Applied Acoustics 148 (2019) 276–288

Contents lists available at ScienceDirect

Applied Acoustics

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

A novel wheezing detection approach based on constrained non-negative matrix factorization

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

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 20 September 2018
Received in revised form 3 December 2018
Accepted 27 December 2018

Keywords:
Detection
Non-negative matrix factorization (NMF)
Divergence
Wheezing
Smoothness
Sparseness

ABSTRACT

The early wheezing detection is still a challenging task in biomedical signal processing because the presence of wheeze sounds often indicate respiratory diseases from airway obstructions. Currently, most of the first clinical examinations to detect any airway obstructions are carried out using auscultation. However, a high percentage of diagnoses are misdiagnosed since they are highly dependent on the physician's training in the wheezing detection, especially in noisy environments in which weak wheeze sounds can be masked by louder respiratory sounds. In this work, we propose a novel wheezing detection approach, based on Constrained Non-negative Matrix Factorization, that uses two-stage cascade: separation and detection. The novelty of the separation stage is to model wheeze and respiratory sounds as reliably as possible that they can be observed in the nature incorporating constraints (sparseness and smoothness) into the NMF factorization. Once the estimated wheezing and respiratory signal are obtained from the separation stage, the detection contribution is based on the use of the Kullback-Leibler divergence to discriminate between wheezing and respiratory areas. The experiments have been conducted using three different datasets composed of healthy or unhealthy patients. First, an optimization process is applied to obtain the optimal parameters of the separation stage. Finally, the performance of the wheezing detection of the proposed method is evaluated taking into account other state-of-the-art methods. Experimental results report that i) the proposed method outperforms recent state-of-the-art wheezing detection approaches showing a robust wheezing detection performance even evaluating noisy environments and ii) the ability of the proposal to reliably detect healthy patients.

1. Introduction

Ambient Assisted Living (AAL) based on acoustic event detection (AED) is currently a challenging topic that has attracted the attention of the signal processing community in the last years [1–4]. In fact, the analysis of respiratory sounds to diagnose patients suffering from lung diseases can be considered as a specific AED task applied in the field of biomedical signal processing [5,6] because it attempts to maximize the reliability of diagnoses by reducing the degree of subjectivity provided by physicians.

Specifically, the appearance of wheezing is an indicator of breathing problems shown by several obstructive pulmonary diseases, such as asthma, bronchiolitis, bronchitis or bronchiectasis, which affect people of all ages worldwide. These respiratory diseases are mainly caused by airway obstructions when air moves through narrowed or swelling breathing tubes in the lungs, reducing the amount of air that can pass through the air passages [7]. As an example, in 2015, there were nearly 400,000 deaths from asthma, most of which occurred in low- and middle-income countries according to the World Health Organization (WHO) [8]. As a consequence, the wheezing detection is one of the most challenging tasks in biomedical engineering and bio-signal processing research community.

Focusing on wheezing context, the aim of auscultation is to listen wheeze sounds emitted in the human breathing process using simple medical instrumentation such as stethoscopes. Currently, most of the first clinical examinations to detect respiratory diseases from airway obstructions are carried out by means of auscultation since it is a non-invasive, low-cost, easy-to-perform, patient-friendly and fast method regardless of age [9]. However, a high percentage of diagnoses are misdiagnosed since they are highly dependent on the physician's experience and acoustic training in the auscultation process. Therefore, it would be crucial to develop a robust wheezing detection system in order to early prevent complications resulting from misdiagnosis or undetected airway obstruction symptoms.

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

https://doi.org/10.1016/j.apacoust.2018.12.035
0003-682X/

It is well-known the temporal and spectral interference between wheeze and respiratory sounds during the inspiration and/or expiration of the breathing stages. The temporal interference is caused because wheeze sounds and respiratory sounds are simultaneously generated by the same airflow through the bronchial tree of the lungs. Instead, the spectral interference is due to the spectral overlapping problem that occurs in the spectral bands in which both types of sounds are active. Respiratory sounds (RS) are represented by a wide-band spectrum where most of the energy is concentrated in the frequency band 60 Hz-1000 Hz [10]. However, wheezing or wheeze sounds (WS) can be defined as continuous adventitious sounds that show a pitched sound generated on during the stages of the human breathing process. Specifically, WS show sinusoidal waveforms in time domain, a set of narrowband spectral peaks forming frequency lines over time (spectral trajectories) that superimpose on normal respiratory sounds in the frequency domain, whose pitch frequency is located in the spectral range of 100 Hz–1000 Hz with duration longer than 100 ms according to Computerized Respiratory Sound Analysis (CORSA) [11–13] as can be seen in Fig. 1. Specifically, WS can be classified into two main categories: monophonic and polyphonic. Monophonic WS are only composed of one fundamental frequency (pitch) but polyphonic WS are composed of one pitch and its harmonically-related frequencies that are located at approximately integer multiple of its associated pitch. In this paper, the term RS only refers to that sounds emitted by the lungs due to the breathing process without WS. The term WS only refers to that sounds composed of wheezing without RS. The term mixture refers to a signal composed of RS and WS.

Over the last two decades, many works have been reported to achieve wheezing detection system. Although several methods can be found in the state-of-the-art literature, which are based on different approaches such as auditory modelling [14], entropy [15], neural networks [16–18], wavelet transform [19,20], tonal index [21,22], mel-frequency cepstral coefficients (MFCC) [23–25] and classifiers [26–29], the most widely used methods to detect wheeze sounds are based on the extracted information provided by the analysis of the spectral peaks of the spectrum [30–35]. Alic et al. [31] modified the searching of spectral peaks using wavelet denoising in order to remove the noise in spectrum. In [32], an enhanced wheeze detector is developed which automatically locates and identifies wheezing-episodes during breath sound recordings based on the subtraction of the underlying breath sound, detection and classification of the detected peaks as wheezes and non-wheezes. Bahoura [24] reported pattern recognition methods to classify RS and WS using features based on Fourier transform, linear predictive coding, wavelet transform and Mel-frequency cepstral coefficients (MFCC) in combination with the classification methods based on vector quantization, Gaussian mixture models (GMM) and artificial neural networks. Wisniewski and Zielinski [21] analysed wheezes detection in normal breath sound as a problem of detection of multi-tones in colored noise using a set of robust descriptors. In [26], a two-layer pattern recognition system architecture for asthma wheezing detection is developed. The first layer consists of two SVM classifiers specifically designed as a cascade stacked in parallel using features based on MFCC. The second layer is realized using a digital detection threshold, with the aim of improving the process of wheezing detection. Shaharum et al. [28] detected wheezing to classify different levels of asthma severity using a feature extraction based on MFCC in combination with the classification method based on the K-nearest neighbour (KNN) algorithm. Recently, Oletic and Vilas [29] presented a wheezing detector by iteratively detecting individual wheezing frequency lines by modelling them using HMM. Its robustness was dominantly influenced by the robustness of frequency line tracking, the method of temporal localization of frequency line and the implementation of the line subtraction. Nevertheless, the main drawback of the most wheezing detection methods is the assumption that spectral peaks of WS are louder than the RS with which they are acoustically mixed [36,30,13,31,20,37,21,38,27].

Because non-negative matrix factorization has the ability to find hidden spectral patterns [39,43] by means of parts-based representation with non-negativity of the data, we propose a novel constrained non-negative matrix factorization approach to model and detect the presence of wheezing, locating the temporal intervals in which WS are active when they are mixed with RS in mono-channel audio mixtures. As far as the authors knowledge extends, non-negative matrix factorization approach has never been applied before to wheezing detection. The main contribution of this work is a robust wheezing detector based on the estimated respiratory and wheezing signals obtained by the proposed constrained NMF. Thus, the proposed method is able to detect and separate the wheezing and respiratory signals from the mixture. In fact, the mixture is decomposed using a cost function to reconstruct the mixture adding typical spectro-temporal behaviors observed in most WS and RS in real life. We assume that WS can be modeled using sparseness in frequency since WS usually show one or a few harmonically-related narrow peaks in the spectrum. We assume that RS are represented as a wide-band spectrum that

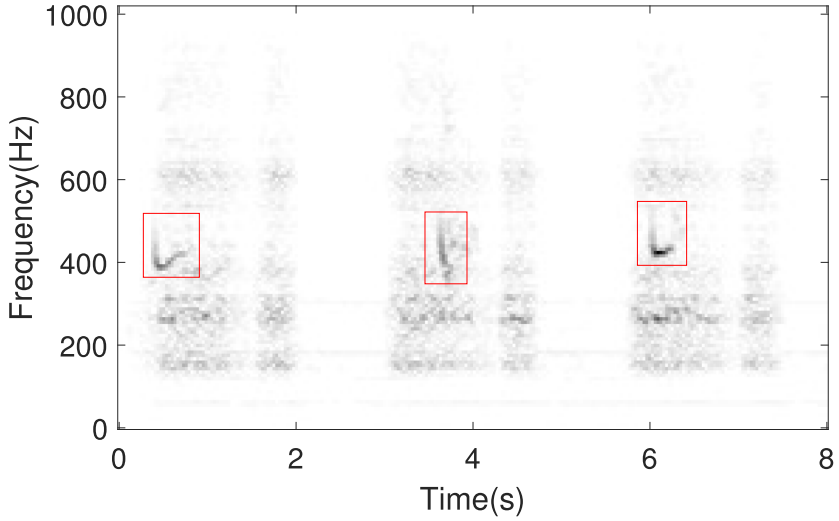

Fig. 1. Spectrogram of a mixture composed of three wheeze sounds (red rectangles) and respiratory sounds. A darker grey colour represents higher energy of each frequency.

shows smoothness (slow variation) in time and frequency. Specifically, RS are the only sounds of the mixture that is not overlapped in some areas (particularly, in the areas of the respiratory cycle where wheezing is inactive). Therefore, we propose the use of the Kullback-Leibler divergence applied to the input spectrogram and the estimated respiratory spectrogram to discriminate between wheezing and respiratory temporal intervals (areas). The Kullback-Leibler divergence will have a very small value in areas where only the RS are present. However, the Kullback-Leibler divergence will have a high value in areas where RS and WS are active. Our proposal is a completely blind method because does not require information about the number of sound sources neither training of the sounds to detect.

