UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

CARLOS FABIEL BUBLITZ

A Non-intrusive OSA Severity Estimation for CPAP Therapy Screening based on Snoring Acoustical Analysis

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Advisor: Prof. Dr. Juergen Rochol
Coadvisor: Prof. Dr. Cristiano Bonato Both

Porto Alegre
October 2018

CIP — CATALOGING-IN-PUBLICATION

Bublitz, Carlos Fabiel

A Non-intrusive OSA Severity Estimation for CPAP Therapy Screening based on Snoring Acoustical Analysis / Carlos Fabiel Bublitz. – Porto Alegre: PPGC da UFRGS, 2018.

77 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2018. Advisor: Juergen Rochol; Coadvisor: Cristiano Bonato Both.

1. OSA. 2. Snoring events. 3. Acoustical Analysis. 4. Machine Learning. 5. Digital Signal Processing. 6. Mobile Health. I. Rochol, Juergen. II. Both, Cristiano Bonato. III. Título.

"If I have seen farther than others,

it is because I stood on the shoulders of giants."

— SIR ISAAC NEWTON

ACKNOWLEDGMENTS

I dedicate this work to my family who even distant always believed in me and gave me the necessary support.

I would like to thank everyone who supported me in some way to complete this work (colleagues, doctors and technicians of the sleep laboratory, UFCSPA students, and professors). Especially, for my supervisors: Prof. Juergen Rochol and Prof. Cristiano Both and our research colleague Profa. Ana Carolina Ribeiro-Teixeira. Thank you for all the support along the way.

ABSTRACT

Obstructive Sleep Apnea (OSA) is characterized by repeated episodes of partial (hypopnea) or complete (apnea) obstruction of the upper airway during sleep. The clinical effects of OSA are related to the cumulative effects of exposure to periodic asphyxia and sleep fragmentation caused by apneas and hypopneas, such as an increased risk of hypertension, nocturnal dysrhythmias, ventricular failure, myocardial infarction, and stroke. The current gold standard for diagnosing OSA is the overnight Polysomnography (PSG), which requires a full-night sleep laboratory stay, attached to different biological sensors and under the supervision of a technician. Besides the discomfort caused by the invasive sensors, the necessity of a clinical setting and highly specialized infrastructure results in a long waiting list in sleep laboratories and high costs, thus restricting the access to diagnosis and treatment. To improve monitoring of OSA evolution, access to diagnosis and treatment follow up, we propose a Mobile Health (mHealth) solution to take advantage of the smartphone capabilities to deploy a non-invasive OSA severity estimation. We make use of the audio recorded through a smartphone to automatically detect snoring events throughout the night and through the analysis of such events estimate patient's necessity for Continuous Positive Airway Pressure (CPAP) therapy. For that, we have divided our solution into two phases: (*i*) a completely unsupervised solution to automatically detect the snoring events in an uncontrolled environment and (*ii*) the analysis of acoustical features of the snoring events for OSA severity estimation. In the first phase, we can prove the viability of recording the audio and detect the snoring events using a smartphone under an environment susceptible to random noises. In the second phase, we show that a set of global acoustical features from the snoring events can predict the patient's need for the CPAP therapy. Our proposed solution was evaluated in an uncontrolled (patient's home) and controlled (sleep laboratory center) environment, reaching satisfactory results in snoring events detection and patient's classification according to the need for CPAP therapy.

Keywords: OSA. Snoring events. Acoustical Analysis. Machine Learning. Digital Signal Processing. Mobile Health.

Uma Estimativa Não Invasiva da gravidade da SAHOS para a Recomendação do Tratamento de CPAP baseado em uma Análise Acústica dos Eventos de Ronco

RESUMO

A Sindrome da Apneia Hipopneia obstrutiva do sono (SAHOS) é caracterizada por episódios repetidos de obstrução parcial (hipopneia) ou completa (apneia) das vias aéreas superiores durante o sono. Os efeitos clínicos da SAHOS estão relacionados aos efeitos cumulativos da exposição à asfixia periódica e à fragmentação do sono causada por apneias e hipopnéias, como o aumento do risco de hipertensão, disritmias noturnas, insuficiência ventricular, infarto do miocárdio e acidente vascular cerebral. O padrão ouro para o diagnóstico de SAHOS é a Polissonografia (PSG), na qual requer que o paciente permaneça durante a noite inteira no laboratorio de sono, conectado a diferentes sensores biológicos e sob a supervisão de um técnico. Além do desconforto causado pelos sensores invasivos, a necessidade de um ambiente clínico e infraestrutura altamente especializada resulta em uma longa lista de espera nos laboratórios do sono e altos custos, restringindo assim o acesso ao diagnóstico e tratamento. Para melhorar o monitoramento da evolução da OSA, o acesso ao diagnóstico e o acompanhamento do tratamento, propõe-se uma solução baseada em *Mobile Health* (mHealth) para utilizar os recursos do smartphone a fim de desenvolver uma estimativa não invasiva da gravidade da SAHOS. Utiliza-se o áudio gravado através de um smartphone para detectar automaticamente os eventos de ronco durante a noite e, através da análise desses eventos, estimar a necessidade do paciente para o tratamento de Pressão Positiva Contínua nas Vias Aéreas (CPAP). Para isso, dividimos nossa solução em duas fases: (*i*) uma solução completamente não supervisionada para detectar automaticamente os eventos de ronco em um ambiente não controlado e (*ii*) a análise das características acústicas dos eventos de ronco para estimativa de gravidade da SAHOS. Na primeira fase, podemos comprovar a viabilidade de gravar o áudio e detectar os eventos de ronco usando um smartphone em um ambiente suscetível a ruídos aleatórios. Na segunda fase, mostramos que um conjunto de características acústicas globais dos eventos de ronco pode prever a necessidade do paciente para o tratamento com CPAP. Nossa solução proposta foi avaliada em ambiente não controlado (domicílio do paciente) e controlado (laboratório do sono), atingindo resultados satisfatórios na detecção de eventos de ronco e classificação do paciente de acordo com a necessidade de tratamento com CPAP.

Palavras-chave: SAHOS, Eventos de ronco, Análise Acústica, Aprendizagem de Máquina, Processamento Digital de Sinais, Mobile Health.

LIST OF ABBREVIATIONS AND ACRONYMS

AHI	*Apnea Hypopnea Index*
BMI	*Body Index Mass*
CART	*Classification and Regression Tree*
CPAP	*Continuous Positive Airway Pressure*
DTFT	*Discrete Time Fourier Transform*
EM	*Expectation-Maximization*
FCM	*Fuzzy C-Means*
GMM	*Gaussian Mixture Models*
KHM	*K-Harmonic Means*
LPC	*Linear Prediction Coefficients*
MFCC	*Mel-Frequency Cepstrum Coefficients*
OSA	*Obstructive Sleep Apnea*
PCA	*Principal Component Analysis*
PSG	*Polysomnography*

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Obstructive Sleep Apnea (OSA) is characterized by repeated episodes of partial (hypopnea) or complete (apnea) obstruction of the upper airway during sleep (JR; ROGERS, 1996). This obstruction results in progressive asphyxia, which increasingly stimulates breathing efforts against the collapsed airway, until the person is awakened. The clinical effects of OSA are related to the cumulative effects of exposure to periodic asphyxia and sleep fragmentation caused by apneas and hypopneas (WIEGAND; ZWILLICH, 1994). Patients often have excessive daytime sleepiness (COLT; HAAS; RICH, 1991) and an increased risk of hypertension, nocturnal dysrhythmias, ventricular failure, myocardial infarction, and stroke (YOUNG; PEPPARD; GOTTLIEB, 2002). OSA severity is determined by the Apnea-Hypopnea-Index (AHI), which is calculated as the average number of apnea and hypopnea events per hour of sleep.

The current gold standard for diagnosing OSA is the overnight Polysomnography (PSG) exam (SATEIA, 2014). PSG requires a full-night sleep laboratory stay, attached to different biological sensors and under the supervision of a technician. Besides the discomfort caused by the invasive sensors, the necessity of a clinical setting and highly specialized infrastructure results in a long waiting list in sleep laboratories and high costs, thus restricting the access to diagnosis and treatment (FLEMONS et al., 2004). For this reason, in most cases, PSG is performed at most once to every other patient, and treatment follow-up is only performed through reports by patients and their partners (PANOSSIAN; AVIDAN, 2009). Patients diagnosed with moderate to severe OSA are referred, in most cases, for treatment with Continuous Positive Airway Pressure (CPAP), which consists of a ventilation system device which supplies the user constant positive air pressure to prevent airway collapse and hence obstructive respiratory events. CPAP improves breathing and decreases obstructive events (OU et al., 2015).

On the one hand, one has the gold standard diagnosis tool which is highly inaccessible, and on the other hand, one has an effective treatment for patients suffering from a chronic respiratory severe dysfunction. However, the treatment may only be prescribed to such patients once they have been diagnosed. We propose to fill in the gap between access to diagnosis and treatment through a Mobile Health (mHealth) solution. mHealth takes advantage of the smartphone capabilities to deploy healthcare applications to a broad population. mHealth patients can benefit from more expedite diagnosis and continuous treatment monitoring. The preventive approach of mHealth systems promotes cost reduction for governments and healthcare companies and contributes to patient's wellness. Moreover, smartphone sensors are being used in a broad range of healthcare applications. One such application is the use of the built-in microphone for snoring events detection (HAO; XING; ZHOU, 2013).

Snoring has long been investigated as a potential indicator of OSA (MAIMON; HANLY, 2010). These acoustic events closely tied to respiration bring valuable information about the apnea and hypopnea events during the night and have a unique advantage over other physiological signals, namely, the possibility of acquisition through a non-contact microphone, *i.e.*, non-intrusively. The acoustical features extracted from snoring events can reveal helpful information about the upper airway system and obstructions sites, aiding in the sleep disorder severity estimation.

To improve the patient's access to treatment, we propose a non-invasive solution for OSA severity estimation using the built-in microphone of a smartphone. Instead of estimating AHI, we propose a screening of patients into two groups: those in need of CPAP treatment and those not. Specifically, we make use of the audio recorded through a smartphone to automatically detect snoring events throughout the night and through the analysis of such events estimate patient's necessity for CPAP therapy. For that, we have divided our solution into two phases: (*i*) a completely unsupervised solution to automatically detect the snoring events in an uncontrolled environment and (*ii*) the analysis of acoustical features of the snoring events for OSA severity estimation. In the first phase, we can prove the viability of recording the audio and detect the snoring events using a smartphone under an environment susceptible to random noises. In the second phase, we show that a set of global acoustical features from the snoring events can predict the patient's need for the CPAP therapy.

For snoring events detection, we have proposed a completely unsupervised solution for segmentation and classification of events using a clustering algorithm to discriminate sound events according to two classes: snore and non-snore. We consider an audio data set collected in an uncontrolled environment setting for mHealth. We introduce a statistical approach to the clustering problem using the Expectation-Maximization (EM) algorithm for Gaussian Mixture Models (GMM) to cluster the data set according to the probability of each data point (segmented event) of belonging to one of two normal distribution components, each of which is self-consistently constructed. The results obtained show that the EM algorithm produces better results on clustering the data, as compared to other clustering algorithms such as Fuzzy C-Means (FCM) and K-Harmonic

Means (KHM), reaching satisfactory accuracy rates (91.3% for simple snores and 79.7% for OSA snores, on average).

For the classification of patients according to the necessity for CPAP Therapy, we have proposed a decision tree classifier based on a set of relevant features extracted from the snoring events detected during the night. The study was performed with a population of 113 patients that went through a PSG exam and had their night's sleep audio recorded concomitantly. To map the acoustical features relevant to patient classification, we propose a decision tree inference process to perform a feature selection and patient classification based on a large set of 79 proposal features. The decision tree allows selecting the most relevant features according to their ability to separate patients with benign snoring or mild OSA (no need for CPAP) from patients with moderate or severe OSA (in need of CPAP), according to the PSG report. The results obtained show that the classifier was able to reach an accuracy rate of 80% in the estimation of necessity to CPAP therapy.

Our main contributions are: (*i*) the proposal of an unsupervised solution for snoring detection in the mHealth context, (*ii*) the application of a statistical approach analysis (the EM algorithm) to the clustering of sound events, (*iii*) the design of an accurate event segmentation method to identify the snore candidate events boundaries, (*iv*) the design of an automatic cluster labeling according to snore and non-snore, and (*v*) the selection of the most relevant acoustic features for patient's classification and underlying relations between these acoustic features and OSA events.

The remaining of this dissertation is organized as follows. In Chapter 2, we present the background and review related work. We show basic concepts related to clinical aspects, digital signal processing, and machine learning. Additionally, we review some essential related work in snoring events detection and OSA severity estimation. In Chapter 3, we introduce our solution for a complete unsupervised detection and segmentation of snoring events and OSA severity estimation for CPAP therapy, and in Chapter 4 we present our experimental evaluation and associated results. Finally, Chapter 5 is devoted to final remarks and future work.

2 BACKGROUND AND RELATED WORK

This chapter presents some concepts laying the ground to the original results of the dissertation. First, a brief overview of OSA is performed and the clinical motivation introduced. In the sequence, a short introduction to the vocal tract speech production describes how upper airway obstructions modulate snoring events produced. We further present the fundamentals of digital signal processing and machine learning, which give support to our scientific research hypothesis. Finally, related work is shown and discussed, emphasizing the major contributions achieved by the present work to the state-of-the-art in the field.

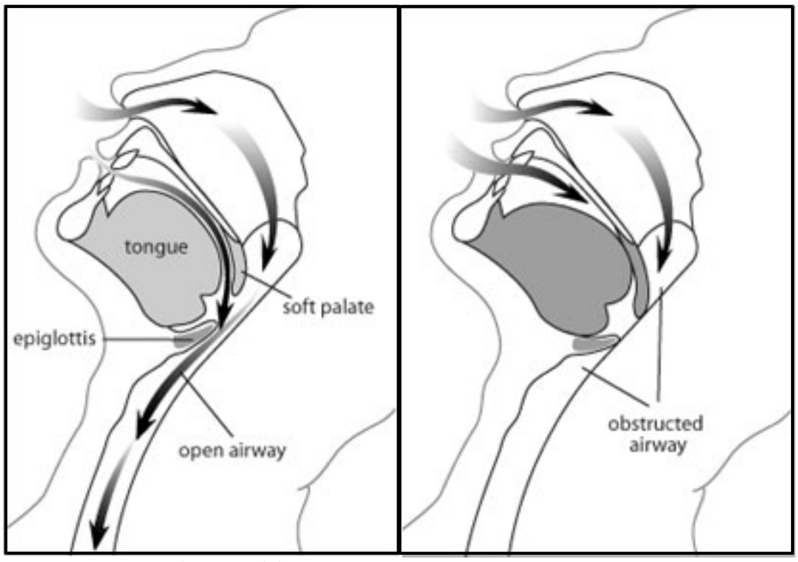
2.1 Obstructive Sleep Apnea

Obstructive Sleep Apnea is a common disorder characterized by repetitive episodes of complete (apneas) or partial (hypopneas) breathing cessations during sleep (SPICUZZA; CARUSO; MARIA, 2015). These breathing cessations are caused by the narrowing of the upper-airway system, which is a result of the musculature relaxation during periods of deep sleep stages. The narrowing can be associated with different factors, such as obesity, neck circumference, nasal congestion, craniofacial and upper-airway structure anomalies, smoking and alcohol consumption (YOUNG; SKATRUD; PEPPARD, 2004). Figure 2.1 illustrates a comparison between a normal breathing and an OSA obstruction event. During moments of normal breathing, we can observe the opened airway with no resistance to the airflow. In the sequence, we observe the obstructed airway caused by the relaxation of the soft palate and tongue musculature, thus restricting the passage of air.

The obstructive events cause progressive asphyxia that increasingly stimulates breathing efforts against the collapsed airway, typically until the person is awakened. As a result, patients suffer from poor sleep quality, daytime sleepiness, and reduced cognitive performance (COLT; HAAS; RICH, 1991). Repetitive episodes of apneas and intermittent hypoxia also elicit an increased risk of hypertension, nocturnal dysrhythmias, ventricular failure, myocardial infarction, and stroke (YOUNG; PEPPARD; GOTTLIEB, 2002). The gold standard method for diagnosing the OSA is the PSG, which requires a full-night sleep laboratory stay, attached to different biological sensors and under the supervision of a technician.

The presence and severity of OSA are most commonly defined by the frequency

Figure 2.1: OSA Airflow Obstruction



Normal Breathing OSA Event

Source: Deller, Hansen e Proakis (1999).

of apneas and hypopneas per hour of sleep (AHI index). Table 2.1 presents OSA classifications as a function of AHI. Patients with less than 5 (five) obstructive events per hours are considered normal. This class includes patients with primary snoring which can be a result of a simple airway anomaly and does not result in significant drops in oxygenation. Patients with mild OSA ($5 < \text{AHI} \leq 15$) have a wider variety of treatment options including, for example, weight loss, physical exercises, avoidance of alcohol for 4-6 hours before bedtime and sleeping on one's side rather than on the stomach or back (TUOMILEHTO et al., 2009). Patients with moderate ($15 < \text{AHI} \leq 30$) to severe ($\text{AHI} > 30$) OSA require a more intrusive treatment, in most cases, with CPAP therapy (OU et al., 2015). CPAP is a device that sends a continuous flow of air to the airway, through a mask, avoiding upper-airway narrowing or collapse. The treatment is usually expensive and uncomfortable. However, this procedure reaches satisfactory results in decreasing the number of obstructive events during the night.

Table 2.1: OSA Severity Classification

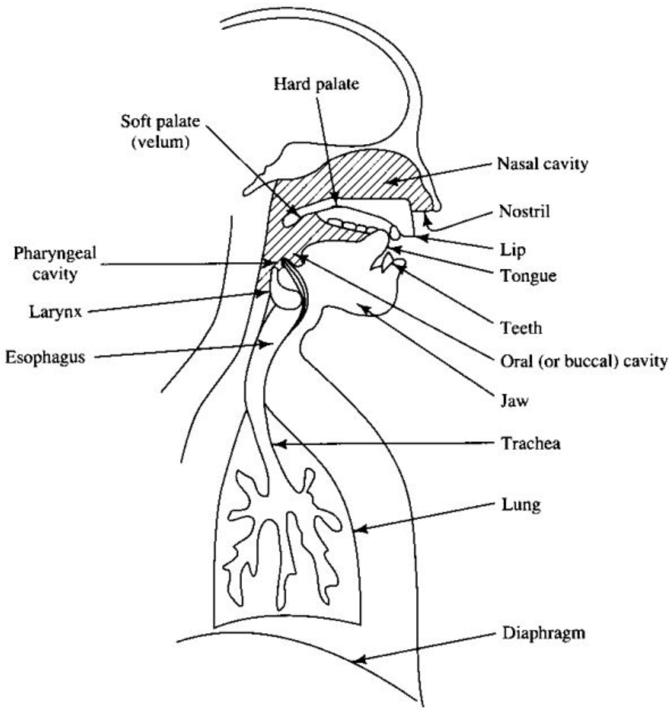
Diagnosis	**AHI**	**Severity**
Normal	<5	N/A
OSA	5-15	Mild
	16-30	Moderate
	>30	Severe

Source: by author (2018).

2.2 Vocal Tract and Snoring Signal Production

The vocal tract is the air passage above the larynx which extends to the mouth. The vocal tract can be divided into the oral cavity and the nasal cavity. The oral cavity refers to the mouth including, among other structures, the lips, tongue, hard and soft palate. The nasal cavity refers to the nose. The upper cavities of the pharynx, mouth, and nose are called the resonating cavities, while the parts of the vocal tract that can be used to produce sounds are called articulators, and can be subdivided into active (those that move, *e.g.*, tongue) and passive (those that are fixed, *e.g.*, hard palate). Most sounds are produced with at least one active and passive articulator. Figure 2.2, shows the complete structure of the vocal tract.

