Noname manuscript No.
(will be inserted by the editor)

Parallel source separation system for heart and lung sounds

**A.J. Muñoz-Montoro · D. Suarez-Dou ·
R. Cortina · F.J. Canadas-Quesada
E.F. Combarro**

Received: date / Accepted: date

Abstract In this paper, we propose a parallel source separation system designed to extract heart and lung sounds from single-channel mixtures. The proposed system is based on a non-negative matrix factorization (NMF) approach and a clustering strategy together with a soft-masking filtering. Furthermore, we propose an offline and online implementation of the framework which can be applied in many real-time scenarios, such as the extraction of clinical parameters, remote auscultation, breath sound analysis, etc. Experimental results show that it is possible to achieve fast execution times, which enable a real-time behavior, combining parallel and high-performance techniques.

Keywords Non-negative matrix factorization (NMF) · Heart · Lung · Real time · Parallel computing · Sound source separation · Single-channel

A.J. Muñoz-Montoro
Department of Telecommunication Engineering, Universidad de Jaén, Spain
E-mail: jmontoro@ujaen.es

D. Suarez-Dou
Department of Computer Science, Universidad de Oviedo, Spain
E-mail: suarezddavid@uniovi.es

R. Cortina
Department of Computer Science, Universidad de Oviedo, Spain
E-mail: raquel@uniovi.es

F.J. Canadas-Quesada
Department of Telecommunication Engineering, Universidad de Jaén, Spain
E-mail: fcanadas@ujaen.es

E.F. Combarro
Department of Computer Science, University of Oviedo, Spain
E-mail: efernandezca@uniovi.es

1 Introduction

The classical stethoscope is an acoustic medical device invented in France in 1816 by René Laënnec. Stethoscopes are mainly used for examining cardiac and respiratory functions of the human body. This process is called auscultation and it remains an integral and important part of clinical medicine. Inspecting the sounds of the body in a non-invasive way provides an important source of clinical information, since cardiac and respiratory sounds potentially contain useful information about the status of the heart and lungs. Unfortunately, cardiac sounds interfere with respiratory sounds, and vice versa, what makes the analysis difficult by using a classical stethoscope. In this regard, removing the acoustic interference between heart and lungs is a challenging task, since both sound sources are simultaneously active in time and frequency domain.

In biomedical research, this task is often treated as a blind source separation problem [9,24,17,6] and has become an important and hot research topic. Early methods were based on combinations of low-pass filters (LPF) and high-pass filters (HPF) [13,7] with the aim of extracting heart and lung sounds. The main drawback of these approaches is the significant information that is removed in the filtering process. During the last decades, different techniques have been used to perform the source separation in single-channel mixtures.

A common approach to this type of problem is based on independent component analysis (ICA) [9,3,19,15], in which the underlying source signals are constrained to be statistically independent. These systems often use sensors to obtain signals from different points of the human chest. Unfortunately, ICA-based methods are not robust in noisy scenarios and do not consider the non-stationary nature of the heart sound in the modeling of the problem, what degrades the obtained results. Other methods in the literature based on nonlocal means (NLM) [26,20] use redundant information of heart sounds to remove the non-target sounds at different instants. Nevertheless, these systems entail high computational costs. Some methods perform the separation using wavelet transform (WT) based filter [16]. In WT based filter approaches, an adaptive separation of desired signal is achieved through an iterative wavelet decomposition-reconstruction process based on either hard or soft thresholding of the WT coefficients at each iteration. However, these approaches are not robust by evaluating real signals due to the thresholding process [25,6]. More recent methods are based on non-negative matrix factorization (NMF) [17,23,6]. These perform a decomposition process using different constraints to discriminate both sources.

In this paper, we propose a parallel framework that addresses the separation of cardiac and respiratory sounds from a single observation mixture. Here, we have developed an efficient system which decomposes the input mixture by combining a NMF approach and a clustering strategy together with a soft-masking filtering. The proposed solution is suitable for different scenarios and applications. Unlike previous NMF-based methods [6], the main contribution of this work is the development of a framework that allows to perform the separation for both offline and online approaches, but always with a real-time applicability. In this sense, our offline solution is suitable for scenarios in which the source separation is required with a very short delay just after the recording of the audio sample. From a medical point of view, this application could be interesting for the estimation of clinical parameters based on the separation of cardiorespiratory sounds. The tested scenarios show that it is possible to reach this real-time behavior combining parallel and high-performance techniques. On the other hand, we propose an online variant of the developed system to perform directly the separation in real time. This online implementation would enable the development of electronic stethoscopes capable of providing doctors with clean heart/lung sounds while monitoring patients.

According to the best of our knowledge, there has not yet been presented a holistic, flexible and free system that addresses this problem on parallel shared-memory systems. As a proof of concept, some experiments are carried out on a dataset of real-world audio samples, showing reliable results in terms of sound separation.

The structure of the article is organized as follows. In Sect. 2, we present the proposed framework and describe the main function of the stages that compose it. In Sect. 3, we analyze the complexity of the proposed system. Sect. 4 describes the online variant of the system. The evaluation setup and the experimental results are shown in Sect. 5. Finally, conclusions are summarized in Sect. 6.

2 Framework description

2.1 System overview

In this work, we present a parallel system to separate sounds emitted by the heart from those emitted by the lungs in single-channel mixtures. In particular, we propose a practical and versatile framework that can be used for different real-time applications, such as extraction of clinical parameters, remote auscultation, breath sound analysis, etc. For this purpose, we have developed an efficient and fast implementation that is able to perform the decomposition of the input mixture by combining a NMF approach and a clustering strategy. As a result, we propose a software solution that satisfies two essential requirements: mobility and real time. Therefore, our design takes into account the low memory resources and low computational power of cheap and handheld devices, what can allow an easy implementation in the medical services. This has been possible using and deeply exploiting the possibilities offered by parallel architectures.