The remainder of this paper is organized as follows. In Section 2, a description about the fundamentals of non-negative matrix factorization is briefly presented. We subsequently propose a robust method to detect wheeze sounds in Section 3. In Section 4, an optimization of the main parameters of the separation stage is carried out in order to maximize the detection performance of the proposed method. Finally, we conclude in Section 5 and provide perspectives on further research.

2. Non-negative matrix factorization

Recently, non-negative Matrix Factorization (NMF) or unconstrained NMF [39,40] has attracted a lot of attention in the field of biomedical signal processing [41–44] because it provides parts-based representation of the most representative objects by imposing non-negative constraints that allow only additive combinations of the input data. Specifically, NMF factorizes the input magnitude spectrogram $\mathbf{X}_{F,T}$ of a mixture signal $x(t)$ into the product of two non-negative estimated matrices, basis matrix $\hat{\mathbf{B}}_{F,K}$ and activation matrix $\hat{\mathbf{A}}_{K,T}$ (see Eq. (1)),

$$\mathbf{X}_{F,T} \approx \hat{\mathbf{X}}_{F,T} = \hat{\mathbf{B}}_{F,K}\hat{\mathbf{A}}_{K,T} \tag{1}$$

where $\hat{\mathbf{X}}_{F,T}$ is the estimated or reconstructed spectrogram and the variables F, T and K represent the number of frequency bins, the number of time frames and the rank or the number of components (generally, $FK + KT \ll FT$ in order to reduce the dimensions of the data). The columns of the estimated basis matrix $\hat{\mathbf{B}}$ are basis functions (or spectral patterns) and the rows of the estimated activation matrix $\hat{\mathbf{A}}$ represent the temporal intervals in which the previous basis functions are active. The NMF factorization is performed by minimizing a cost function $D\left(\mathbf{X}|\hat{\mathbf{X}}\right)$,

$$D\left(\mathbf{X}|\hat{\mathbf{X}}\right) = \sum_{f=1}^{F}\sum_{t=1}^{T} d\left(X_{f,t}|\widehat{X}_{f,t}\right) \tag{2}$$

where $d(i,j)$ is a function of two scalar variables i, j. Some of the most widely used cost functions applied to audio processing are the Euclidean distance, the generalized Kullback-Leibler divergence and the Itakura-Saito divergence [45]. The cost function $D\left(\mathbf{X}|\hat{\mathbf{X}}\right)$ is minimized, ensuring the non-negativity of the bases and the activations using an iterative algorithm based on multiplicative update rules [45]. Given a parameter $\mathbf{Z}$, its multiplicative update rules are obtained calculating the partial derivatives of the cost function as follows,

$$\mathbf{Z} = \mathbf{Z} \odot \frac{\left[\frac{\partial D\left(\mathbf{X}|\hat{\mathbf{X}}\right)}{\partial \mathbf{Z}}\right]^{-}}{\left[\frac{\partial D\left(\mathbf{X}|\hat{\mathbf{X}}\right)}{\partial \mathbf{Z}}\right]^{+}} \tag{3}$$

where $\odot$ is the element-wise multiplication.

However, the main problem of the NMF is the trade-off between signal reconstruction and physical interpretation of the factorized parts-based objects. In other words, the non-negativity can only ensure convergence to local minima that implies the reconstruction of the mixture but the solution may not make physical sense as these type of sounds can be found in real life [46,47]. The previous problem can be overcome adding some prior information into the factorization procedure. This prior information can be added by means of constraints. As result, constraints help to find better local minima incorporating physical sense to the basis or activations into the factorization procedure. Next, two specific constraints used in our proposal (detailed in Section 3) are briefly explained: sparseness and smoothness.

2.1. Sparseness

In general terms, sparseness ψ means that the sources can be considered inactive most of the time or frequency [48,49]. Therefore, sparseness can be applied to the estimated NMF basis or activation matrices. Temporal sparseness $\psi\left(\hat{\mathbf{A}}\right)$, applied to estimated activation matrix $\hat{\mathbf{A}}$, assumes that the sources are inactive most of the time so, a high cost is assigned to nonzero activations. Spectral sparseness $\psi\left(\hat{\mathbf{B}}\right)$, applied to estimated basis matrix $\hat{\mathbf{B}}$, assumes that the sources are inactive most of the frequency so, a high cost is assigned to nonzero bases. As a result, sparseness constraint ψ often obtains better local minima in the NMF decomposition, minimizing the presence of false activations or bases. For example, considering the spectral sparseness,

$$D\left(\mathbf{X}|\hat{\mathbf{X}}\right) = D_{*}\left(\mathbf{X}|\hat{\mathbf{X}}\right) + \alpha\psi\left(\hat{\mathbf{B}}\right) \tag{4}$$

where $D_{*}\left(\mathbf{X}|\hat{\mathbf{X}}\right)$ is the reconstruction cost function to be minimized (i.e. Euclidean, Kullback-Leibler, ...) and $\psi\left(\hat{\mathbf{B}}\right)$ penalizes nonzero bases as previously mentioned. Specifically, one of the most used penalty term is the L^1-norm $\psi\left(\hat{\mathbf{B}}\right) = \left\|\hat{\mathbf{B}}\right\|_1$ as proposed in [49] because it was demonstrated to be less sensitive to changes of the parameter α that controls the importance of the constraint in the factorization process. In a similar way, the temporal sparseness $\psi\left(\hat{\mathbf{A}}\right)$ can be calculated but now taking into account the estimated activations matrix $\hat{\mathbf{A}}$.

2.2. Smoothness

Generally, smoothness ϕ means how continuous or smooth are the spectral or temporal changes related to a source [49]. Therefore, smoothness can be applied to the estimated NMF basis or activation matrices. Temporal smoothness $\phi\left(\hat{\mathbf{A}}\right)$, applied to estimated activation matrix $\hat{\mathbf{A}}$, reports how slow the amplitude variations over time are. Spectral smoothness $\phi\left(\hat{\mathbf{B}}\right)$, applied to estimated basis matrix $\hat{\mathbf{B}}$, indicates how slow the amplitude variations over frequency are. For example, consider temporal smoothness $\phi\left(\hat{\mathbf{A}}\right)$,

$$D\left(\mathbf{X}|\hat{\mathbf{X}}\right) = D_{*}\left(\mathbf{X}|\hat{\mathbf{X}}\right) + \lambda\phi\left(\hat{\mathbf{A}}\right) \tag{5}$$

where $D_{*}\left(\mathbf{X}|\hat{\mathbf{X}}\right)$ is the cost function, $\phi\left(\hat{\mathbf{A}}\right)$ is the function that penalizes abrupt temporal changes and parameter λ controls the importance of the smoothness constraint. According [49], temporal smoothness $\phi\left(\hat{\mathbf{A}}\right)$ of the components is enforced by assigning a

high cost to large changes in time between the activations $\widehat{A}_{f,t}$ and $\widehat{A}_{f,t-1}$ in adjacent frames as follows,

$$\phi\left(\hat{\mathbf{A}}\right)=\sum_{f=1}^{K}\frac{1}{\sigma_f^2}\sum_{t=2}^{T}\left(\widehat{A}_{f,t}-\widehat{A}_{f,t-1}\right)^2 \tag{6}$$

where the activations are normalised by their standard deviation $\sigma_f=\sqrt{\frac{1}{T}\sum_{t=1}^{T}\widehat{A}_{f,t}^2}$ to prevent the numerical scale of the activations from affecting the cost [49,46]. In a similar way, the spectral smoothness $\phi\left(\hat{\mathbf{B}}\right)$ is calculated but considering the spectral changes between the bases $\widehat{B}_{f,t}$ and $\widehat{B}_{f-1,t}$ in adjacent bins as follows,

$$\phi\left(\hat{\mathbf{B}}\right)=\sum_{t=1}^{K}\frac{1}{\sigma_t^2}\sum_{f=2}^{F}\left(\widehat{B}_{f,t}-\widehat{B}_{f-1,t}\right)^2 \tag{7}$$

where the bases are normalised by their standard deviation $\sigma_t=\sqrt{\frac{1}{F}\sum_{f=1}^{F}\widehat{B}_{f,t}^2}$ to prevent the numerical scale of the bases from affecting the cost [49,46].

3. The proposed method

The main problem to detect wheeze sounds from mixture is that both wheeze sounds and respiratory sounds occur simultaneously in frequency and time domain. Therefore, the goal of the proposed method is to improve the wheezing detection applying wheeze/respiratory sound separation. For this purpose, our proposed method consists of two-stage cascade: wheeze/respiratory sound separation (Stage I) and wheezing detection (Stage II). A block diagram of the proposed method is shown in Fig. 2.

3.1. Stage I. Wheeze/respiratory sound separation

The mixture $x(t)$ of unhealthy patients is composed of wheeze sounds $x_w(t)$ and respiratory sounds $x_r(t)$. We assume that the mixture of these sounds is additive $x(t)=x_r(t)+x_w(t)$. The input magnitude spectrogram $\mathbf{X}$ of the mixture can be represented as $\mathbf{X}=\mathbf{X}_R+\mathbf{X}_W$, being $\mathbf{X}_R$ the magnitude spectrogram of only respiratory sounds and $\mathbf{X}_W$ the magnitude spectrogram of only wheeze sounds. Each magnitude spectrogram, composed of T frames and F frequency bins, has been computed using the magnitude of the Short-Time Fourier Transform (STFT) using a Hamming window of size N with 25% overlap.

