

## Article

# Monophonic and Polyphonic Wheezing Classification Based on Constrained Low-Rank Non-Negative Matrix Factorization

Juan De La Torre Cruz <sup>1,\*</sup>, Francisco Jesús Cañadas Quesada <sup>1</sup>, Nicolás Ruiz Reyes <sup>1</sup>,  
Sebastián García Galán <sup>1</sup>, Julio José Carabias Orti <sup>1</sup> and Gerardo Pérez Chica <sup>2</sup>

<sup>1</sup> Department of Telecommunication Engineering, University of Jaen, Campus Científico-Tecnológico de Linares, Avda. de la Universidad, s/n, Linares, 23700 Jaen, Spain; fcanadas@ujaen.es (F.J.C.Q.); nicolas@ujaen.es (N.R.R.); sgalan@ujaen.es (S.G.G.); carabias@ujaen.es (J.J.C.O.)

<sup>2</sup> Pneumology Clinical Management Unit of the University Hospital of Jaen, Av. del Ejercito Espanol, 10, 23007 Jaen, Spain; gerardo.perez.sspa@juntadeandalucia.es

\* Correspondence: jtorre@ujaen.es

**Abstract:** The appearance of wheezing sounds is widely considered by physicians as a key indicator to detect early pulmonary disorders or even the severity associated with respiratory diseases, as occurs in the case of asthma and chronic obstructive pulmonary disease. From a physician's point of view, monophonic and polyphonic wheezing classification is still a challenging topic in biomedical signal processing since both types of wheezes are sinusoidal in nature. Unlike most of the classification algorithms in which interference caused by normal respiratory sounds is not addressed in depth, our first contribution proposes a novel Constrained Low-Rank Non-negative Matrix Factorization (CL-RNMF) approach, never applied to classification of wheezing as far as the authors' knowledge, which incorporates several constraints (sparseness and smoothness) and a low-rank configuration to extract the wheezing spectral content, minimizing the acoustic interference from normal respiratory sounds. The second contribution automatically analyzes the harmonic structure of the energy distribution associated with the estimated wheezing spectrogram to classify the type of wheezing. Experimental results report that: (i) the proposed method outperforms the most recent and relevant state-of-the-art wheezing classification method by approximately 8% in accuracy; (ii) unlike state-of-the-art methods based on classifiers, the proposed method uses an unsupervised approach that does not require any training.

**Keywords:** monophonic; polyphonic; wheezing; non-negative matrix factorization; spectral pattern; spectrogram; constraint; low-rank; asthma; chronic obstructive pulmonary disease



**Citation:** De La Torre Cruz, J.; Cañadas Quesada, F.J.; Ruiz Reyes, N.; García Galán, S.; Carabias Orti, J.J.; Pérez Chica, G. Monophonic and Polyphonic Wheezing Classification Based on Constrained Low-Rank Non-Negative Matrix Factorization. *Sensors* **2021**, *21*, 1661. <https://doi.org/10.3390/s21051661>

Academic Editors: Carlos Gómez and Raúl Alcaraz

Received: 25 January 2021

Accepted: 22 February 2021

Published: 28 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Chronic Respiratory Diseases (CRDs) are increasingly a huge and growing public health problem due to their high prevalence, high morbidity and mortality, and socio-economic cost. CRDs can be defined as disorders of the airways and other physiological structures of the respiratory system [1]. Some of the most common and relevant CRDs are asthma and Chronic Obstructive Pulmonary Disease (COPD). According to the World Health Organization (WHO), there were 417,918 deaths due to asthma at the global level in 2016 [2] and approximately three million people die from COPD every year, which is 6% of all deaths worldwide [3]. Although chronic diseases currently have no medical cure, early detection can lead to appropriate treatment when the disease is in its early stages, thus improving people's quality of life.

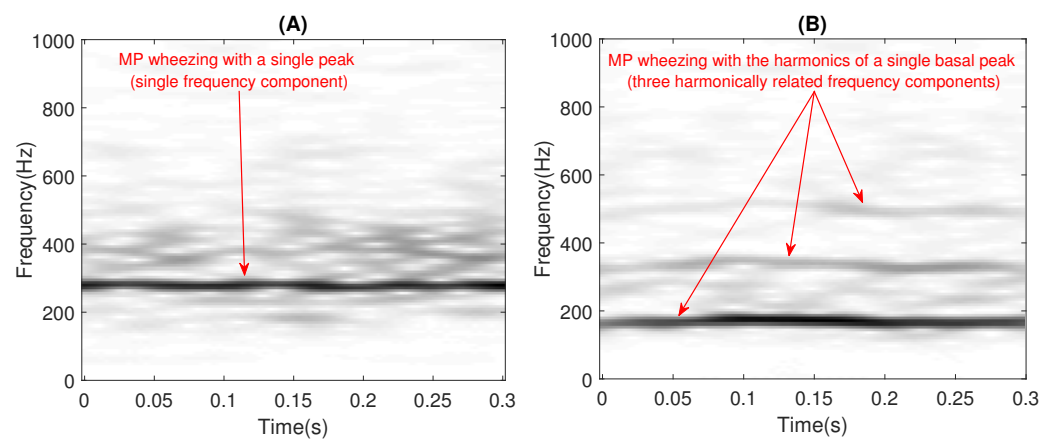
The auscultation examination is considered a widely used method of detecting CRDs because it is a non-invasive, inexpensive, easy, comfortable, and fast method regardless of age [4]. However, the auscultation process has several limitations that reduce the reliability of the diagnosis: (i) high subjectivity conditioned by the physician's training to recognize and interpret the sounds captured by the stethoscope [5,6]; (ii) the discrimination between

adventitious sounds with similar characteristics, such as monophonic and polyphonic wheezing sounds, is a harder task to perform by means of auscultation [7]; and (iii) normal respiratory sounds and adventitious sounds (abnormal and indicative of a lung disorder) are simultaneously mixed in the time and frequency domain, complicating the physician's analysis of the valuable clinical information contained in adventitious sounds [5,8,9]. Considering the above, a misdiagnosis is the main cause of the patient returning to the health center with a worsening of the disease that was not detected in the first medical examination performed by auscultation, so in recent years, it has become crucial to develop novel approaches to help physicians provide reliable diagnoses applied to lung disorders, with the implicit fact of reducing health care costs [10,11].

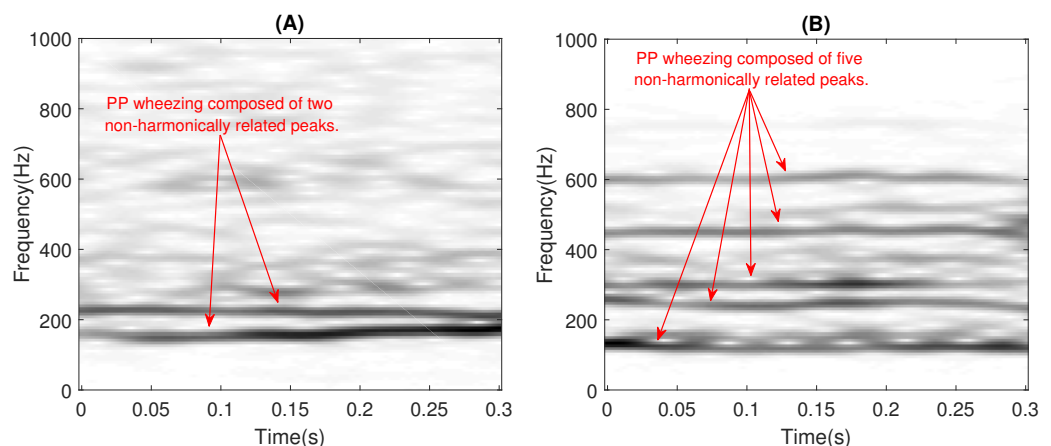
In general, the sounds generated during breathing can be classified into two main categories: normal respiratory sounds and adventitious sounds. Both sounds are mixed in the time-frequency domain as they are simultaneously generated by the same air flow through the bronchial tree of the lungs and also share part of the spectral bands in which they are active [5,9]. Normal respiratory sounds are represented by a wideband spectrum where most of the energy is concentrated in the frequency band of 60 Hz–1000 Hz [12]. Adventitious sounds can be classified into two categories: discontinuous and continuous sounds. Discontinuous Adventitious Sounds (DASs) are characterized by a short duration of less than 25 ms, such as coarse and fine crackles [13]. Continuous Adventitious Sounds (CASs) are characterized by a long duration of more than 100 ms, such as rhonchi, stridor, and wheezing [14]. In recent years, several works have been published that carried out an exhaustive review of lung acoustic measurements [15] and signal processing methods [16,17] applied to adventitious sounds, most of them focused on detection [16–23] and classification tasks [16,17,24–29].

Focusing on wheezing or wheeze sounds, the guidelines established by Computerized Respiratory Sound Analysis (CORSA) [5,30] define them as a pitch located between 100 Hz and 1000 Hz whose duration is greater than 100 ms, displaying trajectories of narrowband spectral peaks over time. The appearance of wheezing is widely considered by doctors as a clue to be able to detect either respiratory diseases or the severity associated with CRDs early, as occurs in the case of asthma and COPD [31,32]. For this reason, many research efforts have been applied in biomedical signal processing in order to develop reliable methods for early wheezing detection. In this sense, many wheezing detection algorithms, based on different approaches, can be found in the state-of-the-art literature: the Autoregressive (AR) model [33], auditory modeling [34], entropy [35], Neural Networks (NN) [36,37], wavelet transform [38,39], tonal index [40,41], Mel-Frequency Cepstral Coefficients (MFCCs) [42,43], Gaussian Mixture Models (GMMs) [44,45], spectral peaks identification [46–48], the Hidden Markov Model (HMM) [49], and recently, Non-negative Matrix Factorization (NMF) [9,50,51].

In addition, wheezing can be classified into two main categories according to the spectral behavior [52]: (i) wheezes that occur with a single peak or with the harmonics associated with that single basal peak are called Monophonic (MP) wheezes (as can be seen in Figure 1); and (ii) wheezes that occur with variable peaks that differ in harmonics are called Polyphonic (PP) wheezes (as can be seen in Figure 2). The scientific interest in the field of biomedical sound signal processing in automatically performing this classification lies in the fact that PP wheezes are usually caused by the pathology of small airways and MP wheezes are caused by the pathology of larger airways [53]. In fact, several studies [4,54–56] have shown that MP and PP wheezes exhibit distinctive physiological and pathological characteristics: (i) in physiological analysis, MP wheezes are caused by a single bronchial narrowing, while PP wheezes are caused by multiple central bronchial compression; and (ii) in pathological analysis, MP wheezes are an indicator of the presence of asthma, while PP wheezes can be considered as a sound marker of COPD.



**Figure 1.** Time-frequency representation of two examples of Monophonic (MP) wheezing: (A) with a single basal peak; (B) with the harmonics of a single basal peak. Note that the frequency components are harmonically related in (B).



**Figure 2.** Time-frequency representation of two examples of Polyphonic (PP) wheezing: (A) with two non-harmonically related peaks; (B) with five non-harmonically related peaks. Note that the frequency components are not harmonically related in the case of PP wheezing.

Despite advances in the analysis of respiratory sounds, MP/PP wheezing classification is a critical step in the diagnosis of asthma [4,54,55] and COPD diseases [54–56], so it is still a challenging topic in biomedical signal processing [7] since both types of wheezes are sinusoidal in nature. Although there are relatively few works [7,18,47,57–60] in which the analysis of MP/PP wheezing is treated, the only works focused on the task of classifying MP/PP wheezing in depth are [7,57,59,60] to our knowledge. All these MP/PP wheezing classification approaches are based on the feature extraction and classifier configuration. Ulukaya et al. [7] proposed to extract a single feature, the Peak Energy Ratio (PER), from a RAtional Dilation Wavelet Transform (RADWT) to discriminate between MP and PP wheezes. Specifically, PER is obtained from the first and second peak with the highest energy of all sub-bands of the wavelet coefficients (considering that the second peak is not consecutive to the first one). Moreover, the authors applied a robust evaluation methodology in which most of the relevant feature extraction methods [57,59,60] were evaluated using some of the most popular classifiers (SVM, KNN, and ELM) and Leave-One-Out (LOO) cross-validation schemes. The results reported that the proposed method, based on only one feature (PER), obtained the best MP/PP wheezing classification performance showing an accuracy equal to 86%.

