Applied Acoustics 161 (2020) 107188

Contents lists available at ScienceDirect

Applied Acoustics

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

A constrained tonal semi-supervised non-negative matrix factorization to classify presence/absence of wheezing in respiratory sounds

J. Torre-Cruz *, F. Canadas-Quesada, S. García-Galán, N. Ruiz-Reyes, P. Vera-Candeas, J. Carabias-Orti

Department of Telecommunication Engineering, University of Jaen, Campus Cientifico-Tecnologico de Linares, Avda. de la Universidad, s/n, 23700 Linares, Jaen, Spain

ARTICLE INFO

Article history:
Received 31 July 2019
Received in revised form 8 November 2019
Accepted 15 December 2019

Keywords:
Non-negative matrix factorization (NMF)
Divergence
Wheezing
Smoothness
Monophonic constraint
Spectral trajectories

ABSTRACT

From a clinical point of view, the detection of wheezing presence in respiratory sounds is a challenging task for early identification of pulmonary diseases since wheezing is the main manifestation associated to airway obstruction. In this article, we propose a novel method to detect the presence or absence of wheeze sounds in breath recordings in order to increase the reliability of the subjective diagnosis provided by the physician in the auscultation process. Specifically, it is assumed an unhealthy subject when wheeze sounds can be detected during breathing. The proposed method consists of three stages. The first stage attempts to estimate the spectral interval, band of interest (BOI), that shows the highest probability to find wheeze sounds. In the second stage, a constrained tonal semi-supervised non-negative matrix factorization (NMF) approach is applied to obtain spectral patterns that models the periodic or tonal nature typically shown by wheeze sounds. The third stage analyzes the estimated wheezing spectrogram based on the smoothness of the spectral trajectories from the most significant energy previously factorized in the BOI. Our system has been evaluated and compared to other state-of-the-art methods, yielding competitive results in the wheezing presence detection in respiratory sounds.

1. Introduction

To this day, auscultation is still the main technique used in most of the health centres in order to obtain the first clinical diagnosis of the status of the respiratory system of the subjects. From a clinical point of view, this technique shows several advantages since it is non-invasive, low-cost, easy-to-perform and subject-friendly that avoids long-time subjects monitoring [1]. The more important auscultation diagnosis becomes in low-income countries, e.g. Sub-Saharan Africa, due to weak health centres in which there are often no more sophisticated and expensive diagnostic tools such as chest radiographs. However, a correct diagnosis derived from auscultation depends largely on the experience and acoustic training of the physician to interpret what is hearing because a conventional stethoscope tends to attenuate or amplify sounds within the spectral range in which wheeze sounds appear [2,3]. For this reason, many research efforts are applied in biomedical signal processing to develop a reliable method for the early wheezing occurrence detection since it is associated with respiratory obstruction observed in different pulmonary diseases such as asthma, chronic obstructive pulmonary disease (COPD) or even pneumonia, causing narrowing and spasms in the small airways of lungs [4,5]. World Health Organization (WHO) estimated 383,000 deaths due to asthma and 1.4 million deaths of children under five years old worldwide due to pneumonia during 2015 [6,7].

The sounds emitted during breathing are generally classified into normal breath sounds and adventitious sounds (AS) such as wheezing, crackles, pleural rubs and stridor. Normal breath sounds (RS), emitted by healthy lungs, are characterized by a wideband spectrum where most of the energy is concentrated in the frequency band 60 Hz–1600 Hz [8]. Wheeze sounds (WS), emitted by unhealthy lungs, are defined as continuous sinusoidal waveforms in time characterized by a dominant fundamental frequency (pitch). As a result, WS show a set of narrowband spectral peaks forming frequency lines over time (spectral trajectories) that superimpose on RS in the frequency domain but there is still no agreement on a single frequency and duration of wheezes in the literature. This fact increases the rate of misdiagnosis because the pitch of a wheeze can vary which implies that it can be located in another range of the spectrum [9]. According to American Thoracic Society (ATS), WS are defined as a pitch higher than 400 Hz whose duration is longer than 250 ms [10]. According to Computerized Respiratory Sound Analysis (CORSA) guidelines, WS are defined as a pitch higher than 100 Hz with duration longer than

* Corresponding author.
E-mail address: jtorre@ujaen.es (J. Torre-Cruz).

https://doi.org/10.1016/j.apacoust.2019.107188
0003-682X/

100 ms [11] as shown in the time-frequency representation of Fig. 1A. This paper assumes the following facts: i) the term mixture refers to any single-channel input signal which can be composed only of RS (healthy subject) or RS mixed with WS (unhealthy subject); ii) the term RS indicates the no presence of any type of AS (see Fig. 1B), as a result, RS is equivalent to non-wheeze sounds; iii) the term WS indicates the presence of WS (see Fig. 1A); iv) the minimum temporal duration for WS has been selected as 100 ms using the CORSA standard [2]; and finally, v) the pitch range associated to wheezing is not fixed a priori since it will be estimated from each mixture.

Several algorithms to detect the wheezing occurrence in respiratory sounds have been published in the last decades using various approaches: Autoregressive (AR) model [12], Auditory modelling [13], Entropy [14], Linear analysis [15], Autocorrelation [16], Neural networks (NN) [5,17], Wavelet transform [4], Instantaneous frequency (IF) analysis [18], Tonal index [19], Mel-frequency cepstral coefficients (MFCC) [20,21], Gaussian Mixture Models (GMM) [22,23], Image processing [24], Spectral peaks identification [25,26], Hidden Markov model (HMM) [27], Empirical mode decomposition (EMD) [3,28]and Non-negative matrix factorization (NMF) [29].

In [4], it is reported a modeling for wheezing and normal breath sounds in the wavelet packet domain using generalized gaussian distributions. Kochetov et al. [17] proposed wheeze detection using convolutional neural networks. Taplidou and Hadjileontiadis [25] located wheezing based on a smoothing procedure that estimates the trend of the frequency content at each time using mean filtering. Wisniewski and Zielinski [19] combined feature extraction from MPEG-7 and MPEG-2 standards using support vector machine (SVM). To discriminate wheeze and non-wheeze sounds, Mazic et al. [20] developed a two-layer pattern recognition using two SVM classifiers designed as a cascade stacked in parallel using MFCC features and thresholding. In [21], it applied a feature extraction based on MFCC and K-nearest neighbour (KNN) to classify different levels of asthma severity shown by wheeze sounds. Mendes et al. [26] identified wheezing sounds using a temporal Gaussian regularization and a reduction of the false positives based on the morphological opening by reconstruction operator. Lozano et al. [28] detected not only wheezing but any continuous adventitious sounds (CAS) using EMD, features extracted from IF sequences and SVM. Recently, our previous study [29] proposed a two-stage cascade system for wheezing detection. In the first stage, a constrained NMF is designed integrating smoothness and sparseness into the decomposition. In the second stage, the Kullback–Leibler divergence is applied to discriminate between wheeze and non-wheeze sounds frames.

The aim of the proposed method is to automatically detect the presence of wheezing sounds in breath recordings avoiding the subject's return to the health centre with a worsening of the airway obstruction, manifested by wheeze sounds, that was not detected early in the first clinical examination performed by auscultation. Three novel contributions are proposed by authors. Unlike most wheezing detection algorithms in which the pitch range of the wheeze is set a priori, our first contribution, band of interest (BOI), estimates the spectral range in which the probability to find wheeze sounds is maximum. As a second contribution, we present a constrained tonal semi-supervised NMF that factorizes into spectral patterns that correctly model the tonal nature, by narrowband peaks, shown by WS in the estimated BOI in order to separate WS and RS. From the separated wheezing spectrogram, we propose an intuitive method to classify the subject's condition as healthy (absence of wheezing) or unhealthy (presence of wheezing) analyzing the temporal smoothness of the spectral trajectories defined by most significant energy factorized in the estimated BOI.

The rest of this paper is structured as follows. Section 1.1 briefly reviews non-negative matrix factorization (NMF). The proposed method is described in Section 2. Section 3 details and discusses the experimental evaluation. Finally, Section 4 concludes the paper.

1.1. Non-negative matrix factorization

Non-negative matrix factorization (NMF) [30] is a decomposition technique that has attracted a lot of efforts in different fields of biomedical signal processing in the last years [31,32]because NMF is a powerful tool to learn additive part-based spectrograms of the most representative sources by imposing non-negative constraint. The aim of NMF can be defined as follows: given a single-channel mixture $x(t)$ with magnitude spectrogram $\mathbf{X} \in \mathbb{R}_{+}^{F \times T}$ with non-negative entries, NMF approximates $\mathbf{X}$ into the product of two non-negative matrices, basis matrix $\mathbf{B} \in \mathbb{R}_{+}^{F \times K}$ and activation matrix $\mathbf{A} \in \mathbb{R}_{+}^{K \times T}$ as shown in Eq. (1),

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{BA} \tag{1}$$

obtaining the estimated spectrogram $\hat{\mathbf{X}}$, the number of frequency bins F, the number of time frames T and the number of components K (K is usually chosen such that $K(F+T) \ll FT$ in order to reduce the data dimension). Therefore, $\mathbf{B}$ can be interpreted as a dictionary or a set of spectral patterns that codes the frequency information associated to the sources active in the spectrogram and $\mathbf{A}$ as a matrix of activations that indicates the activity of each spectral basis in a given time frame.

The NMF factorization is generally calculated by minimizing a cost function $D\left(\mathbf{X}|\hat{\mathbf{X}}\right)$ that penalizes the error reconstruction between $\mathbf{X}$ and $\hat{\mathbf{X}}$. Considering the most used cost functions such

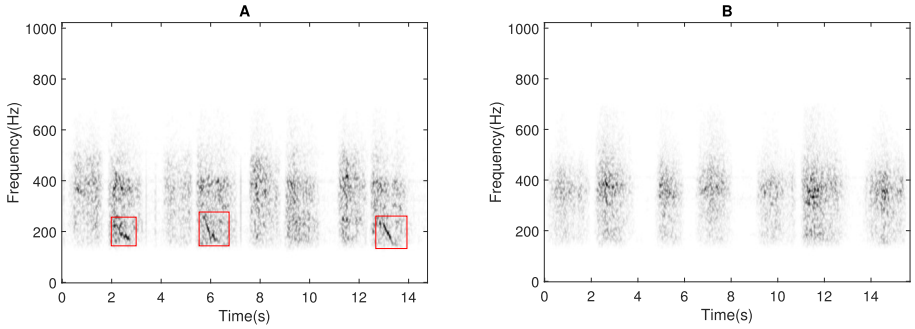

Fig. 1. Time-Frequency representation (spectrogram) of breathing recording. A) Spectrogram of a mixture (unhealthy subject) composed of three wheeze sounds (red rectangles) and normal breath sounds. B) Spectrogram of a mixture (healthy subject) composed only of normal breath sounds. A darker grey colour represents higher energy of each frequency.

as Euclidean distance, the generalized Kullback-Leibler divergence, the Itakura-Saito divergence and the Cauchy distribution [33,34], in this paper, we have used the generalized Kullback-Liebler divergence $D_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}\right)$ because it is non-increasing and ensures the non-negativity of **B** and **A** using multiplicative update rules [30] and moreover, recent previous works [29,31]have obtained competitive results in biomedical signal processing,

$$D_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}\right)=\sum_{f,t}\left(X_{f,t}\log\frac{X_{f,t}}{\widehat{X}_{f,t}}-X_{f,t}+\widehat{X}_{f,t}\right) \tag{2}$$

The main disadvantage of the NMF factorization is its inability to reconstruct each isolated source because NMF only ensures the reconstruction of the whole mixture but it does not guarantee parts-based objects with physical interpretation as occurs in real world [35]. In order to find an optimal NMF factorization, some of the most widely used solutions in audio processing to add physical meaning to the NMF bases and activations are the following: i) adding prior information of the sources into the NMF decomposition using additional constraints [36] and ii) train or characterize spectral patterns of target sounds to be separated [37].