In order to make the proposed model independent of the size and scale of the input spectrogram $\mathbf{X}$, a normalization process is required. Thus, the normalized magnitude spectrogram $\mathbf{X}_n$ is computed as follows,

$$\mathbf{X}_n=\frac{\mathbf{X}}{\left(\frac{\sum_{f=1}^{F}\sum_{t=1}^{T}X_{f,t}}{FT}\right)} \tag{8}$$

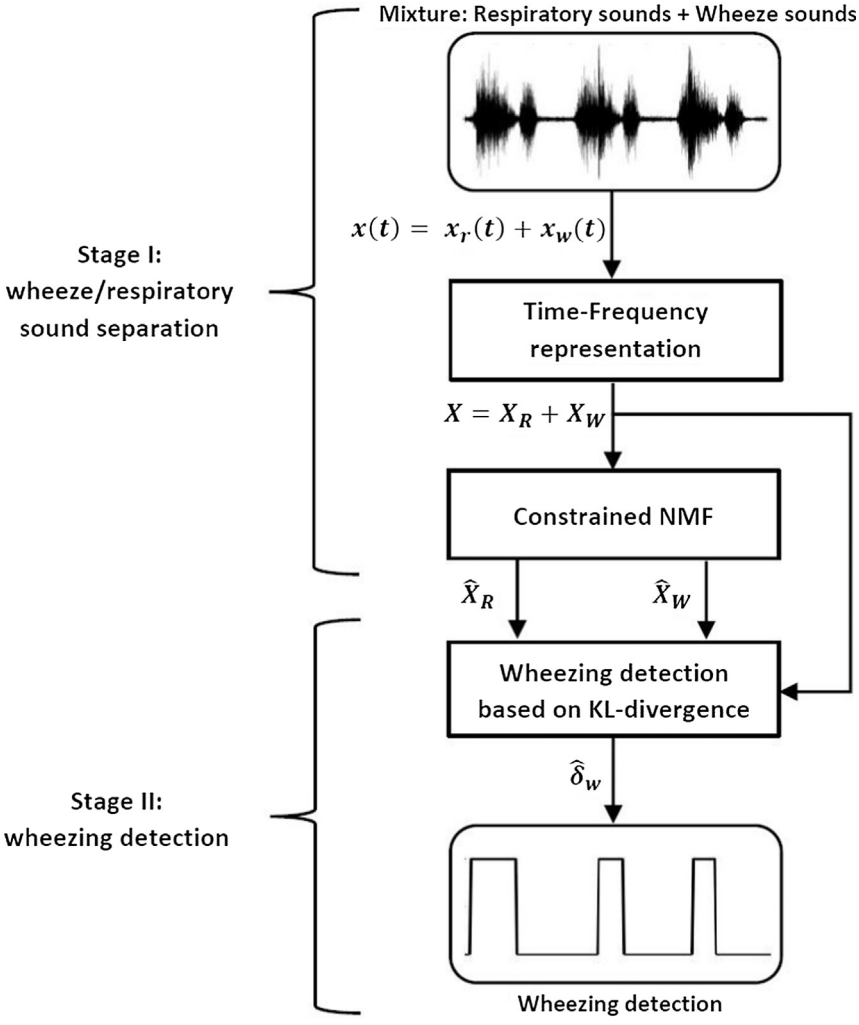

Fig. 2. Block-scheme of the proposed method.

The goal of this stage is to separate wheeze sounds and respiratory sounds. For this purpose, an objective function is defined to decompose a mixture normalized spectrogram $\mathbf{X}_n$ into two separated or estimated spectrograms, $\widehat{\mathbf{X}}_R$ (only respiratory sounds spectrogram) and $\widehat{\mathbf{X}}_W$ (only wheeze sounds spectrogram). The factorization of each spectrogram depends on the estimated basis matrix $\hat{\mathbf{B}}$ (frequency characteristics) and the estimated activation matrix $\hat{\mathbf{A}}$ (temporal characteristics).

$$\mathbf{X}_n = \mathbf{X}_R + \mathbf{X}_W = \mathbf{B}_R\mathbf{A}_R + \mathbf{B}_W\mathbf{A}_W \approx \hat{\mathbf{X}}_n = \hat{\mathbf{X}}_R + \hat{\mathbf{X}}_W = \hat{\mathbf{B}}_R\hat{\mathbf{A}}_R + \hat{\mathbf{B}}_W\hat{\mathbf{A}}_W \tag{9}$$

where $\mathbf{X}_R, \mathbf{X}_W$ are the original spectrograms of the respiratory and wheeze sounds; $\mathbf{B}_R, \mathbf{A}_R$ are the original basis and activation matrix of the respiratory sounds; $\mathbf{B}_W, \mathbf{A}_W$ are the original basis and activation matrix of the wheeze sounds; $\widehat{\mathbf{X}}_n$ is the estimated or reconstructed normalized spectrogram of the mixture; $\hat{\mathbf{B}}_R, \hat{\mathbf{A}}_R$ are the estimated basis and activation matrix of the respiratory sounds; $\hat{\mathbf{B}}_W, \hat{\mathbf{A}}_W$ are the estimated basis and activation matrix of the wheeze sounds. The number of respiratory and wheezing components will be denoted as K_r and K_w, respectively. All of these matrices are non-negative.

The goal of the NMF is to estimate the basis $\left(\hat{\mathbf{B}}_R, \hat{\mathbf{B}}_W\right)$ and activation matrices $\left(\hat{\mathbf{A}}_R, \hat{\mathbf{A}}_W\right)$ minimizing the reconstruction error between the input spectrogram $\mathbf{X}_n$ and the estimated spectrogram $\hat{\mathbf{X}}_n$. In this work, we propose to minimize the Kullback–Leibler divergence cost function $D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)$ (see Eq. (10)), because it has been successfully applied in audio signal processing [45,49].

$$D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right) = \sum_{f=1}^{F}\sum_{t=1}^{T} X_{n_{f,t}} \log\frac{X_{n_{f,t}}}{\widehat{X}_{n_{f,t}}} - X_{n_{f,t}} + \widehat{X}_{n_{f,t}} \tag{10}$$

As previously mentioned, NMF can only reconstruct the spectrogram of the input mixture but ensuring the convergence of the function to local minima. However, these local minima cannot discriminate between wheeze and respiratory bases so, the wheeze/respiratory sound separation is not successfully performed. In order to find a better NMF decomposition that shows spectro-temporal features of the WS and RS as can be observed in real life, we propose to incorporate constraints, sparseness and smoothness, into the NMF decomposition. Thus, we assume that RS can be considered smooth in time (slow variation of the magnitude spectrogram along time) and frequency (wideband spectrum). However, WS can be considered sparse in frequency because monophonic WS or polyphonic WS is characterized by one or more than one narrowband spectral peaks.

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

$$D\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right) = D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right) + \lambda_B\phi\left(\hat{\mathbf{B}}_R\right) + \lambda_A\phi\left(\hat{\mathbf{A}}_R\right) + \alpha_B\psi\left(\hat{\mathbf{B}}_W\right) \tag{11}$$

where λ_B defines the weight of the spectral smoothness $\phi\left(\hat{\mathbf{B}}_R\right)$ applied to only estimated respiratory basis matrix $\hat{\mathbf{B}}_R$, λ_A represents the weight of the temporal smoothness $\phi\left(\hat{\mathbf{A}}_R\right)$ applied to only estimated respiratory activation matrix $\hat{\mathbf{A}}_R$ and α_B is the weight of the spectral sparseness $\psi\left(\hat{\mathbf{B}}_R\right)$ applied to only estimated wheezing basis matrix $\hat{\mathbf{B}}_W$.

The estimated respiratory basis matrix $\hat{\mathbf{B}}_R$ (see Eq. (12)) and the estimated respiratory activation matrix $\hat{\mathbf{A}}_R$ (see Eq. (13)) can be obtained by applying a gradient descent algorithm [45] based on multiplicative update rules (see Eq. (3)). The equations of each term of the respiratory multiplicative update rules can be found in the Appendix A.

$$\hat{\mathbf{B}}_R = \hat{\mathbf{B}}_R \odot \frac{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_R}\right]^- + \lambda_B\left[\frac{\partial\phi\left(\hat{\mathbf{B}}_R\right)}{\partial \hat{\mathbf{B}}_R}\right]^-}{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_R}\right]^+ + \lambda_B\left[\frac{\partial\phi\left(\hat{\mathbf{B}}_R\right)}{\partial \hat{\mathbf{B}}_R}\right]^+} \tag{12}$$

$$\hat{\mathbf{A}}_R = \hat{\mathbf{A}}_R \odot \frac{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_R}\right]^- + \lambda_A\left[\frac{\partial\phi\left(\hat{\mathbf{A}}_R\right)}{\partial \hat{\mathbf{A}}_R}\right]^-}{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_R}\right]^+ + \lambda_A\left[\frac{\partial\phi\left(\hat{\mathbf{A}}_R\right)}{\partial \hat{\mathbf{A}}_R}\right]^+} \tag{13}$$

The estimated wheezing basis matrix $\hat{\mathbf{B}}_W$ (see Eq. (14)) and the estimated wheezing activation matrix $\hat{\mathbf{A}}_W$ (see Eq. (15)) can be obtained by applying a gradient descent algorithm [45] based on multiplicative update rules (see Eq. (3)). The equations of each term of the wheezing multiplicative update rules can be found in the Appendix A.

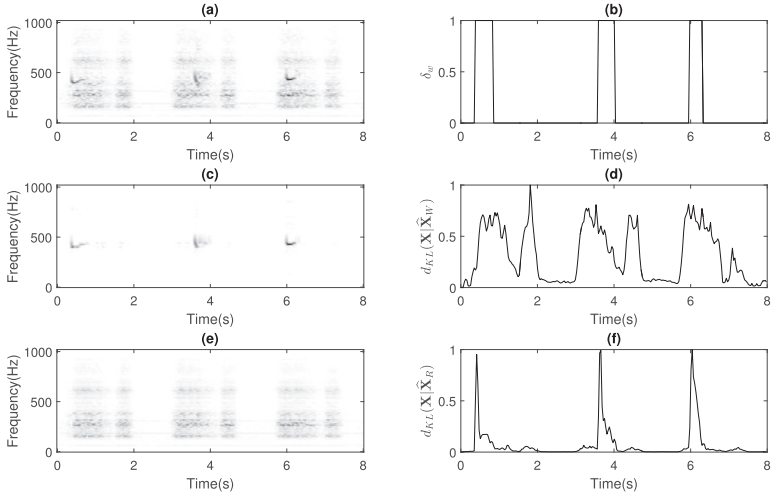

Fig. 3. **(a)** Magnitude spectrogram $\mathbf{X}$ of a mixture $x(t)$. **(b)** The ideal wheezing detection δ_w. **(c)** The estimated wheezing spectrogram $\hat{\mathbf{X}}_W$. **(d)** $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_W\right)$. **(e)** The estimated respiratory spectrogram $\hat{\mathbf{X}}_R$. **(f)** $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$. Note that $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_W\right)$ and $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ have been normalized to adjust the values between 0 and 1.

