GMM-based Classifiers for the Automatic Detection of Obstructive Sleep Apnea

J-A. Gómez- García[1], J-L. Blanco-Murillo[2], J-I. Godino-Llorente[1], L.A. Hernández Gómez[2] and G. Castellanos-Domínguez[3]

[1]*Bioingeniería y Optoelectrónica group (BYO), Universidad Politécnica de Madrid, 28030, Madrid, Spain*

[2]*Signal Processing Applications Group (GAPS), Universidad Politécnica de Madrid, 28040, Madrid, Spain*

[3]*Procesamiento y Reconocimiento de Señal group (PRS), Manizales, Colombia*

{*jorge.gomez.garcia, ignacio.godino*}*@upm.es, jlblanco@gaps.ssr.upm.es, cgcastellanosd@unal.edu.co*

Keywords: GMM, Supervector, GSV, Nuisance Attribute Projection, Pattern Recognition

Abstract: The aim of automatic pathological voice detection systems is to serve as tools, to medical specialists, for a more objective, less invasive and improved diagnosis of diseases. In this respect, the gold standard for those ~~system~~ include the usage of a optimized representation of the spectral envelope, either based on cepstral coefficients from the mel-scaled Fourier spectral envelope (Mel-Frequency Cepstral Coefficients) or from an all-pole estimation (Linear Prediction Coding Cepstral Coefficients) for characterization, and Gaussian Mixture Models for posterior classification. However, the study of recently proposed GMM-based classifiers as well as Nuisance mitigation techniques, such as those employed in speaker recognition, has not been widely considered in pathology detection labours. The present work aims at testing whether or not ~~the employment of~~ such speaker recognition tools might contribute to improve system performance in pathology detection systems, specifically in the automatic detection of Obstructive Sleep Apnea. The testing procedure ~~employs an Obstructive Sleep Apnea~~ database, in conjunction with GMM-based classifiers looking for a better performance. The results show that an improved performance might be obtained by using such approach.

1 INTRODUCTION

The Obstructive Sleep Apnea (OSA) is a highly prevalent disease affecting an estimated 2-4% of male population between the ages of 40-60 (Puertas et al., 2005). It is characterized by recurrent episodes of sleep-related collapses of the Upper Airway (UA) at the level of the pharynx, and is usually associated to loud snoring and increased daytime sleepiness. It is a serious threat to an individual's health as it is known to be a risk factor for hypertension and, possibly, cardiovascular diseases (Coccagna et al., 2006). Recently it has also been related to traffic accidents caused by somnolent drivers (Puertas et al., 2005; Coccagna et al., 2006; Lloberes et al., 2000), and to a poor quality of life and impaired work performance.

OSA is usually detected on the basis of an analysis of the patients characteristic history and a physical examination, though, a full overnight sleep study is required to confirm diagnosis. However, it involves the recording of neuroelectrophysical and cadiorespiratory variables (ECG), as well as complex and intensive post-processing of collected data. Despite the noticeable accuracy of such procedure (up to 90% (Penzel et al., 2002)), it turns out to be quite expensive and time-consuming, and patients usually have to be in waiting lists for several years. Those issues have motivated the research of early diagnosis tools which aim for more advantageous diagnosis of the pathology (Alcázar et al., 2009). For instance in (Fox et al., 1989), acoustic cues to the automatic detection of OSA were found. Particularly, several articulatory, phonation and resonance characteristic were identified when comparing voices from OSA patients with those from healthy ones. With that in mind, it might be reasonable to consider the automatic detection of OSA by means of recorded voice signals.