2. Materials and methods

Our proposal attempts to classify the subject's condition, that is, healthy or unhealthy. We assume as a healthy subject that subject in which the absence of wheezing sounds has been detected. We assume as an unhealthy subject that subject in which the presence of wheezing sounds has been detected. For this purpose, the proposed method (see Fig. 2) is composed of three stages: Estimation of the Spectral Band Of Interest (Stage I), Wheezing/Normal breath Sound Separation (Stage II) and Wheezing presence/absence classification (Stage III).

2.1. Mixing model and time-frequency representation

As previously mentioned in Section 1, the term mixture $x(t)$ refers to any single-channel input signal: i) $x(t)=r(t)$ only composed of RS signal $r(t)$, that is, a healthy subject; or ii) $x(t)=r(t)+s(t)$ composed of RS signal $r(t)$ mixed in an approximately linear manner with WS signal $s(t)$, that is, an unhealthy subject.

The magnitude spectrogram **X** of a mixture $x(t)$ is composed of T frames, F frequency bins and a set of time–frequency units $X_{f,t}$, being $f=1,\ldots,F$ and $t=1,\ldots,T$. Each unit $X_{f,t}$ is defined by the f^{th} frequency bin at the t^{th} frame and is calculated from the magnitude of the Short-Time Fourier Transform (STFT) using a Hamming windows of N samples with 25% overlap. A normalization is applied to **X** in order to be independent of the size and scale where the normalized magnitude spectrogram $\overline{\mathbf{X}}$ is computed as follows,

$$\overline{\mathbf{X}}=\frac{\mathbf{X}}{\left(\frac{\sum_{f,t}X_{f,t}}{FT}\right)} \tag{3}$$

To avoid complex nomenclature throughout the paper, the variable **X** is hereinafter referred to the normalized magnitude spectrogram previously computed in Eq. (3).

2.2. Stage I: estimation of the spectral band of interest

The goal of this stage is to estimate the spectral interval, defined as band of interest (BOI), in which the probability to find wheeze sounds is maximum. The estimation of BOI is performed by modelling the tonal nature shown by wheeze sounds in time–frequency domain (e.g., it can be observed that BOI is delimited between 150–300 Hz in the magnitude spectrogram of an unhealthy subject shown in Fig. 1A).

We propose a method combining NMF and the periodicity principle to estimate the BOI assuming that a wheeze sound exhibits a strongly periodic or tonal nature characterized by the presence of narrowband spectral peaks. In order to estimate BOI, this stage can be divided into three steps that are detailed below.

2.2.1. Step I: factorization of the basis matrix applying NMF

The first step attempts to estimate the basis **B** and activations **A** matrix minimizing the reconstruction error defined by the cost function $D_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}\right)$ (see Eq. (2)). Finally, the matrices **B** and **A** are obtained by applying a gradient descent algorithm based on multiplicative update rules [33].

2.2.2. Step II: clustering of bases based on the tonality principle

Here, the goal is to cluster the NMF bases **B** obtained in *Step I* into two groups: (i) NMF bases that show higher periodicity $\mathbf{B}_W$ (see Fig. 3A depicting bases that model the tonal nature typically shown by WS); and (ii) NMF bases that show lower periodicity $\mathbf{B}_R$ (see Fig. 3B representing bases that model the non-tonal nature shown by RS). The higher periodicity tends to concentrate the energy of the NMF bases in narrowband spectral peaks (see Fig. 3C) while the lower periodicity distributes such energy in wideband spectrum (see Fig. 3D). Therefore, we assume that tonality or periodicity implies a spectrum that shows a sparse distribution of energy.

Although a preliminary study related to several sparse descriptors [38] has been analyzed in this work, empirical results indicated that the sparse descriptor Gini index β provided the best classification results of the NMF bases in terms of the degree of periodicity. Besides, these results are supported by recent works [38,39]that have demonstrated the reliability and robustness shown by β to correctly cluster between sparse and non-sparse

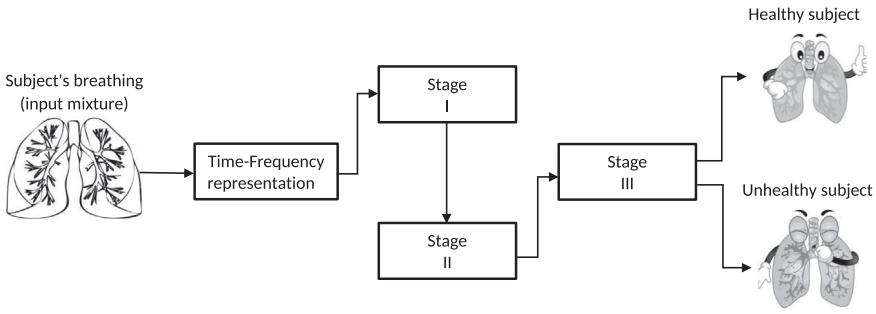

Fig. 2. Block-scheme of the proposed method.

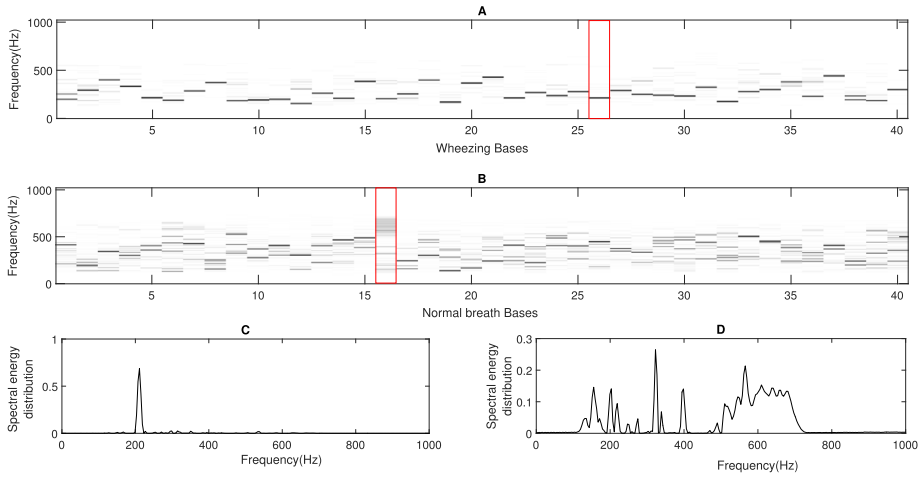

Fig. 3. Example of the NMF bases ($K = 80$ bases) classification between WS and RS considering the magnitude spectrogram belonging to the unhealthy subject shown in Fig. 1A. A) The matrix $\mathbf{B}_W$ is composed of the forty NMF bases with highest spectral tonality (narrowband spectral peaks). B) The matrix $\mathbf{B}_R$ is composed of the forty NMF bases with lowest spectral tonality behavior (wideband). C) The spectral energy distribution (sparse) associated to the most tonal NMF basis, specifically, $\mathbf{B}_W(26)$ (red rectangle displayed in Fig. 3A). D) The spectral energy distribution (non-sparse) of the least tonal NMF basis, specifically, $\mathbf{B}_R(16)$ (red rectangle displayed in Fig. 3B).

distributions. Therefore, we propose to apply the Gini index $\beta(k)$ to calculate the degree of tonality or periodicity for each k^{th} NMF basis of the dictionary $\mathbf{B}$. High values of $\beta(k)$ imply high periodic nature and vice versa. In addition, a remarkable advantage of $\beta(k)$ is that it is a normalized descriptor whose values are ranged between 0 and 1 for each NMF basis.

According to each value of $\beta(k)$, $\mathbf{B}_W$ and $\mathbf{B}_R$ are obtained applying a thresholding process. The aim of this thresholding is not to achieve the optimal separation between WS and RS since it will be obtained in Section 2.3 as shown in Fig. 6. Specifically, this thresholding attempts to locate the BOI identifying two sets of NMF bases, $\mathbf{B}_W$ and $\mathbf{B}_R$, that show different periodic behavior. For this purpose, we propose to apply a threshold ζ_m equal to the median of the sparse values provided by the sparse descriptor $\beta(k)$ for the following two reasons: (i) the median is both a robust threshold against the dispersion of periodicity values and resistant to gross periodicity error by avoiding the problem generated by outliers [40,41]; and (ii) the median achieves to cluster uniformly between two groups, providing half of the NMF bases for both groups $\mathbf{B}_W$ and $\mathbf{B}_R$ (see Eq. (4)). This allows to factorize a reliable set of NMF bases $\mathbf{B}_W$ to locate the BOI in which the probability to find WS is maximum due to the factorization of NMF bases with higher periodicity.

$$\mathbf{B}(k) \rightarrow \begin{cases} \mathbf{B}(k) \in \mathbf{B}_W & \text{if } \beta(k) \geqslant \zeta_m \\ \mathbf{B}(k) \in \mathbf{B}_R & \text{if } \beta(k) < \zeta_m \end{cases} \tag{4}$$

where $k = 1,\ldots,K$.

Next, the estimated wheezing spectrogram $\hat{\mathbf{X}}_W$ used to locate the BOI can be reconstructed as follows,

$$\hat{\mathbf{X}}_W = \mathbf{B}_W \mathbf{A}_W \tag{5}$$

where the wheezing activations matrix $\mathbf{A}_W$ has been factorized using the same set of components clustered in $\mathbf{B}_W$.