$$\hat{\mathbf{B}}_W = \hat{\mathbf{B}}_W \odot \frac{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_W}\right]^- + \alpha_B \left[\frac{\partial \psi\left(\hat{\mathbf{B}}_W\right)}{\partial \hat{\mathbf{B}}_W}\right]^-}{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_W}\right]^+ + \alpha_B \left[\frac{\partial \psi\left(\hat{\mathbf{B}}_W\right)}{\partial \hat{\mathbf{B}}_W}\right]^+} \tag{14}$$

$$\hat{\mathbf{A}}_W = \hat{\mathbf{A}}_W \odot \frac{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_W}\right]^-}{\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_W}\right]^+} \tag{15}$$

The estimated respiratory and wheezing basis and activation matrices are obtained updating the rules until the algorithm converges or reaches a maximum number of iterations M_{iter}. Next, the estimated magnitude spectrograms $\hat{\mathbf{X}}_R$ and $\hat{\mathbf{X}}_W$ can be obtained from the estimated basis and activation matrices (see Eq. (16) and (17)). Finally, the estimated magnitude spectrograms $\hat{\mathbf{X}}_R$ and $\hat{\mathbf{X}}_W$ are denormalized multiplying by the denominator of Eq. (8). The wheeze/respiratory sound separation procedure is summarized in Algorithm 1.

$$\hat{\mathbf{X}}_R = \hat{\mathbf{B}}_R \hat{\mathbf{A}}_R \tag{16}$$

$$\hat{\mathbf{X}}_W = \hat{\mathbf{B}}_W \hat{\mathbf{A}}_W \tag{17}$$

Algorithm 1 Stage I: wheeze/respiratory sound separation procedure

Require: $x(t), K_r, K_w, \lambda_B, \lambda_A, \alpha_B$ and M_{iter}.

1 Compute the normalized magnitude spectrogram $\mathbf{X}_n$ of the mixture signal $x(t)$ using Eq. (8).

2 Initialize $\hat{\mathbf{B}}_R, \hat{\mathbf{A}}_R, \hat{\mathbf{B}}_W$ and $\hat{\mathbf{A}}_W$ with random nonnegative values.

3 Update the estimated respiratory basis matrix $\hat{\mathbf{B}}_R$ using Eq. (12).

4 Update the estimated respiratory activation matrix $\hat{\mathbf{A}}_R$ using Eq. (13).

5 Update the estimated wheezing basis matrix $\hat{\mathbf{B}}_W$ using Eq. (14).

6 Update the estimated wheezing activation matrix $\hat{\mathbf{A}}_W$ using Eq. (15).

7 Repeat steps 3–6 until the algorithm converges (or until the maximum number of iterations M_{iter} is reached).

8 Compute magnitude estimated spectrograms $\hat{\mathbf{X}}_R$ and $\hat{\mathbf{X}}_W$ using Eq. (16) and (17).

9 Denormalize magnitude estimated spectrograms $\hat{\mathbf{X}}_R$ and $\hat{\mathbf{X}}_W$ multiplying by the denominator of Eq. (8).

return $\hat{\mathbf{X}}_R$ and $\hat{\mathbf{X}}_W$

In order to optimize the separation stage (subSection 4.5), the estimated respiratory signal $\hat{x}_r(t)$ and the estimated wheezing signal $\hat{x}_w(t)$ are synthesized using the estimated magnitude spectrograms $\hat{\mathbf{X}}_R$ and $\hat{\mathbf{X}}_W$ with a Wiener filtering [50] and the inverse STFT. Indicate that the Wiener filtering ensures that the reconstruction process is conservative by means of the use of a respiratory $\mathbf{M}_R$ and wheezing $\mathbf{M}_W$ mask. The Wiener masks $\mathbf{M}_R$ ans $\mathbf{M}_W$ represent the relative energy contribution of each source with respect to the energy of the input mixture. These masks are defined as,

$$\mathbf{M}_R = \frac{\left|\hat{\mathbf{X}}_R\right|^2}{\left(\left|\hat{\mathbf{X}}_R\right|^2 + \left|\hat{\mathbf{X}}_W\right|^2\right)} \tag{18}$$

$$\mathbf{M}_W = \frac{\left|\hat{\mathbf{X}}_W\right|^2}{\left(\left|\hat{\mathbf{X}}_R\right|^2 + \left|\hat{\mathbf{X}}_W\right|^2\right)} \tag{19}$$

In order to obtain the estimated complex spectrograms of the separated signals, each mask is multiplied by the complex spectrogram $\mathbf{X}_c$ of the mixture $x(t)$ as follows,

$$\hat{\mathbf{X}}_R = \mathbf{M}_R \odot \mathbf{X}_c \tag{20}$$

$$\hat{\mathbf{X}}_W = \mathbf{M}_W \odot \mathbf{X}_c \tag{21}$$

Finally, the inverse overlap-add STFT is applied to synthesize the estimated respiratory signal $\hat{x}_r(t)$ and the estimated wheezing signal $\hat{x}_w(t)$ in time domain as follows,

$$\hat{x}_r(t) = IDFT\left(\hat{\mathbf{X}}_R\right) \tag{22}$$

$$\hat{x}_w(t) = IDFT\left(\hat{\mathbf{X}}_W\right) \tag{23}$$

3.2. Stage II. Wheezing detection

The goal of this stage is to determine the presence of WS and the temporal intervals or areas in which wheezing is active from the estimated respiratory $\hat{\mathbf{X}}_R$ and wheezing $\hat{\mathbf{X}}_W$ spectrograms obtained in the stage I. Thus, the estimated wheezing detection $\hat{\delta}_w$ is defined (frame-by-frame) to assign the value 1 to the temporal frames where wheezing is active and the value 0 when the wheezing is inactive.

$$\hat{\delta}_w(t) = \begin{cases} 1 & \text{frame with wheezing active} \\ 0 & \text{frame with wheezing inactive} \end{cases} \tag{24}$$

where $t = 1,\ldots T$, being T the number of frames.

Initially, the proposed method performs a preliminary power analysis of the estimated wheezing spectrogram $\widehat{P}_W = |\hat{\mathbf{X}}_W|^2$ to classify whether the input spectrogram $\mathbf{X}$ corresponds to a healthy or unhealthy patient.

In the case of healthy patients, $\widehat{P}_W$ will be equal to 0 because the constrained NMF does not find wheeze bases that can be factorized using the spectral sparseness constraint. Therefore, the output of the wheezing detection stage assigns $\hat{\delta}_w = 0$ to all frames.

In the case of unhealthy patients, $\widehat{P}_W$ will be higher to 0. In this scenario, we propose the use of the Kullback-Leibler divergence $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ frame-by-frame between input spectrogram $\mathbf{X}$ and estimated respiratory spectrogram $\hat{\mathbf{X}}_R$ to discriminate wheezing areas as can be observed in Eq. (25),

$$d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)_t = \sum_{f=1}^{F} X_{f,t} \log \frac{X_{f,t}}{\widehat{X}_{R_{f,t}}} - X_{f,t} + \widehat{X}_{R_{f,t}}, \tag{25}$$

where $t = 1,\ldots T$, being T the number of frames.

The Kullback-Leibler divergence has an interesting property. It is in particular scale-invariant, meaning that low energy components of $\mathbf{X}$ bear the same relative importance as high energy ones. So, although the separation stage provides satisfactory results both in the estimated wheezing (Fig. 2 and respiratory (Fig. 3e) spectrograms, it must be highlighted that the RS are the only sounds of the mixture that is not overlapped in some areas (particularly, in the areas of the respiratory cycle where wheezing is inactive), as can be seen in Fig. 3a. For this reason, $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ is the appropriate way to discriminate the wheezing areas from respiratory areas.

As a result, $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ will have a very small value in areas where only the RS are active. However, in the areas where RS and WS sounds are active, $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ provides a high value (see Fig. 3f).

On the other hand, $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_W\right)$ does not allow to discriminate the wheezing areas, because the WS are always overlapping in the mixture with the RS. For this reason $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_W\right)$ tends to obtain high values, except in the breathing silences (when the airflow is inactive), as can be observed in Fig. 3d.

From $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$, a threshold is defined to determine the wheezing areas. In this paper, we have used the Otsu algorithm [51] to define the threshold ζ_{otsu}. As shown in Fig. 4, the Otsu algorithm allows to define an optimal threshold ζ_{otsu} that clearly differentiates two groups in a histogram, in this case, wheezing and respiratory (non-wheezing) areas. Finally, each frame of the unhealthy mixture $\mathbf{X}$ is labelled as wheezing (wheezing active) or not (wheezing inactive) by means of $\hat{\delta}_w$, combining $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ and the threshold ζ_{otsu} as follows,

$$\hat{\delta}_w(t) = \begin{cases} 1 & \text{if } d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right) \geqslant \zeta_{otsu} \\ 0 & \text{if } d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right) < \zeta_{otsu} \end{cases} \tag{26}$$

where $t = 1,\ldots T$, being T the number of frames.

The wheezing detection procedure is summarized in Algorithm 2.

Algorithm 2 Stage II: wheezing detection procedure

Require: $\mathbf{X}, \widehat{\mathbf{X}}_W$ and $\widehat{\mathbf{X}}_R$.

Compute the power of the estimated wheezing spectrogram $\widehat{P}_W$.

———————— **Healthy patients** ————————

if $\widehat{P}_W = 0$ **then**

return $\hat{\delta}_w = 0, \forall t$.

end if

———————— **Unhealthy patients** ————————

if $\widehat{P}_W \neq 0$ **then**

1 Compute KL-divergence d_{KL} between input spectrogram $\mathbf{X}$ and estimated respiratory spectrogram $\widehat{\mathbf{X}}_R$ using Eq. (25).

2 Compute histogram of $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$.

3 Compute threshold ζ_{otsu}.

4 Combine $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ and ζ_{otsu} to obtain the estimated wheezing detection $\hat{\delta}_w$ using Eq. (26).

return $\hat{\delta}_w$

end if

4. Experimental results

In this section, we asses the potential of the proposed method in the field of wheezing detection applied to mono-channel audio mixtures. As can be seen, an optimization process will be applied to obtain the optimal parameters of its separation stage. Finally, the performance of the wheezing detection of the proposed method is evaluated taking into account other state-of-the-art methods.