Traditionally, the automatic detection of pathologies using voice recordings has been mainly based on the estimation and monitoring of perturbation parameters such as *jitter* and *shimmer*, noise measures as harmonic-to-noise ratio, glottal to noise excitation ratio, normalized noise energy and turbulent noise index, among others spectral domain parameters (Gómez-Vilda et al., 2009) such as Mel Frequency Cepstral Coefficients (MFCC) or Linear Pre-

diction Coding (LPC). All above referred features have been employed for different pathologies, obtaining different results depending on the nature of the problem. For the particular of OSA detection, the representation of the spectral envelope (either obtained from Fourier analysis or linear prediction) has proved to be discriminative for the OSA detection problem (Fernández-Pozo et al., 2009; Blanco-Murillo et al., 2009). On the other hand, for classification purposes, the Gaussian mixture model (GMM) has become the standard method in speaker related applications, and most notably in speaker recognition systems, due to their capability of representing a large class of sample distributions (Reynolds, 2008), its probabilistic framework, its training methods scalable to large data sets, and high-accuracy (Campbell et al., 2006).

Several variations, within the field of speaker recognition, have been proposed for training a suitable GMM. One of such is the *Universal Background Model* (UBM), which aims to represent general speech characteristics, by training a GMM on a large corpus of speech representing the characteristics. In this form, it is possible to derive distinctive specific models coming from this rather general one, which might behave better than a GMM trained directly on the dataset. Other approaches, aim to combine the GMM with *Support Vector Machines* (SVM) (Campbell et al., 2006). For this purposes, the concept of *Supervector* is introduced, and which is usually referred to mapping many small-dimensional vectors into higher-dimensional vectors (for instance, by stacking the mean vectors of adapted GMM) to feed a SVM classifier (Kinnunen and Li, 2009). Additionally, and for improving even further performance of the GMM-SVM based schema, some developments have been done to address the effects linked to the differences between recording sessions due to transmission channel mismatch, additive noise, linguistic content, and speaker variability (Fauve et al., 2007). For instance the *Nuisance Attribute Projection* (NAP) (Solomonoff and Campbell, 2007) removes nuisance attribute-related dimensions in the supervector expansion space via projections, aiming to compensate channel and speaker mismatches.

However, all above referred techniques are mainly employed on speaker recognition tasks, while its use on automatic pathology detection has remained practically unexplored. Having those precedents, the aim of this paper is to explore the usefulness of the aforementioned classifiers in the automatic detection of OSA. The usage of LPC-based cepstral coefficients is supported by the successful results obtained on vowels frames over continuous speech for detection of OSA (Elisha et al., 2011), which are worth testing on sustained speech records.On the other hand, for training the UBM on which all the GMM-based classifiers rely, a voice database recorded by Universidad Politécnica de Madrid is utilized.

The paper is organized as follows: Section 2 presents the theoretical background; Section 3 presents the database, the experimental setup; Section 4 presents the obtained results; finally Section 5 presents the discussions as well as some conclusions of the work.

2 THEORETICAL BACKGROUND

2.1 Classification using Gaussian Mixture models

Having a data vector $\vec{x}$, a *Gaussian Mixture Model* (GMM), defined as a finite mixture of multivariate Gaussian components, is of the form:

$$g(\vec{x}) = \sum_{i=1}^{N} \lambda_i \mathcal{N}(\vec{x}; \vec{m}_i, \vec{\Sigma}_i) \tag{1}$$

where λ_i are mixture weights, and $\mathcal{N}(\cdot)$ are Gaussian density functions, having mean $\vec{m}_i$ and covariances $\vec{\Sigma}_i$ (Campbell et al., 2006).

A general model of speech population might be obtained by training the GMM on a large population. This model is termed Universal Background Model (UBM), and might be employed for generating more specific models. In this way, and within the field of pattern recognition, while considering a binary classification problem, at least two specific models are required for representing the normality (control) and pathology conditions. Consequently, two specific models are adapted by means of Maximum A Posteriori (MAP) adaptation of the UBM means (as it is classically done for speakers' verification). The resulting pair of GMM-UBM adapted models being:

$$g_p(\vec{x}) = \sum_{i=1}^{N} \lambda_i \mathcal{N}(\vec{x}; \vec{m}_i^p, \vec{\Sigma}_i) \tag{2}$$

$$g_n(\vec{x}) = \sum_{i=1}^{N} \lambda_i \mathcal{N}(\vec{x}; \vec{m}_i^n, \vec{\Sigma}_i) \tag{3}$$

where $\vec{m}_i^n$ and $\vec{m}_i^p$ are the adapted means for the normal GMM-UBM model, $g_n(\vec{x})$, and the pathological GMM-UBM model, $g_p(\vec{x})$, respectively.

