Blind Recovery of Cardiac and Respiratory Sounds Using Non-negative Matrix Factorization & Time-Frequency Masking

Ghafoor Shah and Constantinos B. Papadias

Abstract— **Auscultation is an effective noninvasive medical procedure for examining the cardiorespiratory system. However, the cardiac and respiratory acoustic sounds interfere in time as well as in spectral contents, which hampers the diagnostibility of the classical stethoscope. We propose a method for smart auscultation by blindly recovering the original cardiac and respiratory sounds from a single observation mixture. We decompose the spectrogram of the mixture into independent, non-redundant components, by employing non-negative matrix factorization (NMF). To group the decomposed components into original sources, a new unsupervised technique is proposed. Time-frequency masking is used to recover the original sources. This smart auscultation method is successfully applied to actual data collected from different subjects in different clinical settings. Our method demonstrates excellent results even in noisy clinical environments.**

I. INTRODUCTION

Acoustic analysis of the chest sound (which is a mixture of cardiac and respiratory sounds) provides important information in the diagnosis of cardiac and lung conditions. However, the cardiac and respiratory sounds overlap in terms of time-domain and spectral content, which compromises auscultation even in the noise-free clinical environment. Recovery of the original cardiac and respiratory sounds from their mixture can enhance the quality of auscultation. Separation of the cardiac and respiratory sound problem has been investigated as a blind source separation problem in [1]-[5], where two are more observation mixtures are used for the recovery of two signals. However, most of the modern stethoscopes used for chest sound auscultation, can provide only a single observation mixture of the cardiac and respiratory sounds. Conventional time-domain filtering alone, cannot completely separate the two sources from their single observation mixture, because of their overlap in spectral content especially below 200Hz. In [6], fifteen different adaptive methods developed for separating the cardiac sound from respiratory sound are reviewed and the filtering techniques are categorized as linear adaptive filters and filters employing time-frequency based methods. Adaptive filtering does not completely suppress the cardiac sound segments in respiratory sound because of the high non-stationarity of the cardiac sound which makes time alignment of the primary and reference signal difficult in real scenarios. Also the primary and reference input to an adaptive filter must be of the same type for better noise cancellation. It means in the adaptive filtering of the cardiac sound, an acoustic cardiac signal should be used as reference. However, most of the techniques reviewed in [6], use an ECG signal as reference to the cardiac sound. All the studies in [6] were based on the data acquired under ideal conditions, while the potential usefulness of any method rests on its ability to perform in real clinical settings. Recently, separation of the respiratory sound from the cardiac sound using wavelet transform based filtering has been proposed (see e.g. [7]). In wavelet transform based filtering the selection of the decomposition and threshold levels is quite challenging in real scenarios. It is clear from the above that the separation of the cardiac and respiratory sounds from a single observation mixture is a challenging task that needs further investigation.

Matrix decomposition techniques such as non-negative matrix factorization (NMF) [8], have been recently employed in the single channel blind source separation of musical data. The NMF is applied to the magnitude spectrogram in order to produce a low dimensional approximation of the original data, in the form of two non-negative matrices. One matrix having the spectral basis vectors and the second matrix containing time-variant gain information for each basis vector. Different versions of NMF been proposed during the last decade, are reviewed in [9]. Most of the existing techniques incorporate different constraints, reflecting the features in musical data such as temporal structure, harmonic structure etc.. However, the cardiac and respiratory sounds lack the features of the musical data. Moreover, the existing advanced versions of NMF are complex and computationally expensive. On the other hand, the basic version of NMF [8] is simple and computationally efficient. However, there are two main challenges in basic NMF-based blind source separation; 1) there is no systematic mechanism to determine the suitable number of basis vectors, and 2) how to classify and cluster the basis vectors to form the original sources. Different supervised and unsupervised clustering methods have been proposed (see e.g. [10]) for musical data separation which are not suitable in the blind source separation of the cardiac and respiratory sounds.

In this paper we propose a new method to separate the cardiac and respiratory sounds from a single observation mixture, based on NMF and time-frequency masking. We decompose the magnitude spectrogram of the observation mixture into various components using the basic NMF. To cluster the components into the original sources, we propose an unsupervised clustering technique.

G. Shah is with the Broadband Wireless and Sensor Networks Group (BWiSE), Athens Information Technology (AIT), GR 19002, Athens, Greece(e-mail: ghasha@ait.edu.gr). He is also with the Department of Electronic Systems, Aalborg University (AAU), 9220, Aalborg East, Denmark.