However, none of the state-of-the-art methods consider the interference generated by normal respiratory sounds that can affect the MP/PP wheezing classification task. In this work, our proposal is based on the Non-negative Matrix Factorization (NMF)

approach in order to classify MP/PP wheezing sounds according to the harmonic structure shown by removing the sound interference caused by normal respiratory sounds. The first contribution of this work proposes a novel Constrained Low-Rank Non-negative Matrix Factorization (CL-RNMF) approach, which allows the spectral patterns associated with wheezing sounds to be extracted with the least possible sound interference from normal breath sounds. Specifically, we propose a low-rank configuration using a reduced number of wheeze bases to compact the frequency components into the fewest possible bases for further analysis without loss of relevant wheeze content. In addition, the proposed CL-RNMF approach incorporates a set of constraints to model the spectro-temporal behavior of wheezing and normal respiratory sounds. These constraints help to acoustically isolate the wheezing spectral patterns from normal respiratory sounds. To classify between MP or PP wheezing sounds, the second contribution analyzes the harmonic structure of the previous reduced number of wheezing bases based on the spectral location of the wheezing components, rather than the energy of their components.

The structure of the paper is as follows. Section 2 briefly reviews the principles of non-negative matrix factorization, focusing on the standard approach and some regularizations used to model the properties of the sounds of interest. The proposed MP/PP wheezing classification method is presented in Section 3. Section 4 details and discusses the experimental evaluation. Finally, we conclude in Section 5 and provide perspectives on further research.

## 2. Theoretical Background

### 2.1. Non-Negative Matrix Factorization

Non-negative Matrix Factorization (NMF) or standard NMF [61,62] is a decomposition technique that has attracted special attention in different fields of biomedical signal processing in the last few years [63,64]. Previous works showed the efficiency of the NMF approach at detecting [9,50,51] and improving the audio quality of wheezing [65,66]. In general terms, NMF can be defined as an unsupervised learning tool used for linear representation of non-negative two-Dimensional (2D) data where its main advantage is to reduce the dimensionality of a large amount of data in order to find hidden structures by means of part-based representation with non-negative patterns. From a mixture signal  $x(t)$ , its magnitude spectrogram  $\mathbf{X} \in \mathbb{R}_+^{F \times T}$  is obtained by means of the Short-Time Fourier Transform (STFT) applying a window function (e.g., Hamming or Hann) and inter-window overlap to increase the temporal resolution,  $F$  being the number of frequency bins and  $T$  the number of time frames. Here, standard NMF decomposes the magnitude spectrogram  $\mathbf{X}$  into the product of two non-negative matrices: spectral basis matrix (patterns)  $\mathbf{B} \in \mathbb{R}_+^{F \times K}$  and temporal activation matrix (weights)  $\mathbf{A} \in \mathbb{R}_+^{K \times T}$ ,  $K$  being the rank or the number of components (spectral bases),

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{B}\mathbf{A} \quad (1)$$

where  $\hat{\mathbf{X}} \in \mathbb{R}_+^{F \times T}$  is the estimated spectrogram. Each column of the basis matrix  $\mathbf{B}$  defines a spectral pattern that describes the spectral behavior of an active sound event in the input spectrogram  $\mathbf{X}$ . Each row of the activation matrix  $\mathbf{A}$  represents a temporal gain for a spectral pattern. In other words, the matrix  $\mathbf{B}$  provides a dictionary composed of  $K$  spectral bases, and the matrix  $\mathbf{A}$  defines the weight with which the different spectral bases appear along the temporal frames. Due to the non-negativity property, NMF underlies an additive linear interpolation model that results in the so-called part-based representation [61].

The decomposition or factorization of the input magnitude spectrogram  $\mathbf{X}$  into the product  $\mathbf{B}\mathbf{A}$  is usually sought minimizing a defined scalar-valued divergence,

$$\arg \min_{\mathbf{B}, \mathbf{A}} D(\mathbf{X} | \mathbf{B}\mathbf{A}) \quad \mathbf{B}, \mathbf{A} \geq 0 \quad (2)$$

This divergence function measures the error made in the approximation of the observed spectrogram  $\mathbf{X}$  and the reconstruction  $\mathbf{B}\mathbf{A}$ . Typically, the divergence is computed entry-wise:

$$D(\mathbf{X}|\hat{\mathbf{X}}) = D(\mathbf{X}|\mathbf{BA}) = \sum_{f=1}^F \sum_{t=1}^T d(X_{f,t}|\hat{X}_{f,t}) \quad (3)$$

where  $d(i, j)$  is a function of two scalar variables  $i, j$ . It is often called the cost function and is a positive function of  $i \in \mathbb{R}_+$  given  $j \in \mathbb{R}_+$  with a single minimum for  $i = j$ . Some of the most popular cost functions are the Euclidean distance, the generalized Kullback–Leibler divergence, the Itakura–Saito divergence, and the Cauchy distribution [67,68]. In this paper, we propose to minimize the generalized Kullback–Leibler divergence  $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$  (see Equation (4)) because previous works [9,50,51,63,65,66] obtained promising results in biomedical signal processing since  $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$  provides a scale-invariant factorization, that is low energy sound components of  $\mathbf{X}$  bear the same relative importance as high energy ones in the decomposition process.

$$D_{KL}(\mathbf{X}|\hat{\mathbf{X}}) = \sum_{f=1}^F \sum_{t=1}^T X_{f,t} \log \frac{X_{f,t}}{\hat{X}_{f,t}} - X_{f,t} + \hat{X}_{f,t} \quad (4)$$

The most popular minimization method to solve the problem shown in Equation (2) is based on the so-called multiplicative update rules, initially proposed by Lee and Seung [61]. This method obtains the basis and activation matrices, minimizing the Kullback–Leibler divergence function  $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$  and ensuring the non-negativity of the estimated matrices. These rules are obtained directly from the negative and positive terms of the partial derivative of the divergence function  $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$  with respect to the parameters  $\mathbf{B}$  and  $\mathbf{A}$ ,

$$\mathbf{B} \leftarrow \mathbf{B} \odot \frac{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}} \right]^-}{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}} \right]^+} = \mathbf{B} \odot \left( (\mathbf{X} \oslash \mathbf{BA}) \mathbf{A}^T \oslash ([\mathbf{1}] \mathbf{A}^T) \right) \quad (5)$$

$$\mathbf{A} \leftarrow \mathbf{A} \odot \frac{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}} \right]^-}{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}} \right]^+} = \mathbf{A} \odot \left( \mathbf{B}^T (\mathbf{X} \oslash \mathbf{BA}) \oslash (\mathbf{B}^T [\mathbf{1}]) \right) \quad (6)$$

where  $[\mathbf{1}] \in \mathbb{R}_+^{F \times T}$  represents an all-ones matrix,  $^T$  is the transpose operator,  $\odot$  is the element-wise multiplication, and  $\oslash$  is the element-wise division. This procedure always maintains the non-negativity of both parameters, since the used terms in the updating are also non-negative.

As previously described, NMF models the magnitude spectrogram of an input mixture signal as a product of a basis matrix and an activation matrix with the only constraint of the element-wise non-negativity of all matrices. Under this constraint, the aim is to minimize the cost function of the reconstruction error. However, the main problem of NMF is the trade-off between signal reconstruction and the physical interpretation of the factorized part-based objects. In other words, this non-negativity of the parameters does not guarantee a meaningful part-based representation when dealing with real-world mixture signals [69,70]. Several properties can be used to improve the uniqueness of the local minima obtained by NMF, incorporating physical meaning into the basis functions and activations. In particular, these properties can be implemented using regularizations, which are added to the global cost function in the factorization model. The main constraints, sparseness and smoothness, used in this paper to model the spectro-temporal behavior of wheezing and normal respiratory sounds are briefly described below.

## 2.2. Spectral Sparseness

Spectral sparseness  $\psi(\mathbf{B})$  denotes that, for each source, most of its frequencies are zero or close to zero [71,72]. This constraint enforces that only a few frequency bins predominate

in each spectral basis, whilst the other bins are canceled. It is implemented by incorporating a penalty term into the NMF objective function. In practice, the  $L^1$ -norm is often used because it was demonstrated to be less sensitive to changes of the parameter that controls the importance of the constraint in the factorization process. Then, the optimization problem can be expressed as,

$$\arg \min_{\mathbf{B}, \mathbf{A}} D(\mathbf{X}|\mathbf{B}\mathbf{A}) + \alpha \|\mathbf{B}\|_1 \quad \mathbf{B}, \mathbf{A} \geq 0 \quad (7)$$

where  $\alpha$  is the weight parameter that adjusts the influence of the constraint.

### 2.3. Temporal/Spectral Smoothness

Generally, smoothness  $\phi$  means how continuous or smooth the spectral or temporal changes related to a source are [72]. Smoothness constraints have been defined for both activation  $\mathbf{A}$  and basis functions  $\mathbf{B}$  and added to the global cost function as penalty terms as follows,

$$\arg \min_{\mathbf{B}, \mathbf{A}} D(\mathbf{X}|\mathbf{B}\mathbf{A}) + \lambda \phi(\mathbf{A}) + \beta \phi(\mathbf{B}) \quad \mathbf{B}, \mathbf{A} \geq 0 \quad (8)$$

where  $\phi(\mathbf{A})$  and  $\phi(\mathbf{B})$  are the functions that penalize non-smooth temporal activations or spectral patterns and the parameters  $\lambda$  and  $\beta$  control the effect of the regularization in the decomposition procedure.

Temporal smoothness (also known as smooth activations)  $\phi(\mathbf{A})$ , applied to the estimated activation matrix  $\mathbf{A}$ , reports how slow the amplitude variations over time are. In other words, temporal smoothness accounts for the fact that real-world sounds usually have a temporal structure, and their acoustic characteristics vary slowly as a function of time. In [72], the authors proposed to model the temporal smoothness regularization  $\phi(\mathbf{A})$  by applying a high cost to large changes produced between adjacent frames in the activation matrix  $\mathbf{A}$  as follows,

$$\phi(\mathbf{A}) = \sum_{k=1}^K \frac{1}{\sigma_k^2} \sum_{t=2}^T (A_{k,t} - A_{k,t-1})^2 \quad (9)$$

where  $\sigma_k = \sqrt{\frac{1}{T} \sum_{t=1}^T A_{k,t}^2}$  indicates the standard deviation used to normalize the activation functions. This normalization provides that the cost of regularization is independent of the numerical scale of activation [69,72].

Spectral smoothness (also known as smooth basis)  $\phi(\mathbf{B})$ , applied to the estimated basis matrix  $\mathbf{B}$ , measures how fast the amplitude changes along the frequency axis, that is it allows modeling the behavior of those sounds that are represented by a wideband spectrum. In [69,73], the authors proposed to model the spectral smoothness regularization  $\phi(\mathbf{B})$  by applying a high cost to large changes produced between adjacent bins in the basis matrix  $\mathbf{B}$  as follows,

$$\phi(\mathbf{B}) = \sum_{k=1}^K \frac{1}{\sigma_k^2} \sum_{f=2}^F (B_{f,k} - B_{f-1,k})^2 \quad (10)$$

where  $\sigma_k = \sqrt{\frac{1}{F} \sum_{f=1}^F B_{f,k}^2}$  represents the standard deviation used to normalize the basis functions. This normalization achieves that the cost of regularization is independent of the numerical scale of the basis [69,73].

