Analyzing GMMs to characterize resonance anomalies in speakers suffering from apnoea

José Luis Blanco [1], *Rubén Fernández*[1], *David Pardo*[1], *Álvaro Sigüenza*[1], *Luis A. Hernández*[1], *José Alcázar*[2]

[1] Signal, Systems & RadioCommunications Department. Universidad Politécnica de Madrid, Spain
[2] Respiratory Department. Hospital Torrecardenas, Almeria, Spain
jlblanco@gaps.ssr.upm.es

Abstract

Past research on the speech of apnoea patients has revealed that resonance anomalies are among the most distinguishing traits for these speakers. This paper presents an approach to characterize these peculiarities using GMMs and distance measures between distributions. We report the findings obtained with two analytical procedures, working with a purpose-designed speech database of both healthy and apnoea-suffering patients. First, we validate the database to guarantee that the models trained are able to describe the acoustic space in a way that may reveal differences between groups. Then we study abnormal nasalization in apnoea patients by considering vowels in nasal and non-nasal phonetic contexts. Our results confirm that there are differences between the groups, and that statistical modelling techniques can be used to describe this factor. Results further suggest that it would be possible to design an automatic classifier using such discriminative information.

Index Terms: Obstructive Sleep Apnoea (OSA), Gaussian Mixture Models (GMMs), Abnormal Resonance

1. Introduction

Obstructive sleep apnoea (OSA) is a highly prevalent disease [1] which affects an estimated 2-4% of middle-aged adults. It is characterized by recurrent episodes of sleep-related collapse of the upper airway at the level of the pharynx, 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 (cardiovascular diseases, traffic accidents, etc). It 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 –a conventional *Polysomnography* which involves the recording of neurophisiological and cardiorespiratory variables (ECG)– is usually needed to confirm presence of the disorder. This diagnostic procedure is expensive and time-consuming, and patients must often endure long waiting lists before the test is carried out.

Alternative methods for early diagnosis of apnoea patients would be greatly beneficial, primarily if they allow a significant reduction of the time-to-diagnosis. Speech-based methods for OSA detection are promising in this respect by virtue of their non-intrusive nature and their potential to provide quantitative data relatively quickly. Since the upper airways are affected by OSA disease, it seems reasonable to consider whether there are any distinctive speech signal patterns associated with OSA. Research in this area has begun to produce evidence supporting this idea. Much valuable information can be found in Fox and Monoson's work [2], a perceptual study in which skilled judges compared the voices of apnoea patients with those of a control group (referred to as "healthy" subjects). They observed certain peculiarities in the voices of OSA patients, such as abnormal resonance (the work we present here focuses mainly on this factor) and both articulation and phonation abnormalities. These anomalies, rather consistently present in OSA speakers and absent in speakers without the condition, open the path to explore automatic methods to discriminate between both kinds of voices, and thus help in the early diagnosis of obstructive sleep apnoea.

Identifying the distinctive features of OSA speech requires a dedicated effort to design and collect a consistent database that allows contrasting speech data from OSA-suffering and healthy speakers, highlighting those elements of speech in which the reported OSA-related anomalies are commonly found. This requires recording a purpose-designed speech corpus. Our corpus design follows phonetic and linguistic criteria derived from the previous work of Fox and Monoson [2], and it also incorporates data from a preliminary database described in [3].

Other relevant literature delves into certain specific aspects of the acoustic analysis of OSA speaker voices. For instance, interested readers will find in [4] an excellent description, from a physiological point of view, of vocal tract resonances in OSA adults. The study condensed in Fiz et al. [5] is also useful background work for our purposes, as they focus, as we do, on both apnoea disease and vowel sounds. However, while they consider direct inspection of the spectral representation of the collected data, we apply generative statistical modelling techniques based on *Gaussian Mixture Models* (GMMs) to describe acoustical spaces (those of specific sounds, speakers or speaker groups) conveniently for the purpose of characterizing the voice of apnoea speakers. Following this approach further recognition or classification tasks can be performed based on the likelihood that a given unknown sound or utterance was generated by a trained model, similarly to what is done in Automatic Speaker Recognition systems (ASR). In previous work [6] excellent classification rates were achieved by modelling short-time speech spectrum information with cepstral coefficients and using GMM-based classification techniques.