C. B. Papadias is with the Broadband Wireless and Sensor Networks Group (BWiSE), Athens Information Technology (AIT), GR 19002, Athens, Greece(e-mail: papadias@ait.edu.gr).

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

The remainder of the paper is organized as follows. Section II discusses the proposed method; Section III demonstrates experimental results while conclusions and future works are given in Section IV.

II. Proposed Method

In this Section, we describe a novel and computationally efficient method for separating the cardiac and respiratory sounds from a single observation mixture. Our method comprises of three different phases; 1) a *decomposition phase*, which is the decomposition of mixture into independent components based on NMF; 2) a *clustering phase*, where similar components are grouped to form the original sources, and 3) a *reconstruction phase*, where the original sources are recovered from the spectrogram of the mixture using time-frequency masking.

A. Mixing Model

Based on the fact that most of the modern stethoscopes provide a single simultaneous observation of the chest sound, we assume the following instantaneous mixing model for the cardiac and respiratory sound signals:

$$x[m] = \sum_i a_i s_i[m] + \eta[m], \tag{1}$$

where, $x[m]$ represents the observation mixture and $s_i[m]$, a_i represent the i^{th} source and its amplitude, respectively. $i \in \{c, r\}$, where c, r, represent the cardiac and respiratory sound signal domains, respectively and $\eta[m]$ represents white Gaussian noise.

B. Non-negative Matrix Factorization

Non-negative matrix factorization is a useful tool that is employed in a variety of signal processing applications. NMF gives parts-based decomposition and imposes the only constraint of non-negativity. Efficient algorithms for NMF computations have been developed in [11]. In NMF, given an $F \times T$ non-negative matrix V, we wish to approximate V by the factors W and H as

$$V \approx WH, \tag{2}$$

where, W is a $F \times K$ and H is a $K \times T$ non-negative matrices, and K is chosen to be smaller than both T & F. The objective of the NMF is finding a pair of W and H such that the reconstruction error is minimized. The following two cost functions are mostly used for minimizing reconstruction error. The first cost function which is the squared Euclidean distance between V and WH is defined as

$$D_{EUD} = ||V - WH||^2 = \sum_{tf} (V_{tf} - (WH)_{tf})^2, \tag{3}$$

where, t and f represent time index and frequency bin respectively. The second cost function which is the divergence between V and WH is given as

$$D_{KL} = \sum_{tf} (V_{tf} \log \frac{V_{tf}}{(WH)_{tf}} - V_{tf} + (WH)_{tf}). \tag{4}$$

The lower bound of both the measures (3) and (4) is zero and it is optimized if $V = WH$. The recursive updates which converge to local minima are given as

$$W \leftarrow W \bullet \frac{VH^T}{WHH^T}, \quad H \leftarrow H \bullet \frac{W^T V}{W^T WH}, \tag{5}$$

$$W \leftarrow W \bullet \frac{\frac{V}{WH} H^T}{1.H^T}, \quad H \leftarrow H \bullet \frac{W^T \frac{V}{WH}}{W^T.1}, \tag{6}$$

where, $D \bullet E$ denotes element-wise multiplication, $\frac{D}{E}$ denotes element-wise division and 1 is a matrix with all elements unity. The update rules (5) and (6) corresponds to cost functions defined in (3) and (4) respectively. (2) can be re-written as

$$V \approx \sum_{k=1}^{K} w_k h_k, \tag{7}$$

where, w_k represents the k^{th} column of W, h_k represents the k^{th} row of H i. e., $W = \{w_1, w_2, \ldots w_K\}$, $H = \{h_1, h_2, \ldots h_K\}^T$ and T denotes transpose.

C. Decomposition Phase

NMF was originally developed for image processing as a two-dimensional (2D) image can be regarded as a matrix. The time-domain signals which consist of positive as well as negative values are not suitable for NMF. However, NMF can be applied to the magnitude spectrogram of the corresponding signals. The spectrogram of a signal, can be calculated by dividing the time-domain signal into small frames using a suitable window function, and performing the discrete-time Fourier transform on each frame. The discrete-time short time Fourier transform (STFT) of a time-domain signal $s[m]$ is given as

$$S(n, \omega) = \sum_{m=-\infty}^{\infty} s[m] \psi[n - m] e^{-j\omega m}, \tag{8}$$

where, $\psi[n - m]$ is a suitable time window and (n, ω) represents a time-frequency index. Using (8), we can write (1) as

