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*

Keywords: GMM, Supervector, GMM-SVM, Obstructive Sleep Apnea, OSA.

Abstract: The aim of automatic pathological voice detection systems is to support a more objective, less invasive diagnosis of diseases. Those detection systems mostly employ an optimized representation of the spectral envelope; whereas for classification, Gaussian Mixture Models are typically used. However, the study of Gaussian Mixture Models-based classifiers as well as Nuisance mitigation techniques, such as those employed in speaker recognition, has not been widely considered in pathology detection tasks. The present work aims at considering whether such tools might improve system performance in detection of pathologies, particularly for the Obstructive Sleep Apnea. Having this in mind, the present paper employs Linear Prediction Coding Coefficients, in conjunction with Gaussian Mixture Model-based classifiers for the detection of Obstructive Sleep Apnea, in a database containing the sustained phonation of vowel $/a/$. The obtained results demonstrate subtle improvements compared to using baseline automatic detection system.

1 INTRODUCTION

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), characterized by recurrent episodes of sleep-related collapses of the Upper Airway at the level of the pharynx. OSA is usually associated to loud snoring, increased daytime sleepiness, poor quality of life and impaired work performance (Puertas et al., 2005).

OSA is usually detected on the basis of the analysis of the patients history and physical examinations. Nevertheless, a full overnight sleep study involving the recording of physiological variables, as well as complex post-processing of collected data, is required to confirm diagnosis. This procedures is expensive and time-consuming, and patients usually have to be in waiting lists for 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.

The automatic detection of pathologies using voice recordings relies on the estimation of parameters such as *jitter* and *shimmer*, noise measures, among others spectral parameters such as Mel Frequency Cepstral Coefficients (MFCC) or Linear Prediction Coding (LPC). Above features have been employed for different pathologies, obtaining different results depending on the nature of the problem. In particular, for OSA detection, the representation of the spectral envelope (either from Fourier analysis or linear prediction) has proved to be discriminative (Fernández-Pozo et al., 2009). On the other hand, for classification purposes, the Gaussian mixture model (GMM) has become the standard method in speech applications, and most notably in speaker recognition systems, due to, among others, its probabilistic framework, and high-accuracy (Campbell et al., 2006).

Several variations, within the field of speaker recognition, have been proposed for improving the performance of the GMM classifiers. Some of them, which are to be introduced in the next section, are

Gómez-García J., Blanco-Murillo J., Godino-Llorente J., A. Hernández Gómez L. and Castellanos-Domínguez G..
GMM-based Classifiers for the Automatic Detection of Obstructive Sleep Apnea.
DOI: 10.5220/0004252503640367
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2013), pages 364-367
ISBN: 978-989-8565-36-5

the Universal Background Models (UBM), GMM mixed with Support Vector Machines (GMM-SVM) or GMM-SVM with nuisance removal. However, aforementioned techniques are mainly employed on speaker recognition tasks, while its use on automatic pathology detection is still under study. Having those precedents, the aim of this paper is to explore the usefulness of the above classifiers for the automatic detection of OSA when employing LPC features. The usage of LPC is supported by the previous studies in the same task using continuous speech (Elisha et al., 2011). Moreover, and unlike other works in the same topic (Blanco-Murillo et al., 2011a; Blanco-Murillo et al., 2011b), this paper relies merely on the discrimination capability of the sustained phonation of vowel $/a/$. This, compared to the usage of continuous speech, restricts the phonetic information that might be obtained, but turns out to be less complex, while attaining certain advantages such as immunity to speaking rate, dialect and intonation of the speakers (Fernández-Pozo et al., 2009).

The paper is organized as follows: Section 2 presents the theoretical background; Section 3 presents 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

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

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

where λ_i are mixture weights, and $\mathcal{N}(\cdot)$ are Gaussian density functions, having mean $\vec{\boldsymbol{\mu}}_i$ and covariances $\boldsymbol{\Sigma}_i$.