The block diagram of the proposed framework is depicted in Fig. 1. As can be observed, the full system has been decomposed into four main stages: signal representation, signal decomposition, clustering and reconstruction. In the following subsections, we detail and describe the main function of each stage.

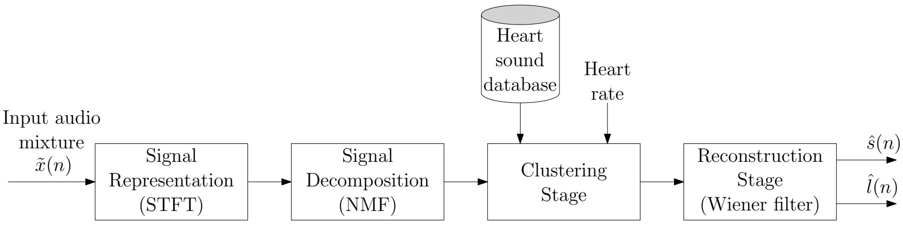

Heart sound database
Heart rate
Input audio mixture $\tilde{x}(n)$
Signal Representation (STFT)
Signal Decomposition (NMF)
Clustering Stage
Reconstruction Stage (Wiener filter)
$\hat{s}(n)$
$\hat{l}(n)$

Fig. 1 Block diagram of the proposed framework.

2.2 Signal representation

The problem considered in this paper is to separate the signals generated by the heart and the lung using a single microphone. Therefore, the observed signal $\tilde{x}(n)$ can be expressed as

$$\tilde{x}(n) = \tilde{s}(n) + \tilde{l}(n), \tag{1}$$

where $\tilde{s}(n)$ is the heart signal and $\tilde{l}(n)$ is the lung signal. Considering the linear problem in Eq. 1, the short-time Fourier transform (STFT) of $\tilde{x}(n)$ can be written as

$$x(f,t) = s(f,t) + l(f,t),$$

where $x(f,t)$, $s(f,t)$ and $l(f,t)$ denote the time-frequency spectrograms of $\tilde{x}(n)$, $\tilde{s}(n)$ and $\tilde{l}(n)$, respectively. Here, $f \in [1,F]$ and $t \in [1,T]$. Collecting the F frequency bins and T time frames, we define the magnitude spectrogram matrices $\mathbf{X} \in \mathbb{R}^{F\times T}$, $\mathbf{S} \in \mathbb{R}^{F\times T}$ and $\mathbf{L} \in \mathbb{R}^{F\times T}$, where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ and $\mathbf{x}_t = [|x(1,t)|, \ldots, |x(F,t)|]^{\mathrm{T}}$. $\mathbf{S}$ and $\mathbf{L}$ are defined similarly to $\mathbf{X}$.

2.3 Signal decomposition

2.3.1 NMF-based signal model

Traditional NMF-based approaches perform the source separation by modelling the magnitude spectrogram $\mathbf{X}$ of the input mixture as

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{W}\mathbf{H}, \tag{2}$$

where $\hat{\mathbf{X}} \in \mathbb{R}^{F\times T}$ is the estimated matrix, $\mathbf{H} \in \mathbb{R}^{K\times T}$ is the activations matrix, $\mathbf{W} \in \mathbb{R}^{F\times K}$ is the basis matrix and K denotes the number of bases.

The proposed method attempts to separate both heart and lung sounds using a NMF-based decomposition and clustering the heart and lung bases obtained in the factorization process. Thus, the estimated heart and lung signals, $\hat{\mathbf{S}} \in \mathbb{R}^{F\times T}$ and $\hat{\mathbf{L}} \in \mathbb{R}^{F\times T}$, can be obtained as

$$\hat{\mathbf{S}} = \ddot{\mathbf{W}}\ddot{\mathbf{H}}, \qquad \hat{\mathbf{L}} = \bar{\mathbf{W}}\bar{\mathbf{H}},$$
$$\hat{\mathbf{X}} = \hat{\mathbf{S}} + \hat{\mathbf{L}} = \underbrace{\begin{bmatrix} \ddot{\mathbf{W}} & \bar{\mathbf{W}} \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} \ddot{\mathbf{H}} \\ \bar{\mathbf{H}} \end{bmatrix}}_{\mathbf{H}},$$

where $K = \ddot{K} + \bar{K}$, $\ddot{\mathbf{W}} \in \mathbb{R}^{F\times\ddot{K}}$ and $\ddot{\mathbf{H}} \in \mathbb{R}^{\ddot{K}\times T}$ are the bases and activations matrices related to the heart signal, and $\bar{\mathbf{W}} \in \mathbb{R}^{F\times\bar{K}}$ and $\bar{\mathbf{H}} \in \mathbb{R}^{\bar{K}\times T}$ are the bases and activations matrices related to the lung signal~~, and ^{T} is the transpose operator~~.

2.3.2 *Factorization process*

In this work, we propose a NMF approach to factorize the signal model parameters in Eq. 2 under the nonnegativity restriction. The estimation of these parameters is obtained by minimizing the generalized Kullback-Leibler $D_{KL}(\mathbf{X}|\hat{\mathbf{X}})$ divergence [11] and the sparsity constraint $D_{SS}(\mathbf{X}|\hat{\mathbf{X}})$ [27]. Thus, the global cost function to be minimized can be defined as,

$$D(\mathbf{X}|\hat{\mathbf{X}}) = D_{KL}(\mathbf{X}|\hat{\mathbf{X}}) + D_{SS}(\mathbf{X}|\hat{\mathbf{X}}).$$

In our approach, $D(\mathbf{X}|\hat{\mathbf{X}})$ is minimized by using an iterative approach based on multiplicative update rules. In this way, we have efficiently implemented the following update rules

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{\frac{\mathbf{X}}{\mathbf{WH}}\mathbf{H}^{\mathrm{T}}}{\mathbf{O}\mathbf{H}^{\mathrm{T}}}, \tag{3}$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W}^{\mathrm{T}}\frac{\mathbf{X}}{\mathbf{WH}}}{\mathbf{W}^{\mathrm{T}}\mathbf{O} + \lambda}, \tag{4}$$

where $\odot$ represents the Hadamard (element-wise) product, $\mathbf{W}$ and $\mathbf{H}$ are initialized as random positive matrices and $\mathbf{O} \in \mathbb{N}^{F\times T}$ denotes an all-ones matrix composed of F rows and T columns.