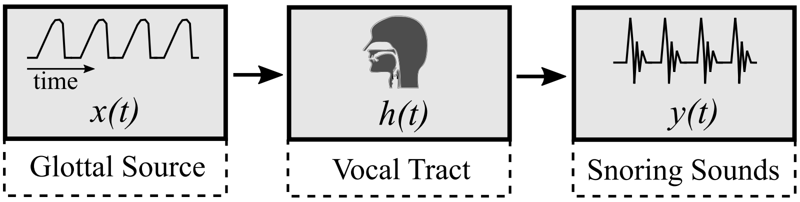
Figure 2.2: Vocal Tract



Source: Deller, Hansen e Proakis (1999).

The source-filter theory of speech production assumes that speech sounds are the response coming from a vocal tract system, where a sound source is fed into and filtered by the resonance characteristics of the vocal tract (HUANG et al., 2001). In Figure 2.3, the sound source $x(t)$ is a glottal sound produced by vocal fold vibration. The fold vibration which is caused by the air passage generates an oscillatory signal. This oscillatory signal is modulated by the vocal tract filter $h(t)$. The vocal tract filter $h(t)$, changes according to the physical changes in the structure. When the vocal tract filter $h(t)$ changes, the resonance characteristics also modify, producing different snoring sounds $y(t)$.

Figure 2.3: Source-filter Theory



Source: by author (2018).

The narrowing of the vocal tract that occurs during an OSA event, changes the resonance characteristics, thus producing snoring sounds with specificities which differ from benign snoring sounds. According to the Source-filter Theory, these changes directly reflect in the snoring events acoustical features. Therefore, the analysis of the acoustical features of snoring events may reveal signs of different levels the vocal tract obstructions.

2.3 Digital Signal Processing

The acoustic signal recorded during the sleep night needs to be digitally processed so one can extract useful information about the OSA severity. A set of digital signal processing techniques in time and frequency domain are used. A brief introduction to such techniques will be given in the following subsections.

2.3.1 Time and Frequency Analysis

The time-domain representation and the frequency-domain are two classical representation of a signal. The time (t) domain express how the signals change over time. Sampling is the process of reduction of a continuous-time signal (analog) to a discrete-time signal (digital). The sampling rate f_s is the number of samples per second. The time interval between samples is called the sampling interval $T_s = 1/f_s$.

The frequency-domain (f) representation of a signal carries information about the signal's magnitude and phase at each frequency. The frequency-domain representation of a signal can be calculated using the Discrete-Time Fourier Transform (DTFT) (HAYKIN; VEEN, 2007). The DTFT is a transformation that maps the signal $x[n]$ in time-domain into a complex valued function of the real variable k, namely:

$$X\left(k\right)=\sum_{n=0}^{N-1}x(n)e^{-j\omega_k n} \tag{2.1}$$

where the signal $X(k)$ is called the spectrum of $x[n]$ and N the number of DFT terms. The spectrum carries information about the frequency distribution of a signal. Digital filters may be applied to a signal's spectrum to cancel out unwanted frequency bands or for noise reduction, for example.

2.3.2 Adaptive Noise Reduction

The signal recorded in an uncontrolled environment suffers from additive random noises. Since one has no access to the instantaneous realization of the contaminating noise, one cannot cancel out noise. However, the average noise spectrum can be subtracted using the statistics of the resulting signal. Such noise reduction can be performed with a Wiener filter, which is an adaptive spectral filter that makes use of a noise template based on a running estimation of the average background noise energy (VASEGHI, 2008). The filter takes as the input a signal $y(m)$, usually a distorted version of desired signal $x(m)$, and produces an output signal $\widehat{x}(m)$, where $\widehat{x}(m)$ is the least-mean-square error estimate for the desired or target signal, $x(m)$. The filter input-output relation is given by

$$\widehat{x}(m)=\sum_{k=0}^{N-1}w_k y(m-k)) \tag{2.2}$$

where m is the discrete-time index, vector $y=[y(m),y(m-1),...,y(m-N-1)]$ is the filter input signal and the vector $w=[w_0,w_1,...,w_{N-1}]$ is the Wiener filter coefficient vector of order N. The filter coefficients w are obtained through a signal noise template, which is calculated as the signal spectral mean of the lowest energy frames of each section $t=0..T$ where T is the number of sections that the signal was divided for noise reduction.

2.4 Machine Learning

Detection of snoring events and OSA severity estimation requires the use of a machine learning classifier which can learns from the different features of the input signal and classifies it according to predefined classes. According to the type of learning, these techniques can be divided into two main classes:

- **Supervised Learning:** uses a priori information about the inputs and expected output. There is a training set containing pairs of features and corresponding classes provided by the instructor. The algorithm builds a model by inferring features of each class in the training set and formulating rules for the classification, which are then applied to an independent test set.
- **Unsupervised Learning:** there is no a priori information about the classes to which inputs belong. The model is created based on observation and discovery. Classes are not defined, so the algorithm needs to observe the data and recognize the patterns by itself.

There is a set of different machine learning algorithms for supervised and unsupervised learning. Each algorithm has its properties and could perform better depending on the type of input data and task involved. Usually, the machine learning task involve tree types: classification, regression, and clustering. Classification aims to predict the outcome of a given sample where the output variable is a category, while regression predicts a real value output. The clustering involves to group samples according to the similarities between each other. The choice of the machine learning algorithm requires understanding the problem and identifying the type of task, and from there choosing the algorithm that better fits the type of input data. Therefore, in the next two subsections we explore the machine learning algorithms chosen for the snoring events detection and patient classification.

2.4.1 Expectation-Maximization

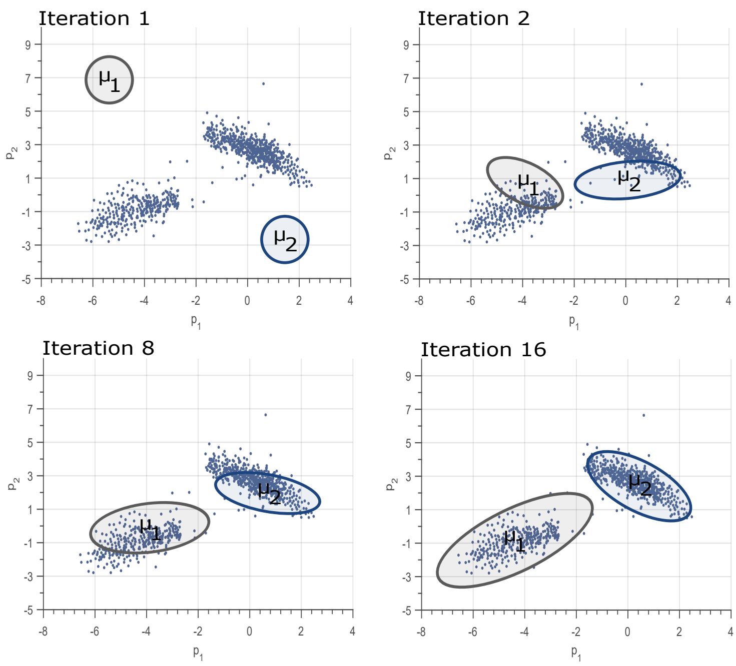
The Expectation-Maximization (EM) is an unsupervised learning algorithm, commonly used for clustering. The input data are defined as coordinates in a d-dimensional space of d features, which each point is defined as an event. Initially, we must set the number of clusters that the data should be classified. EM works through an iterative procedure that starts with some initial estimation of $\theta_j = (\omega_j, \mu_j, \Sigma_j)$ for j distributions, and then proceeds to iteratively update θ_j parameters to maximize the log-likelihood function for the data set:

$$L^{(i)} = \frac{1}{N} \sum_{k=1}^{N} \log \left[\sum_{j=1}^{2} \omega_j^{(i)} \, \phi \left(\mathbf{p}^{(k)} \middle| \mu_j^{(i)}, \Sigma_j^{(i)} \right) \right], \tag{2.3}$$

where $L^{(i)}$ represents the log-likelihood at the i-th iteration, $\phi(\mathbf{x}|\mu, \Sigma)$ is the Gaussian distribution with average $\mu = (\mu_1, \mu_2)$ and covariance matrix $\Sigma_{2\times2}$. The parameters $\{\omega_j\}$ represent the weights given to each of the two Gaussians and $\mathbf{p}^{(k)} = (p_1^{(k)}, p_2^{(k)})$. Each iteration consists of an E-step and an M-step. E-step calculates the membership weights $\omega_j^{(i)}$ for all data points $x(i)$, $1 \leqslant i \leqslant N$ and all mixture components $1 \leqslant j \leqslant K$. M-step uses the membership weights and the data to calculate new parameter values for θ_j.

The maximization is achieved by iteratively updating the set of parameters θ_j until $L^{(i)}$ saturates. Figure 2.4, we can observe an example for the EM iterations. In each iteration, the parameters are adjusted according to the distributions present in the dataset. The process stops when the log-likelihood saturates, resulting in two Gaussian distributions that well fit the data clusters.

Figure 2.4: EM iterations example



Source: by author (2018).

2.4.2 Classification and Regression Tree

The Classification and Regression Tree (CART) (BREIMAN et al., 1984) is a supervised machine learning algorithm commonly used for discovering knowledge from datasets. The goal is to create decision tree models that predicts the value of a target variable by learning simple decision rules inferred from the data features. For that, the rules inferred from the dataset are organized in a tree structure as shown in Figure 2.5. This tree structure is composed of:

- **Root Node:** It represents the entire population, and this further gets divided into two subsets.
- **Splitting:** It is the process of dividing a node into two or more sub-nodes.
- **Decision Node:** When a sub-node splits into further sub-nodes, then it is called the decision node.
- **Leaf:** Nodes that do not split are called Leaf nodes, which represents the resulting classification.

The key to the decision tree CART algorithm is how to define the features for each decision node. The selection of which input feature and the specific split is chosen using the Gini index function, which provides an indication of how pure the leaf nodes are, according to the equation below

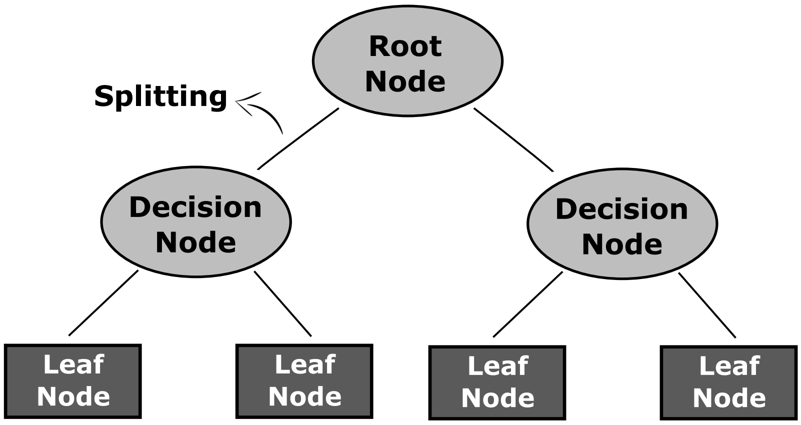
$$G = 1 - \sum_{i=1}^{J} {p_i}^2 \tag{2.4}$$

where G is the Gini index over all classes, p_i is the proportion of training instances with class $i = \{1, 2, ..., J\}$. Gini reaches its minimum (zero) when all cases in the node fall into a single target category. Therefore, the Gini index aims to select the features that better separate the classes (lowest impurity) to construct the decision tree. The tree construction ends using a predefined stopping criterion, such as a minimum of training instances assigned to each leaf node of the tree.

2.4.3 Feature Selection and Dimensionality Reduction

The problem of discovering the optimal set of features to describe the input data is named Feature Selection. Feature Selection is an important problem in machine learning

Figure 2.5: Decision Tree Example



Source: by author (2018).

because, in most cases, there is a lack of a priori information about the relevance of features collected to describe some data, thus leading to inaccurate classifications and also the excessive number of collected features can lead to a classification with high computational cost. It is desirable that irrelevant and redundant features should be removed in the feature selection process and the remaining set should represent enough information for an accurate classification. There is also a lack of information on the combination between different data features, thus encouraging the study of their joint influence, which can be done using a dimensionality reduction technique to extract the essential information form the set of features. Two common techniques for feature selection and dimensionality reduction are:

- **Information Gain:** measures the expected reduction in entropy caused by partitioning the examples according to a given feature. The entropy characterizes the impurity of an arbitrary collection of samples and is given by

$$E = -\sum_{i=1}^{k} p_k \, log_2 \, p_k \tag{2.5}$$

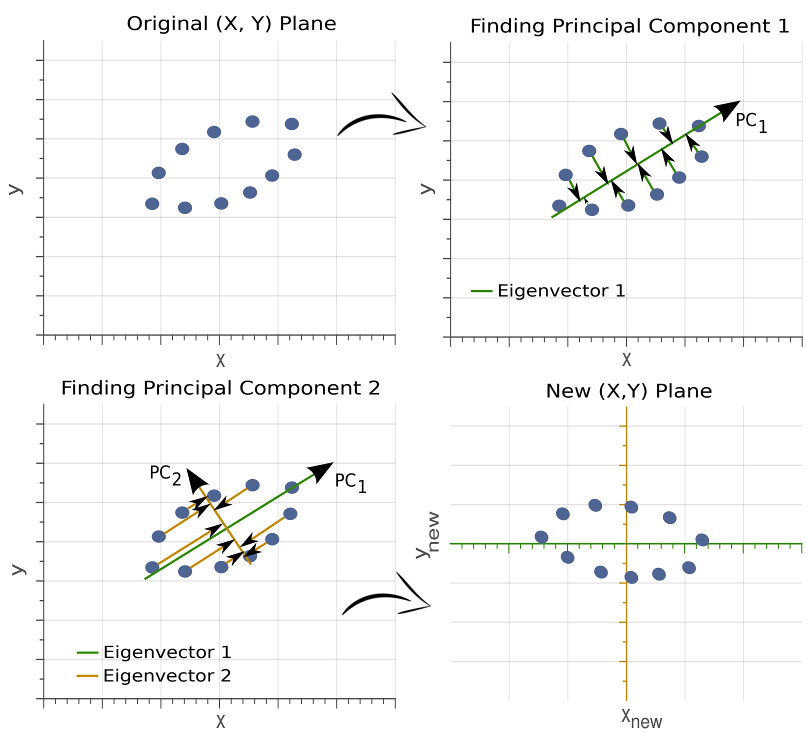
  where p_k is the probability associated with the feature p in class $k = \{1, .., K\}$. The lower entropy, the higher homogeneity of the dataset. The information gain of to the presence of a specific feature to the target classification is given by

$$I = E - \frac{m_L}{m} \, E_L \; - \; \frac{m_R}{m} \, E_R \tag{2.6}$$

  where $E, E_L, andE_R$ represent the entropy for the target classification feature and the entropy for the feature evaluated, and m is the total number of instances of the feature.

- **Principal Component Analysis (PCA):** compute new variables denoted principal components which are obtained as linear combinations of the original features (JOLLIFFE, 2002). The principal components are the directions in data that maximize the variance. For finding the principal components, we can deconstruct the dataset into eigenvectors and eigenvalues as shown in Figure 2.6. Eigenvectors and values exist in pairs: every eigenvector has a corresponding eigenvalue. An eigenvector is a direction, and an eigenvalue is a number, telling you how much variance there is in the data in that direction. The eigenvector with the highest eigenvalue is therefor the principal component.

Figure 2.6: Principal Component Analysis Example



Source: by author (2018).

2.5 Related Work

Recent research efforts have indicated that acoustical analysis of the audio recorded during the night is suitable for the implementation of an automatic snore event detection and OSA severity estimation. However, some key issues remain open. In the next subsections, we show the principal studies on these topics.

2.5.1 Solutions for Automatically Snoring Events Detection

The analysis of acoustic signals aimed at snoring events detection has been performed in different studies. Most of the initiatives performed the recordings in a controlled environment with low levels of noise and high-performance microphones to capture the sound (DAFNA; TARASIUK; ZIGEL, 2013)(DUCKITT; TUOMI; NIESLER, 2006)(WANG et al., 2017). These studies have proven the viability of extracting acoustic features from audio recordings for the segmentation and classification of snoring events. However, the imposed experimental constraints limit the reproducibility in uncontrolled environments, which does not contribute to a solution to the PSG accessibility problem. In order to overcome this limitation, the smartphone started to be seen as an alternative tool to capture the acoustic signal at a patient's home.

A few studies have address the approach of snoring events detection with smartphones as a recording and processing tool (SHIN; CHO, 2014)(HAO; XING; ZHOU, 2013). Hao *et al.* (HAO; XING; ZHOU, 2013) developed an Android application for audio recording and detection of three different events: snoring, body movement, and cough. The classification of the events was based on a decision-tree algorithm. Although their proposed solution is simple and inexpensive regarding processing, the main goal of this study is to assess sleep quality through events counting, and the work does not aim at an acoustic analysis of snoring events for further research. Shin *et al.* (SHIN; CHO, 2014) collected the acoustic signal using a smartphone and performed a formant analysis on the signal within different frequency bands. The authors were able to select the best features for snore classification and applied a quadratic classifier over this set of features for snoring events detection. Even though these studies have obtained good results in snoring detection, all smartphone applications implement supervised learning approaches, which require a considerable data generalization capacity and are prone to overfitting. Furthermore, supervised approaches hinder the automatic character of the solution, which is a necessity when one considers accessibility to a broad population.

Unsupervised learning is an exaction for the automation of snoring events detection solutions. In the literature in the area, only a few works address unsupervised classification, none of which in the context of smartphone platforms (AZARBARZIN; MOUSSAVI, 2011)(MA et al., 2015). These works apply clustering algorithms such as FCM and KHM, which use as a similarity criterion the distance between events in feature space. We verified that this criterion significantly undermines clustering accuracy. We argue that an appropriate algorithm to the problem at hand should aim at snoring events identification through statistical inference approaches, since snoring events from a person are produced by a single human vocal tract, and as such may be considered as independent identically distributed (*i.i.d.*) events, which ideally will be normally distributed.

The most recent work in the literature is not directly comparable with your research on detecting and segmenting snoring events. Çavuşoğlu *et al.* (ÇAVUŞOĞLU; POETS; URSCHITZ, 2017) applied a *Multi-Layer Perceptron* classifier to detect snoring events in children. The authors restricted the data acquisition to only children and focused on finding the most relevant acoustic features to detect snoring of children.

Analyzing Table 2.2 can compare the related work discussed in this subsection. Note that the related work does not include research about an unsupervised detection and segmentation of snoring events using a smartphone as a recording tool.

Table 2.2: Snoring detection solutions

	Smartphone as recording tool	**Unsupervised learning**	**Uncontrolled environment**	**Segmentation of snoring events**
Dafna *et al.*	x	x	x	x
Duckitt *et al.*	x	x	x	x
Shin *et al.*	✓	x	✓	x
Hao *et al.*	✓	x	✓	x
Azarbarzin *et al.*	x	✓	x	✓
Ma *et al.*	x	✓	x	x

Source: by author (2018).

2.5.2 Solutions for OSA Severity Estimation

The state-of-the-art on estimating the OSA severity using audio signal addresses two different approaches: intra-snore and inter-snore events analysis. Intra-snore events refer to the analysis of features extracted from each snoring event. Levartovsky *et al.* (LEVARTOVSKY et al., 2016) investigated the energy intensity of snore events in correlation with AHI. The authors found a weak correlation between the energy intensity and AHI. The authors found better results using the frequency centroid. Emoto *et al.* (EMOTO et al., 2011) explored the power spectra of snores in three different bands: low, middle and high-frequency band. The authors defend that the narrowing of the upper airways during OSA events results in an upward shift of snore frequencies. The study was performed using a high performance microphone concomitantly with the PSG, and the results show a high prevalence of middle and high-frequency band in OSA snorers.