4.1. Dataset

Three datasets have been used in the experiments of the proposed method. The dataset E1 has been used in the optimization of the separation stage. The datasets T1 and T2 have been used in the wheezing detection testing. It must be highlighted that the optimization dataset is not a part of the test datasets (T1 and T2) in order to validate the results. The datasets E1, T1 and T2 can be seen in Table 1.

To optimize the parameters of the separation stage, the dataset E1 have been created mixing only WS 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 (in which wheezing is inactive) obtained from widely used Internet pulmonary data sources [29,52–59]. This dataset is composed of 48 mixtures which show a signal-to-noise ratio (SNR) between 0 dB and 9 dB, with duration between 5 and 24 s, with a total of 92 wheezing and 154 respiratory cycles.

The dataset T1 is the same dataset evaluated in [29] which has been directly shared from its authors. This dataset is composed of 16 recordings of healthy and unhealthy patients, with duration between 4 and 51 s. Specifically, the first part of the dataset consisted of 8 recordings from unhealthy patients, containing a total of 36 intervals of intermittent wheezing dispersed in 31 respiratory cycles. The second part of the dataset consisted of 8 recording from healthy patients, composed of 40 respiratory cycles of normal respiratory sounds.

In order to evaluate the robustness of the proposed detection method, we have created three datasets T2H (SNR = 5 dB), T2M

Table 1
Characteristics of datasets.

Identifier	Type of patients	SNR(dB)
	Optimization dataset	
E1	Unhealthy	[0–9]
	Test datasets	
T1	Unhealthy/Healthy	[2–8]
T2H	Unhealthy	5
T2M	Unhealthy	0
T2L	Unhealthy	−5

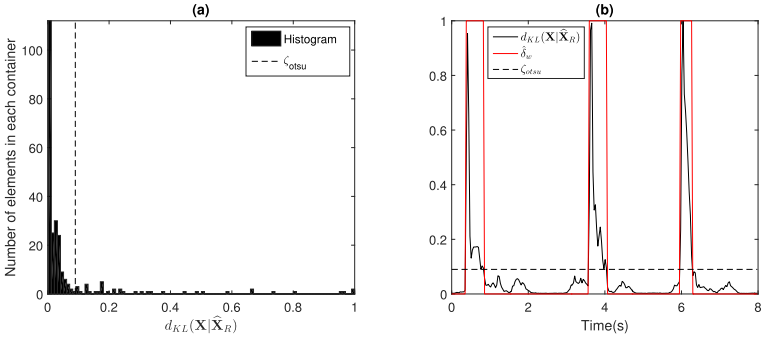

Fig. 4. **(a)** Histogram applied to $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ and threshold ζ_{otsu}. **(b)** Estimated wheezing detection $\hat{\delta}_w$ using $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ and ζ_{otsu}. Note that $d_{KL}\left(\mathbf{X}|\hat{\mathbf{X}}_R\right)$ has been normalized to adjust the values between 0 and 1.

(SNR = 0 dB) and T2L (SNR = −5 dB) with different signal-to-noise ratio (SNR) from the dataset T2 as can be seen in Table 1. The dataset T2 has been generated using a similar procedure and the same Internet data sources used in the creation of the dataset E1. However, the mixtures created in this datasets are not the same as that used in dataset E1 to provide valid results. Specifically, the datasets T2H, T2M and T2L are composed of 16 mixtures with duration between 7 and 22 s, with a total of 41 wheezing and 63 respiratory cycles.

4.2. Initializations

All mixtures were band-limited from 100 Hz to 1000 Hz (as previously mentioned, we assume that wheezes are not active below 100 Hz and above 1000 Hz). The other processing parameters are the following ones: sampling rate $fs = 2048Hz$, size of Hamming window $N = 256$ samples, with 25% overlap (temporal resolution of 31.3 ms). In addition, the convergence was empirically achieved after 120 iterations in all proposed NMF factorization. For this reason, the parameter $M_{iter} = 120$ has been used in this work.

4.3. Metrics

4.3.1. Separation

Three metrics are used to optimize the parameters of the separation stage [60,61], which are widely used in the field of sound source separation [43,46,49]: (1) the source-to-distortion ratio (SDR), which provides information on the overall quality of the separation process; (2) the source-to-interferences ratio (SIR), which is a measure of the presence of WS in the respiratory signal and vice versa; and (3) the source-to-artifacts ratio (SAR), which provides information on the artifacts in the separated signal from separation and/or resynthesis. In this paper, the SDR, SIR and SAR metrics have been calculated using the estimated wheezing signal $\hat{x}_w(t)$ obtained in the stage I.

4.3.2. Detection

Three metrics are used to evaluate the performance of the proposed method for the wheezing detection [62], which are widely used in the field of wheezing detection [27–29]: (1) sensitivity (SE), the probability of detecting wheezing frames correctly; (2) specificity (SP), the probability of detecting respiratory (without wheezing) frames correctly; and (3) accuracy (ACC), the probability of detecting wheezing/respiratory frames correctly.

4.4. Algorithms for comparison of wheezing detection results

We have used three recent state-of-the-art wheezing detection methods to evaluate the proposed method: HMMFL [29], TSVM [26] and MKNN [28]. The methods HMMFL, TSVM and MKNN have been implemented in this study, whereas the wheezing detection results provided by HMMFL (applied to the dataset T1) have been directly taken from [29]. The training dataset used by TSVM and MKNN is the same dataset E1 used in the separation optimization by the proposed method.

4.5. Separation optimization

An initial study showed that the detection results depend directly on the separation results. Specifically, the detection results increased considerably when the separation metrics improve. Therefore, an optimization process was motivated by these results to obtain the optimal separation parameters ($K_r, K_w, \lambda_B, \lambda_A$ and α_B) in order to maximize the detection performance of the proposed method.

Preliminary results of the optimization process indicated no significant SDR differences (lower than 0.3 dB) evaluating different number of wheezing K_w and respiratory K_r components, specifically, from 50 to 250 components, but the number of wheezing and respiratory components must be greater than 50 components in order to correctly model the spectral diversity of the wheezing and respiratory spectral patterns. In this work, $K_w = K_r = 150$ have been selected because preliminary results maximized the separation performance, in SDR, SIR and SAR, using this size of components.

Fig. 5 shows the optimization of the parameters λ_B, λ_A and α_B in order to analyse the effect of the smoothness and sparseness constraints in the optimization dataset E1. NMF (specifically, unconstrained NMF in which smoothness and sparseness constraints are disable, i.e., $\lambda_B = \lambda_A = \alpha_B = 0$) achieves a SDR average value of the estimated wheezing signal approximately equal to 2 dB (see Fig. 5a). This fact indicates that NMF does not properly separate WS and RS because the factorization model is not able to correctly discriminate between wheeze and respiratory bases. Fig. 5a confirms the previous claim since the minimum SDR values are reached when the sparseness constraint are disable ($\alpha_B = 0$). For this reason, the sparseness constraint can be considered crucial since it plays a fundamental role in the wheezing spectral model.

Results show that the maximum SDR value is obtained by enabling the smoothness and sparseness constraints simultaneously because these constraints are capable of modeling typical spectral and temporal features observed in RS and WS as can be found in real life. It can be observed that the maximum SDR value, approximately equal to 14 dB in Fig. 5c, is provided by the proposed method enabling the three proposed constraints, that is, $\lambda_B = 0.5, \lambda_A = 1$ and $\alpha_B = 3$ respectively. Highlight that the constrained NMF that uses the proposed constraints, smoothness and sparseness, produces a significant SDR improvement of approximately 12 dB compared to the unconstrained NMF (in which all proposed constraints are inactive). In this manner, the constrained NMF provides that the estimated wheezing signal exhibit common spectro-temporal features shown by WS, attenuating RS and vice versa considering the estimated respiratory signal. As a consequence, the factorized spectrograms $\hat{\mathbf{X}}_W$ and $\hat{\mathbf{X}}_R$ exhibit time-frequency energy distributions in a more accurate way than it can be found in real-world WS or RS. Besides, Fig. 5 shows that the SDR value, independently of α_B, drops when the values of λ_B and λ_A are very low or very high. On the one hand, when λ_B and λ_A values are very low, the signal reconstruction is performed without adding the common wheezing and respiratory spectro-temporal features to the bases/activations obtained from NMF. On the other hand, when λ_B and λ_A values are very high, the separation performance is not appropriate because it does not allow that NMF prioritizes the reconstruction of the WS and RS with the help of the previous constraints.

As shown in Table 2, in order to analyze the importance of each constraint $\phi\left(\hat{\mathbf{B}}_R\right), \phi\left(\hat{\mathbf{A}}_R\right)$ and $\psi\left(\hat{\mathbf{B}}_W\right)$ in the proposed separation method, eight possible parameters setups (enabling or disabling smoothness and sparseness constraints) have been defined. For each setup, the values of the weights λ_B, λ_A and α_B that maximize the optimization results have been selected.

Fig. 6 shows SDR, SIR and SAR results of the estimated wheezing signal for each parameters setup defined in Table 2. Each box represents 48 data points, one for each signal of the test dataset E1: each blue box represents the analysis for SDR values; each red box represents the analysis for SIR values; and each black box represents the analysis for SAR values. The lower and upper lines of each box show the 25th and 75th percentiles for the dataset E1. The line in the middle of each box represents the median value of the dataset E1. The lines extending above and below each box

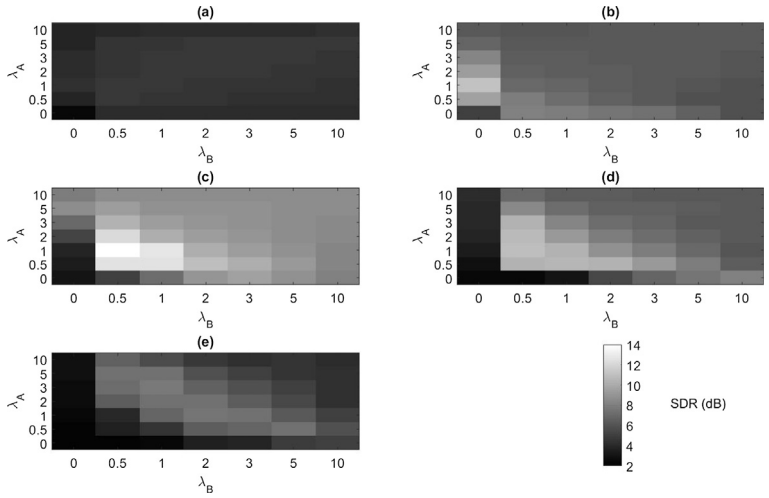

Fig. 5. SDR average results of the estimated wheezing signal provided by the hyperparametric analysis applied to the parameters optimization λ_B, λ_A and α_B in the dataset E1. **(a)** $\alpha_B = 0$, **(b)** $\alpha_B = 1$, **(c)** $\alpha_B = 3$, **(d)** $\alpha_B = 5$ and **(e)** $\alpha_B = 10$. The number of wheezing K_w and respiratory K_r components is equal to 150.