2.4 Clustering stage

This section presents three techniques for clustering the bases obtained in the factorization process previously described. This clustering stage is required to distinguish which bases are related to heart sounds and which to lung sounds. To that end, the following clustering methods exploit both spectral and temporal features of the heart and lung sounds.

2.4.1 Spectral Correlation (SC)

The aim of this clustering method is to compute the spectral similarity between the bases obtained from NMF decomposition and a dictionary of spectral patterns learned from isolated heart signals. In particular, we propose to measure this similarity using the cosine distance in order to compute the spectral correlation between the dictionary entries and the obtained bases. In this sense, the cosine measure as a similarity function can be expressed as

$$SC(i,j) = \frac{\mathbf{w}_i \mathbf{d}_j}{||\mathbf{w}_i||\ ||\mathbf{d}_j||} = \frac{\sum_{f=1}^{F} w_i(f)\ d_j(f)}{\sqrt{\sum_{f=1}^{F} w_i^2(f)}\sqrt{\sum_{f=1}^{F} d_j^2(f)}},$$

where $\mathbf{w}_i$ is the i-th basis of the matrix $\mathbf{W}$ and $\mathbf{d}_j$ is the j-th entry of the dictionary $\mathbf{D}$. It worth noting that in this definition $SC(i,j) \in [0,1]$. Conceptually, a value of 1 indicates that a basis has a strong similarity to a heart sound. On the other hand, the closer value to 0, the less the similarity to a heart sound and the greater the match to a lung sound. Therefore, we propose the following criterion to determine the nature of each basis,

$$\begin{cases} \mathbf{w}_i \in \ddot{\mathbf{W}} & \text{if } s_i \geq U \\ \mathbf{w}_i \in \bar{\mathbf{W}} & \text{if } s_i < U \end{cases}, \tag{5}$$

where $s_i = \max\{SC(i,j)\}$ and U is a predefined threshold.

2.4.2 Roll-Off (RO)

This second clustering method is based on the power spectral density (PSD). The PSD describes the energy distribution of a signal at different frequencies. The PSD of the heart and lung sounds are clearly distributed in different frequency band [8,21,17]. Heart signals concentrate most of their energy in the frequency range $[0-260]$ Hz, while lung signals distribute most of their energy in the frequency range $[260-\frac{f_s}{2}]$ Hz [8,6], being f_s the sampling rate. Thus, analyzing the PSD of each basis $\mathbf{w}_i$ can be useful to determine whether it represents a heart or lung sound.

As in [18], here we classify a basis $\mathbf{w}_i$ as a heart sound if at least 85% of its total energy is concreted in the frequency range $[260-\frac{f_s}{2}]$ Hz. This can be expressed by the following criterion,

$$\begin{cases} \mathbf{w}_i \in \ddot{\mathbf{W}} & \text{if } RO(i) \geq E_T(i) \\ \mathbf{w}_i \in \bar{\mathbf{W}} & \text{if } RO(i) < E_T(i) \end{cases}, \tag{6}$$

where

$$RO(i) = \sum_{t=1}^{T}\sum_{f=1}^{F_0} ||w_i(f)h_i(t)||^2$$

$$E_T(i) = 0.85\sum_{t=1}^{T}\sum_{f=1}^{F} ||w_i(f)h_i(t)||^2$$

with $F_0 = 260$ Hz and $F = \frac{f_s}{2}$.

2.4.3 Temporal correlation (TC)

Finally, this method allows to cluster the bases according to a temporal measure applied to the activations $\mathbf{H}$. Here, we propose to compute the temporal correlation between the heart rate $\mathbf{r} \in \mathbb{R}^T$ and the activations $\mathbf{H}$. In this sense, the estimation of $\mathbf{r}$ is carried out using the information included in the frequency band $[0 - 60]$ Hz of the input spectrogram, where heart and lung sounds do not overlap. First a binary function is built based on the signal energy changes and then temporal peaks that may correspond to a heartbeat are looking for. At the end, a sequence of pulses between 0 and 1 is obtained, where the value 1 indicates the temporal location of the heart sounds. This process generates a temporal pattern $\mathbf{r}$ which can be considered equivalent to the heart rate [4].

Afterwards, a preprocessing over each activation $\mathbf{h}_i$ obtained in the NMF decomposition is required to compute the temporal correlation with respect to the estimated heart rate $\mathbf{r}$. Thus, the activations $\mathbf{H}$ are preprocessed as,

$$h_i(t) = \begin{cases} 1 & \text{if } h_i(t) \geq \frac{\sum_T h_i(t)}{T} \\ 0 & \text{if } h_i(t) < \frac{\sum_T h_i(t)}{T} \end{cases}. \tag{7}$$

In this way, each $\mathbf{h}_i$ is converted to a sequence of 0 and 1 that can be compared to the heart rate.