In the present contribution we test the potential of using GMMs to model and characterize distinctive apnoea voices. In [7] we already successfully applied this method to generic vowel sounds, confirming that there are indeed significant differences between apnoea and ”healthy” group speakers, and that GMM techniques are capable of describing this discriminative information. First we will validate our speech database to guarantee that the heuristic GMM models trained condense enough information to distinguish between both groups. Next, we will focus on abnormal resonances that appear in apnoea speakers, since distinguishing traits for OSA patients have been traditionally sought for in this acoustical aspect. Due to an altered structure of the upper airway, this anomaly should result in an abnormal vocal quality and, in theory, apnoea speakers should produce speech with “inappropriate nasal resonance” [2]. Fox and Monoson’s work on the nasalization characteristics of speakers with sleep apnoea was not conclusive. What they could conclude, however, was that the resonance abnormalities could be perceived either as a form of hyponasality or hypernasality. Perhaps more importantly, speakers with apnoea may exhibit smaller intra-speaker differences between non-nasal and nasal vowels due to this dysfunction (vowels ordinarily acquire either a nasal or a non-nasal quality depending on the presence or absence of adjacent nasal consonants). We expect to shed light on this issue using generative statistical modelling based on GMMs to study this abnormal nasalization in OSA patients. For this purpose we compare the acoustic characteristics of apnoea and healthy voices in nasal and non-nasal vowels using an approximation to the *Kullback-Leibler divergence*.

The remainder of this paper is organized as follows. In Section 2 we present the methodology and experimental setup for our study. Later, in Section 3 experimental results for two different tests are presented. First we will seek validation of the speech apnoea database collected. Secondly, we describe how we used GMMs to study nasalization in speech, comparing the voices of apnoea patients with those in a ‘healthy’ control group. Finally, discussion and conclusions are given in Section 4.

2. Method

2.1. GMM-based Method

*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 [8]. We used the *BECARS* open source tool in our experimental framework [9]. Details on the parameterization and model training for the baseline system now follow.

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. For the task of acoustical space modelling we chose to use 39 standard components: 12 Mel Frecuency Cepstral Coefficients (MFCCs), plus energy, extended with their speed (delta) and acceleration (delta-delta) components. We acknowledge that a representation of the acoustic space that is optimized for the specific aspects we are studying (those related to the resonance anomalies of apnoea speakers) could provide better results, but this would require specific adaptation of the recognition techniques we apply, which is not the intended focus of the work we present here.

After parameterization, statistical pattern recognition can be applied to study or compare voices for specific speech segments. We trained a universal background GMM model (UBM) from phonetically balanced utterances taken from the Albayzin database [10], and used MAP (*Maximum A Posteriori*) adaptation to derive the specific GMMs for the different classes to be trained [8]. This technique increases the robustness of the models especially when sparse speech material is available. Only the means were adapted, as is typically done in speaker verification.

2.2. Distance measure between GMM Models

Approximations to the *Kullback-Leibler divergence* have been widely used for measuring differences between probability distributions. However, many of these distance estimations, while easy to calculate, do not have the properties that the Kullback-Leibler divergence exhibits, and this must be taken into account for their correct interpretation. In the realm of automatic speech processing, when considering various GMM models obtained by MAP adaptation from a common mixture model (UBM), an upper bound to the previous divergence is used as a measure of the distance between the models. This bound can be obtained directly from the Kullback-Leibler divergence.

The Kullback-Leibler divergence for two GMMs, f_1 and f_2, is given on the left side of inequality (1) (it is trivial to see that the inequality holds).

(1)

$$D_{KL}(f_1,f_2)=\int f_1\log\frac{f_1}{f_2}=\int\left(\sum_i a_i f_1^i\right)\log\frac{\left(\sum_i a_i f_1^i\right)}{\left(\sum_i b_i f_2^i\right)}\leq\int\sum_i\left(a_i f_1^i\log\frac{a_i f_1^i}{b_i f_2^i}\right)$$

We call the right side $\tilde{D}(f_1, f_2)$, which is the upper bound we will use (2). Since the GMMs are derived from a common GMM, the weights a_i and b_i are equal. By virtue of the fact that the variances of both GMM probability distributions are equal, and since their components are Gaussian distributions, it can be proved [11] that

(2)

$$D_{KL}(f_1,f_2)\leq\tilde{D}(f_1,f_2)=\sum_i a_i\int f_1^i\log\frac{f_1^i}{f_2^i}=\sum_i\frac{a_i}{2}\left[(\mu_1^{(i)}-\mu_2^{(i)})^T\Sigma_i^{-1}(\mu_1^{(i)}-\mu_2^{(i)})\right]$$

This new distance, which may be interpreted as a weighted sum of the *Mahalanobis* distances between every multidimensional Gaussian distribution in the mixture, has several attractive properties. Most importantly, it has a lower computational cost, especially compared to the Monte-Carlo methods required to accurately estimate the actual Kullback-Leibler divergence. It is also symmetric, so $\tilde{D}(f_1, f_2)=\tilde{D}(f_2, f_1)$, and it is tight to the Kullback-Leibler divergence, as shown in [12].