Table 2
Parameters setups in order to analyze the importance of the constraints $\phi\left(\hat{\mathbf{B}}_R\right)$, $\phi\left(\hat{\mathbf{A}}_R\right)$ and $\psi\left(\hat{\mathbf{B}}_W\right)$. the symbol – indicates constraint disabled. The symbol ✔ indicates constraint enabled.

Identifier	$\phi\left(\hat{B}_R\right)$	$\phi\left(\hat{A}_R\right)$	$\psi\left(\hat{B}_W\right)$	λ_B	λ_A	α_B
S1	–	–	–	0	0	0
S2	✔	–	–	0.5	0	0
S3	–	✔	–	0	1	0
S4	–	–	✔	0	0	1
S5	✔	✔	–	1	1	0
S6	✔	–	✔	3	0	3
S7	–	✔	✔	0	1	1
S8	✔	✔	✔	0.5	1	3

show the extent of the rest of the samples, excluding outliers. Outliers are defined as points that are over 1.5 times the interquartile range from the sample median, which are shown as crosses. Results show, as mentioned above, that the worst results are obtained when all constraints are disabled (Fig. 6: S1). However, a significant improvement is obtained (between 1.5–2.9 dB in SDR, 2.2–6 dB in SIR and 0.5–1.4 dB in SAR) when one of the used constraint is enabled (Fig. 6: S2, S3 and S4). This fact indicates that the smoothness and sparseness constraints makes it possible to efficiently model the common spectro-temporal behavior of the sounds active in the mixture (WS and RS). Results suggest that the sparseness constraint is the constraint that seems to be more significant in the separation performance of the proposed method comparing the cases when only one constraint is active (S2, S3 or S4). Specifically, the setup S4 produces a significant improvement of 2.9 dB in SDR, 6 dB in SIR and 1.4 dB in SAR over the unconstrained NMF. In addition, this setup exceeds the optimization results in the case of enabling the two smoothness constrains S5 that characterize the RS. This seems to indicate that characterizing the WS, using the sparseness constraint, is more important than characterizing RS by means of smoothness constraints to separate both sounds in the proposed method. The best separation results are obtained when the smoothness and sparseness constraints are enabled in the setups S6, S7 and S8. Comparing the setups S7 and S8 in which the spectral sparseness constraint is enabled, results imply that the temporal smoothness constraint S7 is more significant that the spectral smoothness constraint S6. Finally, the optimal separation performance is achieved when all the constraints are enabled in S8. Thus, S8 produces a significant improvement of 11.9 dB in SDR, 14.4 dB in SIR and 7.6 dB in SAR over the unconstrained NMF. As can be observed, Fig. 6 shows the evolution

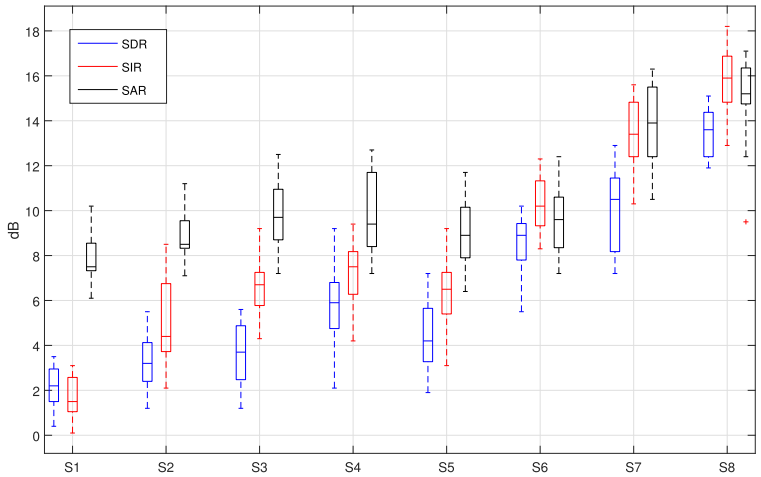

Fig. 6. Optimization results for the different parameters setups defined in Table 2.

of the optimization process from the worst results, obtained using the unconstrained NMF, until the best results achieved enabling all the proposed constraints. This fact confirms that the smoothness and sparseness constraints, used in the proposed method, are able to correctly model the WS and RS, incorporating physical meaning as can be found in nature and real life. Although SDR, SIR and SAR results exhibit an upward behavior, SIR results tend to overcome SDR results. Moreover, SAR results are often higher for all parameter configurations evaluated, indicating that the proposed method avoids introducing artifacts in the separation stage.

Finally, the optimal parameters that maximize the separation results have been used as a starting point of the detection stage to evaluate the wheezing detection performance of the proposed method: number of wheezing K_w and respiratory K_r components equal to 150; weight of the spectral smoothness $\lambda_B = 0.5$; weight of the temporal smoothness $\lambda_A = 1$; and weight of the spectral sparseness $\alpha_B = 3$.

4.6. Detection results

Table 3 shows sensitivity (SE), specificity (SP) and accuracy (ACC) results evaluating the dataset T1 between the proposed method and the previous state-of-the-art methods. Results report that the proposed method is competitive compared to the other evaluated methods. Specifically, the proposed method obtains the best SE and ACC results, 95.71% and 95.86%, respectively. Focusing on the SE metric, the proposed method achieves a significant improvement of approximately 6.4%, 10.4% and 14.85% compared to HMMFL, TSVM and MKNN. This fact indicates that the proposed method is more effective in correctly detecting wheezing frames compared to the state-of-the-art methods evaluated. However, the proposed method obtains the lowest SP result compared to the other methods. This fact suggests that the proposed method tends to provide a higher number of false positives (respiratory frames mistakenly detected as wheezing frames) in order to detect the whole wheezing temporal interval. Moreover, HMMFL obtains a better detection performance compared to TSVM taking into account SE, SP and ACC but TSVM outperforms the detection performance compared to MKNN. Highlight that the proposed method correctly detects all recordings of the dataset T1 corresponding to healthy patients. This fact confirms the reliability of the proposed method to discriminate healthy and unhealthy patients.

Table 4 shows sensitivity (SE), specificity (SP) and accuracy (ACC) results in order to evaluate the wheezing detection robustness of the proposed method and the state-of-the-art methods evaluated using three different SNR datasets (T2H, T2M and T2L). Evaluation indicates that the proposed method provides the best overall detection results compared to the other evaluated methods considering all SNR scenarios evaluated. Specifically, the proposed method outperforms the second best state-of-the-art method (TSVM), in terms of SE and ACC, about 9% and 3% in the dataset T2H, 9% and 4% in the dataset T2M and 14% and 8% in the dataset T2L. However, TSVM method obtains the highest SP results compared to the other methods. It can be observed that SE, SP and ACC detection results of the proposed method decreases an average of approximately 2.2% comparing T2H vs T2M, 1.0% considering T2M vs T2L and 3% considering T2H vs T2L. However, SE, SP and ACC detection results of the rest of the state-of-the-art methods decrease faster when the signal-to-noise ratio drops. Comparing the evaluated datasets with lower (T2H) and higher (T2L) noise: i) the SE reduction of the detection performance is about 3.9% (the proposed method), 8.8% (HMMFL), 8.5% (TSVM) and 8.6% (MKNN); ii) the SP reduction of the detection performance is about 2.8% (the proposed method), 9.7% (HMMFL), 5.6% (TSVM) and 5.7% (MKNN); iii) the ACC reduction of the detection performance is about 2.7% (the proposed method), 8.2% (HMMFL), 8.2% (TSVM) and 6.1% (MKNN). Results demonstrate the higher robustness of the proposed method in noisy environments compared to the other evaluated methods. This robustness shown by the proposed method suggests a greater ability in order to detect the presence of weak WS that can be masked by louder RS. Finally, a remarkable advantage of the proposed method is that it does not depend on any training data set due to its unsupervised nature.

Table 5 shows an analysis of the computational cost for each step of the proposed method, from the input of the mixture signal until the wheezing detection is provided. The number of elementary operations (Multiplications and Additions) and the number of elementary mathematical functions (Functions) such as divisions, square root, exponent, and logarithm are modeled. Note that the parameters on which the computational cost depends are the following: N, F, T, K_r, K_w and M_{iter}, previously defined, and the number of histogram bars H. Results indicate the majority of the cost comes from the Wheeze/respiratory sound separation stage, as can be observed in Table 5. In our experiments, $N = 256$ samples, $F = 256$ bins, $K_r = 150$ components, $K_w = 150$ components, $M_{iter} = 120$ iterations, $H = 100$ bars and $T = 32$ frames per second of mixture signal, obtaining a total computational cost of 966.25 millions of multiplications per second, 1233.5 millions of additions per second and 180.42 millions of elementary mathematical functions per second. The computation cost (in seconds) of the proposed method which has been computed using Matlab on a PC with Intel Core i7-7700HQ CPU of 2.8 GHz and 16 GB of RAM is shown in Fig. 7. It can be observed that the computation cost of the proposed method increases with the size of the temporal duration of the input mixture. However, the processing factor P_F defined as the ratio between the computation cost and the temporal duration of the input mixture decreases with the size of the mixture. This fact guarantees that the computation cost of the proposed method is less than the temporal duration of the input mixture. Specifically, the computation cost of the proposed method to detect wheezing is approximately 4.8 s when mixtures of 30 s, are analyzed. Thus, the proposed method adds complementary

Table 3
Wheezing detection comparison between the proposed method and reference state-of-the-art methods [29] [26] [28] evaluating the dataset T1.