Inter-snore events are related to the global features extracted from the whole set of snoring events. Azarbarzin *et al.* (AZARBARZIN; MOUSSAVI, 2013) hypothesize that snoring sounds vary significantly within a subject depending on the level of obstruction and that this variability is associated with the severity of OSA. To investigate the variability hypothesis, the authors calculated the total variation of a set of features extracted for each individual. Alakuijala *et al.* (ALAKUIJALA; SALMI, 2016) found a positive correlation in periodic snoring sounds during the night OSA severity. Ben-Israel *et al.* (BEN-ISRAEL; TARASIUK; ZIGEL, 2012) combined a set of inter-snore (running variance, apnea phase ratio, and inter-event silence) and intra-snore (mel-cepstability, and pitch density) features in a multivariate linear regression model. The authors found better results combining inter-snore and intra-snore features.

The most recent work in the literature is not directly comparable with your research on estimating the need for CPAP therapy. Akhter *et al.* (AKHTER; ABEYRATNE; SWARNKER, 2017) analyzed acoustical snoring features for estimating the sleep stages in sleep rather than estimating the OSA severity. Kim *et al.* (KIM; KIM; LEE, 2018) proposed a neural network model for detecting the sleep disordered breathing according to the OSA severity. However, the authors combined a sequence of biomarkers for estimating the OSA severity rather than using only the acoustic signal.

Table 2.3 highlights the main differences between the related work on OSA severity estimation. Note that the use of a smartphone as a recording tool and a completely automatic solution is still a challenge to overcome.

Table 2.3: OSA Severity Estimation Solutions

	Smartphone as recording tool	**Complete automat. solution**	**Intra-Snore Features**	**Inter-Snore Features**
Alakuijala *et al.*	x	x	x	✓
Azarbarzin *et al.*	x	✓	x	✓
Ben-Israel *et al.*	x	✓	✓	✓
Emoto *et al.*	x	x	✓	x
Levartovsky *et al.*	x	x	x	✓

Source: by author (2018).

2.5.3 Discussion

As described in the previous sections, some research efforts are investigating the detection of snoring events and OSA severity estimation. However, for the development of a non-intrusive solution that brings together an automatic detection and analysis of snoring events to the context of mHealth, some challenges remain open:

- The viability to record the audio in an uncontrolled environment with a different level of background noises.
- The use of the generic smartphone device as the recording tool.
- A complete unsupervised solution to automatically detect and segment the snoring events.
- A viable and cost-effective OSA severity estimation for improving the access to patient's treatment.
- The understanding of the underlying relations between the acoustic features and OSA events overnight.

Differently, from the related work, we consider all these challenges to propose a complete non-intrusive OSA estimation for CPAP screening. Our proposed solution is described in the next chapters.

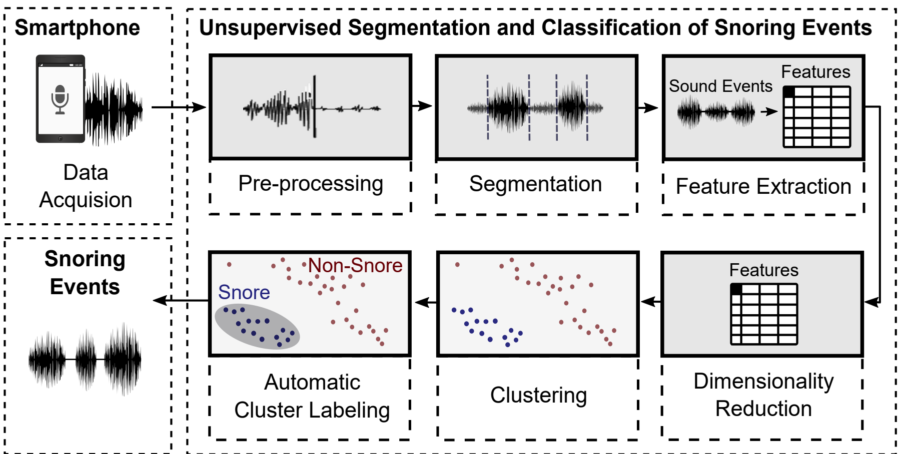
3 OSA SEVERITY ESTIMATION FOR CPAP THERAPY SCREENING

Snoring is characterized by increasing degrees of upper airway obstruction during sleep. We argue that the detection and analysis of acoustical features of these events can bring valuable information about the OSA severity, allowing a nonintrusive CPAP therapy screening. To demonstrate this, we have divided our solution into two phases: (*i*) a completely unsupervised solution to automatically detect the snoring events in an uncontrolled environment and (*ii*) the analysis of acoustical features of the snoring events for CPAP/Non-CPAP classification. In this chapter, we first present and discuss each step for the phase (*i*) and in the sequence for the phase (*ii*).

3.1 Unsupervised Snoring Event Detection

The snoring sounds produced during moments of upper airway obstructions can be recorded for further analysis. However, during the night a variety of random sounds may occur, especially if the recording is performed in an uncontrolled environment. The sounds can range from a simple popping up to a car passing by, so these sound events need to be classified into snore and non-snore events. In order to automatically detect and classify all the sound events, we propose a sequence of steps as can be seen in Figure 3.1. The solution proposed is completely unsupervised due to the fact we are building a mHealth solution to collect and analyze a huge amount of data. We discuss each stage of the overall process in the following subsections.

Figure 3.1: Block diagram of the proposed solution



Source: by author (2018).

3.1.1 Data Acquisition

The acoustic signal is recorded using the built-in microphone of a smartphone and transmitted to the cloud to be processed. Our proposal considers a generic smartphone device, recording in an uncontrolled environment, and an untrained user. These settings may hinder the acoustic quality of the signal but are necessary to achieve accessible and continuous monitoring of the sleep quality, which contributes to the dissemination of sleep self-assessment and patient empowerment. The following steps are cloud processed.

3.1.2 Pre-processing

Noise can significantly degrade the acoustic signal under uncontrolled environments. The background noise is the most common factor degrading the quality and sharpness of the recordings. A noise reduction algorithm needs to be applied to reduce the effects of noise whilst preserving the signal. In our proposal, we chose a Wiener filter, a spectral filter which makes use of a noise template based on a running estimation of the background noise energy (SCALART et al., 1996). This adaptive noise reduction algorithm proposed is explained in details in Section 2.3.2.

3.1.3 Segmentation

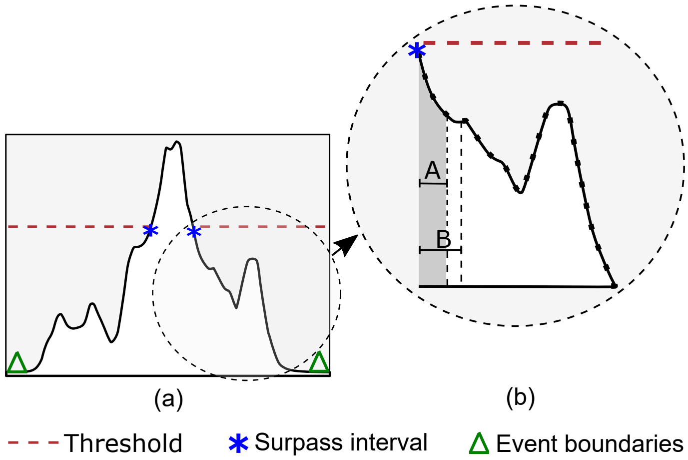
The sound events that occur during the night need to be automatically identified and segmented for future analysis. All detected events are considered as snore candidates. Segmentation proceeds by identifying a snore candidate as an event fulfilling two criteria: (*i*) its energy locally exceeds a certain threshold estimated over the whole night and (*ii*) the event time duration lies within Δt_{min} and Δt_{max}.

A segmentation procedure involves four steps (DAFNA; TARASIUK; ZIGEL, 2013). We designed our segmentation algorithm detailed in the following:

- **Threshold definition:** The energy threshold is computed from the standard deviation of the whole night signal energy. Considering events distribution to be sparse in the whole night signal, we may define the energy threshold as 100 standard deviations. Due to memory constraints, this step is carried on by splitting the signal into sections of length T.

- **Surpass threshold interval:** The estimated threshold is applied to each T long section. The intervals within which the energy signal exceeds the threshold are saved as a candidate event. The threshold can be seen in Figure 3.2 (a).

Figure 3.2: Schematic view of the segmentation stage for one snore event example



Source: by author (2018).

- **Candidate event segmentation:** Having identified possible candidate events, its exact time boundaries, *i.e.,* its beginning and ending times, must be determined. To this end, the area of the shadowed A region, which lies to the right (or left) of the event surpass interval, is computed (see Figure 3.2 (b)). If the relative energy contribution of A region is greater than 0.1%, event ending (or beginning) time is updated accordingly. A region is expanded to B region, iteratively, until the relative energy increment is no longer significant (below 0.1%). Candidate event boundaries (green triangles in Figure 3.2) are properly determined with this algorithm.
- **Fragmentation and duration test (DAFNA; TARASIUK; ZIGEL, 2013):** If two segmented candidate events are close to each other in time, *i.e.,* the ending of the first and beginning of the second are separated by less than δ (of the order of 10^2 ms), the actual event may have been fragmented. In this case, the two candidate events are merged into a single one. Moreover, candidate events duration must be checked to lie within reasonable snoring durations. The time duration of the candidate events is verified to lie within these boundaries, *i.e.*, within Δt_{min} and Δt_{max} (orders of magnitude 10^2 ms and 10^3 ms, respectively), otherwise, the event is discarded.

For each whole night recording, the segmentation stage returns N candidate events, with the respective boundaries for each. Each candidate event is defined by its boundaries, *i.e.*, its initial and final times.

3.1.4 Feature Extraction

A set of features needs to be extracted from each snore candidate for the future events classification. These features can be extracted from time and frequency domains to characterize candidate events individually and are subsequently used to classify candidates as snore or non-snore events during the clustering process. We have generated a set of $m = 75$ features, based on the pool of features proposed by Dafna *et al.* (DAFNA; TARASIUK; ZIGEL, 2013). These set of features are presented in Table 3.1 and is specially designed to cover several acoustic characteristics from within events (denoted intra-event) and comparative between events (denoted inter-event).

For each snore candidate, an $N \times m$ array M is created. Each column i of M contains a column vector of values $(f_i^{(1)}, f_i^{(2)}, \ldots, f_i^{(N)})$, where $f_i^{(k)}$ corresponds to f_i feature value for the kth candidate. Each extracted feature spreads over a particular interval $[\min_k\{f_i^{(k)}\}, \max_k\{f_i^{(k)}\}]$, and normalization should be applied, which rescales the range of all features to lie within $[0, 1]$. This normalization is essential to the variance analysis to be carried on the next stage.

3.1.5 Dimensionality Reduction

A dimensionality reduction technique can be applied to extract the essential information (best features) from the total set of features. From the total set of 75 features, we have selected 10 most important features, with a Mutual Information method for a feature selection, which better describe the snoring events. We have chosen Principal Component Analysis (PCA) to compute new variables denoted principal components which are obtained as linear combinations of the 10 original features (JOLLIFFE, 2002). Therefore, m dimensional set of features extracted can be reduced to only two dimensions, which most explains the data behavior. The first principal component is the linear combination of features having the highest variance in the dataset, which means this component explains the most significant part of the data behavior. The second component has the next highest variance and is subject to the condition that it is uncorrelated, *i.e.,* orthogonal, with the first principal component. Two N vectors are returned in this stage,

Table 3.1: Features extracted from each event

Time Domain Features	
Count	**Feature**
1	Relative energy prior to detected event
2	Relative energy posterior to detected event
3	Rhythm intensity (+- 12 sec)
4	Rhythm period (+- 12 sec)
5	Rhythm period (+- 6 sec)
6	Ratio of relative energy prior and posterior to detected event
7	Normalized area beneath energy envelop
8	Skewness of envelop formation
9	Ratio of areas before and after the peak
10	Total event energy
11	10 seconds before and after period ratio

Frequency Domain Features	
Count	**Feature**
12	First formant frequency
13	First formant magnitude
14-33	20 Mel-Frequency Cepstrum Coefficients (MFCC)
34-53	20 Linear Prediction Coefficients (LPC)
54-57	4 moments of MFCC coefficients
58-61	4 moments of LPC coefficients
62-69	8 subband-frequency distribution
70	Spectral flux
71-74	4 moments of frequency distribution (DFT)
75	Pitch density

Source: by author (2018).

$p_i = (p_i^{(1)}, \ldots, p_i^{(k)}, \ldots, p_i^{(N)})$, with $i \doteq 1, 2$, which contain values for the first ($i = 1$) and second ($i = 2$) principal component of each candidate event.

3.1.6 Clustering

The snoring candidate events must be split into snore and non-snore groups using a clustering algorithm. Candidate event k is defined in the $p_1 - p_2$ plane of principal components as a pair of coordinates $(p_1^{(k)}, p_2^{(k)})$. We propose an EM algorithm for GMM applied to the clustering problem. EM algorithm applied to GMM (MOON, 1996) tries to estimate a set of parameters $\theta_j = (\omega_j, \mu_j, \Sigma_j)$ where $j \doteq 1, 2$ which maximize the log-likelihood function for the set of candidate events:

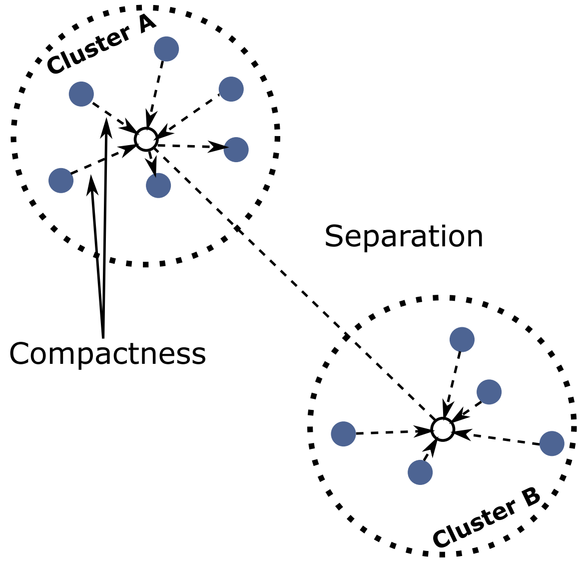
$$L^{(i)} = \frac{1}{N} \sum_{k=1}^{N} \log \left[\sum_{j=1}^{2} \omega_j^{(i)} \, \phi \left(\mathbf{p}^{(k)} \middle| \mu_j^{(i)}, \Sigma_j^{(i)} \right) \right], \tag{3.1}$$

where $L^{(i)}$ represents the log-likelihood at the i-th iteration, $\phi(\mathbf{x}|\mu, \Sigma)$ is the Gaussian distribution with average $\mu = (\mu_1, \mu_2)$ and covariance matrix $\Sigma_{2\times 2}$. The parameters $\{\omega_j\}$ represent the weights given to each of the two Gaussians and $\mathbf{p}^{(k)} = (p_1^{(k)}, p_2^{(k)})$. The maximization is achieved by iteratively updating the set of parameters θ_j until $L^{(i)}$ saturates. The identity of candidate events as snore or non-snore is assessed in a probabilistic manner through the computation of *membership weights*, which are associated to the probability that a given event is generated from Gaussian component 1 or 2. EM iteration is ensured never to decrease the log-likelihood function.

3.1.7 Automatic Cluster Labeling

After constructing the clusters from the data set, each cluster needs to be automatically labeled as snore or non-snore. The most significant distinction between a snore and non-snore cluster is the internal cluster cohesion. The dataset on snore clusters tends to be more compact as compared to non-snore clusters. The algorithm automatically labels as snore the most compact between the two clusters. This labeling is unsupervised, which is essential for a sleep monitoring solution to be able to attain a broad population, demanding a fully automatic mHealth application. A brief illustration of cluster compactness (the internal cohesion) and the separation distance between clusters is given in Figure 3.3.

Figure 3.3: Compactness and separation of events clusters

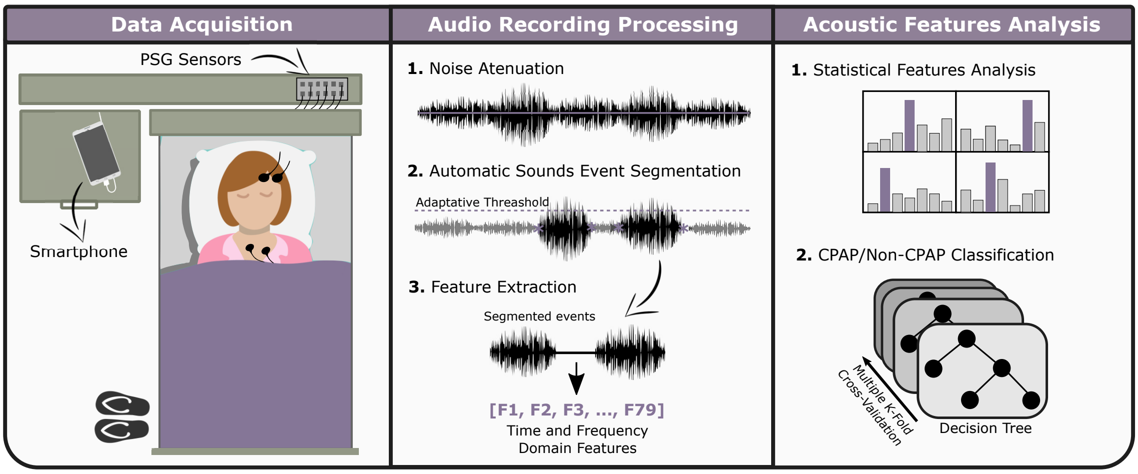


Source: by author (2018).

3.2 Acoustical Analysis of Snoring Events for CPAP/Non-CPAP Classification

The snoring events that occur during the night are the results of the different levels of airway obstructions. These events carry important information about the severity of the OSA obstructive events. A set of acoustic features can be extracted from the snoring events emitted during the night, and a statistical analysis may help to find patterns between these events and the need for CPAP therapy. To automatically infer the patient's need for CPAP, we propose a sequence of steps as shown in Figure 3.4. In the following subsections, we discuss each step proposed.

Figure 3.4: Block diagram of the CPAP/Non-CPAP therapy screening



Source: by author (2018).

3.2.1 Data Acquisition

The acoustic signal is recorded using the built-in microphone of a smartphone. Our proposal considers a generic smartphone device and the recordings were performed in a controlled sleep laboratory environment. The technicians are responsible for starting the audio recording concomitantly to the PSG exam. At the end of the exam, the recordings are stopped, and a backup of the data is performed for the signal processing and analysis.

3.2.2 Audio Recording Processing

The first steps in recording processing are the signal noise reduction and segmentation of snoring events. The signal is pre-processed and the events that occurred during the sleep night are automatically detected and segmented according to the same process proposed in Section 3.1. By the fact the signal acquisition is performed in a controlled environment with low levels of random noises, all events that meet the criteria established by the segmentation algorithm were considered to be snoring events.

After the sound events were automatically detected and segmented from the recording, we must extract features that allow further analysis of the need for CPAP therapy. We propose a set of 79 features extracted in the time and frequency domain, which explore the global acoustical aspects of the sleep night, as shown in Figure 3.2. The set of features studied in Section 3.1 was used as the basis for this new set of global night features that map through the detected events the different aspects of the patient's sleep night. Most of the global night features are obtained through the analysis of variance and average of the features for all events detected during the night. New features were proposed to map behaviors between the events, and not only intra-events. Details of each feature are given in Appendix B.

