat to an individual's health if not treated. The condition is a risk factor for hypertension and, possibly, cardiovascular diseases [2], it is usually related to traffic accidents caused by somnolent drivers [3], and it can lead to a poor quality of life and impaired work performance. At present, the most effective and widespread treatment for OSA is nasal CPAP (*Continuous Positive Airway Pressure*) which prevents apnoea episodes by providing a pneumatic splint to the airway. OSA can be diagnosed on the basis of a characteristic history (snoring, daytime sleepiness) and physical examination (increased neck circumference), but a full overnight sleep study is usually needed to confirm the disorder [1]. The procedure is known as conventional *Polysomnography*, which involves the recording of neuroelectrophisiological and cardiorespiratory variables (ECG). Excellent automatic OSA recognition performance is attainable with this method based on nocturnal ECG recordings. Nevertheless, this diagnostic procedure is expensive and time-consuming, and patients usually have to endure a waiting list of several years before the test is done, since the demand for consultations and diagnostic studies for OSA has recently increased. There is, therefore, a strong need for methods of early diagnosis of apnoea patients in order to reduce these considerable delays.

In our research we investigate the acoustical characteristics of the speech of patients with OSA for the purpose of learning whether severe OSA may be detected using *Automatic Speaker Recognition* techniques (ASR). The acoustic properties of voice from speakers suffering obstructive sleep apnoea are not well understood as not much research has been carried out in this area. However, some studies have suggested that certain abnormalities in phonation, articulation and resonance may be connected to the condition [4]. In order to have a controlled experimental framework to study apnoea voice characterization we collected a speech database [5] designed following linguistic and phonetic criteria. Our work is focused on continuous speech rather than on sustained vowels, the latter being the standard approach in pathological voice analysis [6]. The speech corpus was designed considering previous research in the field as well as an initial manual contrastive study we performed on a small group of healthy subjects and apnoea patients.

After this preliminary acoustic analysis, *GMM-based* Automatic Speaker Recognition techniques [7] were explored trying to model possible peculiarities in apnoea patients' voices. More specifically, our work is mainly focused on the nasality factor, since it has been traditionally identified as an important feature in the acoustic characteristics of apnoea speakers. Successfully detecting traits that prove to be characteristic of the voices of severe apnoea patients by applying such techniques would allow automatic (and rapid) diagnosis of the condition. To our knowledge this study constitutes pioneering research on automatic severe OSA diagnosis using speech processing algorithms on continuous speech. The proposed method is intended as complementary to existing OSA diagnosis methods (e.g. *Polysomnography*) and clinicians' judgment, as an aid for early detection of these cases. We have observed a marked inadequacy of resources that has led to unacceptable waiting periods. Early severe OSA detection can help to increase the efficiency of medical protocols by giving higher priority to more serious cases, thus optimizing both social benefits and medical resources. For instance, patients with severe apnoea have a higher risk of suffering a car accident because of somnolence caused by their condition. Early detection would, therefore, contribute to reducing the risk of suffering a car accident for these patients.

The rest of this document is organized as follows: Section 2 presents the main physiological characteristics of OSA patients and the distinctive acoustic qualities of their voices, based on the previous literature and our initial contrastive study. The speech database used in our experimental work, as well as its design criteria, is explained in Section 3. Section 4 explores the advantages that standard GMMs can bring to apnoea voice characterization. In particular, we describe how we used GMMs to study nasalization in speech, comparing the voices of severe apnoea patients with those in a 'healthy' control group. Next, in Section 5 we present a test we carried out to assess the accuracy of a GMM-based system we developed to classify speakers (apnoea/non-apnoea). Finally, conclusions and a brief outline of future research are given in Section 6.

