Applied Soft Computing 23 (2014) 346–354

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# Detection of severe obstructive sleep apnea through voice analysis

Jordi Solé-Casals [a,*], Cristian Munteanu [b,c], Oriol Capdevila Martín [c], Ferran Barbé [d,e], Carlos Queipo [g], José Amilibia [h], Joaquín Durán-Cantolla [e,f]

[a] Data and Signal Processing Group, University of Vic – Central University of Catalonia, Sagrada Família 7, 08500 Vic, Spain
[b] Universidad de Las Palmas de Gran Canaria, Campus de Tafira, 35017 Las Palmas, Spain
[c] Cooclea, S.L., Mestre Garriga 10, 08500 Vic, Spain
[d] Respiratory Department IRBLleida, Av. Alcalde Rovira Roure, 80, 25198 Lleida, Spain
[e] CIBERES, ISCIII, Sinesio Delgado 20, Madrid, Spain
[f] Sleep Unit, Service of Pneumology, Hospital Txagorritxu, Servicio Vasco de Salud-Osakidetza, José Achótegui s/n, Vitoria-Gasteiz, Spain
[g] Respiratory Department, Sleep Unit – Hospital Universitario Marqués de Valdecilla, Avda. Valdecilla n° 25, 39008 Santander, Spain
[h] Respiratory Department, Sleep Unit – Hospital Universitario de Cruces, Plaza de Cruces, s/n 48903 Barakaldo, Bizkaia, Spain

ARTICLE INFO

*Article history:*
Received 19 October 2012
Received in revised form 4 April 2014
Accepted 16 June 2014
Available online 3 July 2014

*Keywords:*
Obstructive sleep apnea
Voice processing
Genetic algorithms
Feature reduction

ABSTRACT

This paper deals with the potential and limitations of using voice and speech processing to detect Obstructive Sleep Apnea (OSA). An extensive body of voice features has been extracted from patients who present various degrees of OSA as well as healthy controls. We analyse the utility of a reduced set of features for detecting OSA. We apply various feature selection and reduction schemes (statistical ranking, Genetic Algorithms, PCA, LDA) and compare various classifiers (Bayesian Classifiers, kNN, Support Vector Machines, neural networks, Adaboost). *S*-fold crossvalidation performed on 248 subjects shows that in the extreme cases (that is, 127 controls and 121 patients with severe OSA) voice alone is able to discriminate quite well between the presence and absence of OSA. However, this is not the case with mild OSA and healthy snoring patients where voice seems to play a secondary role. We found that the best classification schemes are achieved using a Genetic Algorithm for feature selection/reduction.


## Introduction

Obstructive Sleep Apnea Hypoapnea Syndrome (OSA for short) is a common sleep disorder that manifests itself by daytime sleepiness caused by a cease in breathing occurring repeatedly during sleep, often for a minute or longer and as many as hundreds of times during a single night. OSA is associated with a reduced-calibre upper airway, and repetitive effects of apneas and hypopneas include oxygen desaturation, reductions in intrathoracic pressure, and central nervous system arousals [1]. Diagnosis of the sleep condition is based on the calculation of the apnea–hypopnea index (AHI) which measures the frequency of reductions in airflow associated with upper-airway collapse or narrowing that occurs with the state change from wakefulness to sleep [1]. The gold standard procedure to determine the AHI is polysomnography, however it is a quite costly methodology [2]. No other measure has proven to be superior to AHI in assessing the overall effect of obstructive sleep apnea. Nevertheless, there is no common consensus between laboratories regarding its definition. Other metrics such as the number or frequency of arousals during a night sleep might be considered an equally good indicator of OSA [1]. Thus, seeking alternative methods of diagnosis that are simpler and more cost effective is fully motivated, and in recent years it was advocated that voice may play a central role into detection of OSA syndrome. Preliminary findings on speech disorder in OSA have been reported firstly in [3] employing a rather small sample (39 subjects) and subjective results of acoustic evaluation of voice changes in OSA, followed by a study [4] on a bigger sample (252 patients) giving again only subjective judgement results. An attempt to a more objective evaluation study was given in [5]. To discriminate between OSA patients and controls, the authors apply spectral analysis to vowels, but again the sample taken into account is small (28 subjects). Recently, in [6] and [7] the authors show the importance of using voice as a discriminatory factor for detection of severe sleep apnea employing Gaussian Mixture Models on phrases (in [6]) and on vowels (in [7]). However, the authors recognize the need for a wider training and validation sets. So far, either due to small samples or subjective judgements, it is hard to quantify up to what extent or under what circumstances we might consider voice as a good discrimination measure between OSA and healthy subjects. Recent efforts such as [8] try to model the upper-airway in OSA subjects as compared to controls by employing computational fluid dynamics models, and they conclude that there is a clear tendency to closure of the upper-airway in OSA. As the upper-way coincides in part with the vocal tract, the thinning of the lumen and tendency to closure experienced in OSA do suggest that there may be an identifiable dysfunction in voice also.