In this manner, a log-likelihood decision function is suitable for discriminating between the two hypotheses: speaker belonging to control or OSA

groups. As for the present detection problem and a given test sample $\vec{y}$, the latter is to be evaluated as follows:

$$\Lambda(\vec{y}) = \log(g_p(\vec{y})) - \log(g_n(\vec{y})) \quad (4)$$

2.2 Classification using SVM based Gaussian classifiers

A Support Vector Machine (SVM) is a discriminative binary classifier constructed from sums of a kernel function $\mathcal{K}(\cdot,\cdot)$ such that:

$$f(\vec{x}) = \sum_{i=1}^{L} \alpha_i t_i \mathcal{K}(\vec{x},\vec{x}_i) + d \quad (5)$$

where the t_i are the ideal outputs (-1 or 1), d is a learned constant, $\sum_{i=1}^{L} \alpha_i t_i = 0 | \alpha_i > 0$, and $\vec{x}_i$ are support vectors obtained from the training set by an optimization process.

In order to exploit the discriminative power of the SVM and simultaneously the generalization capabilities of the GMM models, the supervector concept is introduced. A *supervector* $\vec{m}_i$, is a mapping $\psi(\cdot)$, between an utterance and a high-dimensional vector. This supervector is usually achieved by stacking the means of the Gaussian components of the GMM-UBM models (Kinnunen and Li, 2009), permitting to define a linear Kernel:

$$\mathcal{K}(\cdot,\cdot) = \sum_{i=1}^{N} \left(\sqrt{\lambda_i}\vec{\Sigma}_i^{-1/2}\vec{m}_i^n\right)^T \left(\sqrt{\lambda_i}\vec{\Sigma}_i^{-1/2}\vec{m}_i^p\right)^T \quad (6)$$

In this way, the decision function of Eq. (5) for a test sample $\vec{y}$ might be rewritten as follows:

$$f(\vec{y}) = \left(\sum_{i=1}^{L} \alpha_i t_i \psi(\vec{x}_i)\right)^T \psi(\vec{y}) + d = \vec{w}^T \psi(\vec{y}) + d \quad (7)$$

This likelihood function defines a GMM-SVM classifier. However, and in order to increase performance, it might be desirable to remove the directions of undesired variability from supervectors before the SVM training. One technique combating with such drawback is the *Nuisance Attribute Projection* (NAP), which for a given Supervector, $\vec{m}_i$, is as follows:

$$\vec{m}_i = \vec{m}_i - \boldsymbol{U}\left(\boldsymbol{U}^T\vec{m}_i\right) \quad (8)$$

where *pmbU* is an *eigenchannel* matrix, trained using a development dataset with a large number of speakers (Kinnunen and Li, 2009). And where a GMM-SVM-NAP system results after applying this technique.

3 EXPERIMENTAL SETUP

In the experiments described in this contribution two databases were used for the training and testing of the proposed classification scheme.

3.1 Databases

UPM database

The database was recorded by the Universidad Politécnica de Madrid (UPM). It contains 239 normal voices, and 201 pathological voices with a wide variety of organic pathologies (nodules, polyps, edemas, carcinomas, etc.). The dataset contains the sustained phonation of the /aa/ Spanish vowel. The recordings were sampled at 50kHz, 16-bits of resolution, and 2s long. Besides, the recordings were half-band filtered and downsampled to 25kHz. The distribution by gender is: 226 females and 130 males. The age range goes from 9 to 79 years old, with a mean of 35.5 years. To match the same conditions of the OSA database, only the male adults of the normal class are considered. This class is utilized as it provides less variability and a more compact GMM model than the Pathological class, and therefore might provide a better representation of general speaker population.