2 Physiological & Acoustic Characteristics in OSA Speakers

At present neither the articulatory/physiological peculiarities nor the acoustic characteristics of speech in apnoea speakers are well understood. Most of the more valuable information in this area can be found in Fox and Monoson's work [4], a perceptual study in which skilled judges compared the voices of apnoea patients with those of a control group (referred to as "healthy" subjects). The study showed that, although differences between both groups of speakers were found, acoustic cues for these differences are somewhat contradictory and unclear. What did seem to be clear was that the apnoea group had abnormal resonances that might be due to an altered structure or function of the upper airway. Theoretically, such an anomaly should result not only in respiratory but also in speech dysfunction. Consequently, the occurrence of speech disorder in OSA population should be expected, and it could include anomalies in articulation, phonation and resonance:

- ***Articulatory anomalies*:** Fox and Monoson stated that *neuromotor* dysfunction could be found in the sleep apnoea population due to a "*lack of regulated innervations 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. All types of dysarthria affect the articulation of consonants and vowels causing the slurring of speech.
- ***Phonation anomalies*:** These may be due to the heavy snoring of sleep apnoea patients, which can cause inflammation in the upper respiratory system and affect the vocal cords.
- ***Resonance anomalies*:** What seems to be clear is that the apnoea group has abnormal resonances that might be due to an altered structure or function of the upper airway causing *velopharyngeal dysfunction*. This anomaly should, in theory, result in an abnormal vocal quality related to the coupling of the vocal tract with the nasal cavity, and is revealed through two features:
    - Firstly, 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* and *hypernasality*. The former is said to occur when no nasalization is produced when the sound should be nasal. Hypernasality is nasalization during the production of non-nasal (voiced oral) sounds. The interested reader can find an excellent reference in [8]. Fox and Monoson's work on the nasalization characteristics for the sleep apnoea group was not conclusive. What they could conclude was that these resonance abnormalities could have been

associated with vocal tract damping features distinct from airflow in balance between the oral and nasal cavities, affecting speech sound quality. The term applied to this speech disorder is "*cul-de-sac*" resonance, a type of hyponasality that causes the sound to be perceived as if it were resonating in a blind chamber. Perhaps more importantly, it seems that speakers with apnoea may exhibit smaller intra-speaker differences between non-nasal and nasal vowels due to this velopharyngeal dysfunction (vowels ordinarily acquire either a nasal or a non-nasal quality depending on the presence or absence of adjacent nasal consonants).

- Secondly, due to the pharyngeal anomaly, differences in formant values can be expected. This is confirmed in Robb's work [9], 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 apnoea and healthy groups. In particular, the results of the formant frequency analysis showed that F1 and F2 values among the OSA group were generally lower than those in the non-OSA groups. The lower formant values were attributed to greater vocal tract length.

These types of anomalies may occur either in isolation or combined. However, none of them was found to be sufficient on its own to allow accurate assessment of the OSA condition. In fact, all three descriptors were necessary to differentiate and predict whether the subject was in the normal group or in the OSA group.

2.1 Initial Contrastive Acoustic Study

In order to build on the relatively little knowledge available in this area, a preliminary acoustic analysis in an initial version of our apnoea speech database was made. In a related piece of research, Fiz et al. [10] applied spectral analysis on sustained vowels to detect possible apnoea-pathological cases. They used the following acoustic features: maximum frequency of harmonics, mean frequency of harmonics and number of harmonics. They found statistically significant differences between a control group (healthy subjects) and the sleep apnoea group regarding the maximum harmonic frequency for the vowels /i/ and /e/, it being lower for OSA patients. Another piece of research on the acoustic characterization of sustained vowels in apnoea patients using *Linear Predictive Coding* (LPC) can be found in [11]. However, these studies do not investigate all of the possible acoustic peculiarities that may be found in the voices of apnoea patients, since focusing solely on sustained vowels precludes the discovery of acoustic effects that occur in continuous speech only in certain linguistic contexts.

Our contrastive study consisted of a perceptual and visual comparison of frequency representations (mainly spectrographic, pitch, energy and formant analysis) of an initial apnoea and control group speakers uttering a same set of 25 phonetically balanced sentences. Following up on previous research on the acoustic characteristics of OSA speakers, we searched for articulatory and resonance anomalies in apnoea-suffering speakers.