## 3. Proposed Method

The main problem in classifying wheezes from a mixture is that both wheezing sounds and normal respiratory sounds occur simultaneously in the time and frequency domain. Considering the acoustic interference caused by normal respiratory sounds, the proposed signal model is composed of two stages: Modelling and separation of wheezing spectral patterns from normal respiratory sounds based on CL-RNMF (stage I) and Classification between MP/PP wheezing according to its harmonic structure (stage II). In this manner, the

goal of the stage I is to model the spectral patterns that characterize wheezing sounds by isolating them from respiratory interference. The aim of stage II is to analyze the location of the frequency components extracted from the previous stage to determine the type, monophonic or polyphonic, of wheezing according how the wheezing energy is locating in the frequency domain. The flowchart of the proposed method is shown in Figure 3, and details are depicted in Sections 3.1–3.3.

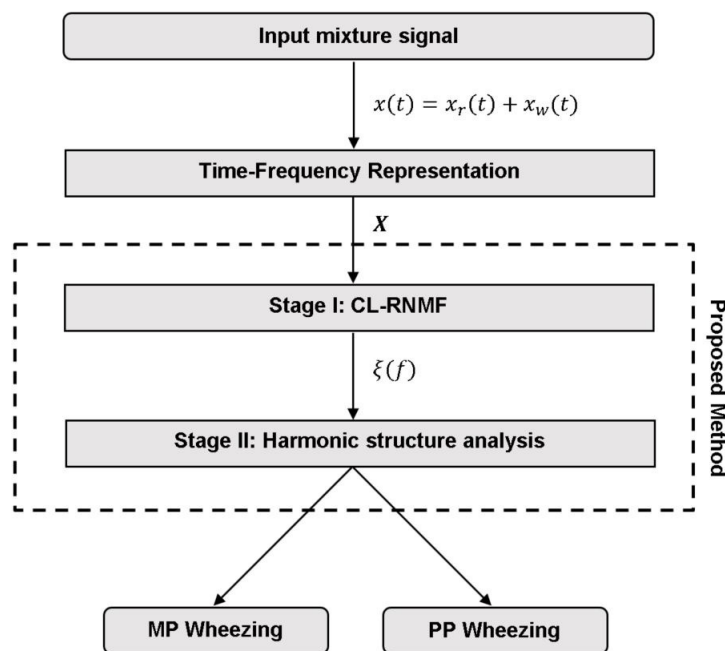


Figure 3. Flowchart of the proposed method.

### 3.1. Time-Frequency Signal Representation

Time-frequency representation by means of spectrograms has been demonstrated to be useful for visualizing the characteristics and behavior of both wheezing and normal respiratory sounds [9,50,51,65,66]. The input mixture signal  $x(t)$  is composed of wheeze sounds  $x_w(t)$  (MP or PP wheezing) and normal respiratory sounds  $x_r(t)$  overlapping in the time and frequency domain. We assume that the mixture of these sounds is additive and can be expressed as  $x(t) = x_r(t) + x_w(t)$ . The input magnitude spectrogram  $\mathbf{X} \in \mathbb{R}_+^{F \times T}$  of the input mixture signal can be represented as  $\mathbf{X} = \mathbf{X}_R + \mathbf{X}_W$ , being  $\mathbf{X}_R \in \mathbb{R}_+^{F \times T}$  the magnitude spectrogram of only respiratory sounds and  $\mathbf{X}_W \in \mathbb{R}_+^{F \times T}$  the magnitude spectrogram of only wheeze sounds. Specifically, each magnitude spectrogram is composed of  $T$  frames,  $F$  frequency bins and a set of time-frequency units  $X_{f,t}$ , being  $f = 1, \dots, F$  and  $t = 1, \dots, T$ . 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 10% overlap. In this work, a normalization process is applied in order to achieve independence regarding the size and scale of the input spectrogram  $\mathbf{X}$ . Thus, the normalized magnitude spectrogram  $\bar{\mathbf{X}}$  is computed as follows,

$$\bar{\mathbf{X}} = \frac{\mathbf{X}}{\left( \frac{\sum_{f=1}^F \sum_{t=1}^T X_{f,t}}{FT} \right)} \quad (11)$$

To avoid complex nomenclature throughout the paper, the variable  $\mathbf{X}$  is hereinafter referred to the normalized magnitude spectrogram previously computed in Equation (11).

### 3.2. Stage I: Constrained Low-Rank Non-Negative Matrix Factorization

As mentioned above, it is common that normal respiratory sounds mask the presence of the wheezing sounds. As a result, this sound mask makes the task of wheezing classification difficult since the spectral patterns associated with normal respiratory sounds can be confused with wheezing spectral content. Therefore, the aim of this stage is to provide a reliable modeling of the different frequency components (spectral patterns) that compose a wheeze, removing any sound interference from normal respiratory sounds. For this purpose, we propose a CL-RNMF approach because, as far as the authors' knowledge, the non-negative matrix factorization approach has never been applied before to MP/PP wheezing classification. In addition, our approach is an unsupervised method because it does not require any training of the sounds to classify. Therefore, the proposed method decomposes a magnitude mixture spectrogram  $\mathbf{X}$  into two estimated spectrograms:  $\hat{\mathbf{X}}_R$  (only normal respiratory sounds without wheezing) and  $\hat{\mathbf{X}}_W$  (only wheeze sounds without normal respiratory sounds). In this manner, each estimated spectrogram can be factorized into the product of its corresponding estimated basis and activation matrices: (i)  $\mathbf{B}_R \in \mathbb{R}_+^{F \times K_r}$  and  $\mathbf{A}_R \in \mathbb{R}_+^{K_r \times T}$  to the factorization of  $\hat{\mathbf{X}}_R$ ,  $K_r$  being the number of respiratory components; and (ii)  $\mathbf{B}_W \in \mathbb{R}_+^{F \times K_w}$  and  $\mathbf{A}_W \in \mathbb{R}_+^{K_w \times T}$  to the factorization of  $\hat{\mathbf{X}}_W$ ,  $K_w$  being the number of wheezing components. The proposed separation model can be formulated with the following objective function,

$$\mathbf{X} \approx \hat{\mathbf{X}}_R + \hat{\mathbf{X}}_W = \mathbf{B}_R \mathbf{A}_R + \mathbf{B}_W \mathbf{A}_W \quad (12)$$

where, considering the non-negative property that characterizes the NMF approach, all the matrices that compose the previous model are non-negative.

As previously mentioned, this stage attempts to ensure that  $\mathbf{B}_W$  contains reliable modeling of the wheezing spectral patterns by means of narrowband spectral peaks that typically characterize the wheeze content. The key assumptions behind the proposed CL-RNMF approach to model wheezing spectral patterns are the following:

- **Low-rank:** The number of wheezing components should be much less than the number of normal respiratory components, that is  $K_w \ll K_r$ . This assumption allows that the number of frequency components can be reduced in the least number of bases possible for their posterior analysis, while normal respiratory sounds are modeled using a higher range of components. Experimental results showed that the best classification performance was obtained when  $2 \leq K_w \leq 6$  and  $K_r \geq 32$ . In particular, when  $K_w = 1$ , the proposed CL-RNMF approach tends to converge very quickly at the expense of losing relevant wheezing content. On the other hand, when  $K_w > 6$ , the spectral wheezing patterns tend to be split into different components of the matrix  $\mathbf{B}_W$ .
- **Constraints:** These characterize wheezing sounds and normal respiratory sounds using opposite restrictions between both sounds. The use of constraints allows isolating the spectral wheezing patterns from the spectral patterns of normal respiratory sounds. Therefore, in order to find a better NMF decomposition that shows spectrotemporal features of the wheezing and normal respiratory sounds as can be observed in the real world, we propose to incorporate sparseness and smoothness into the NMF decomposition process. As shown in Figures 1 and 2, wheezing sounds can be considered sparse in frequency because MP wheezing or PP wheezing is characterized by one or more than one narrowband spectral peak. Moreover, wheezing sounds can be considered smooth or continuous events in time, that is slow variation of the magnitude spectrogram along time. On the other hand, normal respiratory sounds can be considered smooth in frequency, that is they can be modeled assuming wideband spectral patterns. Therefore,  $\mathbf{B}_W$  should contain wheezing spectral patterns composed of one or more than one narrowband spectral peak, depending on the spectral complexity of each wheezing, and  $\mathbf{B}_R$  should be composed of a set of wideband spectral patterns that model the behavior of normal respiratory sounds.



Considering the key assumptions mentioned above, the global objective function  $D(\mathbf{X}|\hat{\mathbf{X}})$  that must be minimized in order to estimate the basis ( $\mathbf{B}_R$ ,  $\mathbf{B}_W$ ) and activation ( $\mathbf{A}_R$ ,  $\mathbf{A}_W$ ) matrices is composed of: (i) the Kullback–Leibler divergence cost function  $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$  to minimize the reconstruction error between the input spectrogram  $\mathbf{X}$  and the estimated spectrogram  $\hat{\mathbf{X}}$ , (ii) the spectral sparseness  $\psi(\mathbf{B}_W)$  and temporal smoothness  $\phi(\mathbf{A}_W)$  restrictions applied to  $\mathbf{B}_W$  and  $\mathbf{A}_W$ , respectively, to model the wheezing spectral patterns, and (iii) the spectral smoothness  $\phi(\mathbf{B}_R)$  restriction applied to  $\mathbf{B}_R$ , to model the spectral patterns of normal respiratory sounds. The global objective function  $D(\mathbf{X}|\hat{\mathbf{X}})$  is detailed as follows,

$$D(\mathbf{X}|\hat{\mathbf{X}}) = D_{KL}(\mathbf{X}|\hat{\mathbf{X}}) + \alpha\psi(\mathbf{B}_W) + \lambda\phi(\mathbf{A}_W) + \beta\phi(\mathbf{B}_R) \quad (13)$$

where the equations of terms  $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$ ,  $\psi(\mathbf{B}_W)$ ,  $\phi(\mathbf{A}_W)$ , and  $\phi(\mathbf{B}_R)$  can be found in Section 2. The parameters  $\alpha$ ,  $\lambda$ , and  $\beta$  define the weight to control the effect of the regularization. Experimental results showed that the best classification performance is obtained when all weights are equal  $\alpha = \lambda = \beta$ , the optimal value being  $\alpha = \lambda = \beta = 0.5$ . Analyzing the sound separation performance of the previous decomposition, we observed empirically that the acoustic interference suffered by wheezing sounds from normal respiratory sounds is minimum, and no significant loss of wheezing content occurs when  $\alpha = \lambda = \beta$ . However, significant losses of wheezing content appear when  $\alpha = \lambda > \beta$ , or significant sound interference by normal respiratory sounds can be observed when  $\alpha = \lambda < \beta$ .