Table 3.2: Global features for the patient's sleep night

Time Domain Features	
Count	**Feature**
1	Number of Snores Events
2	Snoring Time Ratio
3	Variance of Snoring Time Duration
4	Number of Inter Snore Event Silence (>10 and <60 sec)
5	Mean of Inter Snore Events Silence
6-7	Running Energy/Distance Snores Variance
8-9	Variance and Mean of Relative Energy Prior the Detect Event
10-11	Variance and Mean of Relative Energy Posterior the Detect Event
12-13	Variance and Mean of Rhythm Intensity (12 sec)
14-15	Variance and Mean of Rhythm Period (12 sec)
16-17	Variance and Mean of Rhythm Period (6 sec)
18-19	Variance and Mean of Ratio Relative Energy Prior and Posterior
20-21	Variance and Mean of Normalized Area Beneath Energy Envelop
22-23	Variance and Mean of Skewness Envelop Formation
24-25	Variance and Mean of 10 Seconds Before and After Period
26-27	Variance and Mean of Ratio Areas Before and After the Energy Peak
28-29	Variance and Mean of Total Snore Event Energy
Frequency Domain Features	
Count	**Feature**
30	Retro-palatal Ratio
31	Snore Frequency Intercalation
32-33	Variance and Mean of First Formant
34-35	Variance and Mean of Fundamental Frequency
36-43	Variance and Mean of 4 Moments of MFCC Coeficients
44-51	Variance and Mean of 4 Moments of LPC Coeficients
52-59	Variance and Mean of 4 Moments of DFT
60-75	Variance and Mean of [#1 to #8] Subband Frequency
76-77	Variance and Mean of Spectral Flux
78-79	Variance and Mean of Pitch Density

Source: by author (2018).

3.2.3 Classification Model

The patients are divided into two classes: CPAP and Non-CPAP, according to the OSA severity. We propose to apply the CART (Classification and Regression Trees) algorithm to build a decision tree classifier to combine the extracted features from the audio recording to predict the right class for each patient. Decision trees are simple and powerful representations of knowledge and have been widely used in classification as an efficient methods to build classifiers that predict classes based on the attribute values. Through the decision tree, we can perform both a feature selection (evaluate the most important features) and measure the patient's classification performance.

The CART algorithm works selecting input features and split points on those features until a suitable binary tree is constructed. The selection of which input feature and the specific split is chosen using the Gini index function, which indicates how pure the leaf nodes are. Therefore, the Gini index aims to select the features that better separate the classes to construct the decision tree. The tree construction ends using a predefined stopping criterion chosen as a minimum of training instances assigned to each leaf node of the tree. The result obtained is the decision tree, which is used to classify new cases and explore the relationships between the data features.

In this Section, we detailed our proposed solution for the automatic detection of snoring events and classification of patients according to the need for CPAP treatment. In the following Section, is described the methodology and results achieved in each of these stages.

4 EXPERIMENTAL EVALUATION

In this chapter, we present an experimental evaluation of our proposed solution. The evaluation process was divided into two steps. Firstly, we evaluate the performance of the unsupervised detection and classification of snoring events. Secondly, we show the experimental evaluation of the OSA severity estimation for CPAP/Non-CPAP classification.

4.1 Snoring Events Detection Evaluation

In this Section, we present our experimental evaluation for the snoring events detection. Firstly, we show the methodology for the sleep night recording under an uncontrolled environment. Secondly, we discuss the results for our unsupervised snoring events detection.

4.1.1 Methodology for Snoring Events Detection

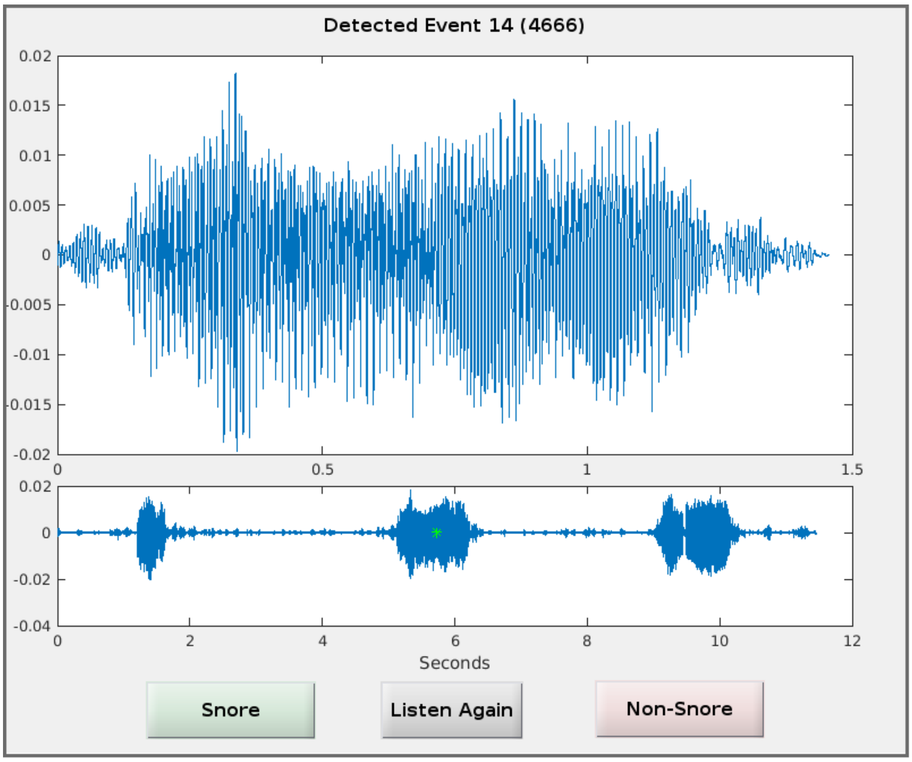
An Android application called *MySleep* was specially designed to acquire and record the acoustic signal from the whole night (6 hours approximately) and to transmit it to the cloud for further processing. The acoustic signal is sampled at a frequency of 44.1 kHz and a bit depth of 16 bits. This sampling rate produce maximum frequency of 20kHz, which is the highest frequency generally audible by humans. The participants were oriented to install the application on their own smartphone and place it at a distance of approximately 50 cm besides the bed before sleep, on a nightstand cleared for other objects as shown in Figure 4.1. In Appendix A is presented more detail about the Android application developed.

Figure 4.1: Uncontrolled sleep recording acquisition

Source: by author (2018).

A trained technician was responsible for listening to all segmented candidate events and manually classify each one as snore or non-snore for the accuracy evaluation of the proposed solution. An interface was specially developed to show the signal of each candidate event and reproduce the sound. Using this interface, it is possible to classify the candidate event as a snore or non-snore event as shown in Figure 4.2.

Figure 4.2: Manual labeling user interface



Source: by author (2018).

Once the signal has been stored in the cloud, we proceed to stages shown in Figure 3.1 for signal processing and event classification. We describe relevant information and parametrization for some of the stages.

- *Pre-processing:* The signal is divided into sections of τ = 10s long, and for each one, a frame energy vector is calculated. The 5 lowest energy frames are elected as noise standards and used to compute the spectral average noise of the section. This average is estimative of the signal noise template and is updated along the night recording (DAFNA; TARASIUK; ZIGEL, 2013). Furthermore, this template is used as input to the Wiener filter (SCALART et al., 1996).
- *Segmentation:* The standard deviation is calculated over the total sum of energies for each T = 10 min section, and the energy threshold is established. For each T long section, an energy vector is calculated using 60 ms frames with 75$\%$ of overlap. After, the resulting signal is tested for threshold surpassing. Having determined the candidate event boundaries, fragmentation and duration tests must be further applied. We chose δ = 200 ms for fragmentation test, Δt_{min} = 600 ms and Δt_{max} = 3500 ms for soring duration test.
- *Dimensionality Reduction:* The 10 features selected through the Mutual Information method are, in order of importance: 6, 1, 29, 61, 65, 2, 36, 63, 70, 7 (count numbers are according to Table 3.1). Using Mutual Information for feature selection involves feeding the algorithm with the manually labeled events classification for a set of events to measure the information quantity, the presence or absence of a feature, contributes to making the correct classification. The 10 features selected in this manner will be applied to all recordings alike. The 10 features are fed to PCA which will then generate a pair of principal components particular to each recording.

4.1.2 Experimental Results

We applied our solution to six-night recordings, totalizing 32.5 hours. The segmentation stage returned a total of 5323 snore candidate events, of which 3574 were manually labeled as snoring. Table 4.1 depicts the actual (manually labeled) numbers of snoring events for simple and OSA snorers and compares the accuracies averaged over nights obtained for an implementation of our solution (EM) and an implementation of FCM, discriminating simple and OSA snorers. For both simple and OSA snorers, one concludes that EM average accuracies are well above FCM corresponding accuracies (over one - either EM or FCM - standard deviations).

Table 4.1: Algorithms comparison for simple and OSA snorers

	Number of Snore Events	**Average Accuracy EM (%)**	**Average Accuracy FCM (%)**
Simple Snorers	2670	91.3 ± 1.0	81.6 ± 8.8
OSA Snorers	904	79.7 ± 0.9	67.7 ± 4.5

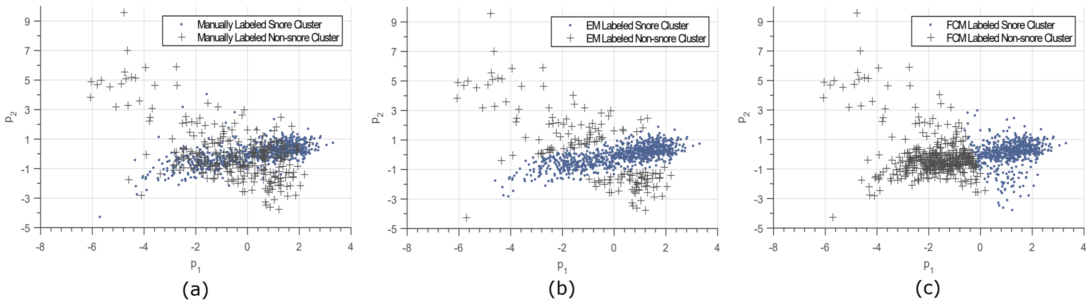
Source: by author (2018).

The explanation for the poor performance of FCM is illustrated in Figure 4.3. We compare the clusters of snore and non-snore as defined by manual labeling (panel (a)), cluster results by our EM solution (panel (b)), and cluster results by an FCM implementation (panel (c)) of the same dataset for one-night recording of an OSA snorer. We see from the panel (a) that the two manually labeled clusters are superimposed in $p_1 - p_2$ plane which, by the PCA analysis, is ensured to span the best principal components for data segregation. This superposition is a typical behavior for overnight sound recordings.

We demonstrate in panel (b) that EM clustering can estimate the correct snoring distribution to an adequate accuracy, which for this specific recording is 81.0%. This accuracy is due to the statistical approach of EM applied to GMM, which assumes a Gaussian distribution of snore and non-snore events, as described in Section 3.1.6. Although we have no reason to expect non-snore events to be normally distributed, the assumption of normal distributed snores is a reasonable one, as explained in Section 2.5.1. Meanwhile, FCM implementation, which clusters events together considering its distances to two iteratively defined centroids, is not able to approximate the right snore/non-snore clusters. In fact, in the specific recording considered in Figure 4.3, the accuracy is 61.3%. We expect all other clustering algorithms based on geometric criteria such as FCM, KHM, etc., to present similarly poor performances, since all of them consider point proximity as similarity criterion.

Figure 4.4 aims to demonstrate that the Gaussian approximation to the snore distribution is an adequate one. The dataset analyzed is the same as in the previous plots. In Figure 4.4, panel (a), we compare manually labeled snoring cluster histogram to the

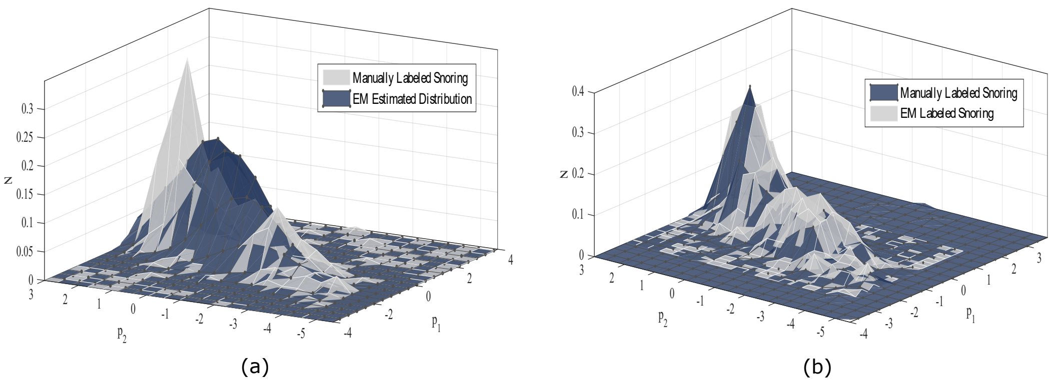
Figure 4.3: Clustering results: comparison between manually labeled, EM labeled and FCM labeled snore and non-snore clusters for one night recording



Source: by author (2018).

same EM estimated Gaussian distribution. Panel (b) depicts histograms of the manually labeled and EM labeled snoring clusters.

Figure 4.4: Clustering results: comparison between EM estimated distribution for snoring events, EM histogram for the snoring events cluster, and manually labeled snoring events histogram



Source: by author (2018).

The ideal behavior of the EM implementation occurs when the Gaussian distributions estimated by the EM algorithm well approximate both the snore and non-snore manually labeled histograms. The non-snore manually labeled histogram will usually not approximate a Gaussian distribution, because events are not identically distributed, since they may be as diverse as a car passing by, a cough, or an object falling on the ground. However, even if non-snore events are not normally distributed, for simple snorers, the manually labeled snore histogram approximates well a Gaussian distribution (with like mean and appropriately defined variance). In such a case, EM is expected to deliver a high accuracy, as is verified from EM average accuracy for simple snorers and its standard deviation (Table 4.1). Below we show that even when OSA snorers are considered, which represents a much more complex situation, EM can perform adequately.

The night recording of an OSA snorer is typically composed of apneic and non-apneic snores (HALEVI et al., 2016). As a result, the manually labeled snoring events histogram is heavy-tailed and almost bimodal (see Figure 4.4, panel (a)). Although both averages $\langle p_1 \rangle$ and $\langle p_2 \rangle$ are very close for the EM Gaussian distribution and manually labeled snores histogram, we see from the panel (a) that the histogram is not well approximated by the EM estimated distribution. Furthermore, we demonstrate in the panel (b) that the estimated EM histogram fits almost perfectly the manually labeled snore events histogram, as we expect from an adequate clustering. This correct estimation is quite remarkable and is only possible because EM can cluster events together by (indirectly) recognizing its "statistical similarity", *i.e.,* the probability of two events being drawn from the same given distribution.

4.2 OSA Severity Estimation Evaluation

In this Section, we present our experimental evaluation for the OSA severity estimation. Firstly, we show the methodology for the sleep night recording in a sleep laboratory. Secondly, we discuss the results for our CPAP/Non-CPAP classification.

4.2.1 Methodology for OSA Severity Estimation

To evaluate our solution for estimating the severity of OSA is necessary to compare the results with the gold standard (PSG). Patients who were submitted to PSG in the sleep laboratory center of São José Hospital from Irmandade da Santa Casa de Misericórdia de Porto Alegre were recruited to participate in this study. Ethics approval was granted by the Institutional Review Board of the Federal University of Health Sciences of Porto Alegre (UFCSPA), Porto Alegre, Brazil (2.230.675) and all procedures were in compliance with the current regulations, and all participants provided written informed consent.

Participants arrived in the sleep laboratory at evening and were normally instrumented for the PSG exam. A smartphone was used to recording the audio of the snoring sounds concomitantly to PSG. The smartphone was placed at a distance of approximately 50 cm besides the bed before the exam start as shown in Figure 4.5. The audio was recorded for each patient during the whole exam (8 hours approximately) and was stored with the medical report produced by the PSG equipment (SONOLAB 632 - Meditron Eletromedicina Ltda) and the physician's evaluation. The recruitment of volunteers lasted three months totaling a set of 113 patients.

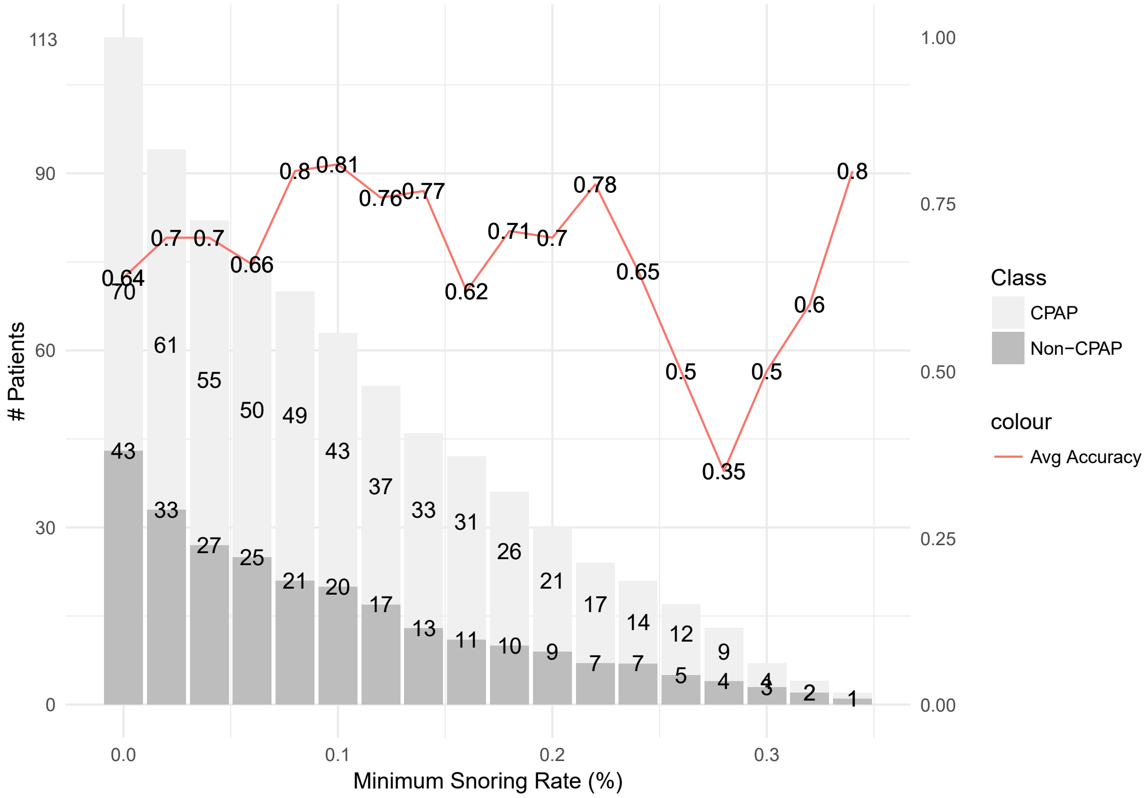
Figure 4.5: Experiment configuration in PSG room

Source: by author (2018).

As described in Section 3.2, our solution is based on the analysis of acoustical features of the snoring events. Therefore, it is imperative that the patient has produced snoring events at night to be able to estimate the severity of the OSA based on these events. To ensure a minimum number of snoring events required for OSA analysis, we have defined a threshold based on the snoring time ratio, which is one of the 79 features extracted from the whole sleep night. The snoring time ratio measures how long the patient remained to snore with the total recording time.

The minimum snoring rate time was varied from zero (includes all recordings) to the maximum value of 0.36 found in our dataset. In Figure 4.6, we can observe that as the minimum snoring rate threshold increase, the number of patients in the data set that satisfy this condition decreases. The average accuracy was also calculated for each minimum snoring rate, which reached a peak of 0.81. To build a dataset that represents the population of CPAP/Non-CPAP patients and also has sufficient snoring events for analysis, the minimum threshold of 0.1 for snoring rate was defined.