2.2.3. Step III: using energy distortion to select the spectral BOI

From the estimated wheezing spectrogram $\hat{\mathbf{X}}_W$, the third step selects the spectral range BOI in which the probability of finding wheeze sounds is maximum. As a result, BOI shows the least spectral energy loss when comparing $\hat{\mathbf{X}}_W$ and $\mathbf{X}$ since $\hat{\mathbf{X}}_W$ roughly models most of the wheeze sounds from the input spectrogram $\mathbf{X}$. To calculate BOI, we propose to use the information provided by the spectral energy distortion $\delta(f)$ (see Eq. (6)) measuring the degree of similarity $\upsilon(f)$ for each frequency bin along all frames between $\mathbf{X}$ and $\hat{\mathbf{X}}_W$ as shown in Eq. (7). A high value of $\upsilon(f_i)$ indicates that the frequency f_i exhibits a high tonal content as typically shown by WS because f_i shows low attenuation in the spectral energy distortion. A low value of $\upsilon(f_i)$ indicates that the frequency f_i exhibits a low tonal content as shown by RS because f_i shows high attenuation in the spectral energy distortion.

$$\delta(f) = \sqrt{\sum_{t=1}^{T} X_{f,t} - \sum_{t=1}^{T} \hat{X}_{W_{f,t}}}, \quad f = 1 \ldots F \tag{6}$$

$$\upsilon(f) = \frac{\sum_{t=1}^{T} \hat{X}_{W_{f,t}}}{\delta(f)}, \quad f = 1 \ldots F \tag{7}$$

Following a conservative strategy to ensure that all wheeze sounds are contained in the BOI, we propose to find the most prominent peak located in $\upsilon(f)$. Note that the prominence of a peak is defined as the minimum vertical distance that the signal must descend on either side of the peak before either climbing back to a level higher than the peak or reaching an endpoint [42]. As shown in Eq. (8), the boundaries (f_{min}, f_{max}) of the BOI can be calculated in terms of the width Δ of the mainlobe related to the most prominent peak previously mentioned. A preliminary analysis indicated that the use of an additional margin $\eta = 50$ Hz from the mainlobe of the most prominent peak achieved the best performance to contain most of the wheeze sounds determined by the estimated BOI.

$$BOI = [f_{min} - f_{max}] \quad f_{min} = min(\Delta) - \eta \quad f_{max} = max(\Delta) + \eta \tag{8}$$

where f_{min} and f_{max} represent the minimum and maximum frequency that defines the spectral interval of the BOI. Fig. 4 shows an example of the procedure described in this step when the magnitude spectrogram corresponds to the unhealthy subject shown in Fig. 1A.

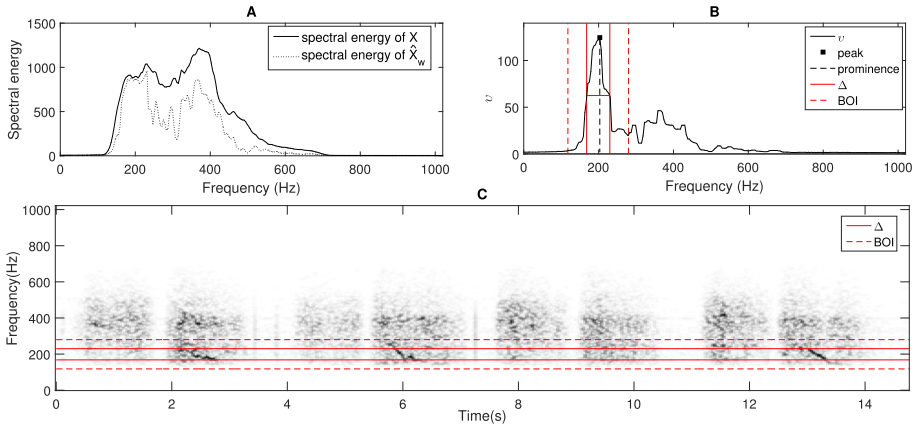

Fig. 4. BOI estimation when the magnitude spectrogram corresponds to the unhealthy subject shown in Fig. 1A. A) Spectral energy of $\mathbf{X}$ and $\hat{\mathbf{X}}_W$. B) Degree of similarity $\upsilon(f)$ and location of the most prominent peak in terms of Δ. C) Magnitude spectrogram shown in Fig. 1A including the frequency range limited by Δ and BOI.

2.3. Stage II: wheezing/normal breath sound separation

The main contribution of this stage is the development of a Wheezing/Normal breath sound separation based on a constrained tonal semi-supervised NMF approach using redundant information from the BOI obtained in the previous stage I. To that end, an objective function is defined to decompose a mixture spectrogram $\mathbf{X}$ into two separated or estimated spectrograms, $\hat{\mathbf{X}}_{R_I}$ (only normal breath sounds spectrogram) and $\hat{\mathbf{X}}_{W_I}$ (only wheeze sounds spectrogram). The factorization model can be defined as follows,

$$\mathbf{X} \approx \hat{\mathbf{X}}_I = \hat{\mathbf{X}}_{R_I} + \hat{\mathbf{X}}_{W_I} = \mathbf{B}_{\mathbf{R_I}} diag(D_R)\mathbf{A}_{\mathbf{R_I}} + \mathbf{B}_{\mathbf{W_I}} diag(D_W)\mathbf{A}_{\mathbf{W_I}} \tag{9}$$

where $\mathbf{B}_{\mathbf{R_I}}, \mathbf{B}_{\mathbf{W_I}}, \mathbf{A}_{\mathbf{R_I}}$ and $\mathbf{A}_{\mathbf{W_I}}$ are the estimated basis and activation matrices of the normal breath sounds and the wheeze sounds and $\hat{\mathbf{X}}_I$ is the estimated magnitude spectrogram estimated at this stage. All of these matrices are non-negative matrices. The number of wheezing and normal breath sounds components is the same and is denoted by Q. The L^2-norm of each column of $\mathbf{B}_{\mathbf{R_I}}$ or $\mathbf{B}_{\mathbf{W_I}}$ is equal to 1.0. The terms D_R and D_W represent vectors with the L^2-norm of each activation component of wheezing and normal breathing, respectively. Therefore, the L^2-norm of each row of $\mathbf{A}_{\mathbf{R_I}}$ or $\mathbf{A}_{\mathbf{W_I}}$ be equal to 1.0 due to the normalization procedure at each iteration. The $diag()$ operator is the diagonal matrix.

To refine and improve the estimated wheezing spectrogram $\hat{\mathbf{X}}_W$ of the stage I, this stage attempts to model the spectral periodic or tonal behavior typically shown by wheeze sounds. We propose to use a constrained tonal semi-supervised NMF approach described below.

2.3.1. A tonal semi-supervised NMF

The tonal semi-supervised NMF is defined as a semi-supervised NMF in which the target source (wheezing) is learned in advanced modelling the spectral tonal or periodic behavior typically shown by WS. The tonal or periodic behavior is modelled using the matrix $\mathbf{B}_{\mathbf{W_I}}$ that is composed of a set of simulated pitches (narrowband peaks) that model all possible tonal sounds active in the estimated BOI. As a result, each q^{th} basis of the dictionary $\mathbf{B}_{\mathbf{W_I}}$ represents a single pitch located at a specific frequency f_q within the BOI. In order to cover all the pitches that can be found in the BOI, each frequency f_q is separated from each other by a value given by the frequency resolution $\tau = 4$ Hz used throughout the model as follows,

$$\mathbf{B}_{\mathbf{W_I}}(q) = G(f - f_q), \quad f_q \in [f_{min} : \tau : f_{max}] \tag{10}$$

where $G(f)$ is the STFT of a sinusoidal signal multiplied by a Hamming window of N samples, f_q is the fundamental frequency of the pitch for the q^{th} NMF basis related to the dictionary $\mathbf{B}_{\mathbf{W_I}}$ and the number of components $Q = \left\lceil \frac{f_{max} - f_{min}}{\tau} \right\rceil$ is obtained in terms of the spectral resolution τ and the BOI (see Fig. 5).

2.3.2. Adding monophonic constraint

We propose to incorporate temporal monophonic constraint $\Omega(\mathbf{A}_{\mathbf{W_I}})$ [43] to the wheezing activation matrix $\mathbf{A}_{\mathbf{W_I}}$ in order to minimize the number of wheezing bases $\mathbf{B}_{\mathbf{W_I}}$ that can be simultaneously activated at each frame. This constraint forces to factorize a more reliable wheezing spectrogram as in typically found in the real-world. Specifically, this constraint $\Omega(\mathbf{A}_{\mathbf{W_I}})$ is computed taking into account the cross-correlations $\mathbf{A}_{\mathbf{W_I}}\mathbf{A}_{\mathbf{W_I}}^T$ detailed in Eq. (11),

$$\Omega(\mathbf{A}_{\mathbf{W_I}}) = \frac{1}{Q(Q-1)} \sum_{q=1}^{Q} \sum_{t=1}^{T} \mathbf{A}_{\mathbf{W_I}}\mathbf{A}_{\mathbf{W_I}}^T - Trace\left(\mathbf{A}_{\mathbf{W_I}}\mathbf{A}_{\mathbf{W_I}}^T\right) \tag{11}$$

where the *Trace* operator computes the sum of diagonal elements of a square matrix. To equilibrate the importance of the monophonic constraint in the decomposition process, $\Omega(\mathbf{A}_{\mathbf{W_I}})$ is weighted by a factor of $\frac{1}{Q(Q-1)}$. Consequently, the cost $\Omega(\mathbf{A}_{\mathbf{W_I}})$ will be equal to 1.0 in the worst case.

The global objective function $D\left(\mathbf{X}|\hat{\mathbf{X}}_I\right)$ that must be minimized taking into account the signal reconstruction, based on the Kullback-Leibler divergence, and the temporal monophonic constraint is detailed as follows,

$$D\left(\mathbf{X}|\hat{\mathbf{X}}_I\right) = D_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_I\right) + \lambda\Omega(\mathbf{A}_{\mathbf{W_I}}) \tag{12}$$

where λ defines the weight of the temporal monophonic constraint $\Omega(\mathbf{A}_{\mathbf{W_I}})$ applied to only estimated wheezing activations matrix $\mathbf{A}_{\mathbf{W_I}}$.

The estimated wheezing activation matrix $\mathbf{A}_{\mathbf{W_I}}$ (see Eq. (13)), the estimated normal breath basis matrix $\mathbf{B}_{\mathbf{R_I}}$ (see Eq. (14)) and the estimated normal breath activation matrix $\mathbf{A}_{\mathbf{R_I}}$ (see Eq. (15)) can be obtained by applying a gradient descent algorithm based on multiplicative update rules. The estimated wheezing basis matrix $\mathbf{B}_{\mathbf{W_I}}$ is fixed and not updated because it has been modeled based on the tonal nature of WS.