Finally, a correlation coefficient is computed for each i-th component of the matrix $\mathbf{H}$ using Eq. 8,

$$TC(i) = \frac{1}{T-1}\left(\sum_{t=1}^{T} \frac{h_i(t) - \mu_{\mathbf{h}_i}}{\sigma_{\mathbf{h}_i}} \frac{r(t) - \mu_{\mathbf{r}}}{\sigma_{\mathbf{r}}}\right), \tag{8}$$

where $\mu_{\mathbf{h}_i}$ and $\sigma_{\mathbf{h}_i}$ represent the mean and standard deviation of the i-th activation of the matrix $\mathbf{H}$, respectively, and $\mu_{\mathbf{r}}$ and $\sigma_{\mathbf{r}}$ indicate the mean and standard deviation of the estimated heart rate $\mathbf{r}$. $TC(i)$ can reach a value in the range $[-1, 1]$, where -1 denotes that the i-th activation belongs to a lung sound and 1 indicates that the i-th activation belongs to a heart sound. Thus, we can define the following criterion

$$\begin{cases} \mathbf{w}_i \in \ddot{\mathbf{W}} & \text{if } TC(i) \geq 0 \\ \mathbf{w}_i \in \bar{\mathbf{W}} & \text{if } TC(i) < 0 \end{cases}. \tag{9}$$

Once detailed these three clustering methods (SC, RO and TC), note that in our implementation all of these clustering strategies are jointly used to classify the bases. We propose to combine all of them to label each $\mathbf{w}_i$ following a conservative strategy, i.e., $\mathbf{w}_i$ is assigned to the matrix $\ddot{\mathbf{W}}$ when it satisfies at least the criterion of one of these three methods (Eq. 5, Eq. 6, Eq. 9). Regarding the evaluation of the clustering process, [6] showed that the best result in terms of source separation is obtained when the three methods are jointly used. Moreover, the authors proved that RO obtains superior results comparing to TC and SC when the clustering is performed using only one method. It is also important to remark that, while RO and SC can be used for both offline and online implementations, TC does not allow an online implementation since the whole input signal is required for computing the heart rate $r(t)$.

2.5 Reconstruction

Finally, the target signals are reconstructed by using a soft-filter strategy. Firstly, the estimated parameters are used to predict the magnitude of heart and lung signal spectrograms by

$$\hat{\mathbf{S}} = \ddot{\mathbf{W}}\ddot{\mathbf{H}}, \qquad \hat{\mathbf{L}} = \bar{\mathbf{W}}\bar{\mathbf{H}}. \tag{10}$$

Then, the source signals $\hat{s}(n)$ and $\hat{l}(n)$ are estimated from the mixture $x(n)$ using a generalized time-frequency mask over the STFT domain. In this sense, we propose to employ a Wiener filtering in order to ensure that the reconstruction process is conservative [14]. The Wiener masks can be computed as the relative energy contribution of each source for each time-frequency bin with respect to the energy of the original mixture,

$$\ddot{\mathbf{V}} = \frac{|\hat{\mathbf{S}}|^2}{|\hat{\mathbf{S}}|^2 + |\hat{\mathbf{L}}|^2}, \qquad \bar{\mathbf{V}} = \frac{|\hat{\mathbf{L}}|^2}{|\hat{\mathbf{S}}|^2 + |\hat{\mathbf{L}}|^2}. \tag{11}$$

Afterwards, Eq. 11 are used to obtain the source magnitude spectrograms $\hat{\mathbf{S}}$ and $\hat{\mathbf{L}}$.

$$\hat{\mathbf{S}} = \ddot{\mathbf{V}} \odot \mathbf{X}, \qquad \hat{\mathbf{L}} = \bar{\mathbf{V}} \odot \mathbf{X}. \tag{12}$$

3 Computational costs

In this section, we study the complexity of the proposed system that has been implemented combining parallel and high-performance techniques with the goal of achieving an application suitable for multi-core architectures and for

Algorithm 1 Proposed system algorithm

```
Calculate $\mathbf{X}$ to obtain the signal representation in the frequency domain using the STFT.
Initialize $\mathbf{W}$ and $\mathbf{H}$ matrix with random non-negative values.
for $iter = 1$ to $N_{\text{iter}}$ do
    Update $\mathbf{W}$ using the Eq. 3.
    Update $\mathbf{H}$ using the Eq. 4.
end while
for $i = 1$ to $K$ do
    Determine if $\mathbf{w}_i \in \ddot{\mathbf{W}}$ or $\mathbf{w}_i \in \bar{\mathbf{W}}$ using Spectral Correlation (see Sec. 2.4.1).
    Determine if $\mathbf{w}_i \in \ddot{\mathbf{W}}$ or $\mathbf{w}_i \in \bar{\mathbf{W}}$ using Roll-Off (see Sec. 2.4.2).
    Determine if $\mathbf{w}_i \in \ddot{\mathbf{W}}$ or $\mathbf{w}_i \in \bar{\mathbf{W}}$ using Temporal Correlation (see Sec. 2.4.3).
end for
Compute $\hat{\mathbf{S}}$ and $\hat{\mathbf{L}}$ using the Eq. 10.
Reconstruct the heart and lung signals in time domain using the inverse of the STFT
  of $\hat{\mathbf{S}}$ and $\hat{\mathbf{L}}$, respectively.
```

real-time scenarios. Next, we analyze the computational cost of the parallel proposal for each stage described in Fig. 1 and summarized in Algorithm 1.

First, the time-frequency spectrogram of the observation is computed following a process of segmenting and windowing of the audio signal into frames, and calculating the fast Fourier transform (FFT) spectrum in each frame. In our implementation we have used the FFTW package [12]. As studied in [2, 1], the temporal complexity of computing the FFT spectrum of one frame is given by $O(N_F \log_2(N_F))$, where N_F is the number of point used in the FFT. Thus, the complexity for computing sequentially the spectrogram of the input signal can be expressed as

$$O\left(T\left(N_F\ \log_2(N_F)\right)\right),$$

where T is the number of frames.