- ***Articulatory anomalies.*** An interesting conclusion from our initial perceptual contrastive study was that, when comparing the distance between the second (F2) and third formant (F3) for the vowel /i/, clear differences between the apnoea and control groups were found. For apnoea speakers the distance was greater, and this was especially clear in diphthongs with /i/ as the stressed vowel, as in the Spanish word "Suiza" (*'suj θa*). This finding may be related to the greater length of the vocal tract of OSA patients [9], but also, and perhaps more importantly, to a characteristically abnormal velopharyngeal opening which may cause a shift in the position of the third formant [12]. Indeed, a lowering of the velum (typical in apnoea speakers) is known to produce higher third formant frequencies.
- ***Resonance anomalies.*** Fox and Monoson state in [4] that a common resonance feature in apnoea patients is abnormal nasality. The presence and the size of one extra low frequency formant can be considered an indicator of nasalization [13], but no perceptual differences between the groups in the overall nasality level could be found. As discussed in previous sections, this could be due to common perceptual difficulties to classify the voice of apnoea speakers as hyponasal or hypernasal. However, we did find differences in both groups (apnoea and non-apnoea) in how nasalization varied from nasal to non-nasal contexts and vice versa. Interestingly, we found variation in nasalization to be smaller for OSA speakers. One hypothesis is that the voices of apnoea speakers have a higher overall nasality level caused by velopharyngeal dysfunction, so differences between oral (no-nasal) and nasal vowels are smaller than normal because the oral vowels are also nasalized. An explanation for this could be that apnoea speakers have weaker control over the velopharyngeal mechanism, which may cause difficulty in changing nasality levels, whether absolute nasalization level is high or low. These hypotheses are intriguing and we will delve deeper into them later.

3 Apnoea Database

The database was recorded in the Respiratory Department at *Hospital Clínico Universitario of Málaga*, Spain. It contains the readings and facial images of 80 male subjects; half of them suffer from severe sleep apnoea (AHI > 30), and the other half are either healthy subjects or only have mild OSA (AHI < 10). Subjects in both groups have similar physical characteristics such as age and Body Mass Index (BMI). The speech material for the apnoea group was recorded and collected in two different sessions: one just before being diagnosed and the other after several months under CPAP treatment. This allows studying the evolution of apnoea voice characteristics for a particular patient before and after treatment.

3.1 Speech and Image Collection

Speech was recorded using a sampling frequency of 48 kHz in an acoustically isolated booth. The recording equipment consisted of a standard laptop computer with a conventional sound card equipped with a *SP500 Plantronics* headset microphone with A/D conversion and digital data exchange through a USB-port.

Additionally, for each subject in the database, two facial images (frontal and lateral views) were collected under controlled illumination conditions and over a flat white background. A conventional digital camera was used to obtain images in 24-bit RGB format, without compression and with 2272x1704 resolution. We decided to collect visual information because OSA is usually associated with a variable combination of different anatomic factors (e.g., narrowing of the upper airway, distinctive craniofacial and pharyngeal dimensions, etc) [14]. Although precise maxillary morphology analysis in OSA cases has been done using radiological analysis of lateral views, *Cephalometrics*, simple visual inspections are also considered as a first step when evaluating patients under clinical suspicion of suffering OSA [1]. Visual examination of patients includes searching for distinctive features of the facial morphology of OSA such as a short neck, characteristic mandibular distances and alterations, obesity, etc. These considerations motivated us to include photographs of the patients' faces in the database, for future research toward simple and cost-efficient automatic diagnosis of severe OSA patients using image processing techniques. To our knowledge, no research has ever been carried out to find facial features that may be extracted by image processing of both frontal and lateral views, and which may help diagnose severe apnoea cases.

3.2 Speech Corpus

In this sub-section, we describe the apnoea speaker database we designed with the goal of covering all the relevant linguistic/phonetic contexts in which physiological OSA-related peculiarities could have a greater impact. These peculiarities include the articulatory, phonation and resonance anomalies revealed in the previous research review (see Section 2). As we pointed out in the introduction, the central aim of our study is to apply speech processing techniques to automatically detect OSA-related traits in continuous speech. We believed that analysing continuous speech may well afford greater possibilities than working with sustained vowels because certain traits of pathological voice patterns, and in particular those of OSA patients, could then be detected in different sound categories (i.e. nasals, fricatives, etc) and also in the co-articulation between adjacent sound units [6].