Algorithm	SE (%)	SP (%)	ACC (%)
Proposed Method	**95.71**	93.02	**95.86**
HMMFL	89.34	**96.28**	94.91
TSVM	85.32	95.36	90.59
MKNN	80.86	93.27	88.48

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

Table 4
Robustness wheezing detection performance comparison between the proposed method and reference state-of-the-art methods [29] [26] [28] evaluating the datasets T2H, T2M and T2L.

Algorithm	SE (%)	SP (%)	ACC (%)
Dataset T2H (SNR = 5 dB)			
Proposed Method	**99.48**	90.77	**97.41**
HMMFL	79.5	88.62	80.06
TSVM	90.38	**95.52**	94.66
MKNN	82.33	90.84	88.12
Dataset T2M (SNR = 0 dB)			
Proposed Method	**97.27**	88.60	**95.06**
HMMFL	73.83	85.86	76.86
TSVM	88.60	**92.46**	91.29
MKNN	78.68	87.62	86.94
Dataset T2L (SNR = −5 dB)			
Proposed Method	**95.57**	87.97	**94.70**
HMMFL	70.72	78.94	71.81
TSVM	81.93	**89.96**	86.45
MKNN	73.77	85.12	82.01

Table 5
Computational complexity of each step of the proposed method.

Steps	Multiplications	Additions	Functions
	Stage I: wheeze/respiratory sound separation procedure (Algorithm 1)		
$\mathbf{X}_n$	$T2N\log_2(N)$	$T3N\log_2(N)$	
$\hat{\mathbf{B}}_R$	$M_{iter}(9FK_r + FT(1+K_r) + K_r)$	$M_{iter}(FK_r(7+T))$	$M_{iter}(F(4K_r+T))$
$\hat{\mathbf{A}}_R$	$M_{iter}(9TK_r + FT(1+K_r) + K_r)$	$M_{iter}(TK_r(7+F))$	$M_{iter}(T(4K_r+F))$
$\hat{\mathbf{B}}_W$	$M_{iter}(6FK_w + FT(1+K_w) + 3K_w)$	$M_{iter}(FK_w(4+T))$	$M_{iter}(2K_w(1+F)+FT)$
$\hat{\mathbf{A}}_W$	$M_{iter}(K_wT(1+F)+FT)$	$M_{iter}(K_wT(F-1))$	$M_{iter}(K_wT)$
$\widehat{\mathbf{X}}_n$	$M_{iter}(FT(K_w+K_r))$	$M_{iter}(FT(2K_w+2K_r-3))$	
$\widehat{\mathbf{X}}_R$	FTK_r	FTK_r	
$\widehat{\mathbf{X}}_W$	FTK_w	FTK_w	
	Stage II: wheezing detection procedure (Algorithm 2)		
$\widehat{P}_W$	FT	FT	
$d_{KL}\left(X\middle\vert\widehat{X}_R\right)$	FT	$2FT$	$2FT$
$hist_{d_{KL}}$		HT	
ζ_{otsu}	$2H+H^2$	$2H+H^2$	$H+H^2$
$\hat{\delta}_w$		T	

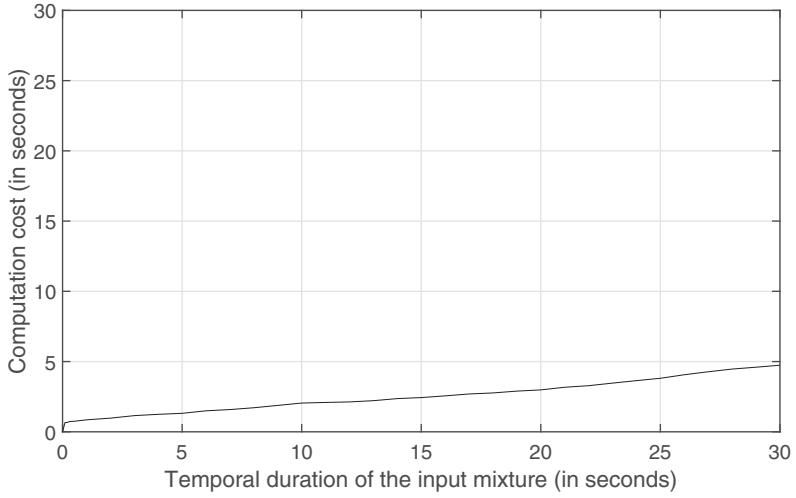

Fig. 7. Computation cost of the proposed method.

information for the physician diagnostic without slowing down the normal course of each physician consultation.

5. Conclusions and future work

In this paper, we propose a novel constrained non-negative matrix factorization approach to detect the presence of wheezing, locating the temporal intervals in which WS are active when they are mixed with RS in mono-channel audio mixtures. As far as the authors knowledge extends, non-negative matrix factorization approach has never been applied before to wheezing detection. The main contribution of the separation stage is to design a NMF framework adding typical spectro-temporal behaviors observed in most WS and RS in real life. The main contribution of the detection stage is the use of the Kullback-Leibler divergence applied to the estimated respiratory spectrogram obtained from constrained NMF to discriminate between wheezing and respiratory temporal intervals (areas).

The most relevant conclusions from the experimental results indicate the following: i) the wheezing detection performance of the proposed method is competitive compared to other state-of-the-art methods; ii) the robustness of the proposal is demonstrated because all detection metrics are reduced by a maximum of 3% comparing three datasets with a SNR difference of 5 dB between them. In this manner, results suggest that the proposed method can be an appropriate tool to be applied in noisy environments; and iii) the proposed method achieves a promising rate for detecting correctly between healthy and unhealthy patients due to wheezing.

Future work will focus on two directions: (i) novel spectro-temporal features to correctly discriminate between wheezing and non-wheezing (respiratory) bases/activations from NMF approaches, and (ii) alternative constrained non-negative matrix factorization approaches not applied to the whole input mixture spectrogram but in temporal segments for wheezing real-time detection.

Acknowledgment

The authors would like to thank Dr. Dinko Oletic and Dr. Vedran Bilas for sharing their test dataset. This work was supported by the Spanish Ministry of Economy and Competitiveness under Project TEC2015-67387-C4-2-R.

Appendix A. Terms of the multiplicative update rules

Here, each of the terms belonging to the respiratory multiplicative update rules are detailed:

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_R}\right]^{-} = \left[\hat{\mathbf{X}}_n^{-1} \odot \mathbf{X}_n\right]\hat{\mathbf{A}}_R' \tag{A.1}$$

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_R}\right]^{+} = \hat{\mathbf{A}}_R' \tag{A.2}$$

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_R}\right]^{-} = \hat{\mathbf{B}}_R'\left[\hat{\mathbf{X}}_n^{-1} \odot \mathbf{X}_n\right] \tag{A.3}$$

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_R}\right]^{+} = \hat{\mathbf{B}}_R' \tag{A.4}$$

$$\left[\frac{\partial \phi\left(\hat{\mathbf{B}}_R\right)}{\partial \hat{\mathbf{B}}_R}\right]^{+}_{f,k_r} = \frac{4F\widehat{B}_{R_{f,k_r}}}{\sum_{j=1}^{F}\widehat{B}^2_{R_{j,k_r}}} \tag{A.5}$$

$$\left[\frac{\partial \phi\left(\hat{\mathbf{A}}_R\right)}{\partial \hat{\mathbf{A}}_R}\right]^{+}_{k_r,t} = \frac{4T\widehat{A}_{R_{k_r,t}}}{\sum_{i=1}^{T}\widehat{A}^2_{R_{k_r,i}}} \tag{A.6}$$

$$\left[\frac{\partial \phi\left(\hat{\mathbf{B}}_R\right)}{\partial \hat{\mathbf{B}}_R}\right]^{-}_{f,k_r} = 2F\left[\frac{\left(\widehat{B}_{R_{f-1,k_r}} + \widehat{B}_{R_{f+1,k_r}}\right)}{\sum_{j=1}^{F}\widehat{B}^2_{R_{j,k_r}}}\right] + \frac{2F\widehat{B}_{R_{f,k_r}}\sum_{j=2}^{F}\left(\widehat{B}_{R_{j,k_r}} - \widehat{B}_{R_{j-1,k_r}}\right)^2}{\left(\sum_{j=1}^{F}\widehat{B}^2_{R_{j,k_r}}\right)^2} \tag{A.7}$$

$$\left[\frac{\partial \phi\left(\hat{\mathbf{A}}_R\right)}{\partial \hat{\mathbf{A}}_R}\right]^{-}_{k_r,t} = 2T\left[\frac{\left(\widehat{A}_{R_{k_r,t-1}} + \widehat{A}_{R_{k_r,t+1}}\right)}{\sum_{i=1}^{T}\widehat{A}^2_{R_{k_r,i}}}\right] + \frac{2T\widehat{A}_{R_{k_r,t}}\sum_{i=2}^{T}\left(\widehat{A}_{R_{k_r,i}} - \widehat{A}_{R_{k_r,i-1}}\right)^2}{\left(\sum_{i=1}^{T}\widehat{A}^2_{R_{k_r,i}}\right)^2} \tag{A.8}$$

Here, each of the terms belonging to the wheezing multiplicative update rules are detailed:

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_W}\right]^{-} = \left[\hat{\mathbf{X}}_n^{-1} \odot \mathbf{X}_n\right]\hat{\mathbf{A}}_W' \tag{A.9}$$

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{B}}_W}\right]^{+} = \hat{\mathbf{A}}_W' \tag{A.10}$$

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_W}\right]^{-} = \hat{\mathbf{B}}_W'\left[\hat{\mathbf{X}}_n^{-1} \odot \mathbf{X}_n\right] \tag{A.11}$$

$$\left[\frac{\partial D_{KL}\left(\mathbf{X}_n|\hat{\mathbf{X}}_n\right)}{\partial \hat{\mathbf{A}}_W}\right]^{+} = \hat{\mathbf{B}}_W' \tag{A.12}$$

$$\left[\frac{\partial \psi\left(\hat{\mathbf{B}}_W\right)}{\partial \hat{\mathbf{B}}_W}\right]^{+}_{f,k_w} = \frac{1}{\sqrt{\frac{1}{F}\sum_{j=1}^{F}\widehat{B}^2_{W_{j,k_w}}}} \tag{A.13}$$

$$\left[\frac{\partial \psi\left(\hat{\mathbf{B}}_W\right)}{\partial \hat{\mathbf{B}}_W}\right]^{-}_{f,k_w} = \sqrt{F}\frac{\widehat{B}_{W_{f,k_w}}\sum_{j=1}^{F}\widehat{B}_{W_{j,k_w}}}{\left(\sum_{j=1}^{F}\widehat{B}^2_{W_{j,k_w}}\right)^{\frac{3}{2}}} \tag{A.14}$$

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.apacoust.2018.12.035.