Regarding the parallel design, parallel and worksharing constructors of OpenMP [10] have been exploited. Moreover, some empirical tests have shown that better performance is obtained when the T FFTs are executed in parallel than when the T FFTs are run sequentially using the parallel implementation of the FFTW package, despite the fact that both designs have the same theoretical complexity cost. Thus, we have chosen a coarse-grained parallelism for our implementation. In this sense, the parallel complexity of this stage is given by

$$O\left(\frac{T}{p}\left(N_F\ \log_2(N_F)\right)\right),$$

where p is the number of processors or cores used.

Attending to the factorization stage (see Sect. 2.3.2), the multiplicative update rules of Eq. 3 and Eq. 4 have been implemented considering the following parallelization strategies: (1) calling BLAS [5] Level 1, 2 and 3 routines for vector-vector, vector-matrix and matrix-matrix operations, respectively, and (2) using OpenMP directives where BLAS routines can not be implemented. Both Eqs. 3 and 4 ~~leads~~ lead two matrix-matrix products (computed by calling BLAS subroutine dgemm) with other auxiliary operations of lower computational intensity and vector-matrix. Thereby, the theoretical parallel computational cost can be approximated by

$$O\left(\frac{N_{\text{iter}}N_FTK}{p}\right),$$

where N_{iter} is the number of iterations of the NMF approach and K denotes the number of bases used for the factorization process (see Sect. 2.3).

The computational complexity of the clustering stage (see Sect. 2.4) depends on the complexity of each clustering method. Thus, the sequential complexity of each method can be approximated by

$$\begin{aligned} \textit{Spectral Correlation}: &\quad O\left(KN_FK_D\right), \\ \textit{Roll-off}: &\quad O\left(KN_FT^2\log_2\left(N_FT\right)\right), \\ \textit{Temporal Correlation}: &\quad O\left(KT\right), \end{aligned}$$

where K_D denotes the number of bases of the dictionary of spectral patterns learned from isolated heart signals. Finally, combining these methods and considering the property of sum of the Big O notation, the computational cost of the sequential version is approximately given by

$$O\left(KN_F\left(K_D+T^2\log_2(N_FT)\right)\right).$$

Attending to the parallel version, a fine-grain parallelism design is not suitable for this stage due to low intensity operations, branch divergence, low temporal and spacial locality, etc. After some calibration experiments, we have chosen a coarse-grained parallelism strategy, exploiting the independence of the values of K bases, using OpenMP directives and BLAS routines where possible. Therefore, the parallel theoretical computational cost can be approximated by

$$O\left(\frac{KN_F\left(K_D+T^2\log_2(N_FT)\right)}{p}\right).$$

In the design of the last stage (see Sect. 2.5), two parallelism strategies have been used exploiting the OpenMP routines. First, a fine-grain parallelism is applied for computing the Wiener masks (see Eq. 11). Then, as in previous stages, a coarse-grained parallelism is applied for the signal reconstruction process (see Eq. 12), so that the reconstruction of the heart and lung signals is computed simultaneously, each one sequentially (nested parallelism is deactivated), using the FFTW package. Thus, the computational complexity is given by

$$O\left(N_FT\log_2(N_F)+\frac{N_FT}{p}\right).$$

Finally, analyzing the theoretical computational complexity of the whole system, we can conclude that the signal decomposition and the clustering stages have a theoretical efficiency close to the maximum, while the reconstruction stage tends to zero when the number of cores and the size of the problem grows. However, since this last stage is not particularly expensive from the computational point of view, a high empirical efficiency can be expected, as long as the audio duration is high and/or the number of processors used is suitable for the magnitude of the problem.

4 Proposed online implementation

In this section, we propose an online scheme for our algorithm to separate heart and lungs sounds in real time. As previously commented, an online variant would enable the development of electronic stethoscopes that would perform the separation of cardiorespiratory sounds in real time while patients are being monitored.

The method proposed in Sect. 2 processes the audio signal as a whole, i.e. requires the whole audio from start to end. However, an online system has to process its input frame-by-frame in a serial fashion, i.e. in the order that the audio stream feeds to the algorithm, without having the whole input available from the start. Thus, here we propose some slight changes that have to be carried out over the proposed offline algorithm.

First, the clustering criterion based on the temporal correlation between the activations and the heart rate can not be considered in the clustering process. Note that the whole input signal is required for the computation of the heart rate and for the processing performed in Eq. 7. This is not critical, since [6] showed that removing this criterion of the clustering process degrades the separation results less than 1.5 dB in terms of the source-to-distortion ratio (SDR).

In addition, as ~~not~~ the whole audio is not initially available, the factorization stage must be adapted to perform the separation as the new audio frames arrive in the system. In this way, we propose to follow the windowing strategy presented in [22]. San Juan et al. [22] proposed an algorithm that recalculates efficiently the NMF when a new audio frame is added to the data matrix. This implementation allows to estimate the bases matrix using a fixed set of activations. This set is updated as new audio frames arrive (windowing strategy).

Following these two basic modifications, a real-time implementation of the proposed framework can be achieved. For the sake of brevity, we propose to focus the experimentation on the offline proposal.

5 Evaluation and experimental results

This section presents the experimentation carried out in the evaluation of the proposed system. In this evaluation, we have developed a synthetic database using a single-channel real-world audio mixture of heart and lung sounds sampled at 8 KHz and with a duration of 7 s. In this regard, we have replicated this audio sample to generate a set of audio files with a duration between 7 and 602 s. This set of samples is used to evaluate the performance of the system in terms of efficiency, speedup and limitations.

Regarding the used testbed, we have focused on the NVIDIA Jetson AGX Xavier development kit, which is an embedded system-on-chip (SoC) with an eight-core ARM v8.2 64-bit CPU. Xavier supports different kinds of running modes (configurable with the *NVPModel* command tool). In this way, different power consumption (10 W, 15 W and 30 W), running cores (2, 4, 6 and 8) and CPU frequencies can be selected using *NVPModel*. This setup allows to simulate a wide range of mobile devices such as smartphones, laptops, tablets and other embedded systems under controlled conditions. As the purpose is to test our system on up-to-date embedded systems, we have selected two test modes following the current market trend: Mode 2 (4 cores, 15W) and Mode 7 (8 cores, 30 W).