The speech corpus contains readings of four sentences in Spanish repeated three times by each speaker that include instances of the following specific phonetic contexts that we derived from previous research:

- In relation to ***resonance anomalies***, we designed sentences that allow intra-speaker variation measurements; that is, measuring differential voice features for each speaker, for instance to compare the degree of vowel nasalization within and without nasal contexts.

- With regard to ***phonation anomalies***, we included continuous voiced sounds to measure irregular phonation patterns related to muscular fatigue in apnoea patients.
- Finally, to look at ***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, velar phonemes such as the Spanish velar approximant 'g'. This pharyngeal region has been seen to display physical anomalies in speakers suffering from apnoea. Thus, it is reasonable to suspect that different coarticulatory effects may occur with these phonemes in speakers with and without apnoea.

All the sentences were designed to exhibit a similar melodic structure, and speakers were asked to read them with a specific rhythmic structure under the supervision of an expert. We followed this controlled rhythmic recording procedure hoping to minimise non-relevant inter-speaker linguistic variability. The sentences used were the following, with the different melodic groups underlined separately:

1. **Francia, Suiza y Hungría ya hicieron causa común.**
   'fraN θja 'suj θa i uŋ 'gri a ya j 'θje roŋ 'kaw sa ko 'mun
2. **Julián no vio la manga roja que ellos buscan, en ningún almacén**.
   xu 'ljan no 'βjo la 'maŋ ga 'řo xa ke 'e λoz 'βus kan en niŋ 'gun al ma 'ken
3. **Juan no puso la taza rota que tanto le gusta en el aljibe**.
   xwan no 'pu so la 'ta θa 'řo ta ke 'taN to le 'γus ta en el al 'xi βe
4. **Miguel y Manu llamarán entre ocho y nueve y media.**
   mi 'γel i 'ma nu λa ma 'ran 'eN tre 'o tʃo i 'nwe βe i 'me ðja

The first phrase was taken from the *Albayzin* database, a standard phonetically balanced speech database for Spanish [15]. It was chosen because it contains an interesting sequence of successive /a/ and /i/ vowel sounds.

The second and third phrases, both negative, have a similar grammatical and intonation structure. They are potentially useful for contrastive studies of vowels in different linguistic contexts. Some examples of these contrastive pairs arise from comparing a nasal context, "m**a**ng**a** roja" (*'maŋ ga 'řo xa*), with a neutral context, "t**a**z**a** rota" (*'ta θa 'řo ta*). As we mentioned in the previous section, these contrastive analyses could be very helpful to confirm whether indeed the voices of speakers with apnoea have an altered overall nasal quality and display smaller intra-speaker differences between non-nasal and nasal vowels due to velopharyngeal dysfunction.

The fourth phrase has a single and relatively long melodic group containing mainly voiced sounds. The rationale for this fourth sentence is that apnoea speakers usually show fatigue in the upper airway muscles. Therefore, this sentence may be helpful to discover various anomalies during the sustained generation of voiced sounds. These phonation-related features of segments of harmonic voice can be characterized following any of a number of conventional approaches that use a set of individual measurements such as the Harmonic to Noise Ratio- HNR, periodicity measures and pitch dynamics (e.g. jitter). Finally, with regard to the resonance anomalies found in the literature, one of the possible traits of apnoea speakers is dysarthria. Our sentence can be used to analyse dysarthric voices that typically show differences in vowel space with respect to normal speakers.

4 Apnoea Voice Modelling with GMM

In this section we present an initial experimentation that sheds light on the potential of using *Gaussian Mixture Models* techniques (GMMs) over speech spectra (*cepstral* domain) to discover and model peculiarities in the acoustical signal of apnoea voices. These peculiarities might be related to the perceptually distinguishable traits described in previous research and corroborated in our preceding contrastive study. The main reason for using GMMs over the cepstral domain is related to the great potential this combination of techniques has shown for the modelling of the acoustic space of human speech. For our study we required a good modelling of the anomalies described in Section 2, which we expected to find in OSA patients. Since cepstral coefficients are related with the spectral envelope of speech signals, and therefore with the articulation of sounds, and since GMM training sets can be carefully selected in order to model specific characteristics (for instance in order to consider resonance anomalies in particular), it seems promising to combine all this information in a fused model. We should expect such a model to be useful for describing the acoustic spaces of both the OSA patient group and the healthy group, and for discriminating between them.