Obstructive Sleep Apnoea Database

This database was recorded in the Respiratory Department at Hospital Clíniico Universitario of Málaga, Spain. It contains recordings of 80 male subjects, exhibiting similar physical characteristics including age and Body Mass Index. Half of patients suffer from severe OSA, and half are either healthy or suffer from mild OSA. The recordings were encoded using 16 bits per sample and recorded at a 16kHz sampling rate. Audio files were pre-processed to remove the beginning and ending of each utterance, while keeping the stable sustained segments which lasted for at least 1 second. The speech corpus includes four phonetically balanced sentences specifically designed for the apnoea detection problem, as well as records of sustained $/a/$ vowel (Fernández-Pozo et al., 2008). This latter set of recordings is the only of interest in the present paper.

3.2 Methodology

A general outline of the proposed automatic pathology detection system, is shown on figure 1, while the main stages are described next.

In the *Preprocessing* stage, a minus one-one normalization is considered so that the dynamic range

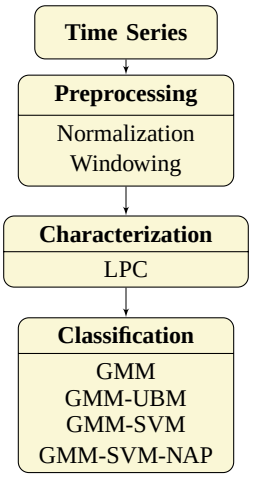

Figure 1: Outline of the automatic voice pathological systempresented on this work

of the recordings remain constant. Additionally it is considered a short time analysis by employing 40ms Hammming windows, overlapped 50%.

Next, in the *characterization* stage a LPC parametrization is employed by using 12,16 and 18 coefficients, to characterize both UPM and OSA database.

Finally, *classification* is performed with the GMM, GMM-UBM, GMM-SVM, and GMM-SVM-NAP systems in a 11-fold crossvalidation scheme. For testing purpose the number of Gaussians is varied between 2 to 20 to prevent overfitting and to keep the same ranges analyzed in (Blanco-Murillo et al., 2011). Besides, and for validation of results, the Classification accuracy, *Sensitivity* (SE), *Specificity* (SP), as well as ROC curves and *Areas Under ROC Curves* (AUC) are employed. Sensitivity and specificity are defined as follows:

$$SE = \frac{TP}{TP+FN}, \quad SP = \frac{TN}{TN+FP}$$

where *TP* - True Positive or Correctly Identified Positive Instances; *FP*- False Positive or Incorrectly Identified Positive Instances; *TN*- True Negative or Correctly Identified Negative Instances; and *FN*- False Negative or Incorrectly Identified Negative Instances.

4 Results

The best operation point obtained for each of the proposed classifier is presented next. The number of LPC parameters is indicated as well as the number of Gaussians for which the best classification rate was obtained. These are both representative of the smoothness of spectral envelope, and the complexity of the trained model, which is related both to the number of dimensions of the feature space and the number of components included in the mixtures.

- *GMM*: 18 LPC, 16 Gaussians
- *GMM-UBM*: 16 LPC, 12 Gaussians.
- *GMM-UBM-SVM*: 16 LPCs, 16 Gaussians
- *GMM-UBM-SVM-NAP*: 18 LPC, 16 Gaussians

Table 1 presents the Classification accuracy, Sensitivity, Specificity, for the OSA detection problem and corresponding to each operation point for which the best classification rate was achieved.

Table 1: Classification Accuracy, Sensitivity and Specificity for the Kay-Elemetrics and OSA databases, by using the GMM, GMM-UBM, GMM-SVM, and GMM-SVM-NAP methodologies.

	Accuracy	SE	SP	AUC
GMM	54 ± 10	0,60	0,5	0,54
GMM-UBM	53 ± 10	0,52	0,53	0,57
GMM-SVM	65 ± 10	0,57	0,75	0,77
GMM-SVM-NAP	62 ± 10	0,57	0,68	0,63

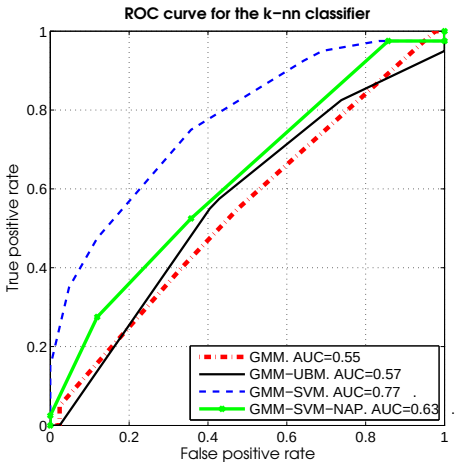

Figure 2: ROC Curve for the best operation point using the Kay-Elemetrics and OSA databases.