## Method

### *Subjects*

We have 376 subjects that undertook this study, both controls (proven healthy subjects) as well as snoring OSA suspects, mild OSA and severe OSA patients, 123 women and 253 men, with ages comprised between 18 and 82. This cross sectional data has been

* Corresponding author. Tel.: +34 93 8815519.
*E-mail address:* jordi.sole@uvic.cat (J. Solé-Casals).

http://dx.doi.org/10.1016/j.asoc.2014.06.017
1568-4946/

pooled from several state hospitals in Spain (namely from Vitoria, Lleida, Cruces and Valdecillas). The diagnosis for each patient was confirmed by specialized medical staff through polysomnography (PSG) or through respiratory polygraphy (RP) whenever PSG was not available. For the present study we consider AHI $\leq$ 5 as controls (healthy subjects) and AHI $\geq$ 30 as severe OSA patients, which is in agreement with the recommendations made by the American Academy of Sleep Medicine [9]. For the purpose of clarity, along the present study, we call these subjects *extreme cases*, while in-between we may have mild OSA, or snoring non-OSA patients. Thus, among the total of 376 available cases we extract a group of 127 controls and a group of 121 severe OSA with the following characteristics.

*Voice database*

Speech was recorded using an AKG Perception 100 condenser microphone, a Digidesing M-box® sound card (Avid Audio), and a sound acquisition software by Pro Tools® (Avid Audio). The microphone was held 20 cm away from the subject's mouth, by a technician designated for this task. The audio signal was sampled at 44.1 kHz with 16 bits per sample, and recording was done for two distinct positions for each subject: upright or seated ('A' position) and supine or stretched ('E' position). Before each recording session, during 3 min the patient was kept as comfortable as possible in order to induce a relaxation feeling as stress is known to affect voice [10]. The room's ambient was kept quiet, in dim comfortable light and no external noise. Each subject was asked to emit the 5 vowels present in Spanish language that are: /a/,/e/,/i/,/o/,/u/ in a sustained fashion for at least 4 s each. Additionally, the patients were asked to utter the following sentence (in Spanish): \De golpe nos quedamos a oscuras\. Between each utterance a silence gap of 2–3 s was enforced through the recording protocol. The reason for using two distinct uttering positions ('A' and 'E') was that as gravity and head position affect differently the vocal tract when seated and when stretched, the sound properties also change [11,12]. Therefore, we add a second source of information per patient besides the utterance in the more common position (seated). To the best of our knowledge, this is the first attempt to detect OSA through voice analysis that uses this idea. All recordings are done by technicians from the sleep units in the 4 hospitals participating in the study, all technicians being "blind" with respect to the outcome of the experiment.

*Voice features*

A total of 253 features per patient where extracted from the utterance of 5 vowels and a sentence in two distinct positions. The rationale behind choosing the following listed features is that most of these measures have been previously employed for detection or characterization of pathological voice. Our working hypothesis is that severe OSA may present abnormalities in the voice production, such as increased nasality, harshness or dullness, which is also in agreement with previous findings (see [3–6]). The features may be grouped as follows.