From Equation (13), the estimated basis matrices ( $\mathbf{B}_W$  and  $\mathbf{B}_R$ ) and activation matrices ( $\mathbf{A}_W$  and  $\mathbf{A}_R$ ) can be obtained by applying a gradient descent algorithm based on multiplicative update rules. Specifically, the multiplicative update rules to learn those matrices can be computed by taking negative and positive terms of the partial derivative of the global objective function  $D(\mathbf{X}|\hat{\mathbf{X}})$  with respect to  $\mathbf{B}_W$ ,  $\mathbf{B}_R$ ,  $\mathbf{A}_W$ , and  $\mathbf{A}_R$ , respectively,

$$\mathbf{B}_W \leftarrow \mathbf{B}_W \odot \frac{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_W} \right]^- + \alpha \left[ \frac{\partial \psi(\mathbf{B}_W)}{\partial \mathbf{B}_W} \right]^-}{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_W} \right]^+ + \alpha \left[ \frac{\partial \psi(\mathbf{B}_W)}{\partial \mathbf{B}_W} \right]^+} \quad (14)$$

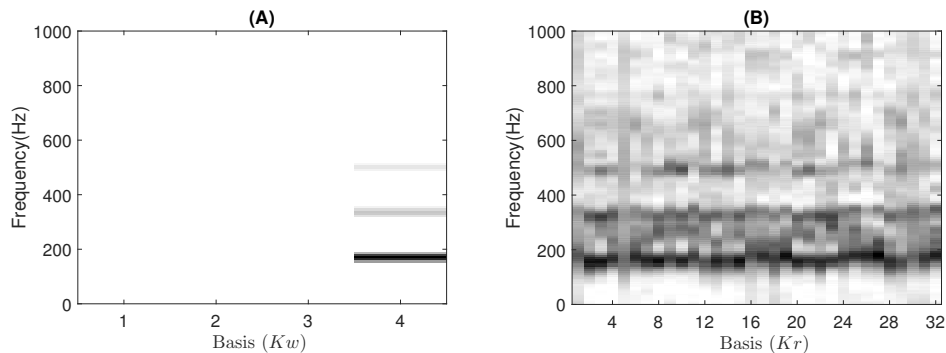
$$\mathbf{B}_R \leftarrow \mathbf{B}_R \odot \frac{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_R} \right]^- + \beta \left[ \frac{\partial \phi(\mathbf{B}_R)}{\partial \mathbf{B}_R} \right]^-}{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_R} \right]^+ + \beta \left[ \frac{\partial \phi(\mathbf{B}_R)}{\partial \mathbf{B}_R} \right]^+} \quad (15)$$

$$\mathbf{A}_W \leftarrow \mathbf{A}_W \odot \frac{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_W} \right]^- + \lambda \left[ \frac{\partial \phi(\mathbf{A}_W)}{\partial \mathbf{A}_W} \right]^-}{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_W} \right]^+ + \lambda \left[ \frac{\partial \phi(\mathbf{A}_W)}{\partial \mathbf{A}_W} \right]^+} \quad (16)$$

$$\mathbf{A}_R \leftarrow \mathbf{A}_R \odot \frac{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_R} \right]^-}{\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_R} \right]^+} \quad (17)$$

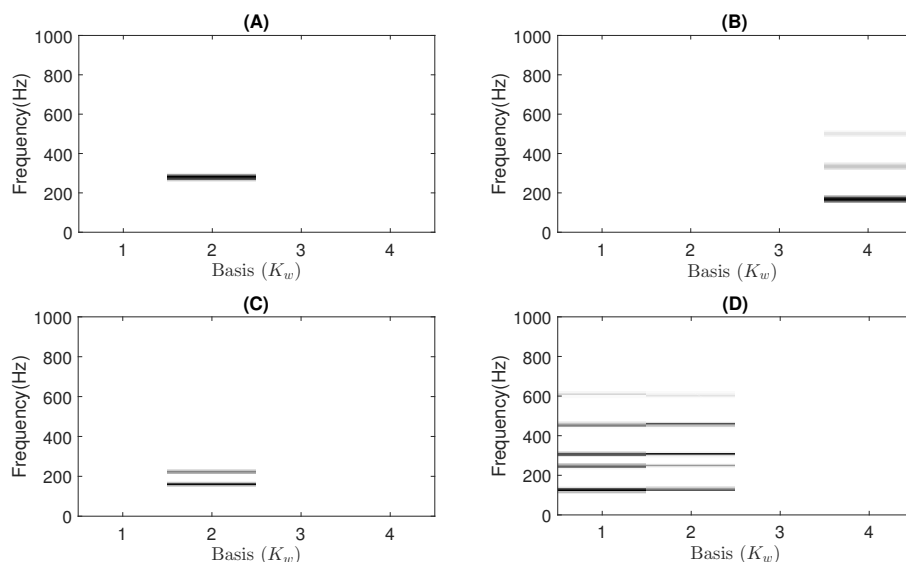
where, for each multiplicative update rule, the division between the negative and positive terms of the partial derivatives is an element-wise division. More details related to the equations of each partial derivative of the multiplication update rules can be found in Appendix A. Finally, the estimated respiratory and wheezing basis ( $\mathbf{B}_W$  and  $\mathbf{B}_R$ ) and activation matrices ( $\mathbf{A}_W$  and  $\mathbf{A}_R$ ) are obtained updating the previous rules until the algorithm converges using  $M$  iterations. Figure 4 shows the estimated matrices  $\mathbf{B}_W$  and  $\mathbf{B}_R$  decomposing the MP wheezing spectrogram shown in Figure 1B. As can be observed, the matrix  $\mathbf{B}_W$  contains spectral patterns that characterize a typical MP wheezing, which are

represented by means of a set of narrowband spectral peaks (or frequency components). In contrast, the estimated matrix  $\mathbf{B}_R$  is composed of a set of wideband spectral patterns that characterize normal respiratory sounds. Therefore, the proposed CL-RNMF approach achieves extracting the wheezing spectral content at the expense of removing normal respiratory sounds.



**Figure 4.** Example of the estimated matrices  $\mathbf{B}_W$  and  $\mathbf{B}_R$  obtained from the proposed CL-RNMF approach, analyzing the MP wheezing spectrogram previously shown in Figure 1B. (A) Although the matrix  $\mathbf{B}_W$  is composed of four spectral bases, the spectral wheezing patterns are compacted into the fourth basis  $\mathbf{B}_W(4)$ . This spectral basis  $\mathbf{B}_W(4)$  is composed of three narrowband spectral peaks. (B) The matrix  $\mathbf{B}_R$  is composed of thirty-two wideband spectral bases.

Experimentally, we found that the proposed CL-RNMF approach tends to compact all the narrowband spectral peaks into a single basis of the matrix  $\mathbf{B}_W$ , as shown in Figure 4A. However, considering that CL-RNMF uses a small set of wheezing components ( $K_w$ ), in some cases, the narrowband spectral peaks are divided into several bases of the same matrix  $\mathbf{B}_W$ . To clarify this issue, Figure 5 shows the matrix  $\mathbf{B}_W$  obtained for the different examples of MP and PP wheezing described in Section 1. As shown in Figure 5D, the energy of the narrowband spectral patterns, which characterizes that PP wheezing, are divided into two bases  $\mathbf{B}_W(1)$  and  $\mathbf{B}_W(2)$  belonging to the matrix  $\mathbf{B}_W$ . In both bases,  $\mathbf{B}_W(1)$  and  $\mathbf{B}_W(2)$ , all narrowband spectral peaks are correctly modeled.

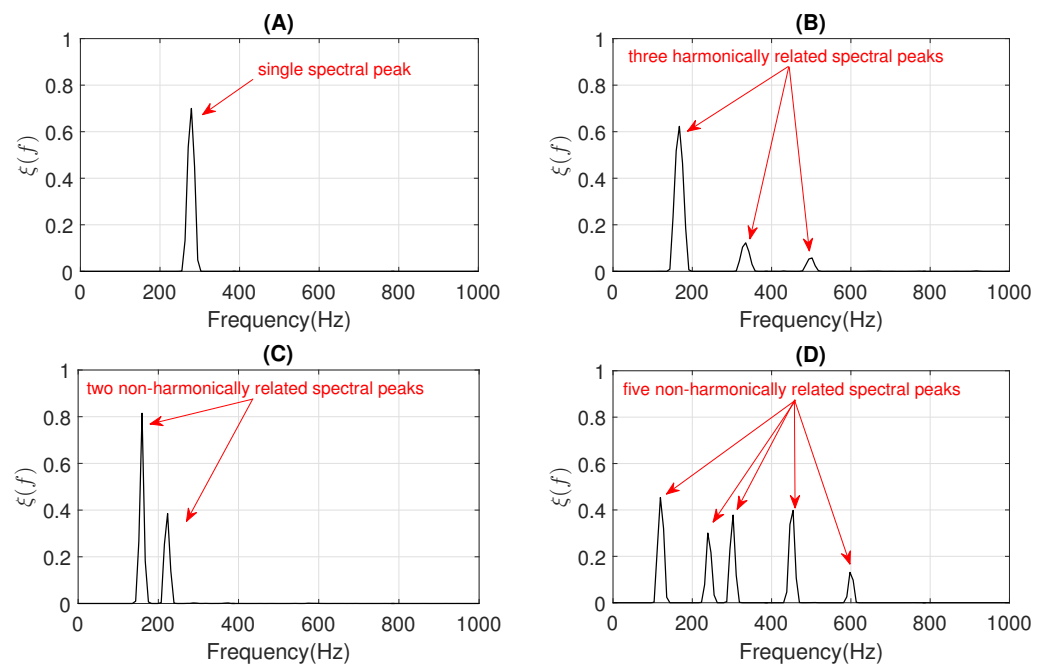


**Figure 5.** The estimated basis matrices  $\mathbf{B}_W$  obtained from CL-RNMF in the examples shown in Section 1. (A)  $\mathbf{B}_W$  for the MP wheezing shown in Figure 1A. (B)  $\mathbf{B}_W$  for the MP wheezing shown in Figure 1B. (C)  $\mathbf{B}_W$  for the PP wheezing shown in Figure 2A. (D)  $\mathbf{B}_W$  for the PP wheezing shown in Figure 2B. The wheezing spectral patterns were compacted into a single basis,  $\mathbf{B}_W(2)$  (in Case (A)),  $\mathbf{B}_W(4)$  (in Case (B)), and  $\mathbf{B}_W(2)$  (in Case (C)). However, the energy of the narrowband spectral peaks was divided into two bases  $\mathbf{B}_W(1)$  and  $\mathbf{B}_W(2)$ , as can be seen in Case (D).

Finally, we propose to obtain the spectral energy distribution  $\zeta(f)$  (see Equation (18)) from the set of bases that compose the matrix  $\mathbf{B}_W$ . This makes it possible to compact the spectral distribution of all narrowband spectral peaks that make up the input MP or PP wheezes to analyze their harmonic structure in Stage II.

$$\zeta(f) = \sum_{k_w=1}^{K_w} \mathbf{B}_{W_{f,k_w}}, f = 1, \dots, F \quad (18)$$

Figure 6 shows the spectral energy distribution  $\zeta(f)$  obtained for the four examples of wheezing shown in Section 1. The pseudocode of this Stage I for the modeling and separation of wheezing spectral patterns based on CL-RNMF is detailed in Algorithm 1.



**Figure 6.** The spectral energy distribution  $\zeta(f)$  provided by CL-RNMF from the estimated basis matrix  $\mathbf{B}_W$  shown in Figure 5: (A) Figure 5A. (B) Figure 5B. (C) Figure 5C. (D) Figure 5D.

---

#### Algorithm 1: CL-RNMF.

---

**Require:**  $x(t)$ ,  $K_r$ ,  $K_w$ ,  $\alpha$ ,  $\beta$ ,  $\lambda$ , and  $M$ .

- 1: Compute the normalized magnitude spectrogram  $\mathbf{X}$  using Equation (11).
  - 2: Initialize  $\mathbf{B}_W$ ,  $\mathbf{B}_R$ ,  $\mathbf{A}_W$ , and  $\mathbf{A}_R$  with random non-negative values.
  - 3: Update the estimated wheezing basis matrix  $\mathbf{B}_W$  using Equation (14).
  - 4: Update the estimated respiratory basis matrix  $\mathbf{B}_R$  using Equation (15).
  - 5: Update the estimated wheezing activations matrix  $\mathbf{A}_W$  using Equation (16).
  - 6: Update the estimated respiratory activations matrix  $\mathbf{A}_R$  using Equation (17).
  - 7: Repeat Steps 3–6 until the algorithm converges (or until the maximum number of iterations  $M$  is reached).
  - 8: Compute the spectral energy distribution  $\zeta(f)$  from  $\mathbf{B}_W$  using Equation (18).
- return**  $\zeta(f)$
- 

### 3.3. Stage II: Harmonic Structure Analysis

The goal of this stage is to classify between MP and PP wheezing by analyzing the spectral energy distribution  $\zeta(f)$  of the different narrowband spectral peaks obtained in the previous stage. Depending on the harmonic structure, wheezing can be classified as MP or PP. Specifically, MP wheezing is composed of a single narrowband spectral peak or the set of harmonically related narrowband spectral peaks. In contrast, PP wheezing is

composed of several non-harmonically related narrowband spectral peaks. For this reason, we propose to obtain the number of narrowband spectral peaks  $\eta$  that can be found from  $\zeta(f)$ . Note that the procedure to detect the spectral peaks is a simple task since, as can be seen in Figure 6, the spectral energy distribution  $\zeta(f)$  from CL-RNMF clearly provides a set of narrowband spectral peaks typically found in wheezing sounds. Once the parameter  $\eta$  is obtained, a preliminary classification of the type of wheezing can be performed as follows,