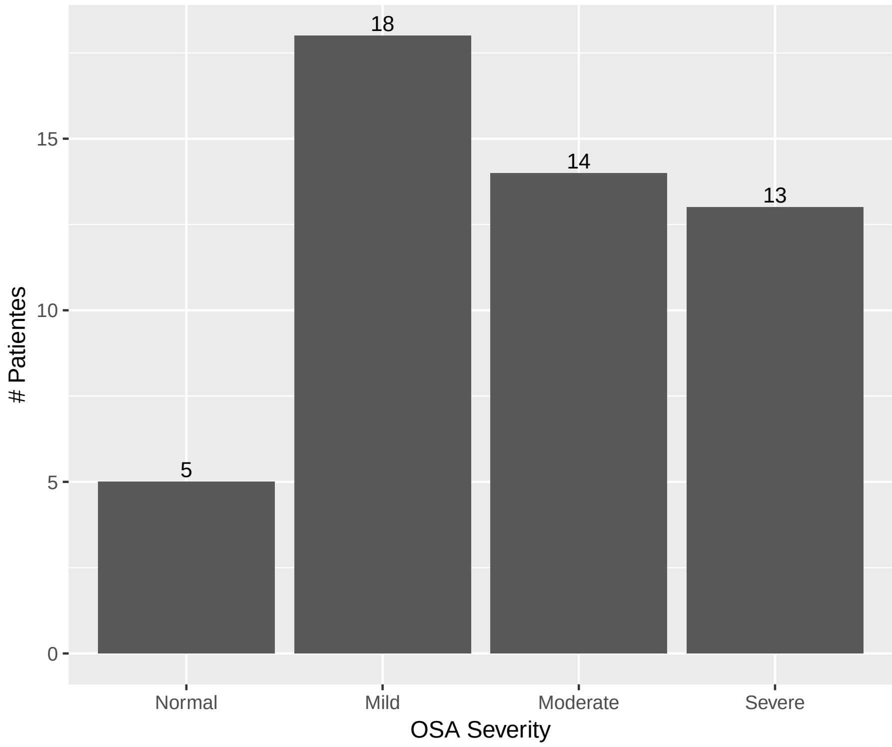
Figure 4.6: Minimum snoring rate vs average accuracy



Source: by author (2018).

Applying the threshold defined above, we reduce our dataset from 113 patient to 63 patients who have sufficient snoring events as input to our classification solution. The other 50 patients are not included in this study. In Figure 4.7, we can observe the severity distribution for the 50 patients not involved in this analysis. Although the amount of snoring events is not adequate for our solution, we cannot assume that these patients do not suffer from OSA only because they produce a small quantity of snoring events. The number of snoring events produced during the night does not correlate with the OSA severity, as we can observe in Table 4.2, patients with less than 10% snoring rate have a lower rate of Apnea/hour and usually a higher rate of Hypopnea events. This behavior is typical of patients that produce a few snoring events during sleep or even did not sleep at all during the PSG, and then, it is necessary to explore different features of the snoring events to be able to estimate the OSA severity for those patients.

Figure 4.7: Severity distribution for the 50 patients not included in the study



Source: by author (2018).

Table 4.2: Patient's OSA severity according the snoring rate

Snoring Rate	# Patients	AHI	Apnea/hour	Hypopnea/hour
<10%	50	26.6 ± 30.9	5.3 ± 11.3	21.2 ± 21.85
>10%	63	31 ± 26.3	7.2 ± 12.8	23.8 ± 19

Source: by author (2018).

The set of 63 patients who produced a satisfactory number of snoring events is composed of 43 CPAP patients and 20 non-CPAP patients. This non-uniform distribution between classes causes training vies which can influence the classifier results and benefit the selection of features that better describe the class with the largest set of data. We have performed tests using 43 CPAP, and 20 Non-CPAP patients and the results favored the predominant class (CPAP). Next, we randomly selected 20 patients from each class (CPAP and Non-CPAP) to create a uniform dataset between CPAP and Non-CPAP patients. In

Table 4.3, can be shown a proper distribution of gender (20 male and 20 female), a similar age group (average age of 56 years), and a slight increase in the Body Index Mass (BMI) between the groups.

Table 4.3: Patient's characteristics in the study group

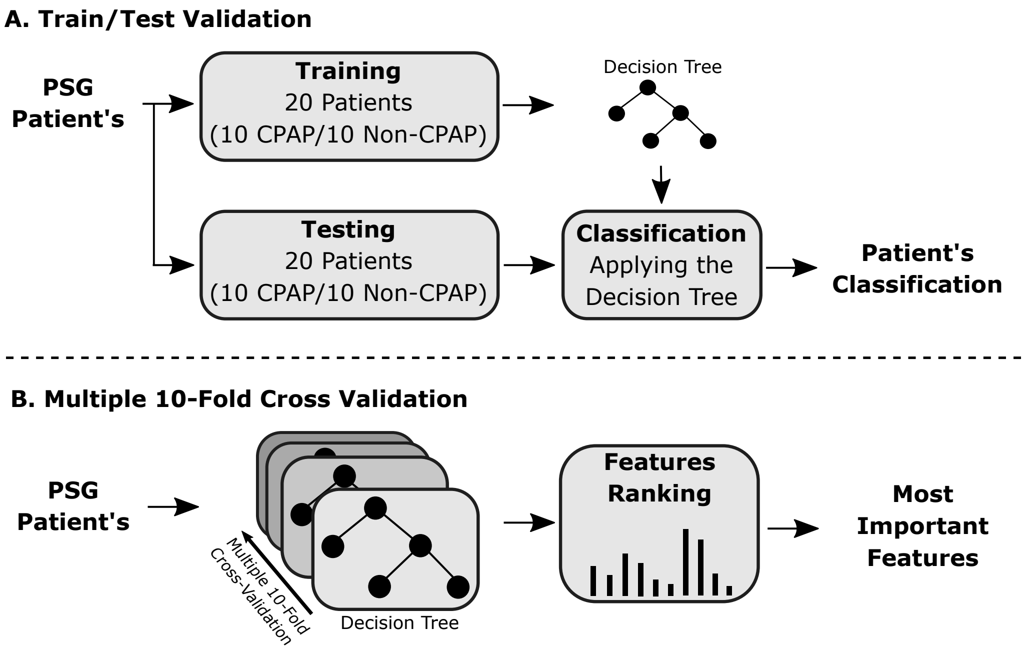
Group	# Patients	Age	BMI	AHI
Non-CPAP	20 (12 female)	53 ± 11.7	24 ± 5.3	7.8 ± 4
CPAP	20 (8 female)	59 ± 12.6	31.8 ± 4	28.2 ± 24.6

Source: by author (2018).

Once the dataset for this study was defined, we performed an analysis on each night recording. A set of 79 acoustic features were extracted from each segmented snoring event. These features map different aspects of the signal regarding time and frequency domain and are widely used in the literature for the detection and analysis of snoring events (BUBLITZ et al., 2017). Our solution is not based on a single analysis of the features of the snoring events detected, but instead, we explore the distribution of these features during the night of sleep. For mapping, the vocal tract instability associated with OSA, for each feature extracted separately from the snoring event, the variance and average for the whole night features was applied. Therefore, it is possible to analyze how often those feature has changed during the night. Our *hypothesis* is in severe apneic patients, where the high vocal tract instability, alters the acoustic features of snoring events at night, obtaining a higher variance than found in normal patients.

To evaluate our proposal we have proposed two different phases as shown in Figure 4.8. In phase *A*, we divided our dataset into two uniform groups regarding CPAP/Non-CPAP patients for training and testing. The training dataset was used to generate the decision tree model, while the testing dataset was used to evaluate the classification performance. From this process, we can select the best features for patients classification and explore the underlying relations between the features. The datasets for training and testing, as well the proportion of CPAP and Non-CPAP patients in each one are equally distributed to avoid a data vies which can benefit the dominant class in training/testing process.

Figure 4.8: Methodologies for CPAP/Non-CPAP classification evaluation



Source: by author (2018).

To guarantee and evaluate the robustness of the decision tree approach and feature selection proposed in phase A, we propose a multiple 10-Fold Cross validation process demonstrated in phase B. This process has the objective of shuffling the input data for the generation of multiple decisions trees. The original dataset are randomly partitioned into 10 equal sized subsamples. A single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The cross-validation process is then repeated 10 times, with each of the subsamples used exactly once as the validation data. This training and validation process was extensively replicated 1000 times to reduce the input data's susceptibility by producing different sets of data for training and validation. We extract the most selected features to patient's classification in the multiples decision tree generated with the objective to check the robustness of the set o features found in phase A.

4.2.2 Experimental Results

We applied our solution to 40 patients, totalizing approximately 321 hours of recording analyzed and 128.408 snoring events detected. For each patient, 2187$\pm$1696 snoring events were automatically extracted from the sleep recording. In Table 4.4 we can see the confusion matrix computed from dataset. The confusion matrix provides an adequate measure of the classification model by showing the number of correct classifications versus the predicted classifications for each class over the dataset. We can observe a large number of correct classification against the number of misclassification (16/4, respectively).

Table 4.4: Confusion Matrix

Real/Predicted	**Non-CPAP**	**CPAP**
Non-CPAP	9	1
CPAP	3	7

Source: by author (2018).

Analyzing Table 4.4, we can calculate some metrics to help the evaluation of the diagnostic agreement of our solution as shown in Table 4.5. Our solution reaches the rate of 80% of agreement with the PSG diagnostic. In other words, we were right in recommend the use or not of CPAP in 80% of the 40 patients. We underestimate the diagnosis in only 15% of the patients, which means that we would wrongly not recommend a CPAP treatment for 3 patients. On the other hand, we overestimate the diagnosis in only 5% which represents the cases where the CPAP treatment was wrongly recommended.

Table 4.5: Diagnostic Agreement

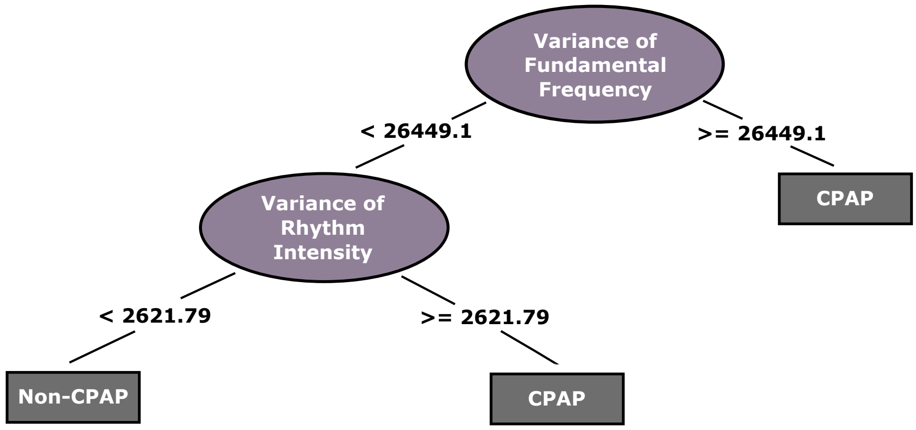
Diagnostic	**%**
Agreement	0.80
Underestimate	0.15
Overestimate	0.05

Source: by author (2018).

These results were obtained from the decision tree learned in the training process described in Section 4.2.1. In Figure 4.9 we can observe the resultant decision tree after training. Two features were selected for the tree nodes: *Variance of Fundamental Frequency* and *Variance of Rhythm Intensity*. These features were selected for their higher capabilities in separating CPAP and Non-CPAP classes. To better understand the reason for selecting these features and what their connections with the OSA, we will detail their properties.

- **Variance of Fundamental Frequency (F78):** We define $f_0(i)$ as the frequency at which the global maximum of the ith snore signal's power spectral density occurs.

Figure 4.9: Decision Tree for CPAP/Non-CPAP classification



Source: by author (2018).

The feature expresses variance of f_0 over all snores in the one-night recording. The Discrete Fourier Transform (DFT) of each snoring event signal $s_i(t)$ was calculated according to $\hat{S}_i(f) = DFT(S_i(t))$. From the frequency domain, the fundamental frequency variance is given by

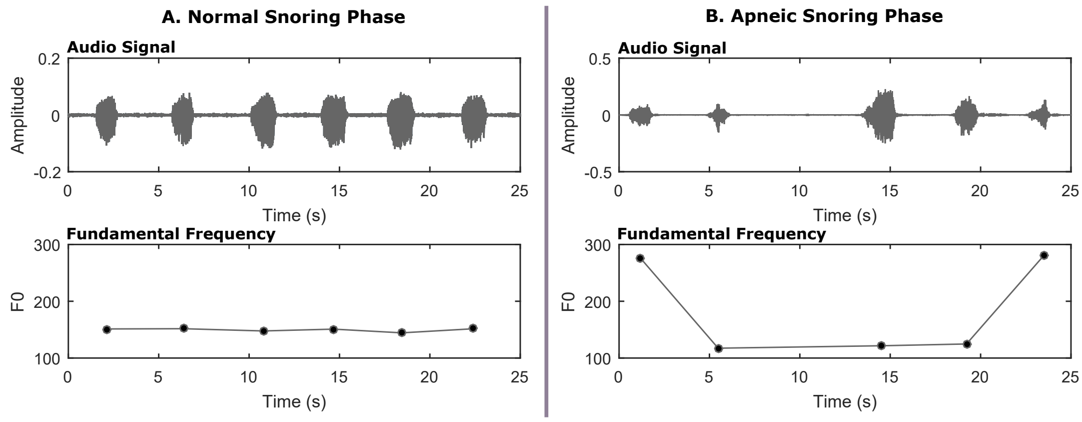
$$\mathbf{F78} = \frac{1}{N_s - 1}\sum_{i=1}^{N_s}(f_0(i) - \bar{f}_0)^2,$$

$$\textbf{where} \quad f_0(i) = f \ : \ \max_f \left(\left| \hat{S}_i(f) \right|^2 \right), \ \textbf{and} \quad \bar{f}_0 = \frac{1}{N_s}\sum_{i=1}^{N_s} f_0(i) \quad (4.1)$$

Relation with OSA: The variance of f_0 over all snores is related to the vocal tract size and shape. The vocal tract structure, which includes mainly the pharynx, nasal and oral cavity, and tongue body, impacts in the formants frequencies emitted during a snoring event sound. OSA patients have alterations in the shape of the vocal tract, due to the different levels of narrowing during sleep. This narrowing impacts in how the frequencies resonate in the vocal tract structure, thus producing a higher variance of the fundamental frequency over all snoring events.

Figure 4.10 illustrates variations in fundamental frequency during a period of normal snoring (without the presence of a respiratory obstruction) and an apneic phase (with increased obstruction level and a short period of respiratory arrest). We can observe that during a normal snoring sequence *(A)*, the fundamental frequency remains practically constant, which demonstrates the higher stability of the vocal tract resonant frequencies. During the more apneic phases *(B)*, there is a higher variance of the fundamental frequency due to the higher changes in the levels of obstruction of the airways. Additionally, a significant decrease in the fundamental frequency in the events that precede and succeed the obstruction period can be noted around the 5-15 minutes. Patients with severe OSA have the tendency to oscillate more frequently between these phases of normal and apneic snoring during the sleep night, resulting in a higher variance of the fundamental frequency over all snoring events.

Figure 4.10: Variance of fundamental frequency feature



Source: by author (2018).

- **Variance of Rhythm Intensity (F50):** Calculate the period intensity feature via autocorrelation, C, of an energy signal interval S_e of 12 seconds, where the detected snoring event is in the middle. The rhythm intensity is calculated as the product of the first peak amplitude value of correlation $C(\tau_p)$ and the normalize square area ($Area$) between the zero-lag autocorrelation and the first correlation peak (τ_p), which can measure the level of the periodicity of the snoring events. A more periodic energy snoring pattern will result in a higher area and, hence, a higher period intensity value. The rhythm intensity variance is given by

$$\textbf{F50} = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} (R_I(i) - \bar{R}_I)^2,$$

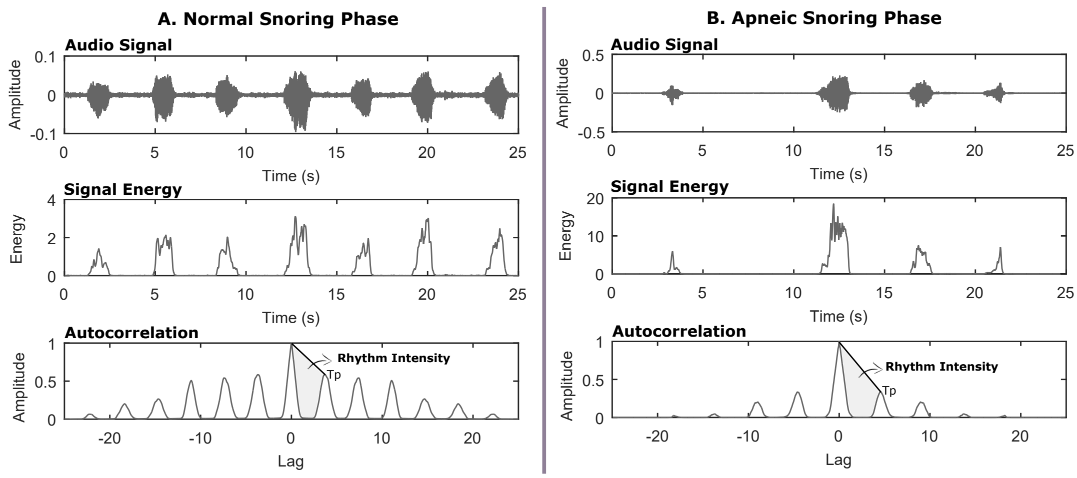
$$\textbf{where} \quad R_I = C(\tau_p) \times Area, \; Area = \frac{1}{\tau_p} \sum_{\tau=0}^{\tau_p} \left(\alpha\tau + 1 - C(\tau)\right)^2,$$

$$\textbf{and} \quad C(n) = \sum_{m=1}^{N-n} S_e(m+n) \times S_e(m) \quad (4.2)$$

Relation with OSA: Usually, the snoring phases during the night are composed of many repeated events according to a specific time period between them. The snoring frequency in the normal snoring patients usually corresponds to the respiratory rate. The rhythm that these events repeat themselves tends to have a small variance since they correspond to respiratory frequency, which is expected not to have significant changes during sleep. In the case of apneic patients, snoring is a sign of narrowing of the upper airway. Breathing cessations during apnea events interrupt the snoring sounds, and it only returns with the efforts to resume breathing. Therefore, the variance of the rhythm intensity of these events is usually higher for apneic patients.

Figure 4.11 illustrates the variation of the signal autocorrelation during a period of normal and apneic snoring. The snoring sequence with a regular period *A* produces a signal autocorrelation with well-defined peaks and a higher area between the central peak and the first peak (rhythm intensity) compared to a signal with the irregular period of events due to breathing obstructions *B*.