$$X(n, \omega) = \sum_i a_i S_i(n, \omega) + \eta(n, \omega). \tag{9}$$

$X(n, \omega)$ represents the spectrogram of the mixture of cardiac and respiratory sounds. The cardiac sound S_c is generally produced by the mechanical activities of the heart. Adults, normally produce two heart sounds, during a single heartbeat, s_1 and s_2. Two more heart sounds s_3 and s_4 also appear sometime during the heartbeat. Similarly, the respiratory sound S_r is also a combination of sounds produced by different parts of the lung during respiration. Therefore, we can rewrite (9) as

$$X(n, \omega) = AB(n, \omega) + \eta(n, \omega), \tag{10}$$

where, $A = \sum_k a_k$, and $B(n, \omega) = \sum_i S_i(n, \omega) = \sum_i \sum_k b_k$, where b_k represents the k^{th} component of source S_i.

As we know that the cardiac and respiratory sounds are composed of different component sounds, therefore, in order to separate these sources, we decompose their mixture into independent and non-redundant components based on NMF.

We apply the basic NMF to the magnitude of the spectrogram (10). The NMF decomposes $X = ||X(n,\omega)||$ (which is a $F \times T$ matrix, shown in Fig. 1) into W which is a $F \times K$, and H which is a $K \times T$ non-negative matrices. Here the idea is to define $K < T$ so that W can be compressed and reduced to its integral components such as W is a matrix having only a set of spectral basis vectors, and H is a matrix containing the weight of each basis vector at each time point. Fig. 1 shows the magnitude spectrogram of a real mixture (of cardiac and respiratory sounds). The relevance of W and H to X is also shown. With a simple

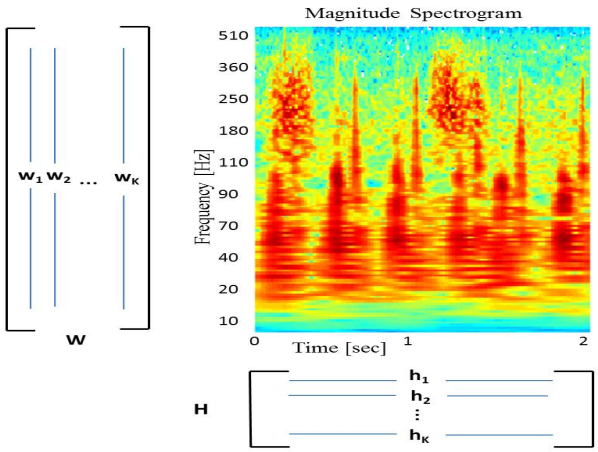

Fig. 1. The magnitude spectrogram X along with its decomposed factors W and H.

example, we demonstrate, how we decompose a mixture into different components. The mixture we use here is an actual observation mixture of the cardiac and respiratory sounds obtained in a clinical setting. The mixture X is decomposed into different components based on NMF, with $K = 8$. The time-domain representation of the decomposed components along with the original mixture are shown in Fig. 2. The following section discusses how to group the different components into original sources.

D. Clustering Phase

Grouping the decomposed components into original sources is the most challenging task in the approach of blind source recovery using basic NMF. Various clustering techniques have been proposed for musical data separation. The existing clustering techniques are not suitable to separate the cardiac and respiratory sounds because of their complex and non-stationary nature. Therefore, we propose a new unsupervised clustering technique for grouping the various independent components into the original sources. We exploit the fact that the spectral interference of the cardiac and respiratory sounds is minimal below 100 Hz, which we call the partial-overlapping region. This partial-overlapping region, which is evident from the spectrogram of the real mixture of the cardiac and respiratory sounds shown in Fig. 1, mostly consists of the cardiac sound. We use this partial-overlapping region as a reference in our unsupervised clustering technique.

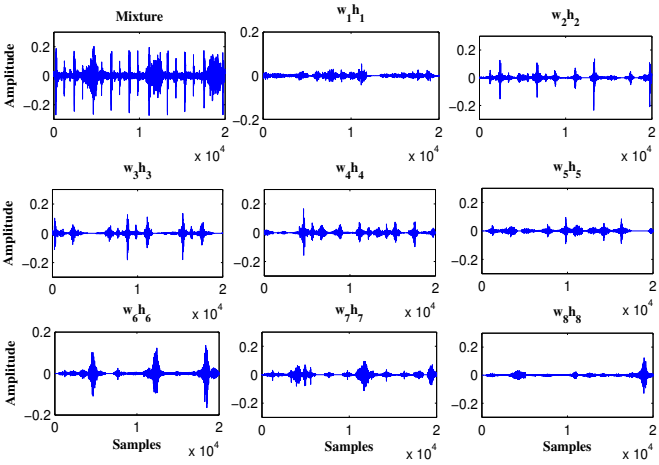

Fig. 2. Time-domain representation of the decomposed components of mixture (1) using NMF.