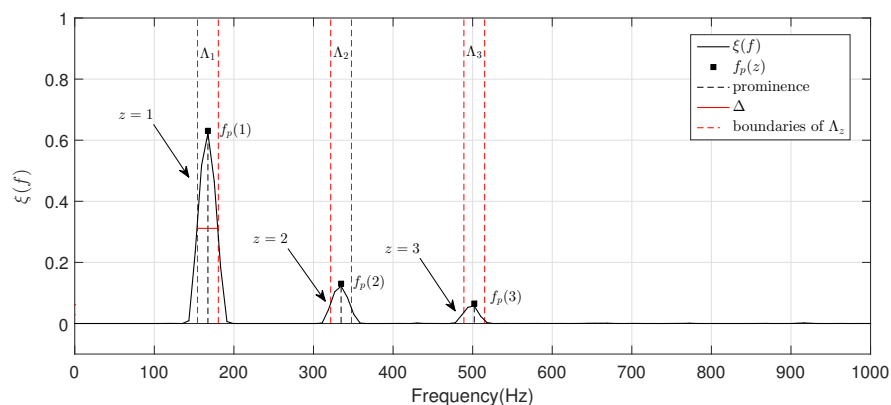
$$\text{Wheezing category} = \begin{cases} \text{MP} & \text{if } \eta = 1 \\ \text{MP or PP} & \text{if } \eta > 1 \end{cases} \quad (19)$$

Wheezing can only be classified as MP when  $\eta = 1$  since a wheezing is composed of a single narrowband spectral peak, as can be seen in Figure 6A. However, a wheezing can be classified as MP or PP when  $\eta > 1$ , depending on the harmonic structure that exists between the different narrowband spectral peaks. Specifically, the wheezing is classified as MP if the set of spectral peaks are harmonically related between them. The wheezing is classified as PP if the spectral peaks are not harmonically related between them. In order to perform the classification between MP and PP wheezing in the case of  $\eta > 1$ , we propose a two-step procedure, as follows:

- The objective of the first step is to locate, in terms of frequency, all the narrowband spectral peaks detected in the previous Stage I. For this, we propose to locate the most prominent frequency  $f_p(z)$  in each spectral peak  $z = 1, \dots, \eta$ . Each value  $f_p(z)$  was calculated using the findpeaks function provided by the MATLAB software [74] due to the satisfactory results obtained in several preliminary analyses performed. Figure 7 shows the location  $f_p(z)$ , in terms of frequency, of each spectral peak for the MP example previously shown in Figure 1B.
- The objective of the second step is to check if the different spectral peaks  $z = 1, \dots, \eta$  are harmonically related or not. We assume that the first spectral peak ( $z = 1$ ) represents the basal peak. Therefore, the wheezing is classified as MP if the rest of spectral peaks ( $z = 2, \dots, \eta$ ) are located in the harmonic frequencies (integer multiple) of the basal peak. Otherwise, the wheezing is classified as PP. From the width  $\Delta$  of the main lobe of the basal peak ( $z = 1$ ) and the value of its most prominent frequency  $f_p(1)$ , the spectral intervals where the possible harmonic frequencies should be located are calculated as follows,

$$\Lambda_z = [zf_p(1) - (\Delta/2), zf_p(1) + (\Delta/2)], z = 1, \dots, \eta \quad (20)$$

where  $[i, j]$  denotes the spectral interval comprised between the lower limit  $i$  and the upper limit  $j$ , in terms of frequency. Specifically,  $\Lambda_1$  represents the spectral interval associated with the basal peak, and  $\Lambda_z$  ( $z = 2, \dots, \eta$ ) corresponds to the spectral intervals where the harmonic frequencies should be located. Note that the width of the main lobe  $\Delta$  was obtained by positioning the reference line beneath the peak at a vertical distance equal to half the peak prominence [74].

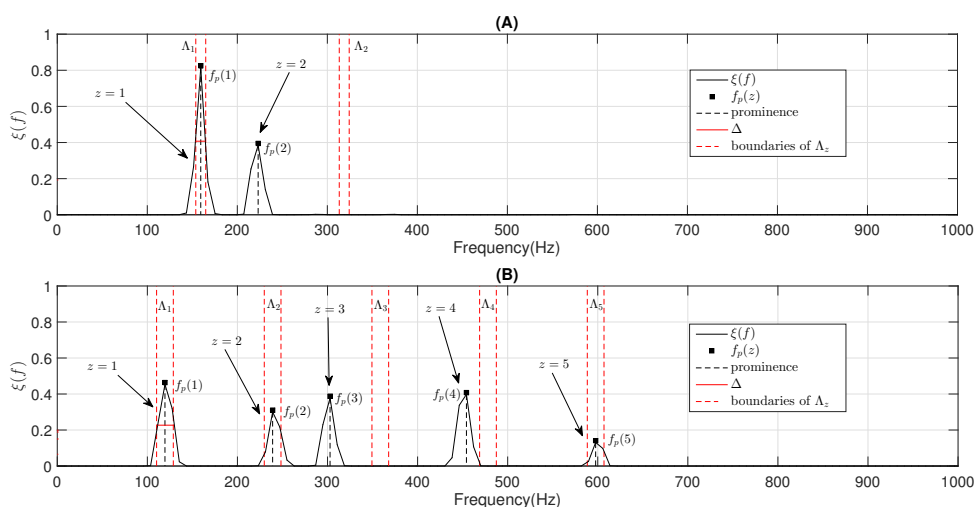


**Figure 7.** Example of the proposed two-step procedure to classify between MP and PP wheezing when  $\eta > 1$  from the example of MP wheezing shown in Figure 1B. Note that the arrows indicate the narrowband spectral peaks that compose the wheezing. In this case, the wheezing is classified as MP because all spectral peaks are harmonically related.

Considering the two-step procedure described above, wheezes that are composed of several narrowband spectral peaks ( $\eta > 1$ ) can be classified as MP or PP as follows,

$$\text{Wheezing category} = \begin{cases} \text{MP} & \text{if } f_p(z) \subseteq \Lambda_z, z = 2, \dots, \eta \\ \text{PP} & \text{otherwise} \end{cases} \quad (21)$$

where  $v \subseteq V$  denotes that element  $v$  is contained in the interval  $V$ . Therefore, when the frequency  $f_p(z)$  of all possible harmonic spectral peaks  $z = 2, \dots, \eta$  is located in the corresponding spectral intervals  $\Lambda_z$ , the wheezing is classified as MP. Otherwise, wheezing is classified as PP because the narrowband spectral peaks that characterize the wheezing are not harmonically related. This occurs when for the frequency  $f_p(z)$ , at least one of the possible harmonic spectral peaks is not located in its corresponding spectral intervals  $\Lambda_z$ . Figure 7 shows an example of the procedure described for MP wheezing composed of a basal peak and two harmonics. Figure 8 shows two examples of the procedure described for two PP wheezing with several non-harmonically related spectral peaks. Finally, the pseudocode of this stage for the classification between MP/PP wheezing according to its harmonic structure is detailed in Algorithm 2.



**Figure 8.** Example of the proposed two-step procedure to classify between MP and PP wheezing when  $\eta > 1$ , considering the two examples of PP wheezing shown in Figure 2. **(A)** Two-step procedure applied to the PP wheezing shown in Figure 2A. **(B)** Two-step procedure applied to the PP wheezing shown in Figure 2B. Note that the arrows indicate the narrowband spectral peaks that compose the wheezing. In this case, both wheezing are classified as PP because not all spectral peaks are harmonically related.

**Algorithm 2:** Harmonic structure analysis.

---

**Require:**  $\xi(f)$ .

- 1: From  $\xi(f)$ , detect the number  $\eta$  of narrowband spectral peaks.

**if**  $\eta = 1$  **then**

**return** Wheezing category = MP

**else**

- 2: Locate the frequency  $f_p(z)$  in each spectral peak  $z = 1, \dots, \eta$ .
- 3: Compute the spectral intervals  $\Lambda_z$  using Equation (20).

**if**  $f_p(z) \subseteq \Lambda_z, z = 2, \dots, \eta$  **then**

**return** Wheezing category = MP

**else**

**return** Wheezing category = PP

**end if**

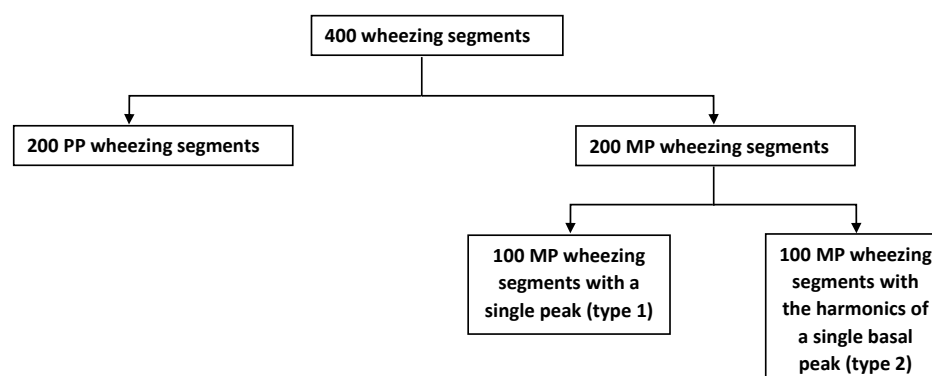
**end if**

---

**4. Experimental Results and Discussion***4.1. Data Collection*

As far as the authors' knowledge, there is no public wheeze database in which wheezing has been labeled as monophonic or polyphonic. For this reason, we received the collaboration of a pneumologist from the University Hospital of Jaén (Spain) to create and label a database according to the wheezing harmonic structure. The database was created by collecting and categorizing a set of recordings from different subjects of the most widely used Internet pulmonary repositories [75–87]. Specifically, all previous recordings were collected from subjects with CRDs (asthma or COPD). Note that the set of recordings selected for this assessment was only composed of normal respiratory sounds and wheezing sounds.

The type of wheezing (MP or PP) was labeled by the pneumologist by means of an acoustic inspection and a visual verification of the spectrogram considering the harmonic structure that distinguishes both types of wheezing. The database consisted of 200 MP and 200 PP wheezing segments, where the duration of each segment was at least 100 ms, to be consistent with the literature. As mentioned above, MP wheezing can show two different harmonic structures: Type 1, wheezes with a single peak, that is only the fundamental frequency component is active; and Type 2, wheezes with the harmonics of a single basal peak, that is both the fundamental frequency component and its frequencies that are harmonically related are active. Therefore, to guarantee the maximum variability of the MP wheezing, the 200 MP wheezing segments were divided into 100 MP wheezing segments with a single peak and 100 MP wheezing segments with the harmonics of a single basal peak. Note that all segments were independent of each other, since each segment corresponded to a different wheezing from the rest. Finally, all segments in the database were sampled at 4096 Hz and had a length between 100 and 700 ms. Figure 9 shows the classification performed on the database created.



**Figure 9.** Scheme of the types of wheezing contained in the database.

#### 4.2. Experimental Setup

To be consistent with the literature, we assumed that wheezing sounds were not active below 100 Hz and above 1000 Hz. For this reason, all segments that compose the database were band-limited from 100 Hz–1000 Hz.

The length of the signal frames was set to  $N = 256$  samples (62.5 ms). This frame size was considered large enough to assume a perfect spectral representation of all wheezing frequency components. The overlap between frames was set to 10% (6.25 ms). To obtain the time-frequency representation, windowing with a Hamming window was applied, and the order of the Discrete Fourier Transform (DFT) was set to  $2N$  frequency bins, similar to [9,50]. This DFT size provides a high enough resolution for modeling the spectral patterns of wheezing sounds and was chosen empirically as a trade-off between achieved quality and complexity. Besides, we empirically observed that the reconstruction error converged after 50 iterations, so the maximum number of iterations for the decomposition was equal to  $M = 50$ .