Concerning the used software, Xavier runs Ubuntu Linux 18.04.1 LTS, the OpenBlas[1] library (release 0.3.9, March 2020), the FFTW[2] library (release 3.3.8, May 2018) and the GNU C Compiler 7 with the specification 4.5 of OpenMP. OpenBLAS is an optimized BLAS library based on GotoBLAS2 1.13 BSD. Note that both packages have been built in our system from source codes. Finally, it should be remarked that the used data type is "*double*" (i.e. IEEE 754 double-precision binary floating-point format).

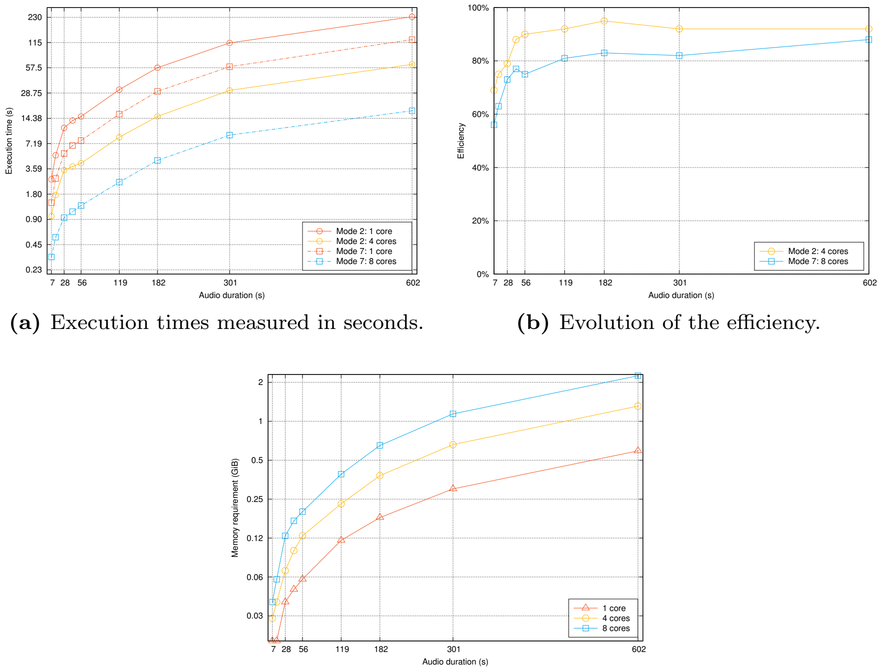
~~In our experiments, we have measured the execution time and memory consumption measured as resident set size (RSS) depending on the duration of the audio mixture for testing the reliability of our proposed framework. Figure 2 summarizes the obtained results.~~

In our experiments, for testing the reliability of our proposed framework, we have measured the execution time and the memory consumption depending on the duration of the audio mixture. The memory consumption is estimated by measuring the resident set size (RSS) using system tools (e.g. calling `system("ps -o rrs | grep ...")` at the end of the target application and before releasing memory). Figure 2 summarizes the obtained results.

Firstly, sequential execution times in Mode 7 are, approximately, reduced by a half compared to Mode 2, which tallies with the possibility to be doubling the CPU frequencies (see Fig. 2a). Parallel execution times in Mode 7 are, approximately, the fourth part of the times in Mode 2, due to the number of running cores and the CPU frequencies. The ratio between the audio signal duration and the required processing time of the system for the sequential approach worst-case scenario is less than 40% and 21% for device operation Modes 2 and 7, respectively. This means that the system can be used in low-latency applications, understanding latency as the delay between receiving the signal and starting to perform the separation. Furthermore, this latency can

[1] `https://www.openblas.net`

[2] `http://www.fftw.org`



(a) Execution times measured in seconds.

(b) Evolution of the efficiency.

(c) Memory requirement (RSS) measured in GiB.

Fig. 2 Experimental results as a function of the audio duration measured in seconds and the operation mode of the NVIDIA AGX Xavier.

be reduced significantly by the parallel approach to less than 14% and 5% for device operation Modes 2 (four cores) and 7 (eight cores), respectively, in the worst-case scenario.

The efficiency of the system (see Fig. 2b) increases significantly as the audio duration grows until 40 s long, when it turns into a slight linear growth. As expected, when the number of cores increases, the efficiencies slightly decrease. Therefore, we can assert that our system scales correctly according to the theoretical complexity estimations when the number of processors and the size of the problem grow.

As expected, Fig. 2c shows that memory consumption increases as the length of the audio becomes longer. Furthermore, owing to the parallelization strategy followed (see Sec. 2.4), memory consumption also grows as the number of cores used increases, since some data structures are replicated according to the number of cores. ~~Consequently, we can affirm that, for any length of audio the parallel approach requires, approximately, the amount of memory used by the sequential approach multiplied by half the number of cores.~~ Consequently, for any length of audio, we can claim that the parallel approach requires in the worst case the amount of memory used by the sequential approach multiplied by the total number of used processors or cores.