First we decompose the partial-overlapping region of the magnitude spectrogram X based on NMF, which generates the factors $\widetilde{H}$ and $\widetilde{W}$. Here $\widetilde{H}$ is an $F \times \widetilde{K}$ matrix, and $\widetilde{W}$ is a $\widetilde{K} \times T$ matrix. Then we try to find the similarities between the mixed and partial-overlapping regions. We propose two different clustering methods.

1) Clustering method 1: Here, we first define a general correlation formula as

$$c_{or} := C_{or}(f_k, g_k) := \frac{\sum f_k g_k}{\sqrt{\sum {f_k}^2}\sqrt{\sum {g_k}^2}} \tag{11}$$

where, f_k and g_k are the vectors having equal lengths, $c_{or} \in [0,1]$ and $c_{or} = 1$ shows maximum correlation where as, $c_{or} = 0$ dictates that f_k and g_k are uncorrelated. Now, on the basis of (11), we calculate two kinds of correlations, which we call *spectral correlation* and *temporal correlation*. The spectral correlation is defined as

$$c_f = C_{or}(w_k, \tilde{w}_k), \tag{12}$$

where, c_f is the correlation between the basis vectors of the mixed and partial-overlapping regions and $c_f \in [0,1]$. Similarly, we also define the temporal correlation as

$$c_t = C_{or}(h_k, \tilde{h}_k). \tag{13}$$

Here, c_t is the correlation between the weight vectors of the mixed and partial-overlapping regions and $c_t \in [0,1]$.

Having defined the similarity criterion, we propose the following algorithm to cluster the similar components into original sources. Our clustering algorithm is divided into three steps:

Step 1) In this step, we initialize the different parameters of the algorithm as follows

$$\begin{cases} c \leftarrow c_f \times c_t \\ \alpha \leftarrow max(c) - \frac{max(c)}{\gamma} \\ \beta \leftarrow min(c) + \frac{min(c)}{\gamma} \end{cases} \tag{14}$$

where, the α and β set thresholds for the different groups,

and $\{\alpha, \beta\} \in [0, 1]$. The parameter $\gamma \geq 1$ and can be found heuristically.

Step 2) In this step we group the different components into the following three groups:

$$G_c = \{W_1, H_1\}, G_r = \{W_2, H_2\}, G_m = \{W_3, H_3\}, \quad (15)$$

where G_c, G_r and G_m represent the components belonging to cardiac, respiratory and mixed sound respectively, and

$$\begin{cases} \{H_1, W_1\} \subseteq \{H, W\}\, s.t. \quad c \geq \alpha \\ \{H_2, W_2\} \subseteq \{H, W\}\, s.t. \quad c \leq \beta \\ \{H_3, W_3\} \subseteq \{H, W\}\, s.t. \quad \beta < c < \alpha, \end{cases} \quad (16)$$

W_l and H_l represent the columns and rows of W and H respectively, and $l \in \{1, 2, 3\}$. It should be noted that $\alpha > \beta$.

Step 3) This is the learning step of the algorithm. Provided that G_m is not empty, we update the different parameters of the algorithm as follows

$$\begin{cases} c \leftarrow c' \\ \alpha \leftarrow max(c) - \frac{max(c)}{\gamma} \\ \beta \leftarrow min(c) + \frac{min(c)}{\gamma} \\ W \leftarrow W_3 \\ H \leftarrow H_3. \end{cases} \quad (17)$$

Here, c' is the correlation between the basis vectors of G_c and G_m, defined as

$$c' = C_{or}(w_{gc}, w_{gm}), \quad (18)$$

where, $w_{gc} \in \{W_1\}$, $w_{gm} \in \{W_3\}$ and $c' \in [0, 1]$. $Step$ 2 and $Step$ 3 are repeated consecutively, until the grouping is completed.