This approach was applied to specific linguistic contexts. In particular, as our apnoea speech database was designed to allow a detailed contrastive analysis of vowels in no-nasal (oral) and nasal phonetic contexts, we focus on reporting perceptual differences related to resonance anomalies that could be perceived as either hyponasality or hypernasality. For this purpose, sub-section 4.2 discusses how GMM techniques can be applied to study these differences in degree of nasalization in different linguistic contexts.

4.1 GMM Training and Testing Protocol

Gaussian Mixture Models (GMMs) and adaptation algorithms are effective and efficient pattern recognition techniques suitable for sparse speech data modelling in Automatic Speaker Recognition systems [7] that we will apply for apnoea voice modelling. In our experimental framework, the *BECARS* open source tool [16] was used. Details on parameterization, model training and classification phases for the BECARS baseline system are the following:

- Parameterization consists in extracting information from speech signal. Our speech database was processed using short-time spectral analysis with a 20 ms time frame and a 10 ms delay between frames, which gives a 50% overlap. We chose an appropriate parameterization for the information, using 39 standard components: 12 *Mel Frecuency Cepstral Coefficients* (MFCCs), plus energy, extended with their speed (delta) and acceleration (delta-delta) components.
- GMMs for apnoea and ‘healthy’ control groups were trained as follows. First, a universal background model (UBM) was trained from a phonetically balanced subcorpus of Albayzin Database [15]. Next, speech from apnoea and non-apnoea group speakers was used to train the apnoea and control

group GMM models. Both models were trained using MAP (Maximum a Posteriori) adaptation from the UBM. This technique increases the robustness of the models especially when sparse speech material is available. Only the means were adapted, as is classically done in speaker verification. Obviously, speech data from speakers for the model training was not included in the test set.

- For testing purposes, and in order to increase the number of tests and thus to improve the statistical relevance of our results, the standard *leave-one-out* testing protocol was used. This protocol consists in discarding one sample speaker from the experimental database to train the classifier with the remaining samples. Then the excluded sample is used as the test data. This scheme is repeated until a sufficient number of tests have been performed.

4.2 A Study of Apnoea Speaker Resonance Anomalies using GMMs

With the aim of testing the capabilities of using the GMM-based experimental set-up, we performed an initial study to try to model certain resonance anomalies that have already been described for apnoea speakers in preceding research [4] and revealed in our own contrastive acoustic study. In particular, as our apnoea speech database was designed to allow a detailed contrastive analysis of vowels in oral and nasal phonetic contexts, we used GMM techniques to perform a study to identify differences in degree of nasalization in different linguistic contexts.

For that purpose, we generated 40 sub-groups of apnoea speakers and 40 sub-groups of control speakers discarding one sample for each class from the whole set of pathological and healthy patients. Two GMMs for each *cluster* were trained using speech with all nasalized and non-nasalized vowels from speakers of each group. Both nasal and non-nasal GMMs were trained following the approach described in Section 4.1. MAP adaptation was carried out with a generic vowel UBM trained using Albayzin database [15]. These two nasal/non-nasal GMMs for each sub-group were used to quantify the acoustic differences between nasal and non-nasal contexts in both the apnoea and the control patients. The smaller the difference between the nasal and the non-nasal GMMs the more similar the nasalized and the non-nasalized vowels are. Unusually similar nasal and non-nasal vowels for any group of speakers reveals the presence of resonance anomalies. We took a fast approximation of the *Kullback-Leibler (KL) divergence* for Gaussian Mixture Models [17] as a measure of distance between nasal and non-nasal GMMs. This distance is commonly used in Automatic Speaker Recognition to define cohorts or groups of speakers producing similar sounds.