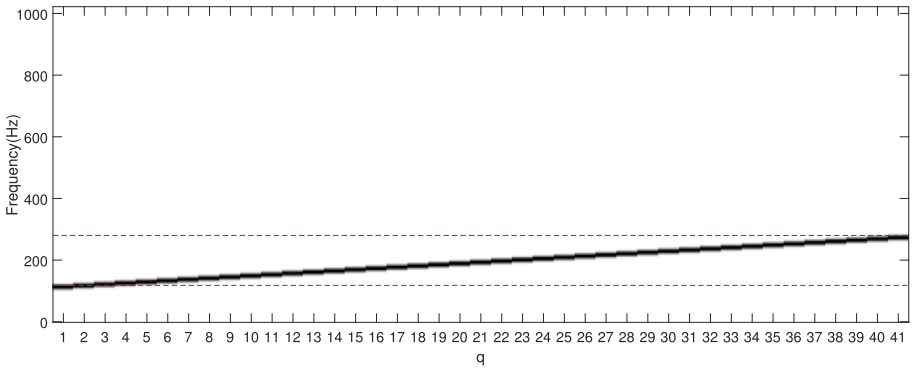

Fig. 5. Dictionary $\mathbf{B}_{\mathbf{W}_I}$ established for the BOI shown in Fig. 4C. The red dotted lines represent the range of the BOI. In this case, the matrix $\mathbf{B}_{\mathbf{W}_I}$ is composed by $Q = 41$ wheezing bases because the BOI, previously obtained in the stage I (Fig. 4C), is spectrally located between $f_{min} = 118$ Hz and $f_{max} = 280$ Hz and the spectral resolution $\tau = 4$ Hz.

$$\mathbf{A}_{\mathbf{W}_I} \leftarrow \mathbf{A}_{\mathbf{W}_I} \odot \frac{\left(\mathbf{B}_{\mathbf{W}_I} diag(D_W)\right)^T \left(\mathbf{X} \text{ø} \hat{\mathbf{X}}_I\right) + 2\lambda \frac{1}{Q(Q-1)} \mathbf{A}_{\mathbf{W}_I}}{\left(\mathbf{B}_{\mathbf{W}_I} diag(D_W)\right)^T \mathbf{1}_{F,T} + 2\lambda \frac{1}{Q(Q-1)} \mathbf{1}_{Q,Q} \mathbf{A}_{\mathbf{W}_I}} \tag{13}$$

$$\mathbf{B}_{\mathbf{R}_I} \leftarrow \mathbf{B}_{\mathbf{R}_I} \odot \frac{\left(\mathbf{X} \text{ø} \hat{\mathbf{X}}_I\right) \left(diag(D_R) \mathbf{A}_{\mathbf{R}_I}\right)^T}{\mathbf{1}_{F,T} \left(diag(D_R) \mathbf{A}_{\mathbf{R}_I}\right)^T} \tag{14}$$

$$\mathbf{A}_{\mathbf{R}_I} \leftarrow \mathbf{A}_{\mathbf{R}_I} \odot \frac{\left(\mathbf{B}_{\mathbf{R}_I} diag(D_R)\right)^T \left(\mathbf{X} \text{ø} \hat{\mathbf{X}}_I\right)}{\left(\mathbf{B}_{\mathbf{R}_I} diag(D_R)\right)^T \mathbf{1}_{F,T}} \tag{15}$$

where $\odot$ is the element-wise multiplication, ø is the element-wise division, $\mathbf{1}_{F,T}$ represents a matrix of all-ones composed of F rows and T columns and ^{T} is the transpose operator.

The set of matrices are obtained updating the rules until the algorithm converges or reaches a maximum number of iterations M. Note that the activation matrices $\mathbf{A}_{\mathbf{W}_I}, \mathbf{A}_{\mathbf{R}_I}$ and the basis matrix $\mathbf{B}_{\mathbf{R}_I}$ must be normalized at each iteration. In particular, the normalization process (see Eq. (16)) ensures that the sum of the square elements of each q^{th} column of the basis matrices $\mathbf{B}_{\mathbf{W}_I}, \mathbf{B}_{\mathbf{R}_I}$ equals 1.0, as does the sum of the square elements of each q^{th} row of the activation matrices $\mathbf{A}_{\mathbf{W}_I}, \mathbf{A}_{\mathbf{R}_I}$ [43].

$$\mathbf{G}(q) = \frac{\mathbf{G}(q)}{\sqrt{\sum \mathbf{G}^2(q)}} \tag{16}$$

$$D_J(q) = D_J(q) \sqrt{\sum \mathbf{G}^2(q)} \tag{17}$$

where $(\mathbf{G}, J) = \{(\mathbf{A}_{\mathbf{W}_I}, W), (\mathbf{A}_{\mathbf{R}_I}, R), (\mathbf{B}_{\mathbf{R}_I}, R)\}$ respectively. If we consider the activation matrix $\mathbf{G} = (\mathbf{A}_{\mathbf{W}_I}, \mathbf{A}_{\mathbf{R}_I})$, $\sqrt{\sum \mathbf{G}^2(q)} = \sqrt{\sum_{t=1}^{T} \mathbf{G}^2(q,t)}$. If we consider the basis matrix $\mathbf{G} = \mathbf{B}_{\mathbf{R}_I}$, $\sqrt{\sum \mathbf{G}^2(q)} = \sqrt{\sum_{f=1}^{F} \mathbf{G}^2(f,q)}$. Next, the estimated magnitude spectrograms $\hat{\mathbf{X}}_{R_I}$ and $\hat{\mathbf{X}}_{W_I}$ (see Eq. (18)) can be obtained from the estimated basis and activation matrices. In this work, we focus only on the estimated wheezing spectrogram $\hat{\mathbf{X}}_{W_I}$ to differentiate healthy and unhealthy subjects. The separation procedure is summarized in Algorithm 1.

$$\hat{\mathbf{X}}_{J_I} = \mathbf{B}_{\mathbf{J}_I} diag(D_J) \mathbf{A}_{\mathbf{J}_I} \tag{18}$$

where J = (W or R) as occurs in Eq. (17).

Algorithm 1 Wheezing/Normal breath sound separation procedure

Require: The input magnitude spectrogram $\mathbf{X}$ and its estimated BOI.

1 Create the dictionary $\mathbf{B}_{\mathbf{W}_I}$ using Eq. (10).
2 Initialize $\mathbf{A}_{\mathbf{W}_I}, \mathbf{B}_{\mathbf{R}_I}$ and $\mathbf{A}_{\mathbf{R}_I}$ with random non-negative values.
3 Update $\mathbf{A}_{\mathbf{W}_I}$ using Eq. (13).
4 Normalize the activation matrix $\mathbf{A}_{\mathbf{W}_I}$ to obtain a L^2-norm of each row of the matrix equal to 1.0 using Eq. (16).
5 Update $diag(D_W)$ with the values that were used to normalize each row of the activation matrix $\mathbf{A}_{\mathbf{W}_I}$ using Eq. (17).
6 Update $\mathbf{B}_{\mathbf{R}_I}$ using Eq. (14).
7 Normalize the basis matrix $\mathbf{B}_{\mathbf{R}_I}$ to obtain a L^2-norm of each column of the matrix equal to 1.0 using Eq. (16).
8 Update $diag(D_R)$ with the values that were used to normalize each column of the basis matrix $\mathbf{B}_{\mathbf{R}_I}$ using Eq. (17).
9 Update $\mathbf{A}_{\mathbf{R}_I}$ using Eq. (15).
10 Normalize the activation matrix $\mathbf{A}_{\mathbf{R}_I}$ to obtain a L^2-norm of each row of the matrix equal to 1.0 using Eq. (16).
11 Update $diag(D_R)$ with the values that were used to normalize each row of the activation matrix $\mathbf{A}_{\mathbf{R}_I}$ using Eq. (17).
12 Repeat steps 3–11 until the algorithm converges (or until the maximum number of iterations M is reached).
13 Compute magnitude estimated spectrograms $\hat{\mathbf{X}}_{R_I}$ and $\hat{\mathbf{X}}_{W_I}$ using Eq. (18).

return $\hat{\mathbf{X}}_{W_I}$

Fig. 6 shows a comparison between the estimated wheezing spectrograms $\hat{\mathbf{X}}_W$ (stage I) and $\hat{\mathbf{X}}_{W_I}$ (stage II). It can be seen that $\hat{\mathbf{X}}_{W_I}$ is able to more reliably represent the spectral trajectories formed only by the energy of the wheeze sounds, almost completely eliminating the normal respiratory sounds compared to $\hat{\mathbf{X}}_W$. The facts that motivate this promising improvement shown by $\hat{\mathbf{X}}_{W_I}$ are due to: (i) the modeling of WS using information from

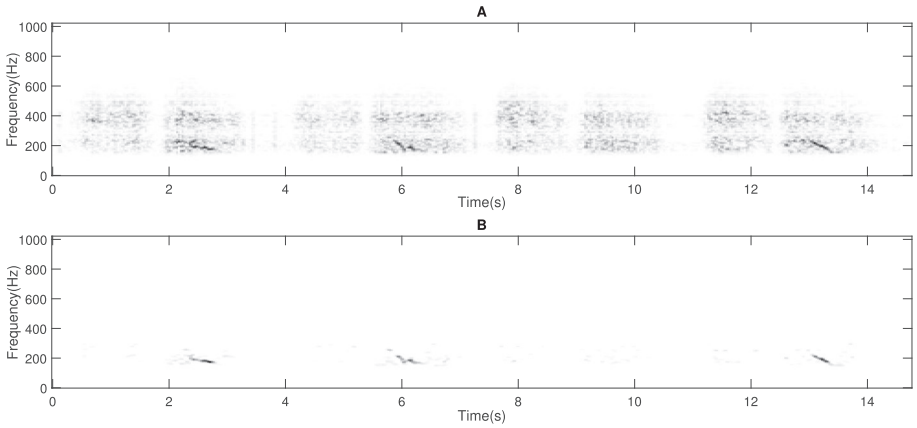

Fig. 6. The estimated wheezing spectrograms related to the unhealthy subject previously shown in Fig. 1A. A) $\hat{\mathbf{X}}_W$ from stage I. B) $\hat{\mathbf{X}}_{W_I}$ from stage II.

$\mathbf{B}_{\mathbf{W_I}}$ and BOI; (ii) the integration of the monophonic constraint into the tonal semi-supervised NMF approach that more accurately models the spectral trajectories generated by WS as they occur in the real world.

2.4. Stage III: wheezing presence/absence classification

As previously mentioned, WS can be considered pitched sounds that display spectral trajectories over time. The main goal of this stage is to define the subject's condition, that is, healthy or unhealthy motivated by the presence or absence of WS. For that purpose, we intend to take advantage of the temporal smoothness information extracted from the most significant spectral peaks forming the previous trajectories factorized in the estimated wheezing spectrogram $\hat{\mathbf{X}}_{W_I}$. The authors assume that the smoothness modelled by spectral trajectories from WS found in $\hat{\mathbf{X}}_{W_I}$ is smoother compared to RS. The reason is because the spectral trajectories from WS tend to show an organized structure in the time–frequency domain (see Fig. 7A) but the spectral peaks that compose the spectral trajectories from RS tend to be randomly distributed in the time–frequency domain without showing any organized structure due to the noisy nature shown by RS (see Fig. 7B). Next, this stage is detailed which is divided into two steps.