By training a GMM on a large speech corpus, covering most speech characteristics, a general model or UBM is obtained. In this form, it is possible to derive (adapt) specific models (GMM-UBM) coming from this rather general UBM, and which might behave better than a GMM trained directly on the dataset. Considering a binary classification problem; two specific models are required for representing the healthy (control) and pathology conditions. Those models are adapted by means of Maximum A Posteriori (MAP) adaptation of the UBM means (as it is classically done for speakers' verification), and are as follows:

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

$$g_n(\vec{\boldsymbol{x}}) = \sum_{i=1}^{G} \lambda_i \mathcal{N}(\vec{\boldsymbol{x}};\vec{\boldsymbol{\mu}}_i^n,\boldsymbol{\Sigma}_i) \tag{2b}$$

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

The log-likelihood decision function chosen for discriminating if $\vec{\boldsymbol{y}}$ belongs to the OSA class is:

$$\Lambda(\vec{\boldsymbol{y}}) = \log(g_p(\vec{\boldsymbol{y}})) - \log(g_n(\vec{\boldsymbol{y}})) \tag{3}$$

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

$$f(\vec{\boldsymbol{x}}) = \sum_{i=1}^{L} \alpha_i t_i \mathcal{K}(\vec{\boldsymbol{x}},\vec{\boldsymbol{x}}_i) + d \tag{4}$$

where t_i are ideal outputs (-1 or 1), α_i are weights such that $\sum_{i=1}^{L} \alpha_i t_i = 0 | \alpha_i > 0$; d is a learned constant; and $\vec{\boldsymbol{x}}_i$ are the L support vectors obtained from a training set by an optimization process. In order to exploit the discriminative power of the SVM and simultaneously the generalization capabilities of the GMM, the *supervectors* are introduced. A supervector $\vec{\boldsymbol{m}}_i$, is a mapping $\psi(\cdot)$, between an utterance and a high-dimensional vector, which is usually formed by stacking the mean vectors of GMM-UBM models (Kinnunen and Li, 2009). By defining $\vec{\boldsymbol{m}}_i^n$ and $\vec{\boldsymbol{m}}_i^p$, as the supervectors for the models of equation (2), a linear Kernel might be considered:

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

Therefore, the decision function of Eq. (4), for a test sample $\vec{\boldsymbol{y}}$, might be rewritten:

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

In addition, some methods have been proposed to increase performance of GMM-SVM systems, by removing the directions of undesired variability in supervectors before the SVM training. One of such is the *Nuisance Attribute Projection* (NAP), which for a given supervector, $\vec{\boldsymbol{m}}_i$ is as follows:

$$\vec{\boldsymbol{m}}_i' = \vec{\boldsymbol{m}}_i - \boldsymbol{U}\left(\boldsymbol{U}^T \vec{\boldsymbol{m}}_i\right)$$

where $\boldsymbol{U}$ is an *eigenchannel* matrix, trained using a development dataset (Kinnunen and Li, 2009). The resulting $\vec{\boldsymbol{m}}_i'$ forms a GMM-SVM-NAP.

3 EXPERIMENTAL SETUP

3.1 Databases

Obstructive Sleep Apnea Database: Was recorded at Hospital Clínico Universitario de Málaga, Spain. It contains recordings of 80 male subjects, with similar age and Body Mass Index. Half of them suffer from severe OSA, and the other half are either healthy or suffer from mild OSA. Recordings were collected at 16kHz and 16 bits per second. The speech corpus includes four sentences, as well as recordings of the sustained vowel $/a/$ (Fernández-Pozo et al., 2009). The latter set is the only of interest for this paper.

UPM Database: Was recorded by Universidad Politécnica de Madrid. It contains 239 normal voices, and 201 pathological voices with a wide variety of organic pathologies (nodules, polyps, etc.). It contains the sustained phonation of the $/a/$ vowel. The distribution by gender is: 226 females and 130 males. The age range goes from 9 to 79 years. The recordings were sampled at 50kHz, 16-bits of resolution, and 2s long, however they were half-band filtered and downsampled to 25kHz. In order to match the same conditions of the OSA database, resampling to 16kHz was performed, while considering only *male adults* from the normal class (101 recordings).

3.2 Methodology

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

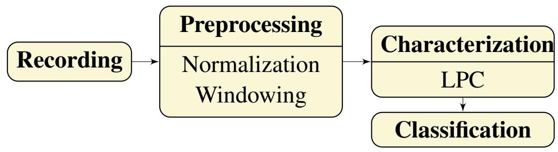

Figure 1: Outline of the automatic voice pathological system presented on this work.

In the *Preprocessing* stage, a minus one-one normalization is utilized. Also, framing and windowing is performed, by employing 40ms Hammming windows, overlapped 50%.

Next, in the *characterization* stage a LPC parametrization is considered by using 12, 16 and 18 coefficients for both UPM and OSA databases.