As can be seen in Figure 1, the distance between nasal and non-nasal vowel GMMs was significantly larger for the control group speakers than for the speakers with severe apnoea. This interesting result confirms that the margin of acoustic variation for vowels articulated in nasal vs. non-nasal phonetic contexts is narrower than normal in speakers with severe apnoea. It also validates the GMM approach as a powerful speech processing and classification technique for research on OSA voice characterization and the detection of OSA speakers.

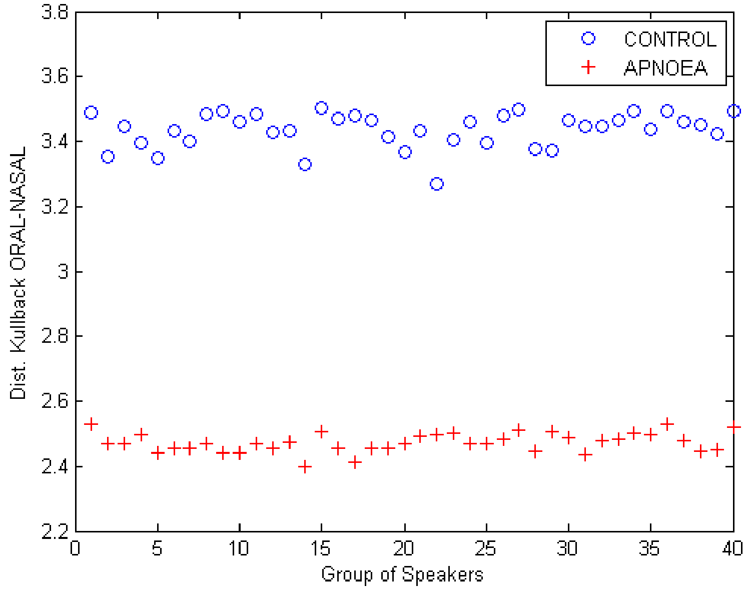

Fig. 1. Kullback-distance differences between nasal and non-nasal models for clusters of severe apnoea (*plus*) and 'healthy' control (*circles*) speakers.

5 Automatic Diagnosis of Severe Apnoea using GMMs

As we have suggested in the previous section, GMM-based ASR techniques can discriminate some of the resonance anomalies of apnoea speakers that have already been described in the literature. Thus it seems reasonable to explore the possibilities of applying GMM-Based Speaker Recognition techniques for the automatic diagnosis of severe apnoea. This method could be suitable for keeping track of the progress of voice dysfunction in OSA patients, it is easy-to-use, fast, non-invasive and much cheaper than traditional alternatives. While we do not suggest it should replace current OSA diagnosis methods, we believe it can be a great aid for early detection of severe apnoea cases.

A speaker verification system is a supervised classification system capable of discriminating between two classes of speech signals (usually 'genuine' and 'impostor' speakers). For our present purposes the classes are not defined by reference to any particular speaker. Rather, we generated a general severe sleep apnoea class and a control class (speech from healthy subjects) by grouping together all the training data from speakers of each class. Thus a GMM model was trained for all the speakers belonging to the apnoea class and another GMM model for those speakers in the control class. Following a similar approach to that of other pathological voice assessment studies [17], GMMs representing the pathological and healthy classes were built as follows:

- The apnoea and control GMMs were trained from the generic UBM relying on MAP adaptation and the standard *leave-one-out* technique, similarly to how we described above (Section 4.1).
- During the apnoea/non-apnoea detection phase an input speech signal corresponding to the whole utterance of the speaker to be diagnosed is presented to the system. The parameterised speech is then processed with each apnoea and control GMM generating two likelihood scores. From these two scores an apnoea/control decision is made according to a decision threshold adjusted beforehand as a trade-off to achieve acceptable rates of both failure to detect apnoea voices (false negative) or falsely classifying healthy cases as apnoea voices (false positive).

Table 1 shows the correct classification rates we obtained when we applied the GMM control/pathological voice classification approach to our speech apnoea database [5]. We see that the overall correct classification rate was 81.25 %.