Finally, note that the performance of the proposed method depends on the initial values with which the basis matrices  $\mathbf{B}_W$ ,  $\mathbf{B}_R$  and the activation matrices  $\mathbf{A}_W$ ,  $\mathbf{A}_R$  are initialized. Although the obtained results are not dispersed and keep the same behavior, in order to overcome this issue, we ran the proposed method five times for each segment that composed the database, and the results shown in this paper are averaged values.

#### 4.3. Evaluation Metrics

The Accuracy rates (ACC) were used to evaluate the performance of the proposed method, which are commonly used in the field of wheezing classification [7]. In order to provide a fair evaluation of the classification performance obtained by the proposed method and the state-of-the-art algorithms, the following accuracy rates were proposed: (i)  $ACC_G$  is the ability to correctly classify a wheezing segment as MP or PP; (ii)  $ACC_P$  represents the ability to correctly classify a wheezing segment as PP; (iii)  $ACC_M$  corresponds to the ability to correctly classify a wheezing segment as MP; (iv)  $ACC_{M1}$  indicates the ability to correctly classify a wheezing segment as MP Type 1; and (v)  $ACC_{M2}$  reports the ability to correctly classify a wheezing segment as MP Type 2. The terms used in Equations (22)–(26) are described in Table 1.

$$ACC_G = \frac{(TP + TM)}{(TP + TM + FP + FM)} \quad (22)$$

$$ACC_P = \frac{TP}{(TP + FP)} \quad (23)$$

$$ACC_M = \frac{TM}{(TM + FM)} \quad (24)$$

$$ACC_{M1} = \frac{TM1}{(TM1 + FM1)} \quad (25)$$

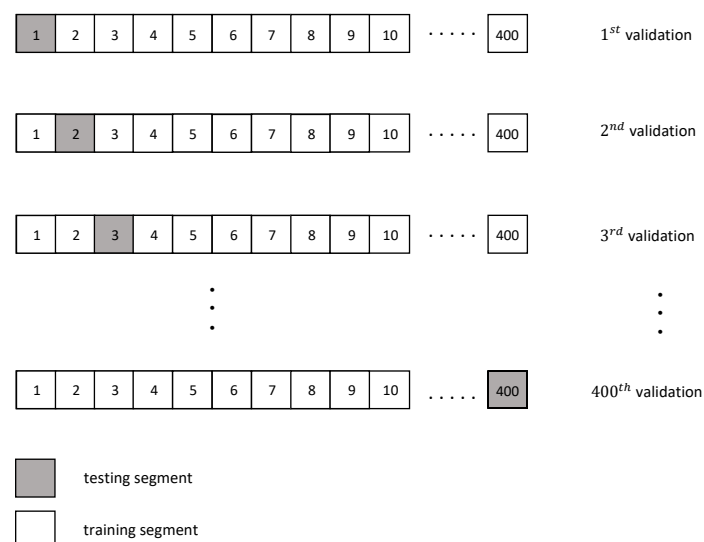
$$ACC_{M2} = \frac{TM2}{(TM2 + FM2)} \quad (26)$$

**Table 1.** Definition of the terms that appear in the metrics detailed in Equations (22)–(26).

Terms	Definitions
$TP$ (True PP)	PP wheezing segments correctly classified
$TM$ (True MP)	MP wheezing segments correctly classified
$FP$ (False PP)	PP wheezing segments misclassified as MP
$FM$ (False MP)	MP wheezing segments misclassified as PP
$TM1$ (True MP Type 1)	MP Type 1 wheezing segments correctly classified
$TM2$ (True MP Type 2)	MP Type 2 wheezing segments correctly classified
$FM1$ (False MP Type 1)	MP Type 1 wheezing segments misclassified as PP
$FM2$ (False MP Type 2)	MP Type 2 wheezing segments misclassified as PP

#### 4.4. State-of-the-Art Method for Comparison

In order to measure the MP/PP classification performance of the proposal, we used the most recent and relevant state-of-the-art algorithm [7], denoted as UPER in this paper. The method UPER was implemented strictly following the instructions specified by the authors in [7]. Firstly, the values of the metric PER were obtained using the 19th parameter set ( $p = 10$ ,  $q = 11$ ,  $s = 7$ , and  $J = 45$ ) in the RADWT model. Then, three classifiers, Support Vector Machine (SVM) with the Radial Basis Function kernel (RBF kernel), K-Nearest Neighbor (KNN), and Extreme Learning Machine (ELM) were applied to the PER features. The classification performance of UPER was obtained in Leave-One-Out (LOO) cross-validation schemes with the SVM, KNN, and ELM classifiers. Specifically, LOO cross-validation is a particular case of Leave-p-Out (LPO) cross-validation with  $p = 1$ . Therefore, the LOO scheme involves using one observation as the validation set and the remaining observations as the training set. This is repeated in all ways to cut the database into a validation set of one observation and a training set. Considering the database evaluated in this work (400 segments in total), the LOO cross-validation scheme has 400 possible combinations of validation in which the training set is composed of 399 segments, and only one segment is tested, as can be observed in Figure 10). Results shown in this paper for all classifiers are the average values obtained from the 400 possible validation combinations.

**Figure 10.** LOO cross-validation scheme for the database described in this paper.

#### 4.5. Accuracy Results

In this section, we evaluate the MP/PP classification performance between the proposed method and UPER [7]. A remarkable distinction between the two methods is that the proposed method is completely unsupervised or blind (no training), but the method UPER depends on the training database.



Table 2 shows the MP/PP classification results, in terms of the accuracy rates, evaluating the database described in Section 4.1. Results provided by UPER, considering the three classifier versions (SVM, KNN, and ELM), were obtained by applying an LOO cross-validation scheme as was previously described in Section 4.4. Results report that the proposed method provides the best overall MP/PP classification results compared to UPER considering all evaluated metrics. Focusing on the different accuracy rates, the following can be observed:

- the improvement, in terms of  $ACC_G$ , of the proposed method is about 8.25% UPER (SVM), 12% UPER (KNN), and 10.5% UPER (ELM).
- the improvement, in terms of  $ACC_P$ , of the proposed method is about 4% UPER (SVM), 7.1% UPER (KNN), and 5.5% UPER (ELM).
- the improvement, in terms of  $ACC_M$ , of the proposed method is about 12.5% UPER (SVM), 17% UPER (KNN), and 15.5% UPER (ELM).
- the improvement, in terms of  $ACC_{M1}$ , of the proposed method is about 5% UPER (SVM), 10% UPER (KNN), and 8% UPER (ELM).
- the improvement, in terms of  $ACC_{M2}$ , of the proposed method is about 20% UPER (SVM), 24% UPER (KNN), and 23% UPER (ELM).

**Table 2.** Comparative ACC results between the proposed method and UPER.

Algorithm	$ACC_G$	$ACC_P$	$ACC_M$	$ACC_{M1}$	$ACC_{M2}$
Proposed Method	<b>92%</b>	<b>91.5%</b>	<b>92.5%</b>	<b>91%</b>	<b>94%</b>
UPER (SVM) [7]	83.75%	87.5%	80%	86%	74%
UPER (KNN) [7]	80%	84.4%	75.5%	81%	70%
UPER (ELM) [7]	81.5%	86%	77%	83%	71%

The main advantage of UPER is that it only uses one feature (PER value) to discriminate between MP and PP wheezing. As shown in Table 2, the SVM classifier obtains the best classification performance in the method UPER. Specifically, the classifier SVM achieves an improvement of 2.25% (KNN) and 1.5% (KNN), in terms of  $ACC_G$ . These results are consistent with those obtained by the authors in [7], confirming that the SVM classifier with the RBF kernel obtains the best classification performance when the number of features (only one PER value) is small [88].

Performing an empirical analysis of the proposed method and UPER, the following observations were extracted:

- (i) Due to the time-frequency overlapping problem, normal respiratory sounds often mask wheezing sounds, hiding relevant medical information [5]. While the proposed method (based on CL-RNMF) allows removing as much as possible the acoustic interference from normal respiratory sounds, the method UPER is based on a feature PER obtained from the sub-band energy of the wavelet coefficients, so the presence of normal respiratory sounds interferes in the selection of the optimal sub-bands that really belong to the wheezing components.
- (ii) The method UPER has more difficulty in discriminating between PP and MP wheezing composed by a basal peak and its harmonics since it achieves the worst performance in terms of  $ACC_{M2}$ . The reason is because UPER is based on energy and ignoring the spectral location of the components that model the harmonic behavior of MP wheezing. Results in Table 2 suggest that MP/PP classification based on the spectral location of the harmonic structure as occurs in the proposed method is more reliable than the use of the energy of the wheezing spectral components, as occurs in UPER.

The LOO cross-validation scheme does not show the dependency that classifiers have with the size of the training segments set, since this scheme always uses one segment as the validation set and the remaining segments as the training set. For this reason, we propose to use an LPO cross-validation scheme by varying the size of the training segments set. The LPO scheme requires training and validating the model  $C_p^n$  times, where  $n$  is the number

of segments that compose the database,  $p$  is the number of validation segments, and  $C_p^n$  is the binomial coefficient. As a result, the associated computational cost can be excessive. In order to overcome this issue, we limited the number of iterations of the LPO scheme to 500. Furthermore, the same number of MP and PP wheezes was selected for both training and validation sets in each iteration. Specifically, we used four LPO schemes: (i)  $p = 80$  uses 80% of the total segments as the training set in each iteration; (ii)  $p = 160$  uses 60% of the total segments as the training set in each iteration; (iii)  $p = 240$  uses 40% of the total segments as the training set in each iteration; and (iv)  $p = 320$  uses 20% of the total segments as the training set in each iteration. Considering all the instructions described above, Table 3 shows the MP/PP classification results, in terms of  $ACC_G$ , obtained by UPER using its three classifier versions (SVM, KNN, and ELM) in order to assess its dependence on the training set size. Comparing the LOO scheme with the LPO scheme ( $p = 320$ ), the  $ACC_G$  reduction of the classification performance is about 7.5% (SVM), 8.25% (KNN), and 6.25% (ELM). Results report that the PER feature allows distinguishing between MP and PP wheezing even when the training set size is reduced. In addition, the ELM classifier shows less dependence on the training database size compared to SVM and KNN.

**Table 3.** Comparative results, in terms of  $ACC_G$ , between the three classifier versions (SVM, KNN, and ELM) of the method UPER using four Leave-p-Out (LPO) cross-validation schemes.

Scheme	Training Set	Validate Set	SVM	KNN	ELM
LOO	399 (99.75%)	1 (0.25%)	<b>83.75%</b>	80%	81.5%
LPO ( $p = 80$ )	320 (80%)	80 (20%)	<b>81.5%</b>	79.25%	80%
LPO ( $p = 160$ )	240 (60%)	160 (40%)	<b>80.5%</b>	77.75%	79.5%
LPO ( $p = 240$ )	160 (40%)	240 (60%)	<b>78.25%</b>	74.75%	77.25%
LPO ( $p = 320$ )	80 (20%)	320 (80%)	<b>76.25%</b>	71.75%	75.25%

## 5. Conclusions and Future Work

In this paper, we present a novel Constrained Low-rank Non-negative Matrix Factorization (CL-RNMF) approach to classify monophonic and polyphonic wheezing sounds according to their harmonic structure. The first contribution of this work proposes a CL-RNMF framework that allows extracting the spectral patterns that characterize wheezing sounds with the least possible interference from normal respiratory sounds. Specifically, a low-rank configuration with a reduced number of wheezing bases is presented to compact its frequency components in the least number of bases possible for their posterior analysis. In addition, CL-RNMF uses a set of constraints to model the spectro-temporal behavior of wheezing and normal respiratory sounds. As far as the authors' knowledge, the non-negative matrix factorization approach has never been applied before to MP/PP wheezing classification. The second contribution analyzes the harmonic structure of the energy distribution from the estimated wheezing spectrogram provided by CL-RNMF to determine the type of wheezing, allowing a more efficient classification based on the location of the wheezing frequency components, rather than the energy of their components.