2) Clustering Method 2: In this method, we try to find the correlation between the decomposed components of the mixture and partial-overlapping region. The correlation function is defined as

$$C_R = \frac{\sum\limits_{n,\omega} Y_i(n,\omega) Z(n,\omega)}{\sqrt{\sum\limits_{n,\omega} Y_i^2(n,\omega)} \sqrt{\sum\limits_{n,\omega} Z^2(n,\omega)}}, \quad (19)$$

where $Y_i(n, \omega) = ||w_i h_i||$, $Z(n, \omega) = ||\widetilde{W}\widetilde{H}||$ and $C_R \in [0, 1]$. The clustering algorithm for this method is the same as discussed in Section II-D.1.

Annotating the different component, we can approximate the magnitude of the original sources as

$$X_i \approx \sum_{p,q} w_p h_q, \quad (20)$$

where w_p and h_q represent the p^{th} basis and q^{th} weight corresponding to the i^{th} source. We discuss the reconstruction of these sounds in the following section.

E. Reconstruction Phase

Once the magnitude spectrogram is approximated into original sources, the corresponding phases can also be approximated using the original spectrogram. An alternative approach is to generate a time-frequency mask for each source and apply the corresponding mask to the original spectrogram, to recover the original sources. We construct a time-frequency mask as

$$M_i = \begin{cases} 1 \ \forall X_i > X_j, \ j \in \{r, c\}, j \neq i \\ 0, \ otherwise. \end{cases} \quad (21)$$

The idea of time-frequency masking is based on the assumption that cardiac and respiratory sound signals are sparse [3], which means that over a small time-frequency region only one source dominates. The time-frequency mask (21) is applied to the spectrogram of the mixture (9) to recover the original sources as

$$S_i(n, \omega) = M_i \bullet X(n, \omega). \quad (22)$$

The inverse short-time Fourier transform (ISTFT) is used to convert the original sources back into the time-domain.The latter approach provides better results as compared to the former.

F. Summary

The three phases of the proposed method are summarized as:

- Decomposition Phase
  1) Generate the spectrogram of mixture using STFT
  2) Decompose the magnitude spectrogram of the mixture into H and W using NMF
- Clustering Phase
  1) Decompose the partial-overlapping region of the spectrogram into $\tilde{H}$ and $\tilde{W}$ using NMF
  2) Find the spectral and temporal similarities
  3) Initialize the algorithm parameters
  4) Perform grouping
  5) Update the algorithm parameters and go to step (4) until grouping is completed
- Reconstruction Phase
  1) Generate a time-frequency mask corresponding to each source
  2) Recover the original sources by applying the time-frequency mask to the spectrogram of the mixture
  3) Take the ISTFT to convert the recovered sources back into time-domain

III. Results and Discussions

A. Experimental Setup

We performed experiments on different sets of clinical data. The clinical data is taken from an online data base [12]. To measure the performance of the method, we define as performance metric the signal-to-interference ratio (SIR):

$$SIR_i = \frac{\sum\limits_{n,\omega} ||M_i(n,\omega) S_i(n,\omega)||^2}{\sum\limits_{n,\omega} ||M_i(n,\omega) I_i(n,\omega)||^2}, \quad (23)$$

where, $I_i(n, \omega)$ is the interference with the i^{th} source. In our experiments, for STFT representation, *Hanning* window of length 1024 samples was used. The parameter K was varied from 2 to 20, where as $\widetilde{K}$ was varied from 1 to 10. $\gamma = 4$ was used for various experiments. The maximum number of iterations used for NMF was 130.

TABLE I

PERFORMANCE OF THE PROPOSED METHOD

Sample mixture	Recovered sources	SIR (dB)
140_1306519735121_A	s_c	24.76
	s_r	18.21
150_1306776340746_B	s_c	22.13
	s_r	14.14
101_1305030823364_B	s_c	21.91
	s_r	17.64
104_1305032492469_A	s_c	15.72
	s_r	17.13

B. Clinical Data

The data samples taken from [12] were obtained from four different subjects, in noisy clinical settings, using an electronics stethoscope, with a data sampling frequency of 4KHz. A data sample is a real mixture of cardiac and respiratory sounds. Fig. 3 shows the time-domain plots of our experiment, where plot (a) shows the mixture, plots (b) and (c) show the recovered respiratory s_r and cardiac s_c sounds respectively. Plots (d) and (f) compare the recovered and original respiratory sounds in *linear* and *log* scales respectively, where as, plots (e) and (g) compare the recovered and original cardiac sounds in *linear* and *log* scales respectively. Note that the comparison plots are zoomed over a small segment of original and recovered signals for better illustration. The performance metrics of the experiments are given in Table I.