2.4.1. Step I: determining the optimal candidate for wheezing temporal interval

The proposed method is developed to analyze any mixture signal $x(t)$ independently of its temporal length. As a consequence to minimize its computational cost, it is only analyzed the optimal candidate OCW for wheezing temporal interval in which the probability of finding wheezing content is maximum. Considering all WS detected in $\hat{\mathbf{X}}_{W_I}$, OCW can be defined as the wheezing that shows the longest duration and the highest energy, with the time duration being more of a priority than energy. In the event that several candidates CW for wheezing intervals have the same temporal duration, the one with the highest energy is chosen as OCW. For this purpose, the spectral energy $\xi(t)$ from $\hat{\mathbf{X}}_{W_I}$ is calculated frame-by-frame as follows,

$$\xi(t) = \sum_{f=1}^{F} \widehat{X}_{W_{I_{f,t}}} \ , t = 1 \dots T \tag{19}$$

A threshold process is used to define all CW detected in $\hat{\mathbf{X}}_{W_I}$ that contain most of the spectral energy. Specifically, the thresholding ζ_t is based on the Otsu algorithm [44] since it allows to differentiates two groups in a histogram, in this case, intervals candidate for wheezing (CW) and intervals not candidates for wheezing. Next, the candidate CW with the longest duration will be selected as OCW. Once OCW has been chosen, the estimated wheezing spec-

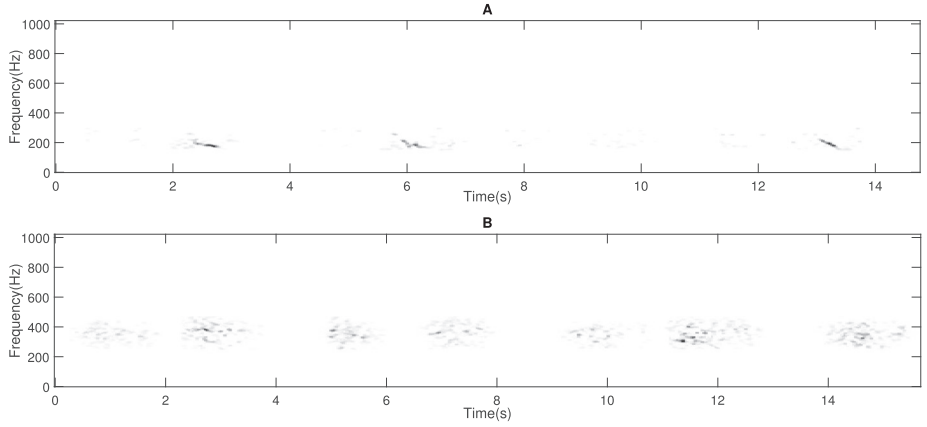

Fig. 7. Comparison between the estimated wheezing spectrograms $\hat{\mathbf{X}}_{W_I}$. A) $\hat{\mathbf{X}}_{W_I}$ related to the unhealthy subject shown in Fig. 1A. B) $\hat{\mathbf{X}}_{W_I}$ related to the healthy subject shown in Fig. 1B.

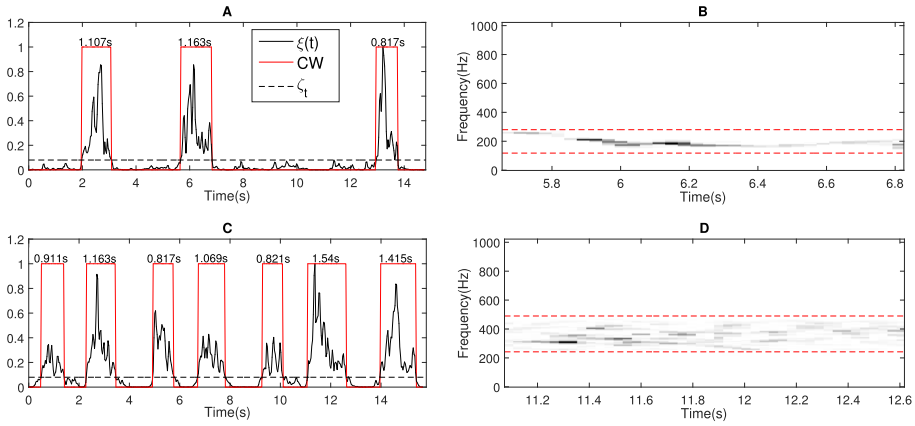

Fig. 8. Estimation of the candidate intervals CW, from the spectrogram $\hat{\mathbf{X}}_{W_l}$, that compete to be the optimal candidate interval OCW. A) CW obtained in the unhealthy subject previously shown in Fig. 1A. B) The spectrogram **Y** related to the OCW chosen from Fig. 8A that shows the highest probability to contain WS. It can be observed that OCW is located in the time interval [5.67s-6.81s]. C) CW obtained in the healthy subject previously shown in Fig. 1B. D) The spectrogram **Y** related to the OCW chosen from Fig. 8C that shows the highest probability to contain WS. It can be observed that OCW is located in the time interval [11.09s-12.61s]. Note that $\xi(t)$ has been normalized to adjust the values between 0 and 1. Both Fig. 8B and Fig. 8D, the red dotted lines display the BOI spectral boundaries associated to each selected interval. Specifically, BOI = [118 Hz–280 Hz] for Fig. 8B and BOI = [242 Hz–490 Hz] associated to Fig. 8D.

trogram $\hat{\mathbf{X}}_{W_l}$ will be limited by the temporal boundaries belonging to OCW in the time domain and the frequency limits belonging to BOI in the frequency domain to locate the portion **Y** of the estimated wheezing spectrogram $\hat{\mathbf{X}}_{W_l}$ where the probability of finding wheezing spectral trajectories is maximum (see Fig. 8).

2.4.2. Step II: smoothness related to the spectral trajectories contained in OCW

To measure the regularity of each spectral trajectory in the time-frequency domain, we propose to determine the spectral smoothness from the set of the most significant spectral peaks detected frame-by-frame. Specifically, the spectral smoothness is obtained calculating the degree of fluctuation σ_i of each spectral trajectory i normalized by the estimated BOI as follows,

$$\sigma_i = \frac{\sum_{z=1}^{Z} \left(S_{i,z} - S_{i,z+1}\right)^2}{BOI} \tag{20}$$

where Z represents the number of frames belonging to the spectrogram **Y** and S is the frequency of the peaks that describe the spectral trajectory i along frames. Empirical results report that the metric σ increases proportionally as the spectral trajectory is more irregular or less smooth. Otherwise, the metric σ decreases proportionally as the spectral trajectory is more regular or smoother in the time-frequency domain. Due to WS are characterized as pitched sounds, we assume that each wheeze sound can be composed by more than one frequency (harmonicity), specifically, by a maximum number of three spectral peaks [45] at each frame: S_1 (lowest frequency), S_2 (intermediate frequency) and S_3 (highest frequency). In order to include the harmonic nature of the WS in the proposed method, we propose to measure the smoothness of each spectral trajectory by means of the set of spectral peaks S_1, S_2 and S_3 as follows,

$$\sigma_G = \sigma_1 + \sigma_2 + \sigma_3 \tag{21}$$

where σ_1, σ_2 and σ_3 denote the degree of fluctuation of each spectral trajectory formed by the previous set of spectral peaks S_1, S_2 and S_3 that have been detected at each frame. When only a single peak is detected in a frame, we assume that this peak refers to the peak of the lowest frequency S_1 since it is common that WS concentrate most of the energy in the peak that show the minimum frequency in a multipitch structure. As a consequence, only frames with two and three spectral peaks detected will be considered for the calculation of σ_2 and σ_3, respectively. Fig. 9 shows the spectral trajectories related to the unhealthy and healthy subject in Fig. 1. In the case of the unhealthy subject (see Fig. 9A), only a single trajectory was detected composed of the spectral peaks S_1, typically found in WS. Besides, it is demonstrated that the continuity of the wheezing trajectories display high smoothness. However, in the case of the healthy subject (see Fig. 9B), three spectral trajectories motivated by the set of peaks of type S_1, S_2 and S_3 were found. It confirms that RS exhibit a high irregularity or discontinuity in time–frequency domain, in other words, a low smoothness because RS tend to locate randomly the peaks in the spectral domain due to the noisy nature of RS.

Our empirical analysis, reported the following facts: (i) the spectral trajectories in the case of unhealthy subjects are distinguished by the continuous nature of WS and the value of $\sigma_G \leq 1$; (ii) the spectral trajectories in the case of healthy subjects are distinguished by the noisy nature of RS and the value of $\sigma_G \gg 1$. For this reason, we set the threshold $\zeta_h = 1$ in order to determine the subject's condition as follows,

$$\text{subject's condition} = \begin{cases} unhealthy & \text{if } \sigma_G \leqslant \zeta_h \\ healthy & \text{if } \sigma_G > \zeta_h \end{cases} \tag{22}$$

In the case of WS compose by more than one spectral peak, the value of the metric σ_G is still significantly low compared to those ones provided by RS. This is because the spectral trajectories originated by the peaks S_1, S_2 and S_3 tend to show an organized structure in the time–frequency domain, i.e. each spectral trajectory will be smooth and continuous due to its wheezing nature and as a consequence, the general metric σ_G will be the sum of 3 small values (σ_1, σ_2 and σ_3).

3. Results and discussion

3.1. Datasets

To evaluate the classification performance of the proposed method, six databases P1, T1, T2H, T2M, T2L and T3 have been used which are detailed in Table 1. In total, these databases provide 4107 s of recording, 114 healthy subjects, 94 unhealthy subjects, 2168 respiratory events (a respiratory event is defined as an inspiration or expiration) and 219 wheezes. Databases P1, T1, T2H, T2M and T2L have been created by collecting a lot of recordings of different subjects of the most widely used Internet pulmonary

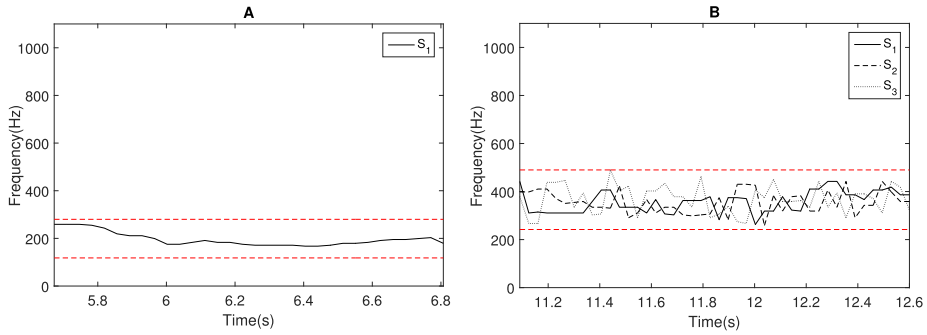

Fig. 9. Spectral trajectories defined by the spectral peaks S_1, S_2, S_3. A) The spectral trajectory ($\sigma_G = 0.45$), related to the unhealthy subject shown in Fig. 8B, that is composed by a single spectral peak S_1. B) The spectral trajectory ($\sigma_G = 98.5$), related to the healthy subject shown in Fig. 8D, that is composed by three types of spectral peaks S_1, S_2 and S_3. Although σ_G in Fig. 9B is significantly higher compared to σ_1 in Fig. 9A, highlight that each smoothness σ_1, σ_2 or σ_3 measured in Fig. 9B is significantly higher compared to the smoothness σ_1 measured in Fig. 9A.

Table 1
Characteristics of each database.