Figure 4.11: Rhythm Intensity Feature

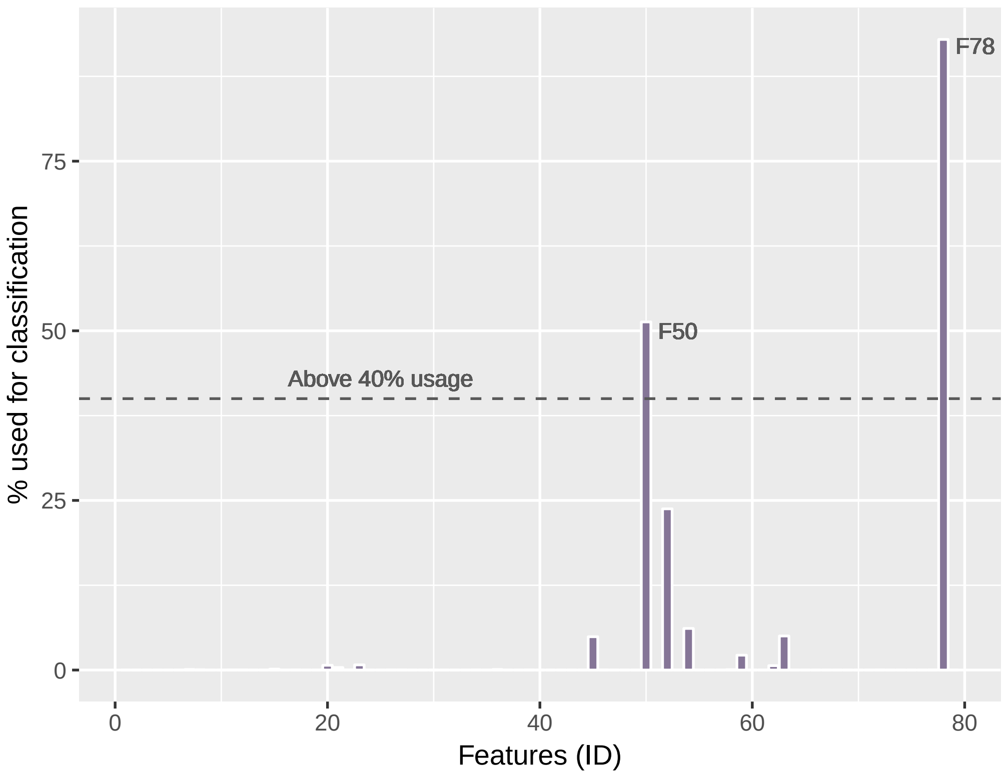


Source: by author (2018).

Multiple 10-Fold Cross Validation:

To guarantee and evaluate the robustness of the decision tree approach regarding the set of most important features extracted from the resulting decision tree, we proposed a multiple 10-Fold cross validation process for shuffle the data and decrease data susceptibility to train/test datasets. After repeating the cross-validation process for 1000 times (each time produce a complete random 10-fold cross validation), it was possible to identify the most used features for patient's classification across all rounds. Figure 4.12, shows the most used features for all decision trees generated. Two features were more widely used for the decision trees: $F78$ (*Variance of Fundamental Frequency*) and $F50$ (*Variance of Rhythm Intensity*). These features are the same found in the first training/testing approach evaluated, which confirms that independently of the data inputs (considering our dataset), we have two features that stand out for the classification of CPAP and Non-CPAP patients.

Figure 4.12: Most used features for patient's classification



Source: by author (2018).

Decision tree features for CPAP/Non-CPAP classification:

Normal and apneic patients have different snoring phases during the sleep night. Generally, the snoring phases coincide with the deeper phases of sleep, *e.g.,* REM, in which a relaxation of the vocal tract musculature produces the snoring events. From the decision tree in Figure 4.9 obtained during the training process, we can infer that a higher variance of the fundamental frequency represents the need for CPAP treatment. The higher fundamental frequency variance shows that the vocal tract suffered different levels of narrowing during the night, altering the resonant frequencies of the vocal tract and consequently, resulting in snoring events with different fundamental frequencies. In this case, the high variance of the fundamental frequency may indicate that the patient presented many apnea events, which altered the frequency spectrum of snoring events at night.

The patients with a lower fundamental frequency variance means that regarding the frequency spectrum the snoring events did not change much at night, even if the patient must have several respiratory arrests, the snoring events emitted between the apnea events did not present major spectral differences capable of estimating the need for CPAP treatment. For these patients, the variance of the rhythm intensity of snoring events was a feature capable of estimating the need for CPAP therapy. Patients with a higher rhythm intensity variance, *e.g.*, snoring events that are emitted in a non-uniform time interval were classified as CPAP patients. This feature can map the moments of pause in respiration that are characteristic of the apnea events. When the snoring events that indicate the narrowing of the airway are interrupted, it may indicate that the vocal tract narrowing reaches its maximum level, that is, the patient does not breathe for a few seconds. The snoring events return with breathing efforts, and the variations of these intervals between the apnea events help to characterize the patients who require CPAP treatment.

Finally, patients with a low fundamental frequency and rhythm intensity variance are classified as Non-CPAP patient. A uniform snoring events fundamental frequency and rhythm intensity during the night can reveal the patients with benign snoring phases. These patients may have minor changes in the upper airways that cause the snoring events, but not necessarily the more severe respiratory obstructions as in the case of severe OSA patients.

4.3 Discussion: Overall Evaluation

We argue that our proposal solution can be a valuable alternative for a preliminary OSA severity estimation and CPAP therapy screening. The use of the smartphone as a patient's data capture device allows a non-intrusive, easily accessible and low cost sensing, which contributes to the treatment access improving. Our solution was proposed and evaluated in two steps: the unsupervised snoring event detection and the acoustical snoring events features analysis for patient's classification.

The unsupervised snoring event detection performs the events classifications in an uncontrolled environment demonstrating the viability to differentiate random noises from snoring events even in the most diverse noise levels. This event detection is a very important archived for the patient's empowerment and access to treatment improving. The EM algorithm was able to well approximate the snoring events distribution, performing significantly better than the FCM algorithm.

The analysis of the acoustic features shows fundamental relations between the acoustic changes in snoring events and OSA events during the night, which contribute to better understanding of the obstructive respiratory events. The set of most important features found through the decision tree model bring relevant information about the acoustical changes of the snoring events during the night of an OSA and non-OSA patient.

5 CONCLUSION

The acoustical analysis of snoring events brings valuable information about the presence and severity of the OSA. Moreover, the detection of snoring events using the smartphone contributes to improving the patient access to treatment by providing an auxiliary tool for screening the need for a clinical diagnosis. This analysis also contributes to the understanding of the relations between the upper airway obstructions and acoustical changes in the snoring events produced.

In this dissertation, we proposed a noninvasive solution for CPAP therapy screening using the built-in microphone of a smartphone. Our solution comprises an unsupervised segmentation and classification of snoring events responsible for process the acoustic signal for events segmentation and clustering into two classes: snore and non-snore, and an acoustical analysis of features extracted from the set of snoring events for the screening of CPAP therapy, separating patients into two classes: CPAP and Non-CPAP. As a result, the set of snoring events produced during the night can be automatically detected and analyzed for estimate the need for treatment and medical follow-up based on a set of acoustic features. Next, we highlight some conclusions and key contributions of our research.

5.1 Summary of Contributions

The main contributions of our research are the following:

- The design of non-invasive acoustical signal acquisition and digital signal processing for automatically detect and segment all the snoring events candidates during a sleep night;
- An unsupervised solution for snoring events classification, using a statistical approach analysis (the EM clustering algorithm) for clustering the candidate events into two classes: snore and non-snore, reaching satisfactory accuracy rates (91.3% for simple snores and 79.7% for OSA snores, on average);
- A CPAP therapy screening based on acoustical features analysis of snoring events using a decision tree approach for patient classification, reaching an accuracy of 80%;
- A meaningful analysis of acoustical features and the underlying relation with the

OSA events, as well as the novel features proposed for patients CPAP therapy screening;

- An evaluation of the unsupervised snoring events detection and patient's CPAP therapy screening using real patient's data collected in uncontrolled (at home) and controlled (at PSG laboratory).

To the best of our knowledge, we were the first to propose a fully automatic solution for detection of snoring events and CPAP therapy screening using a smartphone as the recording tool. This solution was a very important step in the way of proving the viability of the use of fully accessible hardware such as smartphones for improving access to the sleep disorders treatment.

5.2 Final Remarks and Future Work

As part of our future work, we aim to expand our study to a broad population and explore the relations between acoustic features and vocal tract anatomical properties. For that, our future research efforts are the following:

- We aim to evaluate our solution for CPAP therapy screening applied to an uncontrolled environment, expanding the number of patients of the study and measuring the effects of ambient noise on the statistical analysis of features;
- We intend to explore the relations between the patient's anatomical vocal tract properties and the acoustical features of the snoring events, to better understand the mechanisms of snoring production;
- Developing a source-filter model, considering anatomical characteristics of the vocal tract, such as length, internal wall thickness and cross-sectional area, including the position of obstructions;
- Studying the application of the proposed model, as well as the analysis of the acoustic features for the monitoring of other chronic and degenerative diseases, such as *Alzheimer* and *Parkinson*;

REFERENCES

AKHTER, S.; ABEYRATNE, U. R.; SWARNKER, V. Characterizing the nrem/rem sleep specific obstructive sleep apnea severity using snore sounds. In: IEEE. **2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. Korea: IEEE, 2017. p. 2826–2829.

ALAKUIJALA, A.; SALMI, T. Predicting obstructive sleep apnea with periodic snoring sound recorded at home. **Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine**, American Academy of Sleep Medicine, v. 12, n. 7, p. 953–958, 2016.

AZARBARZIN, A.; MOUSSAVI, Z. Snoring sounds variability as a signature of obstructive sleep apnea. **Medical engineering & physics**, Elsevier, v. 35, n. 4, p. 479–485, 2013.

AZARBARZIN, A.; MOUSSAVI, Z. M. Automatic and unsupervised snore sound extraction from respiratory sound signals. **IEEE Transactions on Biomedical Engineering**, v. 58, n. 5, p. 1156–1162, 2011.

BEN-ISRAEL, N.; TARASIUK, A.; ZIGEL, Y. Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults. **Sleep**, Oxford University Press, v. 35, n. 9, p. 1299–1305, 2012.

BREIMAN, L. et al. Classification and regression trees. wadsworth & brooks/cole advanced books & software. **Pacific California**, 1984.

BUBLITZ, C. F. et al. Unsupervised segmentation and classification of snoring events for mobile health. In: **IEEE Global Communications Conference (GLOBECOM)**. Singapore: IEEE, 2017. p. 1–6.

ÇAVUŞOĞLU, M.; POETS, C. F.; URSCHITZ, M. S. Acoustics of snoring and automatic snore sound detection in children. **Physiological measurement**, IOP Publishing, v. 38, n. 11, p. 1919, 2017.

COLT, H. G.; HAAS, H.; RICH, G. B. Hypoxemia vs sleep fragmentation as cause of excessive daytime sleepiness in obstructive sleep apnea. **Chest**, Elsevier, v. 100, n. 6, p. 1542–1548, 1991.

DAFNA, E.; TARASIUK, A.; ZIGEL, Y. Automatic detection of whole night snoring events using non-contact microphone. **PloS one**, Public Library of Science, v. 8, n. 12, p. e84139, 2013.

DUCKITT, W.; TUOMI, S.; NIESLER, T. Automatic detection, segmentation and assessment of snoring from ambient acoustic data. **Physiological measurement**, IOP Publishing, v. 27, n. 10, p. 1047, 2006.

EMOTO, T. et al. High frequency region of the snore spectra carry important information on the disease of sleep apnoea. **Journal of medical engineering & technology**, Taylor & Francis, v. 35, n. 8, p. 425–431, 2011.

FLEMONS, W. W. et al. Access to diagnosis and treatment of patients with suspected sleep apnea. **American journal of respiratory and critical care medicine**, Am Thoracic Soc, v. 169, n. 6, p. 668–672, 2004.

HALEVI, M. et al. Can we discriminate between apnea and hypopnea using audio signals? In: **IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)**. USA: IEEE, 2016. p. 3211–3214.

HAO, T.; XING, G.; ZHOU, G. iSleep: Unobtrusive sleep quality monitoring using smartphones. In: **11th ACM Conference on Embedded Networked Sensor Systems**. USA: ACM, 2013. p. 4.

HAYKIN, S.; VEEN, B. V. **Signals and Systems**. USA: John Wiley & Sons, 2007.

HUANG, X. et al. **Spoken language processing: A guide to theory, algorithm, and system development**. USA: Prentice hall PTR Upper Saddle River, 2001.

JOLLIFFE, I. **Principal component analysis**. USA: Wiley Online Library, 2002.

JR, P. J. S.; ROGERS, R. M. Obstructive sleep apnea. **New England Journal of Medicine**, Mass Medical Soc, v. 334, n. 2, p. 99–104, 1996.

KIM, T.; KIM, J.-W.; LEE, K. Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques. **Biomedical engineering online**, BioMed Central, v. 17, n. 1, p. 16, 2018.

LEVARTOVSKY, A. et al. Breathing and snoring sound characteristics during sleep in adults. **Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine**, American Academy of Sleep Medicine, v. 12, n. 3, p. 375, 2016.

MA, G. et al. Unsupervised snore detection from respiratory sound signals. In: **IEEE International Conference on Digital Signal Processing (DSP)**. Singapore: IEEE, 2015. p. 417–421.

MAIMON, N.; HANLY, P. J. Does snoring intensity correlate with the severity of obstructive sleep apnea? **Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine**, American Academy of Sleep Medicine, v. 6, n. 5, p. 475, 2010.

MOON, T. K. The expectation-maximization algorithm. **IEEE Signal processing magazine**, v. 13, n. 6, p. 47–60, 1996.

OU, Q. et al. Continuous positive airway pressure treatment reduces mortality in elderly patients with moderate to severe obstructive severe sleep apnea: a cohort study. **PLoS One**, Public Library of Science, v. 10, n. 6, p. 10, 2015.

PANOSSIAN, L. A.; AVIDAN, A. Y. Review of sleep disorders. **Medical Clinics of North America**, Elsevier, v. 93, n. 2, p. 407–425, 2009.

SATEIA, M. International classification of sleep disorders: highlights and modifications. **CHEST Journal**, American College of Chest Physicians, v. 146, n. 5, p. 1387–1394, 2014.

SCALART, P. et al. Speech enhancement based on a priori signal to noise estimation. **IEEE International Conference on Acoustics, Speech, and Signal Processing**, v. 2, p. 629–632, 1996.

SHIN, H.; CHO, J. Unconstrained snoring detection using a smartphone during ordinary sleep. **Biomed Eng Online**, v. 13, n. 1, p. 116, 2014.

SPICUZZA, L.; CARUSO, D.; MARIA, G. D. Obstructive sleep apnoea syndrome and its management. **Therapeutic advances in chronic disease**, SAGE Publications Sage UK: London, England, v. 6, n. 5, p. 273–285, 2015.

TUOMILEHTO, H. P. et al. Lifestyle intervention with weight reduction: first-line treatment in mild obstructive sleep apnea. **American journal of respiratory and critical care medicine**, Am Thoracic Soc, v. 179, n. 4, p. 320–327, 2009.

VASEGHI, S. V. **Advanced digital signal processing and noise reduction**. [S.l.]: John Wiley & Sons, 2008.

WANG, C. et al. Automatic snoring sounds detection from sleep sounds via multi-features analysis. **Australasian physical & engineering sciences in medicine**, Springer, v. 40, n. 1, p. 127–135, 2017.

WIEGAND, L.; ZWILLICH, C. W. Obstructive sleep apnea. **Disease-a-Month**, Elsevier, v. 40, n. 4, p. 202–252, 1994.

YOUNG, T.; PEPPARD, P. E.; GOTTLIEB, D. J. Epidemiology of obstructive sleep apnea: a population health perspective. **American journal of respiratory and critical care medicine**, Am Thoracic Soc, v. 165, n. 9, p. 1217–1239, 2002.

YOUNG, T.; SKATRUD, J.; PEPPARD, P. E. Risk factors for obstructive sleep apnea in adults. **Jama**, American Medical Association, v. 291, n. 16, p. 2013–2016, 2004.

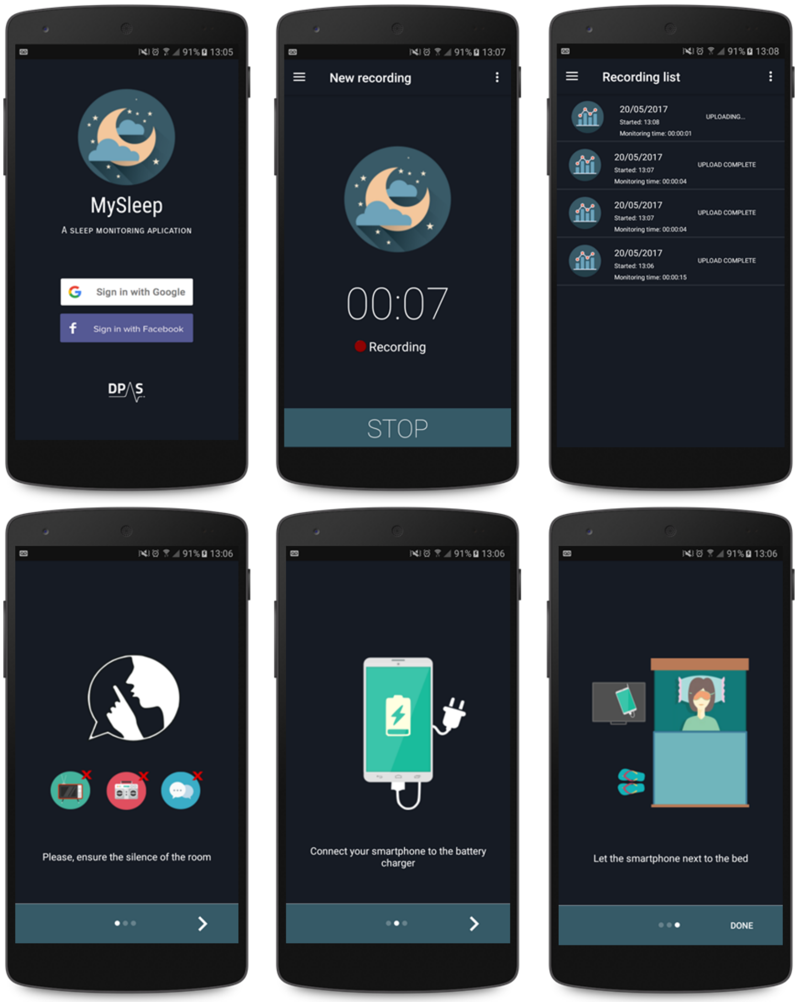
Appendices

AppendixA MYSLEEP: A SLEEP MONITORING APPLICATION

The MySleep application was developed to perform the sleep recording of the volunteers. The audio captured by the smartphone's built-in microphone is automatically transferred to a remote server for future analysis of the snoring events. For example, 6 hours of sleep recording takes approximately 2GB of storage. Storing this file in the internal memory of the smartphone is impractical since the user should have at least 2 GB free. Therefore, the application fragments the recording into small 10-minute files and sends them to the server, where the audio files of the whole night will be merged. The user can finish the recording in the morning or after 8 hours of recording the application automatically finish the recording to prevent the user from forgetting to finalize.

Figure A.1 shows the main screens of the application. The user initially registers using the email account or social network. When a new recording start, the application shows a three-step sequence: *(i)* ensure the silence of the room, *(ii)* connect your smartphone to the battery charger, and *(iii)* the smartphone next to the bed. These steps aim to improve the quality of the captured signal. Finally, it is possible to see the list of recordings that were performed and the status (uploading/upload complete). This application has not been compared to any other sleep applications available for Android devices. The available applications are not comparable because they do not classify the patient regarding the need for CPAP / Non-CPAP therapy.

Figure A.1: Screens of the Android application developed



Source: by author (2018).

AppendixB ACOUSTIC FEATURES

In Table B.1 we can see the set of features extracted from the snoring events and sleep night for each patient. In the sequence, each feature is explained in detail.