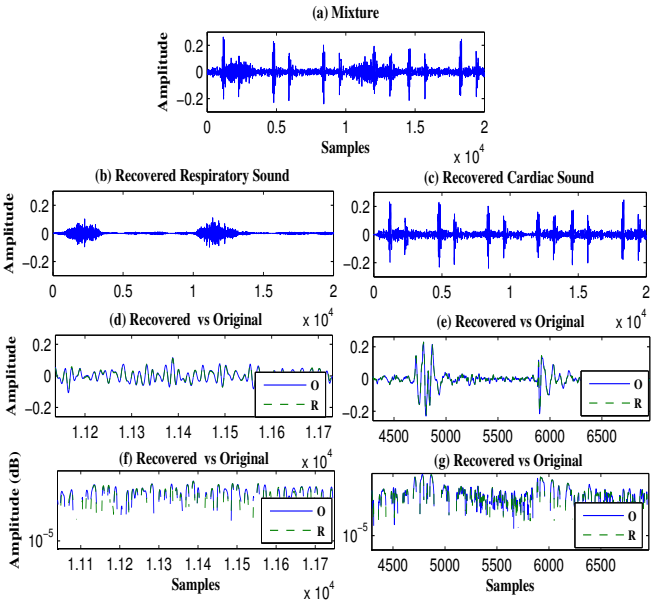

Fig. 3. Recovery of the cardiac and respiratory sounds from a clinical mixture.

IV. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

A novel and computationally efficient method is proposed to separate the cardiac and respiratory sounds from a single observation mixture. The basic NMF is used to decompose the magnitude spectrogram of the mixture into various components. The partial-overlapping region of the spectrogram is used as a reference in the developed unsupervised clustering technique. The method is applied to the actual data recorded in echoic and noisy clinical settings. Excellent recovery of the cardiac and respiratory sounds is achieved with a single observed mixture, thus outperforming, to our best knowledge, other techniques in the literature.

B. Future Works

The main goal of this research is to enhance the quality of auscultation. In this paper, we separated the cardiac and respiratory sounds in normal conditions. While in case of cardiac and respiratory conditions, various pathological sounds are also produced. The next step is to perform analysis of the separated cardiac and respiratory sounds. This includes identification and classification of normal and pathological signals.

ACKNOWLEDGMENTS

The authors would like to thank ELPEN (www.elpen.gr), for partially sponsoring this research work. The opinions expressed in this work do not necessarily reflect those of the sponsor.

REFERENCES

[1] B. Makkiabadi et al., "A new time domain convolutive bss of heart and lung sounds," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 605–608.

[2] K. Hashiodani et al., "Robustly separating sound components in human body based on 2-ch ica and em algorithm with dirichlet distribution," in *Proc. of IEEE-EMBS Int. Conf. on Biomedical and Health Informatics (BHI)*, 2012, pp. 56–59.

[3] G. Shah and C. Papadias., "Separation of cardiorespiratory sounds using time-frequency masking and sparsity," in *Proc. of 18th Int. Conf. on Digital Signal Processing (DSP), Santorini, Greece*, July 2013.

[4] F. Ayari et al., "Lung sound extraction from mixed lung and heart sounds fastica algorithm," in *Proc. of IEEE Mediterranean Electrotechnical Conf. (MELECON), Tunisia*, Mar. 2012, pp. 339– 342.

[5] F.L. Hedayioglu et al., "Separating sources from sequentially acquired mixtures of heart signals," in *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Portugal*, May 2011, pp. 653– 656.

[6] J. Gnitecki and Z. Moussavi, "Separating heart sounds from lung sounds - accurate diagnosis of respiratory disease depends on understanding noises," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 26, no. 1, pp. 20–29, Jan 2007.

[7] A. Misal and G. R. Sinha, "Separation of lung sound from pcg signals using wavelet transform," *Journal of Basic and Applied Physics*, vol. 1, pp. 57–61, Aug 2012.

[8] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[9] Y. Wang et al., "Nonnegative matrix factorization: A comprehensive review," *IEEE Trans. on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, June 2013.

[10] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 15, pp. 1066–1074, 2007.

[11] D. Lee and H. S. Seung, "Algorithms for non-negatvie matrix factorization," *Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

[12] P. Bentley et al., *The PASCAL Classifying Heart Sounds Challenge (CHSC2011) Results*, Std., 2011. [Online]. Available: http://www.peterjbentley.com/heartchallenge/index.html