ID1	ID2	ID3	ID4	ID5	ID6	ID7	ID8	ID9	ID10
P1	96	48/48	5–24	1442	[0–9]	[4–16]	784	[1–8]	92
T1	48	48/0	8–49	1051	–	[6–30]	500	–	0
T2H	16	0/16	7–22	251	5	[6–14]	126	[1–5]	41
T2M	16	0/16	7–22	251	0	[6–14]	126	[1–5]	41
T2L	16	0/16	7–22	251	−5	[6–14]	126	[1–5]	41
T3	48	18/30	4–65	861	[2–8]	[4–20]	506	[3–11]	86

ID1: identifier; ID2: number of recordings; ID3: number of recordings captured from healthy/unhealthy subjects; ID4: the shortest and longest duration, in seconds, captured from recordings; ID5: total duration in seconds; ID6: the lowest and highest SNR, in dB, between WS and RS; ID7: the minimum and maximum number of respiratory events found in the recordings; ID8: the total number of respiratory events; ID9: the minimum and maximum number of wheezes found in the recordings; ID10: the total number of wheezes. The fifth column from T1 does not show SNR because it is only composed of healthy subjects. Analogously, the eighth column from T1 does not show the minimum and maximum number of wheezes found because it does not contain wheeze sounds.

repositories [46–58]. The databases collected consisted of sounds captured from the trachea, anterior, and posterior chest; using either a stethoscope or microphone. These were collected from subjects with different pathologies, including asthma, bronchitis, COPD, and croup. Indicate that the age was not considered as a significant variable because the variations caused by age difference in automatic auscultation is too small to be clinically relevant [59]. Database T3 has been directly shared by authors [27].

Using the visual inspection of spectrograms, datasets P1, T2H, T2M and T2L have been created mixing only WS recordings manually separated (by means of a time-frequency mask applied to the mixture spectrogram to select only the bins of each frame corresponding to wheezing) and only RS recordings (in which wheezing is inactive) obtained from [46–58]. The datasets T2H (SNR = 5 dB), T2M (SNR = 0 dB) and T2L (SNR=-5 dB) represent the same dataset T2 but each of them uses a different signal-to-noise ratio (SNR) between WS and RS. After each manual separation for each dataset recording, the power related to WS and RS are calculated and the signal with the highest power is left fixed while the signal with the lowest power is scaled to obtain the desired SNR in order to avoid audio saturation or distortion in the signal scaling process. However, datasets P1 and T3 preserve the original SNR. More details of dataset T3 can be found in [27].

The previous databases have been evaluated with the following purpose:

- Optimization. P1 optimizes the training of the state-of-the-art methods.
- Testing. T1 assesses the robustness of the proposed method in correctly diagnosing healthy subject.
- Testing. T2H, T2M and T2L assess the robustness of the proposed method when WS are masked by RS.
- Testing. T3 assesses the wheezing detection performance of the proposed method in real medical scenario to decide whether the diagnosis of subject is healthy or unhealthy.

In order to validate the results, it should be noted that each file in one of the following databases P1, T1, T2 and T3 has only been used in that database and has not been used again in the rest of the remaining databases (e.g., P1 is not a part of the rest of datasets T1, T2 and T3).

3.2. Experimental setup

A preliminary study show that the following parameters provide the best trade-off between the classification performance and the computational cost: sampling rate $f_s = 2048$ Hz, Hamming window with $N = 256$ samples length and 25% overlap (temporal resolution of 31.3 ms), a discrete Fourier transform using $2N$ points (frequency resolution of 4 Hz). Furthermore, the convergence of the NMF approaches was empirically achieved after 90 iterations for all signals, so $M = 90$ iterations has been selected. Moreover, a number of components $K = 80$ have been used in the NMF approach of the stage I. Finally the optimal weight of the temporal monophonic constraint is $\lambda = 0.05$.

3.3. State-of-the-art methods for comparison

Three recent state-of-the-art wheezing detection methods have been used to evaluate the performance of the proposed method: TSVM [20], MKNN [21] and EEMD [28]. The state-of-the-art methods have been implemented strictly following the instructions shown in their respective references. MKNN, TSVM and EEMD are supervised approaches using the database P1 for the training

of each of them. However, the proposed method does not use any training due to unsupervised approach. In this work, the subject will be diagnosed as unhealthy when MKNN or TSVM locates a wheezing temporal interval $\geqslant$100 ms. Moreover, EEMD has been implemented without limiting the duration of the input mixture.

3.4. Evaluation metrics

The ability to discriminate between healthy and unhealthy subjects has been evaluated by means of five metrics commonly used in the field of wheezing detection [27–29,60]: i) Sensitivity (SE), the ability to correctly classify a subject as unhealthy analyzing only the recordings corresponding to unhealthy subjects; ii) Specificity (SP), the ability to correctly classify a subject as healthy analyzing only the recordings corresponding to healthy subjects; iii) Positive Predictive Value (PPV), the probability that a subject classified as unhealthy truly has the disease derived from wheezing sounds. Here, recordings of both unhealthy and healthy subjects are analyzed; iv) Negative Predictive Value (NPV), the probability that a subject classified as healthy truly does not have the disease derived from wheezing sounds. Here, recordings of both unhealthy and healthy subjects are analyzed; and v) Accuracy (ACC), the ability to correctly classify a subject as healthy or unhealthy.

$$SE = \frac{TP}{TP + FN} \tag{23}$$

$$SP = \frac{TN}{TN + FP} \tag{24}$$

$$PPV = \frac{TP}{TP + FP} \tag{25}$$

$$NPV = \frac{TN}{TN + FN} \tag{26}$$

$$ACC = \frac{(TP + TN)}{(TP + FP + FN + TN)} \tag{27}$$

In these equations, TP (True Positive) represent the number of unhealthy subjects correctly diagnosed, TN (True Negative) represent the number of healthy subjects correctly diagnosed, FP (False Positive) represent the number of healthy subjects misdiagnosed as unhealthy subjects, and FN (False Negative) represent the number of unhealthy subjects misdiagnosed as healthy subjects.

3.5. Results

In order to assess the performance of the proposed method together with the state-of-the-art methods with respect to the temporal length of the input mixture $x(t)$, we propose to evaluate using three levels of analysis: breathing mixture (first level), respiratory cycle (second level) and respiratory stage (third level). Each next level of analysis is composed of a greater number of audio segments from the input mixture being each segment of shorter duration:

- First level: breathing mixture. The entire input mixture $x(t)$ is analyzed at once (see Fig. 10A). The subject will be diagnosed as unhealthy when at least one wheezing is detected anywhere on the input mixture.
- Second level: respiratory cycle. Each respiratory cycle that composes the input mixture $x(t)$ is analyzed as an independent unit (see Fig. 10B). In this case, the subject will be diagnosed as unhealthy when at least one wheezing is detected in one respiratory cycle among all those that compose the input mixture $x(t)$.
- Third level: respiratory stage. Each respiratory stage (inspiration or expiration) that composes each respiratory cycle is analyzed as an independent unit (see Fig. 10C). In this case, the subject will be diagnosed as unhealthy when at least one wheezing is detected in one respiratory stage among all those that compose the input mixture $x(t)$.

Fig. 11 shows SP results evaluating the database T1 between the proposed method and the aforementioned state-of-the-art methods for the three levels of analysis described above. The database T1 can only be evaluated with the metric SP due to the non-existence of TP and FN since T1 is only composed of audio recordings captured from healthy subjects. Results indicate that the proposed method outperforms the correct diagnosis performance of healthy subject compared to the state-of-the art methods for all levels of analysis. Focusing on the first level of analysis (Breathing mixture level), the proposed method accomplishes a significant improvement of approximately 14.58%, 16.67% and 20.83% versus TSVM, MKNN and EEMD, respectively. This fact seems to suggest that the proposed method is more reliable in correctly diagnosing healthy subjects compared to previous baseline methods, showing a promising robustness. The reason is that the proposed method is the only method of those evaluated that has correctly classified as healthy all recordings from healthy subjects belonging to the database T1 that are evaluated at a breathing mixture level, respiratory cycle level and respiratory stage level.

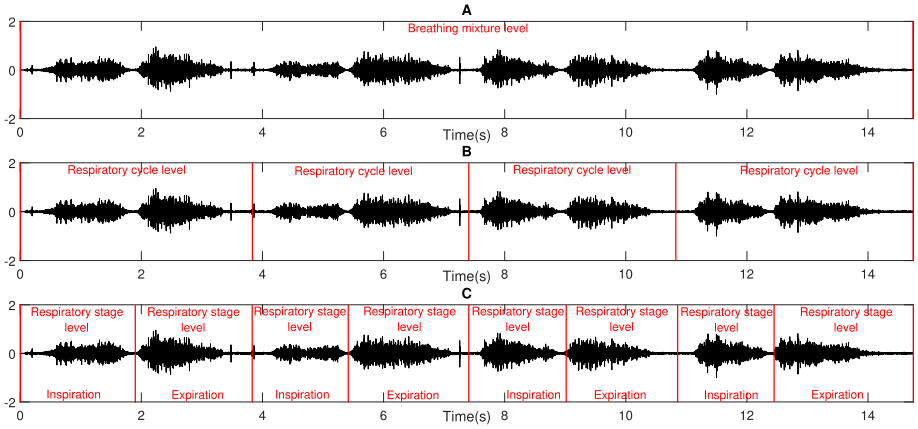

Fig. 10. Different levels of analysis of the input mixture associated to the spectrogram shown in Fig. 1A. A) Breathing mixture level. B) Respiratory cycle level. C) Respiratory stage level. It can be observed a single unit to a breathing mixture level, four units to a respiratory cycle level and finally, eight units to a respiratory stage level.

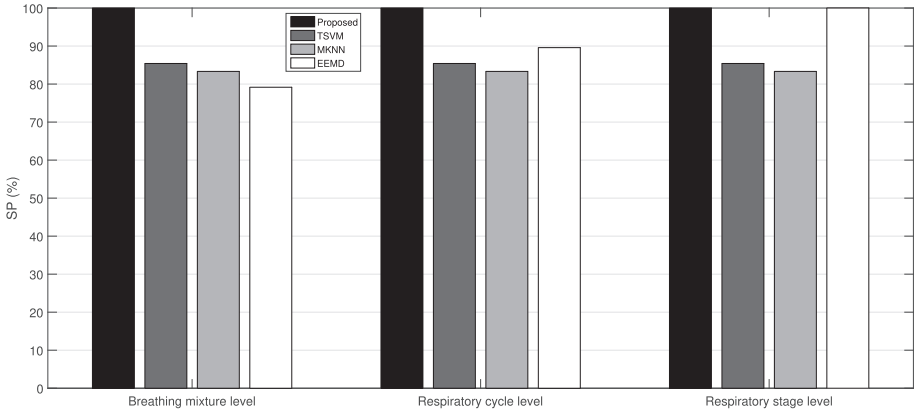

Fig. 11. Comparative SP results between the proposed method and the aforementioned state-of-the-art methods in the database T1 for the three levels of analysis.