Table 1. Correct Classification Rate.

Correct Classification Rate in %	*Control Group*	*Apnoea Group*	*Overall*
	77,50 % (31/40)	**85 % (34/40)**	**81,25 % (65/80)**

We now evaluate the performance of the classifier using the following criteria:

- **Sensitivity**: ratio of correctly classified apnoea-suffering speakers (true positives) to total number of speakers actually diagnosed with severe apnoea.
- **Specificity**: ration of true negatives to total number of speakers diagnosed as not suffering from apnoea.
- **Positive Predictive Value**: ratio of true positives to total number of patients GMM-classified as having a severe apnoea voice.
- **Negative Predictive Value**: ratio of true negatives to total number of patients GMM-classified as *not* having a severe apnoea voice.

Table 2 shows the values we obtained in our test for these measures of accuracy (Fisher's exact test revealed statistically significant, $p<0.0001$):

Table 2. Sensitivity, Specificity, Positive Predictive Value and Negative Predictive Value.

Sensitivity	*Specificity*	*Positive Predictive Value*	*Negative Predictive Value*
77,50 % (31/40)	**85 % (34/40)**	**83,78 % (31/37)**	**79,06 % (34/43)**

Some comments are in order regarding the correct classification rates obtained. The results are encouraging and they show that distinctive apnoea traits can be identified by a GMM based-approach, even when there is relatively little speech material with which to train the system. Furthermore, such promising results were obtained without paying choosing any acoustic parameters in particular on which to base the classification. Better results should be expected with a representation and parameterization of audio data that is optimized for apnoea discrimination. Obviously, our experiments need to be validated with a larger test sample. Nevertheless, our results already give us an idea of the discriminative power of this approach to automatic diagnosis of severe apnoea cases.

6 Conclusions & Future Research

In this paper we have presented pioneering research in the field of automatic diagnosis of severe obstructive sleep apnoea. The acoustic properties of the voices of speakers suffering from OSA were studied and an apnoea speech database was designed attempting to cover all the major linguistic contexts in which these physiological OSA features could have a greater impact. For this purpose we analyzed in depth the possibilities of applying state-of-the-art GMM techniques to the modelling of the peculiar features of the realizations of certain phonemes by apnoea patients. In relation with this issue, we focused on nasality as an important feature in the acoustic characteristics of apnoea speakers. Our state-of-the-art GMM approach has confirmed that there are indeed significant differences between apnoea and control group speakers in terms of relative levels of nasalization between different linguistic contexts. Furthermore, we tested the discriminative power of GMM-based speaker recognition techniques adapted to severe apnoea detection with promising experimental results. A correct classification rate of 81.25 % shows that GMM-based OSA diagnosis could be useful for the preliminary diagnosis of apnoea patients and, which suggests it is worthwhile to continue to explore this area.

Regarding future research, our automatic apnoea assessment needs to be validated with a larger sample from a broader spectrum of population. Furthermore, best results can be expected using a representation of the audio data that is optimized for apnoea discrimination. Regarding the significant differences between apnoea and healthy speakers on the relative nasalization degree between different linguistic contexts, an interesting study would be focused on exploiting this information by means of nasalization measures, in order to the improvement of the GMM-based apnoea/non-apnoea detector. Finally, we mention that future research will also be focused on exploiting physiological OSA features in relevant linguistic contexts in order to explore the discriminating power of each feature using linear discriminant classifiers or calibration tools. We aim to apply these findings to improve the performance of the automatic apnoea diagnosis system.

Acknowledgments. The activities described in this paper were funded by the Spanish Ministry of Science and Technology as part of the TEC2006-13170-C02-02 project. The authors would like to thank the volunteers at *Hospital Clínico Universitario of Málaga*, and to Guillermo Portillo who made the data collection possible.