Table B.1: Global features for the patient's sleep night

Time Domain Features	
Count	**Feature**
1	Number of Snores Events
2	Snoring Time Ratio
3	Variance of Snoring Time Duration
4	Number of Inter Snore Event Silence (>10 and <60 sec)
5	Mean of Inter Snore Events Silence
6-7	Running Energy/Distance Snores Variance
8-9	Variance and Mean of Relative Energy Prior the Detect Event
10-11	Variance and Mean of Relative Energy Posterior the Detect Event
12-13	Variance and Mean of Rhythm Intensity (12 sec)
14-15	Variance and Mean of Rhythm Period (12 sec)
16-17	Variance and Mean of Rhythm Period (6 sec)
18-19	Variance and Mean of Ratio Relative Energy Prior and Posterior
20-21	Variance and Mean of Normalized Area Beneath Energy Envelop
22-23	Variance and Mean of Skewness Envelop Formation
24-25	Variance and Mean of 10 Seconds Before and After Period
26-27	Variance and Mean of Ratio Areas Before and After the Energy Peak
28-29	Variance and Mean of Total Snore Event Energy
Frequency Domain Features	
Count	**Feature**
30	Retro-palatal Ratio
31	Snore Frequency Intercalation
32-33	Variance and Mean of First Formant
34-35	Variance and Mean of Fundamental Frequency
36-43	Variance and Mean of 4 Moments of MFCC Coeficients
44-51	Variance and Mean of 4 Moments of LPC Coeficients
52-59	Variance and Mean of 4 Moments of DFT
60-75	Variance and Mean of [#1 to #8] Subband Frequency
76-77	Variance and Mean of Spectral Flux
78-79	Variance and Mean of Pitch Density

Source: by author (2018).

Number of Snores Events: Number of snoring events detected during the segmentation process.

Snoring Time Ratio: Snoring Time Ratio measure how long the patient remained to snore with the total recording time.

Variance of Snoring Time Duration: Variance of snoring time duration across all detected events during the night.

Number of Inter Snore Event Silence (>10 and < 60): Number of silences detected between two snoring events during the whole night. The silence is considered to be an interval of >10 and <60 seconds.

Mean of Inter Snore Events Silence: Mean of the silence time duration between the snoring events across the night.

Running Energy and Distance Snore Variance: Running energy and distance variance is defined as the within group snore variance. For each patient, the variance across the whole night is calculated, according to the steps below:

1. To quantify inter-snore variability across the night, the total energy of each snore was calculated, as well as the time distance between the events;
2. All the snores were clustered into groups, according to distance from the closest snore in the group: in cases > 1 minute duration between the group and the snore, the snore was ascribed to the next group;
3. The variance of groups energy and distance is calculated.

Variance and Mean of Relative Energy Prior and Posterior the Detected Event: The relative energy prior and posterior the detected event involves calculating the area of the energy signal 10 seconds before and after the event and divide by total event area. This process results in counting the number of similar events before or after the event being tested and is represented by the equations below:

$$RelativeEnergyPrior = \frac{\sum_{m=m_i-10F_r}^{m_i} e_m}{\sum_{m=m_i}^{m_f} e_m} \tag{B.1}$$

$$RelativeEnergyPosterior = \frac{\sum_{m=m_f}^{m_f+10F_r} e_m}{\sum_{m=m_i}^{m_f} e_m} \tag{B.2}$$

where e_m is the signal energy in frame m, m_i, m_f are the initial and final energy frames of the event, respectively, and F_r is the frame rate (e.g. if the energy is calculated in frames of 15 ms, F_r is $1/15$ samples/ms). The variance and mean of relative energy prior and posterior is calculated across the whole snoring events detected during the night.

Variance and Mean of Rhythm Intensity (12 sec): Measure the rhythm intensity detected in a 12-second interval, respectively, when the tested event is in the middle. The period was calculated via autocorrelation over the energy signal associated with the tested interval. The intensity was calculated from the area between the zero-lag autocorrelation and the first peak). The Rhythm Intensity is calculated according to the steps below:

1. Calculate the normalized energy of the 12-second interval;
2. Autocorrelation of the signal interval;
3. Find the first peak of the estimated autocorrelation function;
4. Calculate the period intensity feature (the area between the zero-lag autocorrelation and the first peak).

The variance and mean of the Rhythm Intensity is calculated across the whole snoring events detected during the night.

Variance and Mean of Rhythm Period (6 sec): Measure the period of the rhythm detected in a 6-second interval when the tested event is in the middle. The period was calculated via autocorrelation over the energy signal associated with the tested interval. The Rhythm Period is calculated according to the steps below:

1. Calculate the normalized energy of the 6-second interval;
2. Autocorrelation of the signal interval;
3. Find the first peak of the estimated autocorrelation function;
4. Measure the distance between the zero-lag and the first peak, which represents the Rhythm Period.

The variance and mean of the Rhythm Period is calculated across the whole snoring events detected during the night.

Variance and Mean of Normalized Area Beneath Energy Envelop: Calculate the area beneath the energy envelop of the detected event. The area is calculated according to the steps below:

1. The area using numerical integration via the trapezoidal method;
2. Calculate the window (rectangle area that contains the event);
3. Divide the area by the window.

The variance and mean of the normalized area is calculated across the whole snoring events detected during the night.

Variance and Mean of Skewness Envelop Formation: Calculate the variance and mean of skewness (3rd moment) of the envelop formation from the all snoring events during the night.

Variance and Mean of 10 Seconds Before and After Period Ratio: Calculate the variance and mean of the ratio between the period calculated over 10 seconds before and after the detected event.

1. Normalize the energy signal;
2. Get energy intervals of 10 seconds before and after the event interval;
3. Calculate the autocorrelation of the intervals;
4. Find the first peak (period) in the autocorrelation function for both intervals (before and after);
5. Divide the period calculated in the prior 10 seconds by posterior 10 seconds period.

Variance and Mean of Ratio Areas Before and After Energy Peak: Calculate the variance and mean of the ratio between the areas located prior and after to the maximum energy peak location of the snoring event. The ratio of areas is calculated according to the steps below:

1. Find the maximum energy peak of the detected event;
2. Calculate the area before and after the maximum peak location using numerical integration via the trapezoidal method;
3. Divide the calculated area before by the area after the peak.

Variance and Mean of Total Snore Event Energy: Calculate the variance and mean of the total event energy for all snoring detected.

Retro-Palatal Ratio: Calculate the ratio between the counting of snoring events with the first frequency in the retro-palatal frequency band (> 100 Hz and < 1500 Hz) and the total snoring events counting. The ratio is calculated according to the steps below:

1. Find the first formant frequency for each event detected;
2. Count the number of events in the retro-palatal frequency band;
3. Divide the number of retro-palatal snore event by the total snore events.

Snore Frequency Intercalation: Counting the number of times that the snoring first formant have ranged from the [< 500 Hz] to [> 500 Hz and < 1800 Hz] range.

Variance and Mean of First Frequency Formant: Estimate the first formant frequency and magnitude from the detected event using Linear Predictive Coding (LPC). The variance and mean of the first frequency formant is calculated across the whole snoring events detected during the night.

Variance and Mean of Fundamental Frequency: Calculate the variance and mean of fundamental frequency for each snoring event detected during the night.

Variance and Mean of 4 Moments of MFCC Coefficients: Calculate the variance and mean of the 4 moments (mean, variance, skewness and kurtosis) from the 20 MFCC extracted from the snoring events detected.

Variance and Mean of 4 Moments of LPC Coefficients: Calculate the variance and mean of 4 moments (mean, variance, skewness and kurtosis) from the LPC coefficients from the snoring events detected.

Variance and Mean of 4 Moments of DFT: Calculate the variance and mean of 4 moments (mean, variance, skewness and kurtosis) from the DFT coefficients from the snoring events detected.

Variance and Mean of [#1 to #8] Subband Frequency: Calculate the variance and mean for each 8 sub-bands of the DFT-128 content of the detected snoring event. The sub-bands is calculated according equation below:

$$SubbandFrequencies(i) = \frac{\sum_{k=8i}^{8i+7} |DFT_k|}{\sum_{k=0}^{127} |DFT_k|} \tag{B.3}$$

where DFT_k denotes a 128-coefficient DFT of the snoring event signal, and $i = 1..8$ is the sub-band frequencies.

Variance and Mean of Spectral Flux: Spectral flux is a measure of how quickly the power spectrum of a signal is changing, calculated by comparing the power spectrum for one frame against the power spectrum from the previous frame. It is calculated by the standard deviation of the squared differences of the frame sequence of 40 ms of the event, according to equation below:

$$SpectralFlux(i) = std\left(\sum_{j=1}^{N} [Norm(abs(DFT_{j+1})) - Norm(abs(DFT_j))]^2\right) \tag{B.4}$$

where DFT_j represents the DFT at the frame index j, and N is the total number of frames of the event. The variance and mean of the spectral flux is calculated across the whole snoring events detected during the night.

Variance and Mean of Pitch Density: The pitch density for each snore was calculated as the fraction of the snoring time where the pitch is detectable ($peak(R_{ii}) > 0.5$) over the total snoring time, according the equation below:

$$PitchDensity(i) = \frac{\left(\sum_i^{N_s} bool\left\{peak(R_{ii}) > 0.5\right\}\right)}{N_s} \tag{B.5}$$

where R_{ii} is the autocorrelation function of i^{th} frame and N_s is the number of frames in the s^{th} snore. The variance and mean of the pitch density is calculated across the whole snoring events detected during the night.

AppendixC PUBLISHED PAPER – GLOBECOM 2017

Mobile Health broadens the accessibility to healthcare applications using mobile devices, such as smartphones. These devices can be used for tracking the patient health signals for remote monitoring and treatment follow up. Sleep disorders affect a substantial part of the population and may have serious comorbidities. To improve monitoring of evolution of sleep disorders and/or treatment follow up, we propose an unsupervised segmentation and classification of snoring events (as snore or non-snore) on a Mobile Health context. We apply a statistical analysis through the Expectation-Maximization algorithm applied to Gaussian Mixture Models to cluster the data sets. We evaluate our solution using recordings from simple and Obstructive Sleep Apnea (OSA) snorers. The results show that our proposal is able to classify snoring events achieving average accuracies of 91.3% for simple snorers and 79.7% for OSA snorers.

- **Title –**
  Unsupervised Segmentation and Classification of Snoring Events for Mobile Health
- **Conference –**
  IEEE Global Communications Conference (GLOBECOM 2017)
- **Type –**
  Main track (full-paper)
- **Qualis –**
  A1
- **Date –**
  Dec 04-08, 2017
- **Held at –**
  Singapore
- **Digital Object Identifier (DOI) –**
  <10.1109/GLOCOM.2017.8255031>

Unsupervised Segmentation and Classification of Snoring Events for Mobile Health

Carlos F. Bublitz[1], Ana C. Ribeiro-Teixeira[2], Thatiane A. Pianoschi[2], Juergen Rochol[1], Cristiano B. Both[2]
[1]Federal University of Rio Grande do Sul, Porto Alegre, Brazil
[2]Federal University of Health Sciences of Porto Alegre, Porto Alegre, Brazil
Email: [1]{cfbublitz, juergen}@inf.ufrgs.br,
[2]{anacarolinart, thatiane, cbboth}@ufcspa.edu.br

Abstract—**Mobile Health broadens the accessibility to healthcare applications using mobile devices, such as smartphones. These devices can be used for tracking the patient health signals for remote monitoring and treatment follow up. Sleep disorders affect a substantial part of the population and may have serious comorbidities. To improve monitoring of evolution of sleep disorders and/or treatment follow up, we propose an unsupervised segmentation and classification of snoring events (as snore or non-snore) on a Mobile Health context. We apply a statistical analysis through the Expectation-Maximization algorithm applied to Gaussian Mixture Models to cluster the data sets. We evaluate our solution using recordings from simple and Obstructive Sleep Apnea (OSA) snorers. The results show that our proposal is able to classify snoring events achieving average accuracies of 91.3% for simple snorers and 79.7% for OSA snorers.**

Index Terms—**Mobile Health, Snoring events, EM, Clustering.**

I. Introduction

The latest generations of smartphones have provided more powerful computing capability, larger storage capacity, multiple embedded sensors, and operating systems that encourage the development of applications [1]. Mobile Health (mHealth) takes advantage of these smartphone capabilities to deploy healthcare applications to a wide population. mHealth patients can benefit from more expedite diagnosis and continuous treatment monitoring. The preventive approach of mHealth systems promotes cost reduction for governments and healthcare companies and contributes to patient's wellness. Moreover, smartphone sensors are being used for a broad range of healthcare applications. As an important instance, one may mention the utilization of the built-in microphone for snoring events detection [2].

Sleep has a vital role in good health maintenance and wellbeing throughout an individual's life [3]. Insufficient sleep has severe mental and physiological consequences [4]. Sleep disorders may cause or exacerbate preexisting psychiatric and medical conditions, which can be associated with high rates of depression, anxiety, and impaired daytime functioning, as well as an increased risk of developing high blood pressure, strokes, heart diseases and obesity [5]. The current gold standard for diagnosing sleep disorders is the overnight Polysomnography (PSG) exam [3]. The patient must sleep in a laboratory attached to different sensors under the supervision of a technician. Notwithstanding its efficiency, the necessity of a clinical setting and highly specialized infrastructure results in a long waiting list in sleep laboratories and high costs, thus restricting the access to diagnosis and treatment [6]. For this reason, in most cases, PSG is performed at most once to every other patient, and treatment follow-up is only performed through reports by patients and/or their partners [4].

The study of snoring events can bring valuable information about the sleep quality. It can provide relevant information to the diagnostics of sleep-related respiratory disorders such as Obstructive Sleep Apnea (OSA) [7]. Moreover, snoring events can be detected through the processing and analysis of the audio recorded during one night of sleep [1]. Some studies tried to implement such solution using a high-performance microphone to capture the sound [8]-[9], but only a few research works investigated the detection of snoring events on smartphone platforms [7]. Smartphones are nowadays widely available devices anytime and anywhere. Hence, we argue that a smartphone can lend itself as an auxiliary tool in the accessibility problem to PSG, since it allows for a preliminary monitoring which can aid the screening of patients according to the need for laboratory sleep exams. Moreover, it allows for continuous monitoring of sleep treatments and sleep disorders evolution.

The state-of-the-art on detection of snoring events on smartphones addresses a supervised machine learning process to classify the sound events as snore/non-snore [1][2]. The concern about supervised learning is the limited amount of training data available that cannot well represent the snoring events distribution in the overall population and also the high susceptibility to overfitting [10]. To overcome this limitation, some studies focused on unsupervised learning as a way of creating a model with a greater capacity of data generalization [11][12]. However, these studies have not applied the proposed solutions to the mHealth context.

In this paper, we propose an unsupervised segmentation and classification of snoring events using a clustering algorithm to discriminate sound events according to two classes: snore and non-snore. We consider an audio data set collected in an uncontrolled environment setting for mHealth. We introduce a statistical distribution analysis using the Expectation-Maximization (EM) algorithm for Gaussian Mixture Models (GMM) to cluster the data set according to the probability of each data point of belonging to some normal distribution component. The results obtained show that the EM algorithm produces better results on clustering the data, as compared to other clustering algorithms such as Fuzzy C-Means (FCM) and K Harmonic Means (KHM), reaching satisfactory accuracy rates (91.3% for simple snores and 79.7% for OSA snores, on average). Our main contributions are: (*i*) the proposal of an unsupervised solution for the mHealth context, (*ii*) the application of a statistical distribution analysis (EM algorithm) to the clustering of sound events, (*iii*) the design of an accurate event segmentation method to identify the snore candidate events boundaries, and (*iv*) the design of an automatic cluster labeling according to snore and non-snore.

The remainder of the paper is organized as follows. In Section 2, a brief overview of related work is presented. In Section 3, we describe and explain our proposal solution. In Section 4, results of our solution are presented as well as an evaluation of its implementation. Finally, final remarks and perspectives for future works are presented in Section 5.

II. Related Work

The analysis of acoustic signals aimed at snoring events detection has been performed in different studies. Most of the initiatives performed the recordings in a controlled environment with low levels of noise and high-performance microphones to capture the sound [13][14]. These studies have proven the viability of extracting acoustic features from audio recordings for the segmentation and classification of snoring events. However, the imposed experimental constraints limit the reproducibility in uncontrolled environments, which does not contribute to a solution to the PSG accessibility problem. In order to overcome this limitation, the smartphone started to be seen as an alternative tool to capture the acoustic signal at a patient's home.

The approach of snoring events detection with smartphones as a recording and processing tool has been addressed by a few studies [1][2]. An Android application was developed by Hao *et al.* [2] for audio recording and detection of three different events: snoring, body movement, and cough. The classification of the events was based on a decision-tree algorithm. Although their proposed solution is simple and inexpensive in terms of processing, the main goal of this study is to assess sleep quality through events counting, and the work does not aim at an acoustic analysis of snoring events for further study. Shin *et al.* [1] collected the acoustic signal using a smartphone and performed a formant analysis on the signal within different frequency bands. The authors were able to select the best features for snore classification and applied a quadratic classifier over this set of features for snoring events detection. Even though these studies have obtained good results on snoring detection, all smartphone applications implement supervised learning approaches, which require a considerable data generalization capacity and are prone to overfitting. Furthermore, supervised approaches hinder the automatic character of the solution, which is a necessity when one considers accessibility to a wide population.

Unsupervised learning is an exaction for the automation of snoring events detection solutions. In the literature in the area, only a few works address unsupervised classification, none of which in the context of smartphone platforms [11][12]. These works apply clustering algorithms such as FCM and KHM, which use as similarity criterion the distance between events in feature space. We verified that this criterion significantly undermines clustering accuracy. We argue that an appropriate algorithm to the problem at hand should aim at snoring events identification through statistical inference approaches, since snoring events from a person are produced by a single human vocal tract, and as such may be considered as independent identically distributed (*i.i.d.*) events, which ideally will be normally distributed.

III. Unsupervised Segmentation and Classification of Snoring Events

The unsupervised segmentation and classification of snore events proposed in this paper involve a sequence of steps as can be seen in Figure 1. We discuss each stage of the overall process in the following subsections.

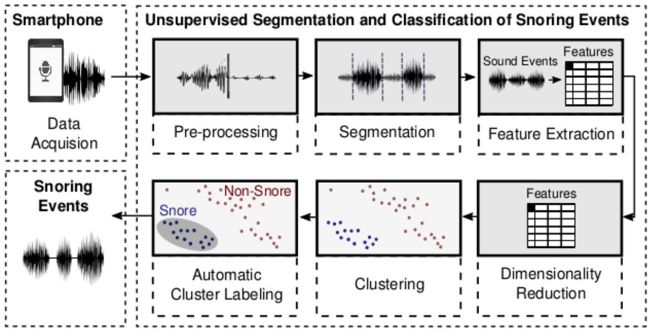


Figure 1: Block diagram of the proposed solution

A. Data Acquision

The acoustic signal is recorded using the built-in microphone of a smartphone and transmitted to the cloud to be processed. Our proposal considers a generic smartphone device, recording in an uncontrolled environment, and an untrained user. These settings may hinder the acoustic quality of the signal but are necessary to achieve an accessible and continuous monitoring of the sleep quality, which contributes to the dissemination of sleep self-assessment and patient empowerment. The following steps are cloud processed.

B. Pre-processing

Noise can significantly degrade the acoustic signal under uncontrolled environments. The background noise is the most common factor degrading the quality and sharpness of the recordings. To reduce the effects of noise whilst preserving the signal, a noise reduction algorithm needs to be applied. In our proposal we chose a Wiener filter, a spectral filter which makes use of a noise template based on a running estimation of the background noise energy [15].

C. Segmentation

The sound events that occur during the night need to be automatically identified and segmented for future analysis. All detected events are considered as snore candidates. Segmentation proceeds by identifying a snore candidate as an event fulfilling two criteria: (*i*) its energy locally exceeds a certain threshold estimated over the whole night and (*ii*) the event time duration lies within Δt_{min} and Δt_{max}.

A segmentation procedure involves four steps [13]. We designed a segmentation algorithm detailed in the following:

- **Threshold definition:** The energy threshold is computed from the standard deviation of the whole night signal energy. Considering events distribution to be sparse in the whole night signal, we may define the energy threshold as 100 standard deviations. Due to memory constraints, this step is carried on by splitting the signal into sections of length T.
- **Surpass threshold interval:** The estimated threshold is applied to each T long section. The intervals within which the energy signal exceeds the threshold are saved as a candidate event. See Figure 2 (a).