It confirms the robustness of the proposed method with respect to the temporal length of the recording to correctly diagnose of healthy subjects. Both, TSVM and MKNN achieve a similar result in the three levels of analysis evaluated but TSVM obtains a slightly better classification performance compared to MKNN. Nevertheless, TSVM and MKNN can be considered robust with the recording duration since are based on the detection of the temporal intervals in which wheezing is active, i.e. MKNN and TSVM determine the wheezing frames and then analyze the duration of each temporal interval detected as wheezing to confirm the presence or absence of them in the recording. In contrast, the performance of the method EEMD highly depends on the length of the recording analyzed. Specifically, the performance of EEMD equals the performance of the proposed method analyzing at respiratory stage level. However, EEMD reduces its classification performance by 20.83% when analyzing at breathing mixture level and 10.42% when analyzing at respiratory cycle level. This behavior is due to the fact that EEMD is based on the extraction of segments that are candidates for being wheezing segments, which are obtained from a set of thresholds that directly depend on the input recording. Specifically, when the duration of the input recording is longer than the respiratory stage, EEMD begins to erroneously detect the wheezing candidate segments that impairs its performance in the classification between presence or absence of WS in breath recordings. While EEMD classifies all the wheeze segments it has extracted, the proposed method only analyzes the segment with the highest probability of finding WS, previously defined as OCW. For this reason, the results of the proposed method are the same in all three levels of analysis, while the EEMD results worsen at the respiratory cycle level and breathing mixture level. The wheezing classification performance against the duration of the recording obtained by the proposed method is a notable advantage, since a correct wheezing diagnosis can be independent of the performance in the detection of the segment at respiratory level, a critical stage for the operation of the EEMD to maximize its classification results.

Fig. 12 shows the SE results by evaluating three databases T2H (SNR = 5 dB), T2M (SNR = 0 dB) and T2L (SNR=−5 dB) with different SNR to compare the robustness of the proposed method in the diagnosis of unhealthy subjects when WS are increasingly masked by RS. The databases T2H, T2M and T2L can only be evaluated with this metric due to the non-existence of TN and FP because these databases are composed exclusively of audio recordings from unhealthy subjects, in other words, each recording contains at least one wheezing. Results show that the proposed method provides the best overall wheezing presence/absence classification results compared to the other evaluated state-of-the-art methods considering all SNR scenarios evaluated for the three levels of analysis. Focusing on the breathing mixture level, it can be observed the following:

- Database T2H: the SE improvement of the proposed method is about 12.5% (TSVM), 12.5% (MKNN) and 18.75% (EEMD).
- Database T2M: the SE improvement of the proposed method is about 18.75% (TSVM), 25% (MKNN) and 25% (EEMD).
- Database T2L: the SE improvement of the proposed method is about 25% (TSVM), 31.25% (MKNN) and 31.25% (EEMD).

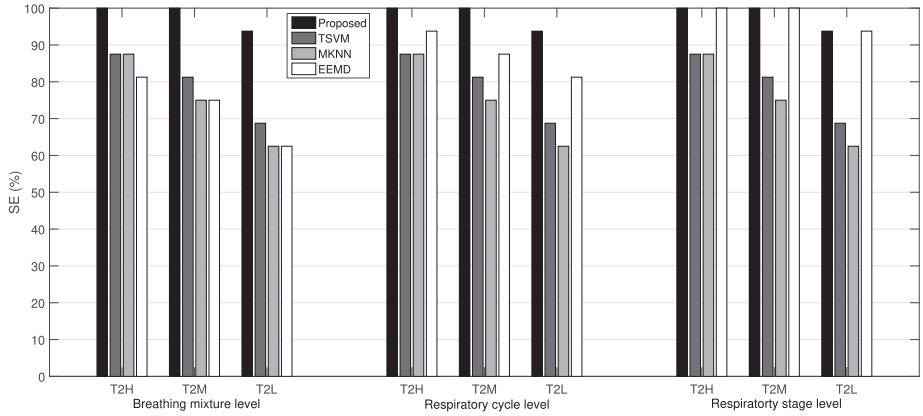

Fig. 12. Comparative SE results between the proposed method and the aforementioned state-of-the-art methods in the databases T2H, T2M and T2L for the three levels of analysis.

Results indicate that the proposed method is more effective and reliable to diagnose unhealthy subjects compared to the other state-of-the-art methods evaluated. It should be emphasized that the proposed method is the only method that correctly classifies, evaluating T2H and T2M, all recordings from unhealthy subjects considering the three levels of analysis. EEMD equals the performance of the proposed method in T2H and T2L but applied in a respiratory stage level. Unlike EEMD, the proposed method again does not need to analyze at the respiratory stage level to maximize its best classification performance. This fact avoids that the performance of the proposed method be dependent on the correct selection of the level of analysis. In addition, it can be clearly seen how EEMD improves its classification performance from Fig. 12 (left) to Fig. 12 (right) as the temporal length of the analyzed recording is reduced to its optimal level, in this case, the respiratory stage level. This is the reason why EEMD obtains worse results compared to the other methods at breathing mixture level but outperforms TSVM and MKNN both respiratory cycle level and respiratory stage level, being this improvement higher at the last level of those mentioned. TSVM achieves the same or better performance than MKNN in any evaluated scenario and analysis level. Highlight that the classification improvement between the proposed method and the rest of the evaluated methods increases as the WS are more difficult to detect because they are masked to a greater extent by RS. This fact seems to demonstrate the greater robustness of the proposed method compared to the assessed state-of-the-art methods. Considering T2L in which an acoustic scenario SNR= −5dB has been modeled, WS are barely audible due to the high interference caused by RS so the wheezing classification process is more complex. Focusing on the decrease in SE results by comparing T2H with T2L at all three levels of analysis, the following can be observed:

- Considering a breathing mixture level: The SE results drop approximately 6.25% (Proposed), 18.75% (TSVM), 25% (MKNN) and 18.75% (EEMD).
- Considering a respiratory cycle level: The SE results drop approximately 6.25% (Proposed), 18.75% (TSVM), 25% (MKNN) and 12.5% (EEMD).
- Considering a respiratory stage level: The SE results drop approximately 6.25% (Proposed), 18.75% (TSVM), 25% (MKNN) and 6.25% (EEMD).

Based on the previous results, it can be said that the wheezing presence/absence classification performance taking into account the proposed method, TSVM and MKNN do not depend on the temporal length of the recording. Although EEMD equals the performance of the proposed method evaluating in a respiratory stage level, only the performance of EEMD worsens significantly from the third level of analysis (respiratory stage level) to the first one (breathing mixture level).

In order to assess the performance of the proposed method in a real medical scenario to detect the presence (unhealthy subject) or absence (healthy subject) of WS, Table 2 shows detailed results evaluating the dataset T3. The proposed method outperforms the classification performance compared to the state-of-the-art methods in terms of SP, SE, ACC, PVP and NVP for the three levels of analysis. The better performance of the proposed method compared to TSVM, MKNN and EEMD suggests that our proposal is competitive to improve the reliability of the subjective diagnosis provided by the physician in the auscultation process to detect the presence of WS. The reason can be due to the fact that the proposed method models common spectro-temporal characteristics that attempt to describe the typical behavior of wheezing sounds as occurs in the nature of the real world. Specifically, a detailed analysis of the proposed method in terms of each evaluation metric indicates the following:

- The performance of the proposed method in terms of SP indicates that 100% of the healthy subjects are correctly classified.
- The performance of the proposed method in terms of SE indicates that 93.3% of the unhealthy subjects are correctly classified.
- The performance of the proposed method in terms of ACC indicates that 95.8% of the healthy and unhealthy subjects are correctly classified.
- The performance of the proposed method in terms of PPV indicates that 100% of the subjects classify as unhealthy subjects are correctly detected and that none of the healthy subjects have been erroneously classified as unhealthy.
- The performance of the proposed method in terms of NPV indicates that 90% of the subjects are correctly classified.

Similarly as occurs in the evaluation of the databases T1 and T2, TSVM and MKNN show their robustness with respect to the temporal length of the recording. TSVM and MKNN are machine learning methods based on the extraction of features and classifiers. However, TSVM obtains better performance compared to MKNN because TSVM consists of two SVM classifiers specifically designed to avoid false positives and false negatives as much as possible. The best performance of the evaluated method is achieved in terms of SP for any level of analysis (the proposed method), in terms of SE for any level of analysis (MKNN and TSVM) and in terms of SE at the first level of analysis and in terms of SP at the second and third levels of analysis (EEMD). The proposed method outperforms the second best state-of-the-art method EEMD in terms of ACC (16.6% in the first level of analysis, 10.4% in the second level of analysis and 2.1% in the third level of analysis) but this ACC improvement is significantly higher, approximately 25%, compared to MKNN and TSVM for the three levels of analysis. Although Table 2 confirms that EEMD can be considered an efficient method to classify wheezing presence/absence, EEMD results shows a significantly decrease of its classification performance when the level of analysis is not the respiratory stage.

4. Conclusions and future work

In this study, we propose a novel method to automatically detect the presence or absence of wheezing sounds in breath recordings. Three novel contributions are proposed by authors. The first contribution, band of interest (BOI), estimates the spectral

Table 2
Comparative results (%) in order to classify healthy/unhealthy subjects evaluating the database T3 for the three levels of analysis.

	Breathing mixture level					Respiratory cycle level					Respiratory stage level				
Method	SP	SE	ACC	PPV	NPV	SP	SE	ACC	PPV	NPV	SP	SE	ACC	PPV	NPV
Proposed	**100**	**93.3**	**95.8**	**100**	**90**	**100**	**93.3**	**95.8**	**100**	**90**	**100**	**93.3**	**95.8**	**100**	**90**
TSVM	55.5	76.6	68.7	74.2	58.8	55.5	76.6	68.7	74.2	58.8	55.5	76.6	68.7	74.2	58.8
MKNN	50	70	62.5	70	50	50	70	62.5	70	50	50	70	62.5	70	50
EEMD	77.8	80	79.2	85.7	70	88.9	83.3	85.4	92.6	76.2	94.4	**93.3**	93.7	96.5	89.5

Each value in bold indicates the highest value obtained in each column.

range in which the probability to find wheeze sounds is maximum. As a second contribution, a constrained tonal semi-supervised NMF is presented to factorize into spectral patterns that correctly model the tonal nature, by narrowband peaks, shown by WS in the estimated BOI. From the separated wheezing spectrogram, we propose an intuitive method to classify healthy or unhealthy subjects analyzing the temporal smoothness of the spectral trajectories defined by the most significant energy decomposed in the estimated BOI.

The main conclusions derived from the experimental results indicate that:

- The proposed method provides the best overall wheezing presence/absence classification results compared to the other state-of-the-art methods considering all databases evaluated. Results suggest that the proposed method obtains a promising performance to improve the reliability of the subjective diagnosis provided by the physician in the auscultation process.
- As the results confirm, one of the main advantages of the proposed method is the robustness with respect to the temporal length of the recording to correctly classify the subject's condition.
- Similar to what occurs in the real world, identifying the presence of wheezing is much more complex in those acoustic scenarios in which wheeze sounds are barely audible due to the high interference caused by normal breath sounds. Comparing the evaluation of the databases T2H and T2L in the first level of analysis, SE results of the proposed method only drop 6.25% while results obtained by the other state-of-the-art methods drop above 18.75%. It suggest that the proposed method is more robust than baseline methods in the assessment of noisy scenarios, specifically, SNR < 5 dB.
- Unlike the state-of-the-art methods evaluated based on machine learning, the proposed method does not depend on any training dataset but the modelling of common spectro-temporal characteristics typically shown in wheezing sounds to be discriminate from normal breath sounds.