Execution times (s)						
Stages	*Separation*	*Clustering*	*Reconstr.*	*Separation*	*Clustering*	*Reconstr.*
Mode 2 (15W)						
Dur. (s)	1 core			4 core		
7	0.39	2.29	0.01	0.21	0.76	0.01
14	0.80	4.39	0.03	0.33	1.41	0.01
28	1.74	9.26	0.05	0.62	2.83	0.03
42	2.84	10.64	0.06	0.93	2.87	0.03
56	3.77	11.25	0.08	1.16	2.99	0.04
119	8.33	23.10	0.14	2.41	6.06	0.08
182	13.21	44.16	0.23	3.81	11.21	0.12
301	22.55	90.92	0.35	6.37	24.43	0.18
602	46.80	185.87	0.70	13.40	49.32	0.38
Mode 7 (30W)						
Dur. (s)	1 core			8 core		
7	0.21	1.20	0.01	0.09	0.22	0.00
14	0.43	2.32	0.01	0.12	0.42	0.01
28	0.93	4.54	0.02	0.21	0.71	0.01
42	1.52	5.26	0.03	0.32	0.78	0.01
56	2.02	5.74	0.04	0.41	0.88	0.02
119	4.45	11.60	0.08	0.77	1.69	0.04
182	7.05	22.93	0.12	1.19	3.26	0.07
301	12.10	46.97	0.19	2.24	6.72	0.10
602	25.03	99.11	0.43	3.88	13.57	0.24

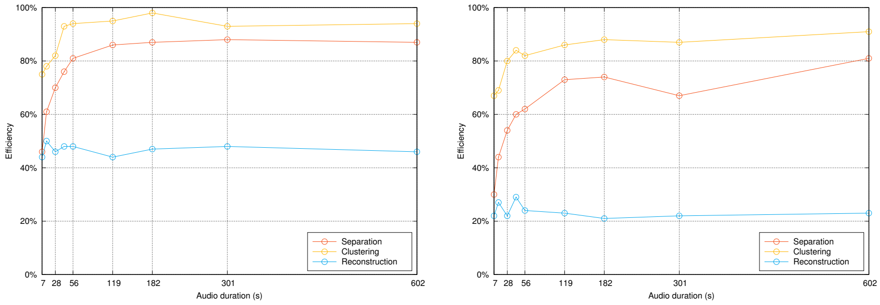
Table 1 Execution times measured in seconds for each stage and operating mode.

Table 1 summarize the execution times of each stage for both Mode 2 and Mode 7. As previously explained in Sect. 3, the computational complexity of the whole system depends mainly on two stages: the clustering stage, which represents approximately 77% of the whole system execution time; and the factorization stage, which represents approximately 22% of the whole time.

According to the efficiency of each stage (see Fig. 3), the factorization and the clustering stages have achieved high results, close to the theoretical maximum estimations, especially for audio duration longer than 50 s. The low values reached by the reconstruction stage do not have a strong impact over the overall system performance, as described in Sect. 3 ~~and it is as reasonably as expected~~. Therefore, it is empirically demonstrated that the obtained results are similar to the theoretical analysis performed.

6 Conclusion

In this paper, we have proposed a parallel framework that performs the separation of cardiac and respiratory sounds from a single-channel mixture. Our system has focused on achieving fast execution times that allows its implementation in real-time scenarios, reaching reliable results in terms of sound



(a) Evolution of the efficiency in Mode 2 (15 W).

(b) Evolution of the efficiency in Mode 7 (30 W).

Fig. 3 Experimental results as a function of the audio duration measured in seconds, the operation mode of the NVIDIA AGX Xavier and the stage of the proposed framework.

separation. Under these conditions, we have developed an efficient and parallel system which decomposes the input mixture by combining a NMF approach and a clustering strategy together with a soft-masking filtering. In this sense, two different approaches have been proposed, an offline and an online variant.

The proposed system has been evaluated using a large database. Experimental results show that reliable results for the cardiorespiratory sounds separation task can be achieved in real time.

Acknowledgements This work was supported by the Regional Ministry of the Principality of Asturias under grant *FC-GRUPIN-IDI/2018/000226*, by the Ministry of Economy, Knowledge and University of the Government of the Junta de Andalucía under project *P18-RT-1994* and by the "Programa Operativo FEDER Andalucía 2014-2020" under project with reference *1257914*.

References

1. Alonso, P., Cortina, R., Rodríguez-Serrano, F.J., Vera-Candeas, P., Alonso-González, M., Ranilla, J.: Parallel online time warping for real-time audio-to-score alignment in multi-core systems. The Journal of Supercomputing **73**(1), 126–138 (2017). DOI 10.1007/s11227-016-1647-5. URL `http://link.springer.com/10.1007/s11227-016-1647-5`
2. Alonso, P., Vera-Candeas, P., Cortina, R., Ranilla, J.: An efficient musical accompaniment parallel system for mobile devices. Journal of Supercomputing **73**(1), 1–11 (2016). DOI 10.1007/s11227-016-1865-x
3. Ayari, F., Ksouri, M., Alouani, A.T.: Lung sound extraction from mixed lung and heart sounds FASTICA algorithm. In: 2012 16th IEEE Mediterranean Electrotechnical Conference, pp. 339–342. IEEE (2012). DOI 10.1109/MELCON.2012.6196444. URL `http://ieeexplore.ieee.org/document/6196444/`
4. Barabasa, C., Jafari, M., Plumbley, M.D.: A robust method for S1/S2 heart sounds detection without ecg reference based on music beat tracking. In: 2012 10th International Symposium on Electronics and Telecommunications, pp. 307–310. IEEE (2012). DOI 10.1109/ISETC.2012.6408110. URL `http://ieeexplore.ieee.org/document/6408110/`
5. Blackford, L.S., Demmel, J., Dongarra, J., Duff, I., Hammarling, S., Henry, G., Heroux, M., Kaufman, L., Lumsdaine, A., Petitet, A., Pozo, R., Remington, K., Whaley, R.C.:

An updated set of basic linear algebra subprograms (blas). ACM Transactions on Mathematical Software **28**, 135–151 (2001)