5 DISCUSSIONS AND CONCLUSIONS

This paper has investigated the usefulness of ~~some~~ GMM-based classifiers, usually employed in speaker recognition, to the issue of automatic detection of OSA. For characterization purposes LPC parametrization has been employed and transformed according to the standard cepstral transformation. LPC coefficients provide uniform resolution across

the frequency axis and straightly focus on spectral resonances which come as response to a linear system. Both aspects may be suitable to characterize articulation and resonance abnormalities identified for OSA speakers (Fox et al., 1989) on sustained speech records in a similar way as it was previously shown for vowels segments (Elisha et al., 2011). Results obtained for the GMM and GMM-UBM classifiers (see 1) can be compared to previous ones on MFCC parameterization (Blanco-Murillo et al., 2011). A slight improvement is actually observed, which reinforces our understanding on OSA characteristic patterns nature.

The influence of the training database on the classification rates achieved by the GMM-UBM scheme was previously addressed in (Blanco-Murillo et al., 2011), where it was concluded that the ~~closest~~ the database used to train the UBM was ~~to the speech content to be finally analyzed~~, the better results were ~~achieved~~ even if the size of the dataset for training the UBM ~~is~~ shorter than that in an alternative database. Nevertheless, by the time those experiments were ~~developed~~, the latter database wasn't available~~, hence~~ it was worth verifying whether the addressed conclusion was correct. According to the results obtained, GMM-based classifiers trained on these databases outperform those for a specific but smaller database tested in that reference. Such result perfectly ~~matched~~ what was concluded in that ~~reference.~~

Moreover, as shown in table 1 the best classification results were obtained when following the GMM-SVM approach, while the worst were obtained for the GMM-UBM scheme. However this result does not hold when comparing the results in terms of the AUC. In this case the best performance remains with the GMM-SVM, while the worst is obtained for the most elementary GMM classifier. On the other hand, the NAP technique, which was intended to minimize the effects of the undesired variability observed in the GMM-SVM classifier and should have provided better performance results, stayed in between the other two. The failure of the NAP method might be due to the difficulty in finding the spurious sources of variability within the supervector space which might have contributed to a reliable improvement in the accuracy rate.

However, despite the results obtained when using the GMM-based classifiers, these are not significantly better than those obtained for the baseline GMM, as over a 10% improvement was reached in the obtained classification accuracy, which might encourage research on using this techniques, within the field of automatic OSA detection. This improvement lies beyond the 10% confidence interval estimated for the classification accuracy.

Also, and since the methodology for a correct discrimination of the OSA phenomena is still an open issue, specially for which is the best set of features that might improve the classification accuracy; the proposed approach suggests that a better performance might be obtained in the basis of more complex classifiers. It also suggests that the usage of the GMM-based classifiers, which are usually employed in speaker recognition, might improve further the efficiency as has been demonstrated before. In (Wang et al., 2011) a GMM-SVM classifier was utilized in the detection of voice pathologies in the Kay-Elemetring database, obtaining a classification improvement when employing a GMM-SVM rather than the classical GMM.

According the obtained results, it might be reasonable to conclude that the usage of the GMM-based classifiers might improve classification accuracy in the discrimination of OSA. There is however, much effort to be put into finding better characterization methodologies capable of correctly distinguishing the normality and pathology phenomena.

ACKNOWLEDGEMENTS

This research was carried out under grants: *TEC2009-14123-C04* from the Spanish Ministry of Education; *AL11-P(I+D)-022* and *Ayudas para la realización del doctorado (RR01/2011)* from Universidad Politécnica de Madrid, Spain; and partially funded by the Spanish Ministry of Science and Innovation as part of the TEC2009-14719-C02-02 (PriorSpeech) project.