Finally, *classification* is carried out with GMM, GMM-UBM, GMM-SVM, and GMM-SVM-NAP, whose parameters where varied between 2 to 20 to keep the same ranges analyzed in (Blanco-Murillo et al., 2011b). For validation of results a 11-fold cross-validation scheme was employed, assuring that recordings from the same patients are not used in both training and testing set. The calculated performance measures were: Classification accuracy (ACC) within some *confidence intervals* (IC), *Sensitivity* (SE), *Specificity* (SP), ROC curves and *Areas Under ROC Curves* (AUC). Assuming 95% confidence, the IC is estimated as $IC = \pm 1.96\sqrt{ACC(1-ACC)/N}$, where N is the total number of classified patterns.

4 RESULTS

The number of LPC parameters and the number of Gaussians for which the best classification rate was obtained, are presented next. Also, Table 1 summarizes ACC, SE, SP, for those parameters, while Figure 2 presents the corresponding ROC curves.

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

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

	ACC $\pm$ IC	SE	SP	AUC
GMM	54 ± 10	0,60	0,50	0,55
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

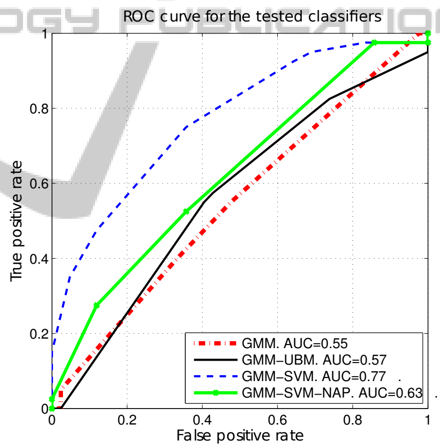

Figure 2: ROC Curve for the combination of LPC and Gaussian parameters using the OSA database.

5 DISCUSSIONS AND CONCLUSIONS

This paper has investigated the usage of GMM-based classifiers, typically employed in speaker recognition, to the issue of automatic OSA detection. LPC-based coefficients were chosen for speech parametrization as they provide uniform resolution across the frequency axis and focus on spectral resonances, which might 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 on vowel segments (Elisha et al., 2011). Obtained results, for the GMM and GMM-UBM classifiers (Table 1) are comparable to previous findings using MFCC parametrization (Blanco-Murillo et al., 2011b), reinforcing our understanding on the role of speech spectral envelope for automatic detection of OSA. In Addition, since the GMM-UBM approach on top of a MFCC parametrization apparently outperforms the same scheme when using LPC, but not when the alternative GMM-SVM scheme is considered, the role of the symbiosis between the features set and the classification scheme is highlighted.

The influence of the training database on the classification rates achieved by the GMM-UBM scheme had been addressed in (Blanco-Murillo et al., 2011b); concluding that better classification results are to be expected when the characteristics of the database used to train the UBM match those of the final classification task. Nevertheless, by the time the experiments in (Blanco-Murillo et al., 2011b) were developed, the UPM database was not available and was worth verifying this conclusion on a LPC-parametrization. The results obtained have shown that GMM-based classifiers trained on these databases outperform those for a specific but smaller database, matching perfectly what had been concluded. Nonetheless, the limitations imposed by the apnea database are hard due to the usage of the $/a/$ sound, which might not be the best choice for OSA-related phenomena.

Moreover, as shown in Table 1 the best classification results were obtained when following the GMM-SVM approach, outperforming the GMM and GMM-UBM schemes (up to 10% absolute improvement, though the large confidence intervals must be taken into account). This same pattern is observed for the AUC. These had already been described in (Wang et al., 2011), and has been verified for the OSA detection on sustained speech. On the other hand, the scheme including NAP technique, which was introduced to minimize the effects of undesired variability observed in the GMM-SVM classifier, was found sit in between the previous. The limited performance of the NAP method might be explained by the difficulty in finding the spurious sources of variability within the supervector space, which should have contributed to an improvement in classification. Since the methodology for a correct discrimination of OSA-related phenomena is still an open issue, specially regarding the selection of the features, accuracy rates may be enhanced in a number of alternative ways. Results in this paper suggest that improvement should be expected on the basis of more complex classifiers and by focusing on spectral resonances analysis.

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., Hernández, L., Fernández, R., and Ramos, D. (2011a). Introducing non-linear analysis into sustained speech characterization to improve sleep apnea detection. *Advances in Nonlinear Speech Processing*, pages 215–223.

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

Elisha, O., Tarasiuk, A., and Zigel, Y. (2011). Detection of obstructive sleep apnea using speech signal analysis. In *MAVEBA*.

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.

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.

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

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*.

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