6. 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. Applied Acoustics **125**, 7–19 (2017). DOI 10.1016/j.apacoust.2017.04.005. URL `https://linkinghub.elsevier.com/retrieve/pii/S0003682X16304923`
7. Charbonneau, G., Racineux, J.L., Sudraud, M., Tuchais, E.: An accurate recording system and its use in breath sounds spectral analysis. Journal of Applied Physiology **55**(4), 1120–1127 (1983). DOI 10.1152/jappl.1983.55.4.1120. URL `https://www.physiology.org/doi/10.1152/jappl.1983.55.4.1120`
8. Charleston-Villalobos, S., Dominguez-Robert, L.F., Gonzalez-Camarena, R., Aljama-Corrales, A.T.: Heart Sounds Interference Cancellation in Lung Sounds. In: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1694–1697. IEEE (2006). DOI 10.1109/IEMBS.2006.259357. URL `http://ieeexplore.ieee.org/document/4462097/`
9. Chien, J.C., Huang, M.C., Lin, Y.D., Chong, F.c.: A Study of Heart Sound and Lung Sound Separation by Independent Component Analysis Technique. In: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5708–5711. IEEE (2006). DOI 10.1109/IEMBS.2006.260223. URL `http://ieeexplore.ieee.org/document/4463102/`
10. Dagum, L., Menon, R.: Openmp: an industry standard api for shared-memory programming. IEEE computational science and engineering **5**(1), 46–55 (1998)
11. Févotte, C., Bertin, N., Durrieu, J.L.: Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. Neural Computation **21**(3), 793–830 (2009). DOI 10.1162/neco.2008.04-08-771. URL `http://www.mitpressjournals.org/doi/10.1162/neco.2008.04-08-771`
12. Frigo, M., Johnson, S.G.: The design and implementation of fftw3. Proceedings of the IEEE **93**(2), 216–231 (2005)
13. Gavriely, N., Palti, Y., Alroy, G.: Spectral characteristics of normal breath sounds. Journal of Applied Physiology **50**(2), 307–314 (1981). DOI 10.1152/jappl.1981.50.2.307. URL `https://www.physiology.org/doi/10.1152/jappl.1981.50.2.307`
14. Grais, E.M., Erdogan, H.: Single channel speech music separation using nonnegative matrix factorization and spectral masks. In: 2011 17th International Conference on Digital Signal Processing (DSP), pp. 1–6. IEEE (2011). DOI 10.1109/ICDSP.2011.6004924. URL `http://ieeexplore.ieee.org/document/6004924/`
15. Hedayioglu, F.L., Jafari, M.G., Mattos, S.S., Plumbley, M.D., Coimbra, M.T.: Separating sources from sequentially acquired mixtures of heart signals. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 653–656. IEEE (2011). DOI 10.1109/ICASSP.2011.5946488. URL `http://ieeexplore.ieee.org/document/5946488/`
16. Hossain, I., Moussavi, Z.: An overview of heart-noise reduction of lung sound using wavelet transform based filter. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), pp. 458–461. IEEE (2003). DOI 10.1109/IEMBS.2003.1279719. URL `http://ieeexplore.ieee.org/document/1279719/`
17. Lin, C., Hasting, E.: Blind source separation of heart and lung sounds based on nonnegative matrix factorization. In: 2013 International Symposium on Intelligent Signal Processing and Communication Systems, pp. 731–736. IEEE (2013). DOI 10.1109/ISPACS.2013.6704646. URL `http://ieeexplore.ieee.org/document/6704646/`
18. Peeters, G.: A large set of audio features for sound description. Tech. rep. (2004). DOI 10.1234/12345678
19. Pourazad, M., Moussavi, Z., Farahmand, F., Ward, R.: Heart Sounds Separation From Lung Sounds Using Independent Component Analysis. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, pp. 2736–2739. IEEE (2005). DOI 10.1109/IEMBS.2005.1617037. URL `http://ieeexplore.ieee.org/document/1617037/`
20. Rudnitskii, A.G.: Using nonlocal means to separate cardiac and respiration sounds. Acoustical Physics **60**(6), 719–726 (2014). DOI 10.1134/S1063771014050121. URL `http://link.springer.com/10.1134/S1063771014050121`

21. Salazar, A.J., Alvarado, C., Lozano, F.E.: System of heart and lung sounds separation for store-and-forward telemedicine applications. Revista Facultad de Ingenieria (64), 175–181 (2012)
22. San Juan, P., Vidal, A., Garcia-Molla, V.: Updating/downdating the Non-Negative Matrix Factorization. Journal of Computational and Applied Mathematics **318**, 59–68 (2017). DOI 10.1016/j.cam.2016.11.048. URL https://www.sciencedirect.com/science/article/pii/S0377042716305908https://linkinghub.elsevier.com/retrieve/pii/S0377042716305908
23. Shah, G., Koch, P., Papadias, C.B.: On the Blind Recovery of Cardiac and Respiratory Sounds. IEEE Journal of Biomedical and Health Informatics **19**(1), 151–157 (2015). DOI 10.1109/JBHI.2014.2349156. URL http://ieeexplore.ieee.org/document/6879427/
24. Shah, G., Papadias, C.: Separation of cardiorespiratory sounds using time-frequency masking and sparsity. In: 2013 18th International Conference on Digital Signal Processing (DSP), pp. 1–6. IEEE (2013). DOI 10.1109/ICDSP.2013.6622792. URL http://ieeexplore.ieee.org/document/6622792/
25. Sibu, T., Nishi, S.: Detection and elimination of heart sound from lung sound based on wavelet multi resolution analysis technique and linear prediction. The International Journal Research Publication's **1**(10) (2012)
26. Tracey, B.H., Miller, E.L.: Nonlocal Means Denoising of ECG Signals. IEEE Transactions on Biomedical Engineering **59**(9), 2383–2386 (2012). DOI 10.1109/TBME.2012.2208964. URL http://ieeexplore.ieee.org/document/6242391/
27. Virtanen, T.: Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria. IEEE Transactions on Audio, Speech and Language Processing **15**(3), 1066–1074 (2007). DOI 10.1109/TASL.2006.885253. URL http://ieeexplore.ieee.org/document/4100700/