Future work will focus on the design of new NMF constraints that model the time–frequency behaviour of different types of wheezing, specifically monophonic and polyphonic wheezing with the aim of assessing the severity of lung disease caused by the appearance of these wheezing sounds.

Conflict of interest

There is no conflict of interest.

Acknowledgments

The authors would like to thank Dr. Vedran Bilas and Dr. Dinko Oletic for sharing their dataset labelled T3 in this manuscript, and the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.apacoust.2019.107188.

References

[1] Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. Ann Thoracic Med 2015;10(3):158.
[2] Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds: advances beyond the stethoscope. Am J Respiratory Crit Med 1997;156(3):974–87.
[3] Lozano-García M, Fiz JA, Martínez-Rivera C, Torrents A, Ruiz-Manzano J, Jané R. Novel approach to continuous adventitious respiratory sound analysis for the assessment of bronchodilator response. PloS One 2017;12(2):e0171455.
[4] Le Cam S, Belghith A, Collet C, Salzenstein F. Wheezing sounds detection using multivariate generalized gaussian distributions. In: IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE; 2009. p. 541–4.
[5] Lin BS, Wu HD, Chen SJ. Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network. J Healthcare Eng 2015;6(4):649–72.
[6] World Health Organization. Asthma,https://www.who.int/news-room/fact-sheets/detail/asthma [Online. Accessed: 2019-25-07].
[7] World Health Organization, Pneumonia,https://www.who.int/maternal_child_adolescent/news_events/news/2011/pneumonia/en/ [Online. Accessed: 2019-25-07].
[8] Salazar AJ, Alvarado C, Lozano FE. System of heart and lung sounds separation for store-and-forward telemedicine applications. Revista Facultad de Ingeniería Universidad de Antioquia 2012;64:175–81.
[9] Lin BS, Lin BS, Wu HD, Chong FC, Chen SJ. Wheeze recognition based on 2d bilateral filtering of spectrogram. Biomed Eng: Appl, Basis Commun 2006;18(03):128–37.
[10] Jain A, Vepa J. Lung sound analysis for wheeze episode detection, in: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE; 2008, pp. 2582–2585.
[11] Sovijarvi A, Dalmasso F, Vanderschoot J, Malmberg L, Righini G, Stoneman S. Definition of terms for applications of respiratory sounds. Eur Respiratory Rev 2000;10(77):597–610.
[12] Bokov P, Mahut B, Flaud P, Delclaux C. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. Comput Biol Med 2016;70:40–50.
[13] Qiu Y, Whittaker A, Lucas M, Anderson K. Automatic wheeze detection based on auditory modelling. Proc Inst Mech Eng, Part H: J Eng Med 2005;219(3):219–27.
[14] Zhang J, Ser W, Yu J, Zhang T. A novel wheeze detection method for wearable monitoring systems. In: International Symposium on Intelligent Ubiquitous Computing and Education, IEEE; 2009. p. 331–34.
[15] Aydore S, Sen I, Kahya YP, Mihcak MK. Classification of respiratory signals by linear analysis. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE; 2009. p. 2617–620.
[16] Emrani S, Gentimis T, Krim H. Persistent homology of delay embeddings and its application to wheeze detection. IEEE Signal Process Lett 2014;21(4):459–63.
[17] Kochetov K, Putin E, Azizov S, Skorobogatov I, Filchenkov A. Wheeze detection using convolutional neural networks. In: EPIA Conference on Artificial Intelligence. Springer; 2017. p. 162–73.
[18] Jin F, Krishnan S, Sattar F. Adventitious sounds identification and extraction using temporal-spectral dominance-based features. IEEE Trans Biomed Eng 2011;58(11):3078–87.
[19] Wisniewski M, Zielinski TP. Joint application of audio spectral envelope and tonality index in an e-asthma monitoring system. IEEE J Biomed Health Inf 2015;19(3):1009–18.
[20] Mazić I, Bonković M, Džaja B. Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in children's respiratory sounds. Biomed Signal Process Control 2015;21:105–18.
[21] Shaharum SM, Sundaraj K, Aniza S, Palaniappan R, Helmy K. Classification of asthma severity levels by wheeze sound analysis. In: IEEE Conference on Systems, Process and Control (ICSPC). IEEE; 2016. p. 172–6.
[22] Bahoura M, Pelletier C. Respiratory sounds classification using gaussian mixture models. In: Canadian Conference on Electrical and Computer Engineering, vol. 3, IEEE; 2004. p. 1309–312.
[23] Mayorga P, Druzgalski C, Morelos R, Gonzalez O, Vidales J. Acoustics based assessment of respiratory diseases using gmm classification. In: Annual International Conference of the IEEE Engineering in Medicine and Biology, IEEE; 2010. p. 6312–316.
[24] Riella R, Nohama P, Maia J. Method for automatic detection of wheezing in lung sounds. Braz J Med Biol Res 2009;42(7):674–84.
[25] Taplidou SA, Hadjileontiadis LJ. Wheeze detection based on time-frequency analysis of breath sounds. Comput Biol Med 2007;37(8):1073–83.
[26] Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, Chouvarda I, Maglaveras N, Tsara V, Teixeira C, Carvalho P, Henriques J, et al. Detection of wheezes using their signature in the spectrogram space and musical features. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE; 2015. pp. 5581–584.
[27] Oletic D, Bilas V. Asthmatic wheeze detection from compressively sensed respiratory sound spectra. IEEE J Biomed Health Inf 2018;22(5):1406–14.
[28] Lozano M, Fiz JA, Jané R. Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency. IEEE J Biomed Health Inf 2015;20(2):486–97.
[29] Torre-Cruz J, Canadas-Quesada F, Carabias-Orti J, Vera-Candeas P, Ruiz-Reyes N. A novel wheezing detection approach based on constrained non-negative matrix factorization. Appl Acoust 2019;148:276–88.
[30] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401(6755):788–91.
[31] Canadas-Quesada F, Ruiz-Reyes N, Carabias-Orti J, Vera-Candeas P, Fuertes-Garcia J. A non-negative matrix factorization approach based on spectro-temporal clustering to extract heart sounds. Appl Acoust 2017;125:7–19.

[32] Dia N, Fontecave Jallon J, Gumery PY, Rivet B. Denoising phonocardiogram signals with non-negative matrix factorization informed by synchronous electrocardiogram. In: 26th European Signal Processing Conference (EUSIPCO). IEEE; 2018. p. 51–5.
[33] Févotte C, Bertin N, Durrieu JL. Nonnegative matrix factorization with the itakura-saito divergence: with application to music analysis. Neural Comput 2009;21(3):793–830.
[34] Liutkus A, Fitzgerald D, Badeau R. Cauchy nonnegative matrix factorization. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), IEEE; 2015. p. 1–5.
[35] Laroche C, Kowalski M, Papadopoulos H, Richard G. A structured nonnegative matrix factorization for source separation. In: 23rd European Signal Processing Conference (EUSIPCO), IEEE; 2015. p. 2033–37.
[36] Canadas-Quesada FJ, Vera-Candeas P, Ruiz-Reyes N, Carabias-Orti J, Cabanas-Molero P. Percussive/harmonic sound separation by non-negative matrix factorization with smoothness/sparseness constraints. J Audio, Speech, Music Process 2014;2014(26):1–17.
[37] Kitamura D, Ono N, Saruwatari H, Takahashi Y, Kondo K. Discriminative and reconstructive basis training for audio source separation with semi-supervised nonnegative matrix factorization. In: IEEE International Workshop on Acoustic Signal Enhancement (IWAENC). IEEE; 2016. p. 1–5.
[38] Hurley N, Rickard S. Comparing measures of sparsity. IEEE Trans Inf Theory 2009;55(10):4723–41.
[39] Feng C, Xiao L, Wei Z. Compressive sensing isar imaging with stepped frequency continuous wave via gini sparsity. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE; 2013. p. 2063–6.
[40] Toh KKV, Isa NAM. Noise adaptive fuzzy switching median filter for salt-and-pepper noise reduction. IEEE Signal Process Lett 2010;17(3):281–4.
[41] Rafii Z, Pardo B. Repeating pattern extraction technique (repet): a simple method for music/voice separation. IEEE Trans Audio, Speech, Language Process 2013;21(1):73–84.
[42] Prominence criterion of a peak according to the MATLAB software,https://es.mathworks.com/help/signal/ref/findpeaks.html?searchHighlight=findpeak&s_tid=doc_srchtitle#buff2uu [Online. Accessed: 2019-25-07].
[43] Cañadas-Quesada FJ, Vera-Candeas P, Martinez-Munoz D, Ruiz-Reyes N, Carabias-Orti JJ, Cabanas-Molero P. Constrained non-negative matrix factorization for score-informed piano music restoration. Digital Signal Process 2016;50:240–57.
[44] Yuan X, Martínez JF, Eckert M, López-Santidrián L. An improved otsu threshold segmentation method for underwater simultaneous localization and mapping-based navigation. Sensors 2016;16(7):1148.
[45] Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: a systematic review. PloS One 2017;12(5):e0177926.
[46] The r.a.l.e. repository.http://www.rale.ca [Online. Accessed: 2019-25-07].
[47] Stethographics lung sound samples.http://www.stethographics.com [Online. Accessed: 2019-25-07).
[48] 3m littmann stethoscopes.https://www.3m.com [Online. Accessed: 2019-25-07].
[49] East tennessee state university pulmonary breath sounds,http://faculty.etsu.edu [Online. Accessed: 2019-25-07].
[50] ICBHI 2017 Challenge.https://bhichallenge.med.auth.gr [Online. Accessed: 2019-25-07].
[51] Lippincott NursingCenter.https://www.nursingcenter.com [Online. Accessed: 2019-25-07].
[52] Thinklabs Digital Stethoscope.https://www.thinklabs.com [Online. Accessed: 2019-25-07].
[53] Thinklabs youtube.https://www.youtube.com/channel/UCzEbKuIze4AI1523_AWiK4w [Online. Accessed: 2019-25-07].
[54] Emedicine/Medscape.https://emedicine.medscape.com/article/1894146-overview#a3 [Online. Accessed: 2019-25-07].
[55] E-learning resources.https://www.ers-education.org/e-learning/reference-database-of-respiratory-sounds.aspx [Online. Accessed: 2019-25-07].
[56] Respiratory wiki.http://respwiki.com/Breath_sounds [Online. Accessed: 2019-25-07].
[57] Easy Auscultation.https://www.easyauscultation.com/lung-sounds-reference-guide [Online. Accessed: 2019-25-07].
[58] Colorado State University.http://www.cvmbs.colostate.edu/clinsci/callan/breath_sounds.htm [Online. Accessed: 2019-25-07].
[59] Gross V, Dittmar A, Penzel T, Schuttler F, Von Wichert P. The relationship between normal lung sounds, age, and gender. Am J Respiratory Crit Care Med 2000;162(3):905–9.
[60] Pramono RXA, Imtiaz SA, Rodriguez-Villegas E. Evaluation of features for classification of wheezes and normal respiratory sounds. PloS One 2019;14(3):e0213659.