REFERENCES

Alcázar, J., Fernández, R., Blanco, J., Hernández, L., López, L., Linde, F., and Torre-Toledano, D. (2009). Automatic speaker recognition techniques: A new tool for sleep apnoea diagnosis. *Am. J. Respir. Crit. Care Med*.

Blanco-Murillo, J. L., Fernández-Pozo, R., Pardo de Vera, D. D., Alvaro Sigüenza, Hernández-Gómez, L. A., and Alcázar-Ramírez, J. (2009). Analyzing GMMs to characterize resonance anomalies in speakers suffering from apnoea. In *INTERSPEECH*, pages 1459–1462.

Blanco-Murillo, J. L., Fernández-Pozo, R., Torre-Toledano, D., Caminero, J., and López, E. (2011). Analyzing training dependencies and posterior fusion in discriminant classification of apnea patients based on sus-

tained and connected speech. In *INTERSPEECH*, pages 3033–3036.

Campbell, W., Campbell, J., Reynolds, D. A., Singer, E., and Torrescarrasquillo, P. (2006). Support vector machines for speaker and language recognition. *Computer Speech & Language*, 20(2-3):210–229.

Coccagna, G., Pollini, A., and Provini, F. (2006). Cardiovascular disorders and obstructive sleep apnea syndrome. *Clinical and Experimental Hypertension*, 28:217–224.

Elisha, O., Tarasiuk, A., and Zigel, Y. (2011). Detection of obstructive sleep apnea using speech signal analysis. In *Proceedings of the 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*.

Fauve, B., Matrouf, D., and Scheffer, N. (2007). State-of-the-art performance in text-independent speaker verification through open-source software. *Audio, Speech, and Language Processing*, 15(7):1960–1968.

Fernández-Pozo, R., Blanco-Murillo, J. L., Hernández-Gómez, L., López-Gonzalo, E., Alcázar Ramírez, J., and Toledano, D. T. (2009). Assessment of severe apnoea through voice analysis, automatic speech, and speaker recognition techniques. *EURASIP J. Adv. Signal Process*, 2009:6:1–6:11.

Fernández-Pozo, R., Hernández-Gómez, L. A., López-Gonzalo, E., Alcázar-Ramírez, J., Portillo, G., and T.Toledano, D. (2008). Design of a multimodal database for research on automatic detection of severe apnoea cases. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, B. M. J. M. J. O. S. P. D. T., editor, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Fox, A. W., Monoson, P. K., and Morgan, C. D. (1989). Speech dysfunction of obstructive sleep apnea. a discriminant analysis of its descriptors. *Chest*, 96(3):589–95.

Gómez-Vilda, P., Fernández-Baillo, R., Rodellar-Biarge, V., Nieto-Lluis, V., Álvarez Marquina, A., Mazaira-Fernández, L., Martínez-Olalla, R., and Godino-Llorente, J. (2009). Glottal source biometrical signature for voice pathology detection. *Speech Communication*, 51(9):759–781.

Kinnunen, T. and Li, H. (2009). An Overview of Text-Independent Speaker Recognition: from Features to Supervectors. *Image Processing*.

Lloberes, P., Levy, G., Descals, C., Sampol, G., Roca, A., Sagales, T., and 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.

Penzel, T., McNames, J., de Chazal, P., Raymond, B., Murray, A., and Moody, G. (2002). Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. *Medical and Biological Engineering and Computing*, 40:402–407.

Puertas, F. J., Pin, G., María, J. M., and Durán, J. (2005). Documento de consenso nacional sobre el síndrome de apneas-hipopneas del sueño. *Grupo Español De Sueño*.

Reynolds, D. A. (2008). Gaussian Mixture Models. *Encyclopedia of Biometric Recognition*, 31(7):1047–64.

Solomonoff, A. and Campbell, W. (2007). Nuisance attribute projection. *Speech Communication*.

Wang, X., Zhang, J., and Yan, Y. (2011). Discrimination between pathological and normal voices using GMM-SVM approach. *Journal of voice : official journal of the Voice Foundation*, 25(1):38–43.