The most relevant conclusions from the experimental results indicate the following: (i) the proposed method provides the best overall performance related to MP/PP wheezing classification compared to the most relevant method of the state-of-the-art; (ii) unlike most state-of-the-art methods based on classifiers, the proposed method is an unsupervised (blind) approach that does not require any training from wheezing sounds; (iii) the proposed method achieves removing most of the interference from normal respiratory sounds; (iv) specific accuracy rates,  $ACC_M$  and  $ACC_P$ , obtained by the proposed method seem to suggest the ability of the proposal to classify both monophonic and polyphonic wheezing sounds correctly.

Future work will be focused on the design of new constraints, to be applied in NMF approaches, that improve the modeling of time-frequency respiratory sound events, analyzing different types of adventitious sounds, such as wheezes and crackles. The objective

of this future research line is to perform an early detection and classification among the different types of adventitious sounds active in the auscultation process in order to maximize the reliability of the diagnosis issued by the physician in the case of pathologies of lung diseases caused by the appearance of such adventitious sounds.

**Author Contributions:** Conceptualization, J.D.L.T.C., F.J.C.Q., N.R.R., and G.P.C.; data curation, J.D.L.T.C. and F.J.C.Q.; formal analysis, J.D.L.T.C., F.J.C.Q., N.R.R., S.G.G., and J.J.C.O.; investigation, J.D.L.T.C. and F.J.C.Q.; methodology, J.D.L.T.C., F.J.C.Q., and N.R.R.; software, J.D.L.T.C., F.J.C.Q., S.G.G., and J.J.C.O.; supervision, F.J.C.Q., N.R.R., S.G.G., and J.J.C.O.; validation, J.D.L.T.C., F.J.C.Q., and G.P.C.; visualization, J.D.L.T.C.; writing, original draft, J.D.L.T.C., F.J.C.Q., N.R.R., S.G.G., J.J.C.O., and G.P.C.; writing, review and editing, J.D.L.T.C., F.J.C.Q., N.R.R., S.G.G., and J.J.C.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Programa Operativo FEDER Andalucía 2014–2020 under the project with Reference 1257914 and the Ministry of Economy, Knowledge and University, Junta de Andalucía under Project P18-RT-1994.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Terms of the Multiplicative Update Rules

Here are each of the terms belonging to the multiplicative update rule to obtain the basis matrix  $\mathbf{B}_W$ :

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_W} \right]^- = (\mathbf{X} \oslash \hat{\mathbf{X}}) \mathbf{A}_W^T \quad (\text{A1})$$

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_W} \right]^+ = ([\mathbf{1}] \mathbf{A}_W^T) \quad (\text{A2})$$

$$\left[ \frac{\partial \psi(\mathbf{B}_W)}{\partial \mathbf{B}_W} \right]_{f,k_w}^- = \sqrt{F} \left( \frac{B_{W_{f,k_w}} \sum_{j=1}^F B_{W_{j,k_w}}}{\left( \sum_{j=1}^F B_{W_{j,k_w}}^2 \right)^{\frac{3}{2}}} \right) \quad (\text{A3})$$

$$\left[ \frac{\partial \psi(\mathbf{B}_W)}{\partial \mathbf{B}_W} \right]_{f,k_w}^+ = \frac{1}{\sqrt{\frac{1}{F} \sum_{j=1}^F B_{W_{j,k_w}}^2}} \quad (\text{A4})$$

Here are each of the terms belonging to the multiplicative update rule to obtain the basis matrix  $\mathbf{B}_R$ :

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_R} \right]^- = (\mathbf{X} \oslash \hat{\mathbf{X}}) \mathbf{A}_R^T \quad (\text{A5})$$

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{B}_R} \right]^+ = ([\mathbf{1}] \mathbf{A}_R^T) \quad (\text{A6})$$

$$\begin{aligned} \left[ \frac{\partial \phi(\mathbf{B}_R)}{\partial \mathbf{B}_R} \right]_{f,k_r}^- &= 2F \left( \frac{(B_{R_{f-1,k_r}} + B_{R_{f+1,k_r}})}{\sum_{j=1}^F B_{R_{j,k_r}}^2} \right)^+ \\ &+ \frac{2FB_{R_{f,k_r}} \sum_{j=2}^F (B_{R_{j,k_r}} - B_{R_{j-1,k_r}})^2}{\left( \sum_{j=1}^F B_{R_{j,k_r}}^2 \right)^2} \end{aligned} \quad (\text{A7})$$

$$\left[ \frac{\partial \phi(\mathbf{B}_R)}{\partial \mathbf{B}_R} \right]_{f,k_r}^+ = \frac{4FB_{R,f,k_r}}{\sum_{j=1}^F B_{R,j,k_r}^2} \quad (\text{A8})$$

Here are each of the terms belonging to the multiplicative update rule to obtain the activations matrix  $\mathbf{A}_W$ :

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_W} \right]^- = \mathbf{B}_W^T (\mathbf{X} \oslash \hat{\mathbf{X}}) \quad (\text{A9})$$

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_W} \right]^+ = (\mathbf{B}_W^T [\mathbf{1}]) \quad (\text{A10})$$

$$\left[ \frac{\partial \phi(\hat{\mathbf{A}}_W)}{\partial \mathbf{A}_W} \right]_{k_w,t}^- = 2T \left( \frac{(A_{W_{k_w,t-1}} + A_{W_{k_w,t+1}})}{\sum_{i=1}^T A_{W_{k_w,i}}^2} \right)^+ + \frac{2TA_{W_{k_w,t}} \sum_{i=2}^T (A_{W_{k_w,i}} - A_{W_{k_w,i-1}})^2}{(\sum_{i=1}^T A_{W_{k_w,i}}^2)^2} \quad (\text{A11})$$

$$\left[ \frac{\partial \phi(\mathbf{A}_W)}{\partial \mathbf{A}_W} \right]_{k_w,t}^+ = \frac{4TA_{W_{k_w,t}}}{\sum_{i=1}^T A_{W_{k_w,i}}^2} \quad (\text{A12})$$

Here are each of the terms belonging to the multiplicative update rule to obtain the activations matrix  $\mathbf{A}_R$ :

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_R} \right]^- = \mathbf{B}_R^T (\mathbf{X} \oslash \hat{\mathbf{X}}) \quad (\text{A13})$$

$$\left[ \frac{\partial D_{KL}(\mathbf{X}|\hat{\mathbf{X}})}{\partial \mathbf{A}_R} \right]^+ = (\mathbf{B}_R^T [\mathbf{1}]) \quad (\text{A14})$$

## References

- World Health Organization. Chronic Respiratory Diseases. Available online: [https://www.who.int/health-topics/chronic-respiratory-diseases#tab=tab\\_1](https://www.who.int/health-topics/chronic-respiratory-diseases#tab=tab_1) (accessed on 30 December 2020).
- World Health Organization, Asthma. Available online: <https://www.who.int/news-room/fact-sheets/detail/asthma> (accessed on 30 December 2020).
- World Health Organization. Chronic Obstructive Pulmonary Disease. Available online: <http://www.emro.who.int/health-topics/chronic-obstructive-pulmonary-disease-copd/index.html> (accessed on 30 December 2020).
- Sarkar, M.; Madabhavi, I.; Niranjana, N.; Dogra, M. Auscultation of the respiratory system. *Ann. Thorac. Med.* **2015**, *10*, 158. [[CrossRef](#)] [[PubMed](#)]
- Pasterkamp, H.; Kraman, S.S.; Wodicka, G.R. Respiratory sounds: Advances beyond the stethoscope. *Am. J. Respir. Crit. Care Med.* **1997**, *156*, 974–987. [[CrossRef](#)]
- Lozano-Garcia, M.; Fiz, J.A.; Martinez-Rivera, C.; Torrents, A.; Ruiz-Manzano, J.; Jane, R. Novel approach to continuous adventitious respiratory sound analysis for the assessment of bronchodilator response. *PLoS ONE* **2017**, *12*, e0171455. [[CrossRef](#)]
- Ulukaya, S.; Serbes, G.; Kahya, Y.P. Wheeze type classification using non-dyadic wavelet transform based optimal energy ratio technique. *Comput. Biol. Med.* **2019**, *104*, 175–182. [[CrossRef](#)] [[PubMed](#)]
- Andrès, E.; Gass, R.; Charloux, A.; Brandt, C.; Hentzler, A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. *J. Med. Life* **2018**, *11*, 89. [[PubMed](#)]
- 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*, 276–288. [[CrossRef](#)]
- Leng, S.; San Tan, R.; Chai, K.T.C.; Wang, C.; Ghista, D.; Zhong, L. The electronic stethoscope. *Biomed. Eng. Online* **2015**, *14*, 1–37. [[CrossRef](#)]
- Sen, I.; Saraclar, M.; Kahya, Y.P. A comparison of SVM and GMM-based classifier configurations for diagnostic classification of pulmonary sounds. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1768–1776. [[CrossRef](#)]

12. Salazar, A.J.; Alvarado, C.; Lozano, F.E. System of heart and lung sounds separation for store-and-forward telemedicine applications. *Rev. Fac. Ingeniería Univ. Antioq.* **2012**, *64*, 175–181.
13. Douros, K.; Grammeniatis, V.; Loukou, I. Crackles and Other Lung Sounds. In *Breath Sounds*; Springer International Publishing: Cham, Switzerland, 2018; Chapter 12; pp. 225–236.
14. Lozano, M.; Fiz, J.A.; Jané, R. Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency. *IEEE J. Biomed. Health Inform.* **2015**, *20*, 486–497. [[CrossRef](#)]
15. Rao, A.; Huynh, E.; Royston, T.J.; Kornblith, A.; Roy, S. Acoustic methods for pulmonary diagnosis. *IEEE Rev. Biomed. Eng.* **2018**, *12*, 221–239. [[CrossRef](#)]
16. Pramono, R.X.A.; Bowyer, S.; Rodriguez-Villegas, E. Automatic adventitious respiratory sound analysis: A systematic review. *PLoS ONE* **2017**, *12*, e0177926. [[CrossRef](#)]
17. Rocha, B.M.; Pessoa, D.; Marques, A.; Carvalho, P.; Paiva, R.P. Automatic Classification of Adventitious Respiratory Sounds: A (Un) Solved Problem? *Sensors* **2021**, *21*, 57. [[CrossRef](#)] [[PubMed](#)]
18. Jin, F.; Krishnan, S.; Sattar, F. Adventitious sounds identification and extraction using temporal–spectral dominance-based features. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 3078–3087. [[PubMed](#)]
19. Gurung, A.; Scrafford, C.G.; Tielsch, J.M.; Levine, O.S.; Checkley, W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: A systematic review and meta-analysis. *Respir. Med.* **2011**, *105*, 1396–1403. [[CrossRef](#)] [[PubMed](#)]
20. Sakai, T.; Kato, M.; Miyahara, S.; Kiyasu, S. Robust detection of adventitious lung sounds in electronic auscultation signals. In Proceedings of the 21st IEEE International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 1993–1996.
21. Liu, X.; Ser, W.; Zhang, J.; Goh, D.Y.T. Detection of adventitious lung sounds using entropy features and a 2-D threshold setting. In Proceedings of the 2015 10th IEEE International Conference on Information, Communications and Signal Processing (ICICS), Singapore, 2–4 December 2015; pp. 1–5.
22. Matsutake, S.; Yamashita, M.; Matsunaga, S. Abnormal-respiration detection by considering correlation of observation of adventitious sounds. In Proceedings of the 2015 23rd IEEE European Signal Processing Conference (EUSIPCO), Nice, France, 31 August–4 September 2015; pp. 634–638.
23. Nakamura, N.; Yamashita, M.; Matsunaga, S. Detection of patients considering observation frequency of continuous and discontinuous adventitious sounds in lung sounds. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 3457–3460.
24. İçer, S.; Genç, Ş. Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds. *Digit. Signal Process.* **2014**, *28*, 18–27. [[CrossRef](#)]
25. Yamashita, M.; Himeshima, M.; Matsunaga, S. Robust classification between normal and abnormal lung sounds using adventitious-sound and heart-sound models. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 4418–4422.
26. Aykanat, M.; Kılıç, Ö.; Kurt, B.; Saryal, S. Classification of lung sounds using convolutional neural networks. *EURASIP J. Image Video Process.* **2017**, *65*, 1–9. [[CrossRef](#)]
27. Bardou, D.; Zhang, K.; Ahmad, S.M. Lung sounds classification using convolutional neural networks. *Artif. Intell. Med.* **2018**, *88*, 58–69. [[CrossRef](#)]
28. Ma, Y.; Xu, X.; Li, Y. LungRN+NL: An improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation. In Proceedings of the Interspeech 2020, 21st Annual Conference of the International Speech Communication Association, Virtual Event, Shanghai, China, 25–29 October 2020; pp. 2902–2906.
29. Demir, F.; Ismael, A.M.; Sengur, A. Classification of Lung Sounds With CNN Model Using Parallel Pooling Structure. *IEEE Access* **2020**, *8*, 105376–105383. [[CrossRef](#)]
30. Sovijarvi, A.; Dalmaso, F.; Vanderschoot, J.; Malmberg, L.; Righini, G.; Stoneman, S. Definition of terms for applications of respiratory sounds. *Eur. Respir. Rev.* **2000**, *10*, 597–610.
31. Meslier, N.; Charbonneau, G.; Racineux, J. Wheezes. *Eur. Respir. J.* **1995**, *8*, 1942–1948. [[CrossRef](#)] [[PubMed](#)]
32. Baughman, R.P.; Loudon, R.G. Lung sound analysis for continuous evaluation of airflow obstruction in asthma. *Chest* **1985**, *88*, 364–368. [[CrossRef](#)]
33. Cortes, S.; Jane, R.; Fiz, J.; Morera, J. Monitoring of wheeze duration during spontaneous respiration in asthmatic patients. In Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Shanghai, China, 17–18 January 2006; pp. 6141–6144.
34. 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*, 219–227. [[CrossRef](#)]
35. Zhang, J.; Ser, W.; Yu, J.; Zhang, T. A novel wheeze detection method for wearable monitoring systems. In Proceedings of the IEEE International Symposium on Intelligent Ubiquitous Computing and Education, Chengdu, China, 15–16 May 2009; pp. 331–334.
36. Lin, B.S.; Wu, H.D.; Chen, S.J. Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network. *J. Healthc. Eng.* **2015**, *6*, 649–672. [[CrossRef](#)]