References

[1] Stowell D, Giannoulis D, Benetos E, Lagrange M, Plumbley MD. Detection and classification of acoustic scenes and events. IEEE Trans Multimedia 2015;17(10):1733–46.

[2] Cobos M, Perez-Solano JJ, Berger LT. Acoustic-Based Technologies for Ambient Assisted Living. In: Introduction to Smart eHealth and eCare Technologies. Boca Raton, USA: Taylor & Francis Group; 2016. p. 159–80.

[3] Alsina-Pages RM, Navarro J, Alias F, Hervas M. homeSound: Real-Time Audio Event Detection Based on High Performance Computing for Behaviour and Surveillance Remote Monitoring, Sensors, vol. 17, no. 4; 2017.

[4] Martin-Morato I, Cobos M, Ferri FJ. Analysis of data fusion techniques for multi-microphone audio event detection in adverse environments. In: 19th IEEE International Workshop on Multimedia Signal Processing (MMSP), Lutton, UK.

[5] Mondal A, Banerjee P, Tang H. A novel feature extraction technique for pulmonary sound analysis based on EMD. Comput Methods Programs Biomed 2018;159:199–209.

[6] Mondal A, Banerjee P, Somkuwar A. Enhancement of lung sounds based on empirical mode decomposition and Fourier transform algorithm. Comput Methods Programs Biomed 2017;139:119–36.

[7] MedlinePlus. Wheezing; 2018. [Online]. Available:https://medlineplus.gov/ency/article/003070.htm.

[8] World Health Organization, Chronic respiratory diseases; 2018. [Online]. Available:http://www.who.int/respiratory/asthma/en/.

[9] Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. Ann Thoracic Med 2015;10(3):158–68.

[10] Lozano F, Salazar A, Alvarado C. System of heart and lung sounds separation for store-and-forward telemedicine applications. Antioquia: Revista Facultad Ingennieria Univ; 2012.

[11] Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds advances beyond stethoscope. Am J Respir Crit Care Med 1997;156:974–87.

[12] Sovijarvi F, Dalmasso F, Vanderschoot J, Malmberg L, Righini G, Stoneman S. "Definition of terms for applications of respiratory sounds. Eur Respir Rev 2000;10:597–610.

[13] Lin B, Lin B, Wu H, Chong F, Chen S. Wheeze recognition based on 2D bilateral filtering of spectrogram. Biomed Eng Appl Basis Commun 2006;18(3).

[14] Qiu Y, Whittaker A, Lucas M, Anderson K. Automatic wheeze detection based on auditory modeling. Proc Inst Mech Eng 2005;219(3):219–27.

[15] Jin F, Sattar F, Goh D. Automatic wheeze detection using histograms of sample entropy. In 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2008. p. 1890–93.

[16] Forkheim K, Scuse D, Pasterkamp H. A comparison of neural network models for wheeze detection. In: WESCANEX 95. Communications, Power, and Computing. Conference Proceedings. IEEE, vol. 1; 1995. p.214–19.

[17] Lin B, Wu H, Chen S. Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network. J Healthcare Eng 2015;6(4):649–72.

[18] Kochetov K, Putin E, Azizov S, Skorobogatov I, Filchenkov A. Wheeze detection using convolutional neural networks. In: Progress in Artificial Intelligence. EPIA 2017. Lecture Notes in Computer Science, 10423. Springer; 2017.

[19] Kandaswamy A, Kumar C, Ramanathan R, Jayaraman S, Malmurugan N. Neural classification of lung sounds using wavelet coefficients. Comput Biol Med 2004;34:523–37.

[20] Le Cam S, Belghith A, Collet Ch, Salzenstein F. Wheezing sounds detection using multivariate generalized gaussian distributions. IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan; 2009.

[21] Wisniewski M, Zielinski T. Tonality detection methods for wheezes recognition system. 19th International Conference on Systems, Signals and Image Processing (IWSSIP); 2012. p 472–75.

[22] Wisniewski M, Zielinski T. 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.

[23] Chien J, Wu H, Chong F, Li C. Wheeze detection using cepstral analysis in Gaussian Mixture Models. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society Models. p. 3168–71.

[24] Bahoura M. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Comput Biol Med 2009;39 (9):824–43.
[25] Mayorga P, Druzgalski C, Morelos R, Gonzalez O, Vidales J. In: IEEE Engineering in Medicine and Biology Annual International Conference of the. p. 6312–6.
[26] Mazic I, Bonkovic M, Dzaja 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.
[27] 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.
[28] Shaharum S, Sundaraj K, Aniza S, Palaniappan R, Helmy K. Classification of asthma severity levels by wheeze sound analysis. IEEE Conference on Systems, Process and Control (ICSPC); 2016. p. 172–76.
[29] Oletic D, Bilas V. Asthmatic wheeze detection from compressively sensed respiratory sound spectra. IEEE J Biomed Health Inf 2017. https://doi.org/10.1109/JBHI.2017.2781135.
[30] Homs-Corbera A, Fiz J, Morera J, Jane R. Time-frequency detection and analysis of wheezes during forced exhalation. IEEE Trans Biomed Eng 2004;51 (1):182–6.
[31] Alic A, Lackovic I, Bilas V, Sersic D, Magjarevic R. A novel approach to wheeze detection. In: World Congress on Medical Physics and Biomedical Engineering 2006. Springer; 2007. p. 963–6.
[32] Taplidou SA, Hadjileontiadis LJ. Wheeze detection based on time-frequency analysis of breath sounds. Comput Biol Med 2007;37(8):1073–83.
[33] Jain A, Vepa J. "Lung sound analysis for wheeze episode detection", 30th. Ann Int IEEE EMBS Conference 2008:2582–5.
[34] 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.
[35] Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, Chouvarda I, Maglaveras N, et al. Detection of wheezes using their signature in the spectrogram space and musical features. In: Proceedings of the, 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC; 2015. p. 5581–84.
[36] Shreur H, Vanderschoot J, Zwinderman A, Dijkman JH J, Sterk P. The effect of methacholine-induced acute airway narrowing on lung sounds in normal and asthmatic subjects. Eur Respir J 1995;8:257–65.
[37] Nagasaka Y. Lung Sounds in Bronchial Asthma. Allergology Int 2012;61:353–63.
[38] Oletic D, Arsenali B, Bilas V. Low-power wearable respiratory sound sensing. Sensors 2014;14:6535–66.
[39] Lee D, Seung H. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401(6755):788–91.
[40] Lee D, Seung S. Algorithms for non-negative matrix factorization. In: Proceedings of Advances in Neural Inf. Process. System; 2000. p. 556–62.
[41] Damon C, Liutkus A, Gramfort A, Essid S. Non-negative matrix factorization for single-channel EEG artifact rejection. In: IEEE International Conference on Acoustics, Speech and Signal Processing. p. 1177–81.
[42] Tsubakida H, Shiratori T, Ishiyama A, Ono Y. Nonnegative matrix factorization common spatial pattern in brain machine interface. In: 3rd International Winter Conference on Brain-Computer Interface. p. 1–4.
[43] 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.
[44] Torre-Cruz J, Canadas-Quesada F, Vera-Candeas P, Montiel-Zafra V, Ruiz-Reyes N. Wheezing sound separation based on constrained non-negative matrix factorization. In: Proc ICBBT 10th International Conference on Bioinformatics and Biomedical Technology. p. 18–24. https://doi.org/10.1145/3232059.3232072.
[45] Févotte C, Bertin N, Durrieu J. Nonnegative matrix factorization with the Itakura-Saito divergence with application to music analysis. Neural Comput 2009;21(3):793–830.
[46] Canadas F, Vera P, Ruiz N, Carabias J, Cabanas P. Percussive harmonic sound separation by non-negative matrix factorization with smoothness-sparseness constraints. J Audio Speech Music Process 2014;2014(26):1–17.
[47] Laroche C, Kowalski M, Papadopoulos H, Richard G. A structured nonnegative matrix factorization for source separation. In: 23rd European Signal Processing Conference (EUSIPCO); 2015. p. 2033–37.
[48] Eggert J, Korner E. Sparse coding and nmf. Neural Networks 2004.
[49] Virtanen T. Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria. IEEE Trans Audio Speech Lang Process 2007;15(3):1066–74.
[50] Parras J, Canadas F, Vera P, Ruiz N. Audio restoration of solo guitar excerpts using a excitation-filter instrument model. Stockholm Music Acoustics Conference jointly with Sound And Music Computing Conference; 2013.
[51] Yuan X, Martnez JF, Echert M, pez-Santidrin L. An improved Otsu threshold segmentation method for underwater simultaneous localization and mapping-based navigation. Sensors 2016;16(7):1148.
[52] The r.a.l.e. repository. [Online]. Available:http://www.rale.ca/.
[53] Stethographics lung sound samples. [Online]. Available:http://www.stethographics.com/.
[54] 3m littmann stethoscopes. [Online]. Available:http://solutions.3m.com/.
[55] East tennessee state university pulmonary breath sounds. [Online]. Available: http://faculty.etsu.edu.
[56] ICBHI 2017 Challenge. [Online]. Available:https://bhichallenge.med.auth.gr/sites/default/.
[57] Lippincott NursingCenter. [Online]. Available:https://www.nursingcenter.com/.
[58] Thinklabs Digital Stethoscope. [Online]. Available:https://www.thinklabs.com/.
[59] Emedicine/Medscape. [Online]. Available:https://emedicine.medscape.com/.
[60] Fevotte C, Gribonval R, Vincent E. BSS_EVAL toolbox user guide – Revision, 2.0, Technical Report 1706, IRISA; (April 2005).
[61] Vicent E, Fevotte C, Gribonval R. Performance measurement in blind audio source separation. IEEE Trans Audio Speech Language Process 2006;14 (4):1462–9.
[62] Theodoridis S, Koutroumbas K. Pattern Recognition. Third edition. Elsevier: Academic Press; 2006.