3. Experiments

In this section we present two experiments that shed light on the potential of using the approach we have already described (GMM-based models and distance measures on the data we have at our disposal) to discover and model peculiarities in the acoustical signal of apnoea voices. First of all, we will try to guarantee that the GMM models built describe the acoustic space accurately. After this preliminary analysis, sub-section 3.2 discusses how GMM techniques can be applied to study the OSA resonance anomalies identified in the previous research review. We will study differences in degree of nasalization in different linguistic contexts.

3.1. Analyzing the Speech Database

All the required data was extracted from the previously mentioned database we collected [3], because, to our knowledge, there were no other available resources we could use for this specific task. The database contains the recordings of 80 Spanish male subjects; half of them suffer from severe sleep apnoea, and the other half are either healthy subjects or have only mild OSA. As we pointed out in the introduction, the database has been designed to expect to cover relevant linguistic/phonetic contexts in which physiological OSA-related peculiarities could have a greater impact. This includes:

- In relation to resonance anomalies, we designed sentences that allow intra-speaker variation measurements; that is, including vowels in different linguistic contexts to measure, for instance, how nasalization varied from nasal to non-nasal contexts (the focus of this study)
- 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; anatomical region has been seen to display physical anomalies in OSA speakers.

Since we needed to consider acoustical features in specific phonetic contexts as we will see, we performed an automatic phonetic segmentation of every utterance in the database using the HTK open-source tool [13]. Using automatic forced alignment avoids the need for costly annotation of the data set by hand. It also guarantees good quality segmentation, which is crucial if we are to distinguish phonemes and phonetic contexts properly.

Once we have collected and segmented our speech database, it is necessary to validate it and to ensure that the heuristic GMM models which we will train condense enough information to distinguish between both groups: patients suffering from OSA and healthy people. The way to do this is by building successive GMM models increasing the size of the data used to train them, and calculating the distances between non-nasal and nasal vowels (using the measure described in Section 2.2), for both healthy people from the control group and OSA-suffering patients. Figure 1 summarises the results we obtained, showing both mean distance and standard deviation for the resulting values across the various experiments we carried out for each size of the data set.

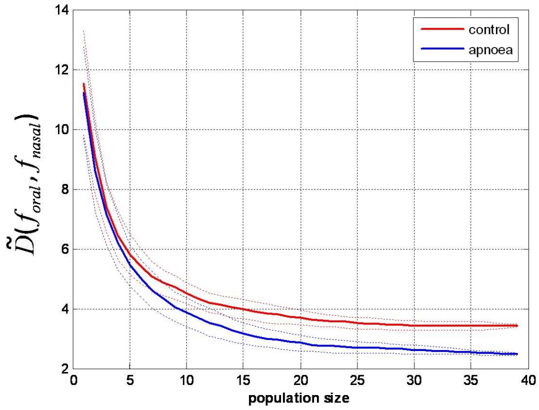

Figure 1: Distance between the acoustic models of vowels in non-nasal and nasal contexts, for both the control group and the apnoea patients, as a function of the size of the training population

As the figure suggests, as the amount of data used to train the new GMM models increases (by varying the number of speakers used), the models converge, with respect to the defined distance, to a common model for each of the classes. In fact the observed convergence is quite stable, as the deviation of the distance between models for the same population size diminishes with it, and is quite small for the final models conveying data from 39 speakers for each class. Given the fact that the distance used is an upper bound of the Kullback-Leibler divergence, that it is tight to the latter, and that our results converge, we can guarantee that the GMMs generated accurately describe the acoustic spaces for the given classes. We can further affirm that they are relatively stable with respect to the training set and that they converge also with respect to the Kullback-Leibler divergence.

3.2. Study of resonance anomalies using GMMs

The previous analysis offers a measure of the distance between inter-class models, i.e. between non-nasal and nasal vowels for both healthy subjects from the control group and patients suffering from apnoea. However, certain other measurements may be useful to characterize the distance between these classes. It would be possible to design a classifier of this information, with its discriminative power in some way associated with such measurements.

We ultimately want to measure the differences between both classes, so it seems reasonable to evaluate the distance between both of them directly, paying no regard to the different linguistic contexts of the vowels the acoustic parameters of which we are analysing, by combining the data for all of the speakers in each group. The distance thus calculated between our two classes (“apnoea” and “healthy”) was 1.66 ± 0.05. This is a reasonable measure of the distance between the trained generative heuristic models of both classes, and it may be taken as a reference of the classification problem. We found the distance to be reasonably stable throughout the set of experiments that we carried out. For this reason we may take it as a reference for the distances we calculate next. We note that we will not make any further mention to the variability observed in this distance measure, the limits of which are represented as dotted lines in Figure 1. For all distances some dispersion obviously exists, but in all cases it is negligible for the models derived for 39 speakers.

The next step is to analyse different linguistic contexts using the same metric. A model was generated for each of the two sub-classes nasal and non-nasal vowels, for each of the subject groups (OSA and non-OSA), and the distances between the subclasses across the subject groups were then estimated. The following table shows the results obtained.