References

1. Puertas, F.J., Pin, G., María, J.M., & Durán, J.: Documento de consenso Nacional sobre el síndrome de Apneas-hipopneas del sueño (SAHS). Grupo Español De Sueño (GES) (2005)
2. Coccagna, G., Pollini, A., & Provini, F.: Cardiovascular disorders and obstructive sleep apnea syndrome. In Clinical and Experimental Hypertension, Vol. 28:217–224 (2006)
3. Lloberes, P., Levy, G., Descals, C., Sampol, G., Roca, A., Sagales T., & de la Calzada, M.D.: Self-reported sleepiness while driving as a risk factor for traffic accidents in patients with obstructive sleep apnoea syndrome and in non-apnoeic snorers. In Respiratory Medicine 94, nº 10:971–976 (2000)
4. Fox, A.W., & Monoson, P.K.: Speech dysfunction of obstructive sleep apnea: A discriminant analysis of its descriptors. In Chest Journal 96(3):589-595 (1993)
5. Fernández R., Hernández L.A., López E., Alcázar J., Portillo G., & Toledano D.T.: Design of a Multimodal Database for Research on Automatic Detection of Severe Apnoea Cases. In Proceedings of 6th Language Resources and Evaluation Conference, LREC, Marrakech, Morocco (2008)
6. Parsa, V., & Jamieson, D.G.: Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech. In Journal of Speech, Language, and Hearing Research, ISSN 1092-4388, Vol. 44, no2:327-339 (2001)
7. Reynolds, D.A., Quatieri, T.F., & Dunn, R.B.: Speaker verification using adapted gaussian mixture models. In Digital Signal Processing, Vol. 10:19-41 (2000)
8. Pruthi, T.: Analysis, vocal-tract modeling and automatic detection of vowel nasalization. Doctor Thesis at the University of Maryland (2007)
9. Robb M., Yates J., & Morgan E.: Vocal Tract Resonance Characteristics of Adults with Obstructive Sleep Apnea. In Acta Otolaryngologica, 117:760-763 (1997)
10. Fiz, J.A., Morera, J., Abad, J., Belsulnces, A., Haro, M., Fiz, J.I., Jane, R., Caminal, P., & Rodenstein, D.: Acoustic analysis of vowel emission in obstructive sleep apnea. In Chest Journal, Vol. 104:1093–1096 (1993)
11. Obrador, A., Haro, M., Álvarez, Ll., Calderón, J.C., Sabater, G., Casamitja, M.T., & Sendra, S.: El análisis de la voz y la pulsioximetria nocturna como técnicas de despistaje del síndrome de apnea-hipopnea en el sueño. Congreso Nacional de Neumología, Tenerife, Spain (2008)
12. Hidalgo, A., & Quilis, M.: Fonética y fonología españolas. Edi. Tirant blanch (2002)
13. Glass, J., & Zue, V.: Detection of nasalized vowels in American English**.** In Proceedings of Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP, Vol. 10:1569-1572. Tampa, USA (1985)
14. Cakirer, B., Hans, M.G., Graham, G., Aylor, J., Tishler, P.V., & Redline, S.: The Relationship between Craniofacial Morphology and Obstructive Sleep Apnea in Whites and in African-Americans. In American Journal of Respiratory and Critical Care Medicine, Vol. 163, Number 4:947-950 (2001)
15. Moreno, A., Poch, D., Bonafonte, A., Lleida, E., Llisterri, J., Mariño, J.B., & Naude, C.: ALBAYZIN Speech Database: Design of the Phonetic Corpus. In Proceedings of Eurospeech, Vol. 1:175-178. Berlin, Germany (1993)
16. Blouet, R., Mokbel, C., Mokbel, H., Sanchez Soto, E., Chollet, G., & Greige, H.: BECARS: a Free Software for Speaker Verification. In Proceedings of The Speaker and Language Recognition Workshop, ODYSSEY, pp 145-148. Toledo, Spain (2004)
17. Do, M.N.: Fast approximation of Kullback-Leibler distance for dependence trees and Hidden Markov Models. In IEEE Signal Processing Letter, Vol. 10: 115-118 (2003)
18. Fredouille, C., Pouchoulin, G., Bonastre, J.F., Azzarello, M., Giovanni, A., & Guio, A.: Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). In Proceeding of 9th European Conference on Speech Communication and Technology, Interspeech, pp:149-152. Lisboa, Portugal (2005)