37. Kochetov, K.; Putin, E.; Azizov, S.; Skorobogatov, I.; Filchenkov, A. Wheeze detection using convolutional neural networks. In *EPIA Conference on Artificial Intelligence*; Springer: Cham, Switzerland, 2017; pp. 162–173.
38. Kandaswamy, A.; Kumar, C.S.; Ramanathan, R.P.; Jayaraman, S.; Malmurugan, N. Neural classification of lung sounds using wavelet coefficients. *Comput. Biol. Med.* **2004**, *34*, 523–537. [[CrossRef](#)]
39. Le Cam, S.; Belghith, A.; Collet, C.; Salzenstein, F. Wheezing sounds detection using multivariate generalized Gaussian distributions. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 541–544.
40. Wisniewski, M.; Zielinski, T.P. Tonality detection methods for wheezes recognition system. In Proceedings of the IEEE 19th International Conference on Systems, Signals and Image Processing (IWSSIP), Vienna, Austria, 11–13 April 2012; pp. 472–475.
41. Wisniewski, M.; Zielinski, T.P. Joint application of audio spectral envelope and tonality index in an e-asthma monitoring system. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1009–1018. [[CrossRef](#)]
42. Chien, J.C.; Wu, H.D.; Chong, F.C.; Li, C.I. Wheeze detection using cepstral analysis in gaussian mixture models. In Proceedings of the 29th IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Lyon, France, 22–26 August 2007; pp. 3168–3171.
43. Bahoura, M. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Comput. Biol. Med.* **2009**, *39*, 824–843. [[CrossRef](#)]
44. Bahoura, M.; Pelletier, C. Respiratory sounds classification using Gaussian mixture models. In Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering, Niagara Falls, ON, Canada, 2–5 May 2004; Volume 3, pp. 1309–1312.
45. Mayorga, P.; Druzgalski, C.; Morelos, R.; Gonzalez, O.; Vidales, J. Acoustics based assessment of respiratory diseases using GMM classification. In Proceedings of the IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 6312–6316.
46. Taplidou, S.A.; Hadjileontiadis, L.J. Wheeze detection based on time-frequency analysis of breath sounds. *Comput. Biol. Med.* **2007**, *37*, 1073–1083. [[CrossRef](#)] [[PubMed](#)]
47. Jain, A.; Vepa, J. Lung sound analysis for wheeze episode detection. In Proceedings of the 30th IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Vancouver, BC, Canada, 20–25 August 2008; pp. 2582–2585.
48. 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 Proceedings of the 37th IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 5581–5584.
49. Oletic, D.; Bilas, V. Asthmatic wheeze detection from compressively sensed respiratory sound spectra. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1406–1414. [[CrossRef](#)]
50. Torre-Cruz, J.; Canadas-Quesada, F.; García-Galán, S.; Ruiz-Reyes, N.; Vera-Candeas, P.; Carabias-Orti, J. A constrained tonal semi-supervised non-negative matrix factorization to classify presence/absence of wheezing in respiratory sounds. *Appl. Acoust.* **2020**, *161*, 107188. [[CrossRef](#)]
51. De La Torre Cruz, J.; Cañadas Quesada, F.J.; Carabias Orti, J.J.; Vera Candeas, P.; Ruiz Reyes, N. Combining a recursive approach via non-negative matrix factorization and Gini index sparsity to improve reliable detection of wheezing sounds. *Expert Syst. Appl.* **2020**, *147*, 113212. [[CrossRef](#)]
52. Nagasaka, Y. Lung Sounds in Bronchial Asthma. *Allergol. Int.* **2012**, *61*, 353–363. [[CrossRef](#)]
53. Mason, R.C.; Murray, J.F.; Nadel, J.A.; Gotway, M.B. *Murray & Nadel's Textbook of Respiratory Medicine E-Book*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2015.
54. Taplidou, S.A.; Hadjileontiadis, L.J. Analysis of wheezes using wavelet higher order spectral features. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1596–1610. [[CrossRef](#)]
55. Forgacs, P. The functional basis of pulmonary sounds. *Chest* **1978**, *73*, 399–405. [[CrossRef](#)]
56. Jácome, C.; Oliveira, A.; Marques, A. Computerized respiratory sounds: A comparison between patients with stable and exacerbated COPD. *Clin. Respir. J.* **2017**, *11*, 612–620. [[CrossRef](#)]
57. Hashemi, A.; Arabalibiek, H.; Agin, K. Classification of wheeze sounds using wavelets and neural networks. In *International Conference on Biomedical Engineering and Technology*; IACSIT Press: Singapore, 2011; Volume 11, pp. 127–131.
58. Naves, R.; Barbosa, B.H.; Ferreira, D.D. Classification of lung sounds using higher-order statistics: A divide-and-conquer approach. *Comput. Methods Programs Biomed.* **2016**, *129*, 12–20. [[CrossRef](#)]
59. Ulukaya, S.; Sen, I.; Kahya, Y.P. A novel method for determination of wheeze type. In Proceedings of the 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, Turkey, 16–19 May 2015; pp. 2001–2004. [[CrossRef](#)]
60. Ulukaya, S.; Sen, I.; Kahya, Y.P. Feature extraction using time-frequency analysis for monophonic-polyphonic wheeze discrimination. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 5412–5415.
61. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [[CrossRef](#)] [[PubMed](#)]
62. Lee, D.D.; Seung, H.S. Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.* **2001**, *23*, 556–562.

63. 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. [[CrossRef](#)]
64. Dia, N.; Fontecave-Jallon, J.; Gumery, P.Y.; Rivet, B. Denoising Phonocardiogram signals with Non-negative Matrix Factorization informed by synchronous Electrocardiogram. In Proceedings of the 2018 26th IEEE European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 51–55.
65. 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 Proceedings of the 2018 10th International Conference on Bioinformatics and Biomedical Technology, Amsterdam, The Netherlands, 18–24 May 2018; pp. 18–24.
66. De La Torre Cruz, J.; Cañadas Quesada, F.J.; Ruiz Reyes, N.; Vera Candeas, P.; Carabias Orti, J.J. Wheezing Sound Separation Based on Informed Inter-Segment Non-Negative Matrix Partial Co-Factorization. *Sensors* **2020**, *20*, 2679. [[CrossRef](#)]
67. Févotte, C.; Bertin, N.; Durrieu, J.L. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Comput.* **2009**, *21*, 793–830. [[CrossRef](#)]
68. Liutkus, A.; Fitzgerald, D.; Badeau, R. Cauchy nonnegative matrix factorization. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 18–21 October 2015; pp. 1–5.
69. Canadas-Quesada, F.J.; 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. *EURASIP J. Audio Speech Music Process.* **2014**, *2014*, 26. [[CrossRef](#)]
70. Laroche, C.; Kowalski, M.; Papadopoulos, H.; Richard, G. A structured nonnegative matrix factorization for source separation. In Proceedings of the 2015 23rd IEEE European Signal Processing Conference (EUSIPCO), Nice, France, 31 August–4 September 2015; pp. 2033–2037.
71. Eggert, J.; Korner, E. Sparse coding and NMF. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 4, pp. 2529–2533.
72. Virtanen, T. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE Trans. Audio Speech Lang. Process.* **2007**, *15*, 1066–1074. [[CrossRef](#)]
73. Marxer, R.; Janer, J. Study of regularizations and constraints in NMF-based drums monaural separation. In Proceedings of the International Conference on Digital Audio Effects Conference (DAFx-13), Maynooth, Ireland, 2–6 September 2013.
74. Prominence Criterion of a Peak According to the MATLAB Software. Available online: [https://es.mathworks.com/help/signal/ref/findpeaks.html?searchHighlight=findpeak&\\_tid=doc\\_srchttitle#buff2uu](https://es.mathworks.com/help/signal/ref/findpeaks.html?searchHighlight=findpeak&_tid=doc_srchttitle#buff2uu) (accessed on 30 December 2020).
75. The r.a.l.e. Repository. Available online: <http://www.rale.ca> (accessed on 30 December 2020).
76. Stethographics Lung Sound Samples. Available online: <http://www.stethographics.com> (accessed on 30 December 2020).
77. 3m Littmann Stethoscopes. Available online: <https://www.3m.com> (accessed on 30 December 2020).
78. East Tennessee State University Pulmonary Breath Sounds. Available online: <http://faculty.etsu.edu> (accessed on 30 December 2020).
79. ICBHI 2017 Challenge. Available online: <https://bhichallenge.med.auth.gr> (accessed on 30 December 2020).
80. Lippincott NursingCenter. Available online: <https://www.nursingcenter.com> (accessed on 30 December 2020).
81. Thinklabs Digital Stethoscope. Available online: <https://www.thinklabs.com> (accessed on 30 December 2020).
82. Thinklabs Youtube. Available online: [https://www.youtube.com/channel/UCzEbKuIze4AI1523\\_AWiK4w](https://www.youtube.com/channel/UCzEbKuIze4AI1523_AWiK4w) (accessed on 30 December 2020).
83. Emedicine/Medscape. Available online: <https://emedicine.medscape.com/article/1894146-overview#a3> (accessed on 30 December 2020).
84. E-Learning Resources. Available online: <https://www.ers-education.org/e-learning/reference-database-of-respiratory-sounds.aspx> (accessed on 30 December 2020).
85. Respiratory Wiki. Available online: [http://respwiki.com/Breath\\_sounds](http://respwiki.com/Breath_sounds) (accessed on 30 December 2020).
86. Easy Auscultation. Available online: <https://www.easyauscultation.com/lung-sounds-reference-guide> (accessed on 30 December 2020).
87. Colorado State University. Available online: [http://www.cvmb.colostate.edu/clinsci/callan/breath\\_sounds.htm](http://www.cvmb.colostate.edu/clinsci/callan/breath_sounds.htm) (accessed on 30 December 2020).
88. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 1–27. [[CrossRef](#)]