*Formant and pitch based*

For each vowel we compute the second formant using the classical algorithm of root finding for the Linear Predictive Coefficient polynomial [13], with a previous octave-jump filtering step. Next, we extract the Mean Frequency (MF), Coefficient of Variation in Frequency (CVF), Jitter Factor (JF), Relative Average Perturbation (RAP), Mean Bandwidth (MBW) and Coefficient of Variation of the Bandwidth (CVBW). Definitions of these measures are given for example in [14,15]. Voice pitch is extracted for each vowel employing an improved autocorrelation method given in [16]. The post-processing octave-jump filtering stage and the features extracted from pitch are exactly the same as in the case of the second formant.

*Time domain analysis*

The time signal (one signal for each vowel and each subject position) yields a set of features that are pitch-synchronous in that we take as a reference signal the pitch extracted in Formant and pitch based section. The features (see [17] for detailed definitions) are the Mean Intensity/Amplitude (MIA), the Coefficient of Variation of the Intensity/Amplitude, the Shimmer of the signal Intensity (SIA) and a measure of the perturbation in the signal amplitude: Amplitude Perturbation Quotient (APQ).

*Voice harshness and turbulence analysis*

The first measure employed is related to the content of harmonics present in voice (versus non-harmonics content, denoted as noise) and is commonly designated as Harmonics to Noise Ratio (HNR). To compute HNR we took a well-established frequency method described in [18] among other more basic variants such as [19,20]. A particularly useful feature as turned-out to be from results obtained (see Results section) is the MHNR: the mean HNR computed at the beginning (approximately the first second) of vowel \a\. Other measures are the Soft Phonation Index (SPI) and the Voice Turbulence Index (VTI). VTI measures the turbulence components caused by incomplete or loose adduction of the vocal folds; SPI evaluates the poorness of high-frequency harmonic components that may be an indication of loosely adducted vocal folds during phonation. In our implementation we compute SPI and VTI according to definitions in [14] but employing the improved algorithm in [18] to calculate the intra-harmonic and inter-harmonic energies present in the voice signal.

*Linear prediction analysis*

Based on a linear predictions analysis on the voice signal, we extracted the Pitch Amplitude (PA) and Spectral Flatness Ratio (SFR) with methods described in [21]. PA measures the dominant peak of the residual signal auto-correlation function, and SFR quantifies the flatness of the residue signal spectrum.

*Dynamical systems analysis*

To account for significant non-linear and non-Gaussian random phenomena present in disordered sustained vowels we employ two features inspired by dynamical system analysis performed on the voice signal. These features were introduced in [22]. The authors apply state-space recurrence analysis to produce an entropy measure Hnorm, and Fractal scaling analysis that yields a measure called Detrended Fluctuation Analysis (DFA).

*LTAS based*

So far, we introduced features computed on sustained vowels. Next, we present features extracted from phrase analysis. The core analysis method of the sentence was Long-Term Average Spectrum (LTAS). In [23–25] the authors focus on the use of LTAS to quantify voice quality, and therefore we find LTAS as a suitable (and quite simple) method for detecting a decline in voice quality for severe OSA. Based on LTAS we extract the following features: the Absolute Spectral Slope (SLOPE_LTAS), statistical measures: spectral centroid (CENTRAL_LTAS), spectral spread (SPREAD_LTAS), spectral skewness (SKEWNESS_LTAS), spectral kurtosis (KURTOSIS_LTAS). Next, we have the spectral roll-off (ROLLOFF_LTAS) which, as the SLOPE_LTAS measure, quantifies the energy decay at higher frequencies. Finally, we have two measures computed on 5 frequency bands of the LTAS: the Spectral Flatness Ratio (SFR1–5) and Spectral Crest (SC1–5); the frequencies bands are: 175–500 Hz, 500–1000 Hz, 1000–2000 Hz, 2000–3000 Hz, and 3000–4000 Hz.