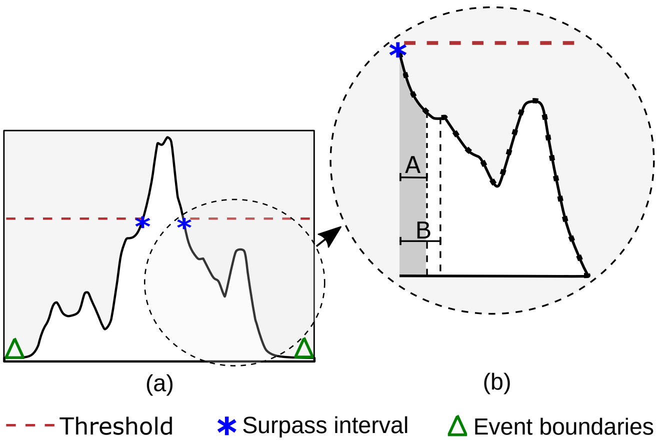


Figure 2: Schematic view of the segmentation stage

- **Candidate event segmentation:** Having identified possible candidate events, its exact time boundaries, *i.e.* its beginning and ending times, must be determined. To this end, the area of the shadowed A region, which lies to the right (or left) of the event surpass interval, is computed (see Figure 2 (b)). If relative energy contribution of A region is greater than 0.1%, event ending (or beginning) time is updated accordingly. A region is expanded to B region, iteratively, until the relative energy increment is no longer significant (below 0.1%). Candidate event boundaries (green triangles in Figure 2) are properly determined with this algorithm.
- **Fragmentation and duration test [13]:** If two segmented candidate events are close to each other in time, *i.e.* the ending of the first and beginning of the second are separated by less than δ (of the order of 10^2 ms), the actual event may have been fragmented. In this case, the two candidate events are merged into a single one. Moreover, candidate events duration must be checked to lie within reasonable snoring durations. The time duration of the candidate events is verified to lie within these boundaries, *i.e.*, within Δt_{min} and Δt_{max} (orders of magnitude 10^2 ms and 10^3 ms, respectively), otherwise, the event is discarded.

For each whole night recording, the segmentation stage returns N candidate events, with the respective boundaries for each. Each candidate event is defined by its boundaries, *i.e.*, its initial and final times.

D. Feature Extraction

A set of features needs to be extracted from each snore candidate. These features can be extracted from time and frequency domains to individually characterize candidate events and are subsequently used to classify candidates as snore or non-snore events during the clustering process. We have generated a set of $m = 75$ features, based on the pool of features proposed by Dafna *et al.* [13]. These set of features are presented in Table I and further analyzed in Section IV. This set of features is specially designed to cover several acoustic characteristics from within events (denoted intra-event) and comparative between events (denoted inter-event).

For each snore candidate an $N \times m$ array M is created. Each column i of M contains a column vector of values $(f_i^{(1)}, f_i^{(2)}, \ldots, f_i^{(N)})$, where $f_i^{(k)}$ corresponds to f_i feature value for the kth candidate. Each extracted feature spreads over a particular interval $[\min_k\{f_i^{(k)}\}, \max_k\{f_i^{(k)}\}]$ and a normalization should be applied, which rescales the range of all features to lie within $[0, 1]$. This normalization is essential to the variance analysis to be carried on in the next stage.

E. Dimensionality Reduction

A dimensionality reduction technique can be applied to extract the most important information from the set of features. From the total set of 75 features, we have selected 10 most important features, with a Mutual Information method for a feature selection. We have chosen Principal Component Analysis (PCA) to compute new variables denoted principal components which are obtained as linear combinations of the 10 original features [16]. Therefore, the m dimensional set of features extracted can be reduced to only two dimensions, suitable for the clustering process. The first principal component is the linear combination of features having the greatest variance in the data set, which means this component explains the largest part of the data behavior. The second component has the next highest variance and is subject to the condition that it is uncorrelated, *i.e.* orthogonal, with the first principal component. Two N vectors are returned in this stage, $p_i = (p_i^{(1)}, \ldots, p_i^{(k)}, \ldots, p_i^{(N)})$, with $i \doteq 1, 2$, which contain values for the first ($i = 1$) and second ($i = 2$) principal component of each candidate event.

F. Clustering

The snoring candidate events must be split into snore and non-snore groups using a clustering algorithm. Candidate

Table I: Features extracted from each event

Count	Feature
Time Domain Features	
1	Relative energy prior to detected event
2	Relative energy posterior to detected event
3	Rhythm intensity (+- 12 sec)
4	Rhythm period (+- 12 sec)
5	Rhythm period (+- 6 sec)
6	Ratio of relative energy prior and posterior to detected event
7	Normalized area beneath energy envelop
8	Skewness of envelop formation
9	Ratio of areas before and after the peak
10	Total event energy
11	10 seconds before and after period ratio
Frequency Domain Features	
12	First formant frequency
13	First formant magnitude
14-33	20 Mel-Frequency Cepstrum Coefficients (MFCC)
34-53	20 Linear Prediction Coefficients (LPC)
54-57	4 moments of MFCC coefficients
58-61	4 moments of LPC coefficients
62-69	8 subband-frequency distribution
70	Spectral flux
71-74	4 moments of frequency distribution (DFT)
75	Pitch density

event k is defined in the $p_1 - p_2$ plane of principal components as a pair of coordinates $(p_1^{(k)}, p_2^{(k)})$. We propose an EM algorithm for GMM applied to the clustering problem. EM algorithm applied to GMM [17] tries to estimate a set of parameters $\theta_j = (\omega_j, \mu_j, \Sigma_j)$ where $j \doteq 1, 2$ which maximize the log-likelihood function for the set of candidate events:

$$L^{(i)} = \frac{1}{N} \sum_{k=1}^{N} \log \left[\sum_{j=1}^{2} \omega_j^{(i)} \, \phi \left(\mathbf{p}^{(k)} \middle| \mu_j^{(i)}, \Sigma_j^{(i)} \right) \right], \quad (1)$$

where $L^{(i)}$ represents the log-likelihood at the i-th iteration, $\phi(\mathbf{x}|\mu, \Sigma)$ is the Gaussian distribution with average $\mu = (\mu_1, \mu_2)$ and covariance matrix $\Sigma_{2\times2}$. The parameters $\{\omega_j\}$ represent the weights given to each of the two Gaussians and $\mathbf{p}^{(k)} = (p_1^{(k)}, p_2^{(k)})$. The maximization is achieved by iteratively updating the set of parameters θ_j until $L^{(i)}$ saturates. The identity of candidate events as snore or non-snore is assessed in a probabilistic manner through the computation of *membership weights*, which are associated to the probability that a given event is generated from Gaussian component 1 or 2. EM iteration is ensured to never decrease the log-likelihood function.

G. Automatic Cluster Labeling

After constructing the clusters from the data set, each cluster needs to be automatically labeled as snore or non-snore. The most significant distinction between a snore and non-snore cluster is the internal cluster cohesion. The data set on snore clusters tends to be more compact as compared to non-snore clusters. The algorithm automatically labels as snore the most compact between the two clusters. This labeling is unsupervised, which is essential for a sleep monitoring solution to be able to attain a wide population, demanding a fully automatic mHealth application.

IV. Evaluation

In this Section, we discuss and evaluate our unsupervised snoring events detection solution for mHealth. We first describe how the proposal solution was implemented and in the sequence, we present and discuss the results obtained.

A. Methodology

An Android application was specially designed to acquire and record the acoustic signal from the whole night (6 hours approximately), and to transmit it to the cloud for further processing. The acoustic signal is sampled at a frequency of 44.1 kHz and a bit depth of 16 bits. The participants were oriented to install the application on their own smartphone and place it at a distance of approximately 50 cm beside the bed before sleep, on a night stand cleared for other objects. A trained technician was responsible for listening to all segmented candidate events and manually classify each one as snore or non-snore for the accuracy evaluation of the proposed solution.

Once the signal has been stored in the cloud, we proceed to stages III-B-III-G for signal processing and event classification. We describe below relevant information and parametrization for some of the stages.

- *Pre-processing:* The signal is divided into sections of τ = 10s long and for each one, a frame energy vector is calculated. The 5 lowest energy frames are elected as noise standards and used to compute the spectral average noise of the section. This average is estimative of the signal noise template and is updated along the night recording [13]. This template is then used as input to the Wiener filter [15].
- *Segmentation:* The standard deviation is calculated over the total sum of energies for each T = 10 min section and the energy threshold is established. For each T long section, an energy vector is calculated using 60 ms frames with 75% of overlap. The resulting signal is then tested for threshold surpassing. Having determined the candidate event boundaries, fragmentation and duration tests must be further applied. We chose δ = 200 ms, Δt_{min} = 600 ms and Δt_{max} = 3500 ms.
- *Dimensionality Reduction:* The 10 features selected through the Mutual Information method are, in order of importance: 6, 1, 29, 61, 65, 2, 36, 63, 70, 7 (count numbers are according to Table I). Using Mutual Information

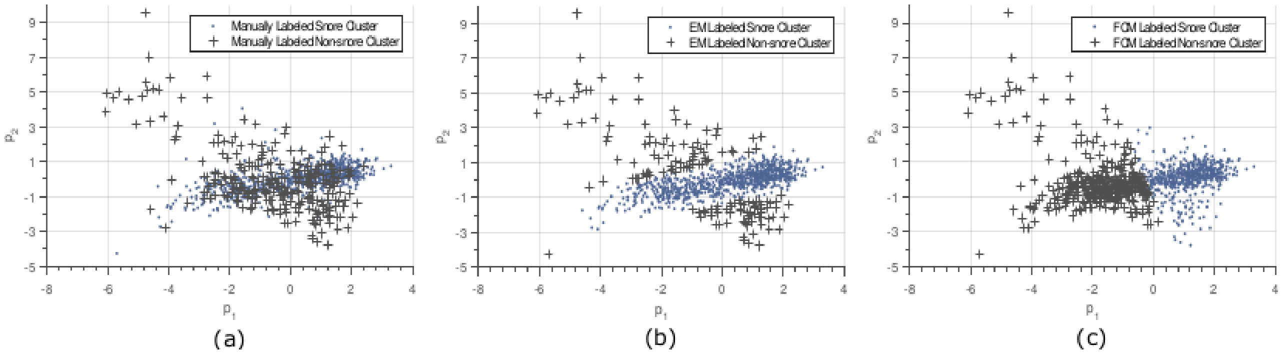


Figure 3: Clustering results: comparison between manually labeled, EM labeled and FCM labeled snore and non-snore clusters for one night recording

for feature selection involves feeding the algorithm with the manually labeled events classification for a set of events to measure how much information the presence or absence of a feature contributes to making the correct classification. The 10 features selected in this manner will be applied to all recordings alike. The 10 features are fed to PCA which will then generate a pair of principal components particular to each recording.

B. Results

We applied our solution to 6 night recordings, totalizing 32.5 hours. The segmentation stage returned a total of 5323 snore candidate events, of which 3574 were manually labeled as snoring. Table II depicts the actual (manually labeled) numbers of snoring events for simple and OSA snorers, and compares the accuracies averaged over nights obtained for an implementation of our solution (EM) and an implementation of FCM, discriminating simple and OSA snorers. For both simple and OSA snorers, one concludes that EM average accuracies are well above FCM corresponding accuracies (over one - either EM or FCM - standard deviations).

Table II: Algorithms comparison for simple and OSA snorers

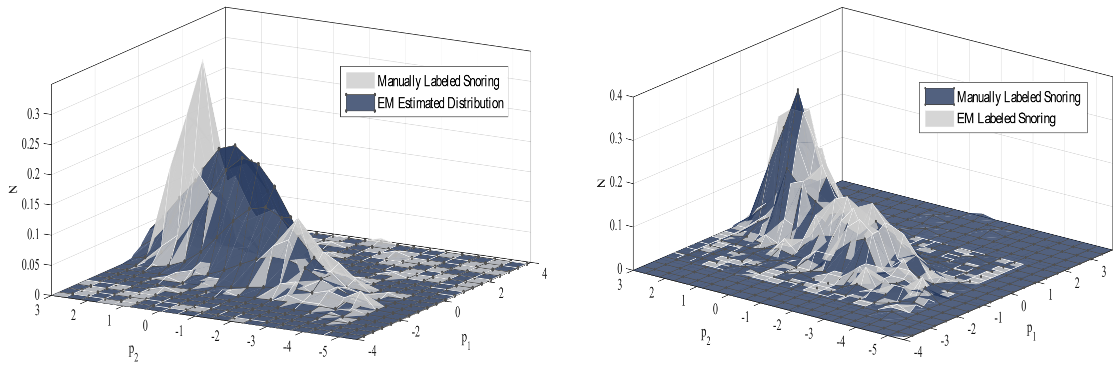
	Number of Snore Events	**Average Accuracy EM (%)**	**Average Accuracy FCM (%)**
Simple Snorers	2670	91.3 ± 1.0	81.6 ± 8.8
OSA Snorers	904	79.7 ± 0.9	67.7 ± 4.5

The explanation for the poor performance of FCM is illustrated in Figure 3. We compare the clusters of snore and non-snore as defined by manual labeling (panel (a)), cluster results by our EM solution (panel (b)), and cluster results by an FCM implementation (panel (c)) of the same data set for one night recording of an OSA snorer. We see from panel (a) that the two manually labeled clusters are superimposed in $p_1 - p_2$ plane which, by the PCA analysis, is ensured to span the best principal components for data segregation. This superposition is a typical behavior for overnight sound recordings.

We demonstrate in panel (b) that EM clustering is able to estimate the true snoring distribution to an adequate accuracy, which for this specific recording is 81.0%. This is due to the statistical approach of EM applied to GMM, which assumes a Gaussian distribution of snore and non-snore events, as described in Section III-F. Although we have no reason to expect non-snore events to be normally distributed, the assumption of normal distributed snores is a reasonable one, as explained in Section II. Meanwhile, FCM implementation, which clusters events together considering its distances to two iteratively defined centroids, is not able to approximate the true snore/non-snore clusters. In fact, in the specific recording considered in Figure 3, the accuracy is 61.3%. We expect all other clustering algorithms based on geometric criteria such as FCM, KHM, etc, to present similarly poor performances, since all of them consider point proximity as similarity criterion.

Figure 4 aims to demonstrate that the Gaussian approximation to the snore distribution is an adequate one. The data set considered is the same as in the previous plots. In Figure 4, panel (a), we compare manually labeled snoring cluster histogram to the same EM estimated Gaussian distribution. Panel (b) depicts histograms of the manually labeled and EM labeled snoring clusters.

The ideal behavior of the EM implementation occurs when the Gaussian distributions estimated by the EM algorithm well approximate both the snore and non-snore manually labeled histograms. The non-snore manually labeled histogram will usually not approximate a Gaussian distribution, because events are not identically distributed, since they may be as diverse as a car passing by, a cough, or an object falling on the ground. However, even if non-snore events are not normally distributed, for simple snorers, the manually labeled snore histogram approximates well a Gaussian distribution (with like mean and appropriately defined variance). In such a case, EM is expected to deliver a high accuracy, as is verified from EM average accuracy for simple snorers and its standard deviation



(a) (b)

Figure 4: Clustering results: comparison between EM estimated distribution for snoring events, EM histogram for the snoring events cluster, and manually labeled snoring events histogram

(Table II). Below we show that even when OSA snorers are considered, which represents a much more complex situation, EM is able to perform adequately.

The night recording of an OSA snorer is typically composed of apneic and non-apneic snores [7]. As a result, the manually labeled snoring events histogram is heavy tailed and almost bimodal (see Figure 4, panel (a)). Although both averages $\langle p_1 \rangle$ and $\langle p_2 \rangle$ are very close for the EM Gaussian distribution and manually labeled snores histogram, we see from panel (a) that the histogram is not well approximated by the EM estimated distribution. Still, we demonstrate in panel (b) that the estimated EM histogram fits almost perfectly the manually labeled snore events histogram, as we expect from an adequate clustering. This is quite remarkable and is only possible because EM is able to cluster events together by (indirectly) recognizing its "statistical similarity", *i.e.* the probability of two events being drawn from the same given distribution.

V. Conclusion

In this paper, we proposed an unsupervised segmentation and classification of snoring events for the mHealth context. The design highlighted a sequence of stages, which process the acoustic signal for events segmentation and clustering into two classes: snore and non-snore. The proposal was evaluated with recordings collected from simple and OSA snorers, reaching satisfactory accuracy results. The EM algorithm was able to well approximate the snoring events distribution, performing significantly better than the FCM algorithm. The results demonstrated that our proposal is a viable solution for the detection of snoring events in uncontrolled environments.

For future works, we intend to expand the number of recordings to investigate new features and behaviors of the OSA snore events. Statistical methods should be applied aiming to establish relations between the distributions of event features and sleep disorders.

References

[1] H. Shin and J. Cho, "Unconstrained snoring detection using a smartphone during ordinary sleep," *Biomed Eng Online*, vol. 13, no. 1, p. 116, 2014.

[2] T. Hao, G. Xing, and G. Zhou, "iSleep: Unobtrusive sleep quality monitoring using smartphones," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, 2013, p. 4.

[3] M. Sateia, "International classification of sleep disorders: highlights and modifications," *CHEST Journal*, vol. 146, no. 5, pp. 1387–1394, 2014.

[4] L. A. Panossian and A. Y. Avidan, "Review of sleep disorders," *Medical Clinics of North America*, vol. 93, no. 2, pp. 407–425, 2009.

[5] L. Wang, A. Cai, J. Zhang, Q. Zhong, R. Wang, J. Chen, and Y. Zhou, "Association of obstructive sleep apnea plus hypertension and prevalent cardiovascular diseases: A cross-sectional study," *Medicine*, vol. 95, no. 39, p. e4691, 2016.

[6] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, "Access to diagnosis and treatment of patients with suspected sleep apnea," *American journal of respiratory and critical care medicine*, vol. 169, no. 6, pp. 668–672, 2004.

[7] M. Halevi, E. Dafna, A. Tarasiuk, and Y. Zigel, "Can we discriminate between apnea and hypopnea using audio signals?" in *IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 3211–3214.

[8] K. Qian, Z. Xu, H. Xu, Y. Wu, and Z. Zhao, "Automatic detection, segmentation and classification of snore related signals from overnight audio recording," *IET Signal Processing*, vol. 9, no. 1, pp. 21–29, 2015.

[9] M. Cavusoglu, M. Kamasak, O. Erogul, T. Ciloglu, Y. Serinagaoglu, and T. Akcam, "An efficient method for snore/nonsnore classification of sleep sounds," *Physiological measurement*, vol. 28, no. 8, p. 841, 2007.

[10] T. Dietterich, "Overfitting and undercomputing in machine learning," *ACM computing surveys (CSUR)*, vol. 27, no. 3, pp. 326–327, 1995.

[11] A. Azarbarzin and Z. M. Moussavi, "Automatic and unsupervised snore sound extraction from respiratory sound signals," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1156–1162, 2011.

[12] G. Ma, B. Xue, H. Hong, X. Zhu, and Z. Wang, "Unsupervised snore detection from respiratory sound signals," in *IEEE International Conference on Digital Signal Processing (DSP)*, 2015, pp. 417–421.

[13] E. Dafna, A. Tarasiuk, and Y. Zigel, "Automatic detection of whole night snoring events using non-contact microphone," *PloS one*, vol. 8, no. 12, p. e84139, 2013.

[14] W. Duckitt, S. Tuomi, and T. Niesler, "Automatic detection, segmentation and assessment of snoring from ambient acoustic data," *Physiological measurement*, vol. 27, no. 10, p. 1047, 2006.

[15] P. Scalart *et al.*, "Speech enhancement based on a priori signal to noise estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 629–632, 1996.

[16] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[17] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal processing magazine*, vol. 13, no. 6, pp. 47–60, 1996.