Table 1: Distance measures between both classes in nasal and non-nasal contexts

		APNOEA	
		Non-nasal	nasal
CONTROL	Non-nasal	1.89	4.24
	nasal	3.93	1.95

Finally, two further measures should be considered, which are none other than the distances between nasal and non-nasal vowels within the acoustic space of each of the subject groups that were shown in Figure 1. For this distance we obtained an estimated value of 3.45 for the control group and of 2.45 for the OSA patients.

4. Discussion & Conclusions

Bearing in mind the useful properties of the metric we have defined, and the limited scattering of the distances obtained, interesting insights can be drawn from the results.

It can be seen in Figure 1 that the distances between models is bigger when comparing nasal and non-nasal vowels from the control group than when comparing the same in the apnoea patient group. This first result supports Fox and Monoson's finding that nasal and non-nasal-context vowels are harder to distinguish in the case of OSA-suffering speakers. It also suggests that we should expect good classification rates with a classifier based on this idea. This result just reflects a difference in intra-group distances between the two types of vowels, but it does not establish any relation between the nasal and non-nasal contexts of both experimental groups. Therefore, we cannot establish, based on this measure alone, whether OSA-suffering speakers tend to hypernasalize or, rather, hyponasalize.

From Table 1 we can see that the non-nasal vowels from the apnoea group are much more similar to vowels in a non-nasal context from the control group than to nasal vowels in this group, and the converse is true for the nasal-context vowels. Therefore, when classifying both groups, non-nasal vowels from the apnoea patients would more likely be confused with non-nasal vowels from the control group, rather than with the nasal ones, and vice versa. But are the abnormal effects present in the speech of apnoea patients associated with one type of vowel rather than the other? Well, we may observe that the estimated distance between the non-nasal vowels of both groups is approximately the same as the distance between the nasal vowels across both groups. Therefore, whatever the phenomena causing this difference between both groups of subjects may be, it affects both non-nasal and nasal vowels to a similar extent. We suspect this may be caused by the abnormal coupling and decoupling of the oral and nasal cavities, which would explain the fact that the effect is observable whether the context is nasal or non-nasal. This particular result fits Fox and Monoson's conclusions well [2]. We are very encouraged by the conclusion that it is possible to use GMM techniques to describe the observable differences between apnoea and healthy speakers

Finally, the distance found between both classes is the smallest of all of the measures estimated (1.66). Therefore,, it seems reasonable to expect better classification results when using the nasal / non-nasal subclasses instead of the overall datasets (which were the ones we used in [6]), assuming there is enough data for a precise heuristic modelling of the distributions by adapting the primitive UBM. We remain confident that applying these findings to improve the performance of automatic apnoea diagnosis using speech processing algorithms on continuous speech is a distinct possibility. However, much work must yet be done to provide a more accurate description of the resonance anomalies observed in patients suffering from OSA.

5. Acknowledgements

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.

6. 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] 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. 1989

[3] Fernandez 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. 2008

[4] Robb M., Yates J., and Morgan E. "Vocal Tract Resonance Characteristics of Adults with Obstructive Sleep Apnea" *Acta Otolaryngologica*, 117, 760--763. 1997

[5] 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*; 104: 1093 – 1096. 1993

[6] Fernandez R., Blanco J.L., Hernández L. A., López E., Alcázar J., & Toledano D. T. "Severe apnoea detection using speaker recognition techniques". In BIOSIGNALS 2009 International Conference on Bio-inspired Systems and Signal Processing, Oporto. 2009

[7] Blanco J.L., Fernández R., Díaz D., Hernández L. A., López E., & Toledano D. T. "Apnoea voice characterization through vowel sounds analysis using generative gaussian mixture models" To appear in 3rd Advanced Voice Function Assessement International Workshop, Madrid. 2009

[8] Reynolds, D.A., Quatieri, T.F., & Dunn, R.B. "Speaker verification using adapted gaussian mixture models". In Digital Signal Processing 10: 19-41. 2000

[9] 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. 2004)

[10] 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 Eurospecch 93. Berlin, Germany, 21-23. Vol. 1 pp. 175-178. 1993

[11] Ramos-Castro, D., Fierrez-Aguilar, J., González-Rodríguez J. & Ortega-García J. "Speaker verification using speaker- and test-dependent fast score normalization". In Pattern Recognition Letters, vol 98-1.90-98. 2007

[12] Do, M. N., "Fast approximation of Kullback-Leibler distance for dependence trees and Hidden Markov Models". IEEE Signal Processing Letter 10, 115-118. 2003

[13] Young, S. (2002) "The HTK Book (for HTK Version 3.2)". First published December 1995, Revised for HTK Version 3.2.