SINGLE CHANNEL SOURCE SEPARATION FOR CONVOLUTIVE MIXTURES WITH APPLICATION TO RESPIRATORY SOUNDS

A. K. Kattepur
INRIA, Rennes, France

F. Jin
School of Electrical& Electronic Engineering, Nanyang Technological University, Singapore

F. Sattar
Faculty of Comp. Science and Infor. Tech, University of Malaya, Malaysia

Keywords: Single channel, Blind Source Separation (BSS), Respiratory Sound (RS), Non-negative Matrix Factorization (NMF).

Abstract: In this paper, we attempt to extend single channel source separation techniques to the separation of respiratory sound (RS) and heart sounds (HS). This single channel recording is analyzed and shown to be a convolutive mixture model. After analyzing the reasons for failure of commonly used blind source separation algorithms, we evaluate the efficacy of non-negative matrix factorization (NMF) techniques for this application. Analysis on simulated single channel convolutive mixtures at various sensor positions has been performed. It indicates an average signal to interference ratio (SIR) improvement of greater than 10 dB for the optimal sensor locations. The corresponding range of received power has been also studied for reliable separation of RS and HS. Finally, the proposed model and the NMF separation performance are demostrated to work well on real RS recordings.

1 INTRODUCTION

Single channel source separation is a problem with considerable interests (Jang and Lee, 2003) where the separation of multiple sources is performed from a single channel recording. Compared to traditional blind source separation (BSS) models (Choi et al., 2005), this provides less diversity than the critically determined case. When the convolutive mixtures are involved, this problem becomes much more complex.

In the case of convolutive mixtures, the observed signals are assumed to be combinations of delayed and filtered versions of the independent components. The task is to estimate the original sources without resorting to *a priori* information about the mixing system. In case of BSS, the prior assumptions of independence and non-Gaussianity of the original signals are used for the separation process (Choi et al., 2005).

In certain applications, only single sensor is preferred due to the ease of data acquisition, and the sensor can only be positioned in one of the optimal locations. For example, single channel recordings of RS is vital for computerized pulmonary auscultation (Cort*é*s et al., 2005). Similarly, the single channel HS recording play an important role in HS analysis for the diagnosis of the heart valve dysfunction and degeneration (Zhao and Wang, 2007). The optimal microphone pick-up location depends on the application and it is either on the suprasternal notch (tracheal sounds), or over the left or right posterior base of the lungs (lung sounds) (Sovijärvi et al., 2000). However, HS and RS show non-stationary behavior and overlapping of frequency contents at all sensor locations (Charleston-Villalobos et al., 2006). Heart beating produces an intrusive quasi-periodic interference that masks the clinical auscultative interpretation of respiratory sound. Therefore, it is crucial to separate both HS and RS signals effectively for accurate diagnosis.

Hence, we propose a convolutive mixing model

K. Kattepur A., Jin F. and Sattar F. (2010).
SINGLE CHANNEL SOURCE SEPARATION FOR CONVOLUTIVE MIXTURESWITH APPLICATION TO RESPIRATORY SOUNDS.
In *Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing*, pages 220-224
DOI: 10.5220/0002712402200224

for the mixture of HS and RS. Simulated mixtures using different mixing indices have been then used to verify the presented model. Finally, we compare the techniques for separating convolutive single channel mixtures based on real recorded sounds captured over the chest and suprasternal notch for the extraction of both clean HS and RS for clinical use.

2 FAILURE ANALYSIS OF CONVENTIONAL BSS ALGORITHMS

Algorithms typically used for BSS like JADE (Cardoso and Souloumiac, 1993), FastICA (Bingham et al., 2000), ACMA (Van der Veen and Paulraj, 1996) and the time-frequency techniques like LI-TIFROM (Abrard and Deville, 2005), have many applications in multi-channel source separation. These include the separation of communication, speech and audio signals.

However, when applied to single channel convolutive mixtures, the above mentioned algorithms fail due to various reasons as shown in Table 1. Therefore, more specific BSS algorithms are required to separate the single channel convolutive mixture of RS and HS. The non-negative matrix factorization (NMF) algorithm satisfying these criteria, thus becomes a suitable choice for this particular application.

Table 1: Failure Analysis of Common BSS algorithms.

Algorithm	Failure Analysis
JADE	Unable to handle under-determined mixtures. Unable to handle convolutive mixtures.
FastICA	Unable to handle single source mixtures. Deteriorated performance for convolutive mixtures.
ACMA	Unable to handle under-determined mixtures. Requires additional cons tant modulus criterion. Deteriorated performance for audio & speech mixtures.
LI-TIFROM	Unable to handle convolutive mix tures. Requires strict sparsity criterion.

3 THE PROPOSED MODEL FOR RESPIRATORY SOUNDS

Respiratory sounds being heard over the large airways/chest are primarily related to vibrations of the upper airway walls/chest and the turbulent airflow, while heart sounds occur mainly due to the valvular activity of the heart. With an approximation, the hypothetical sound sources of HS and RS can be considered to have mutually uncorrelated point sources (Kompis et al., 1998). RS can be acoustically characterized by broad spectrum noise and the presence of small time delay is related to the distance between sound source and microphone (typically 0.03 ms) (Gavriely, 1999). The observed noisy RS signal is considered to be a continuous RS signal interfered by a discontinuous HS signal.

A reliable separation of the signals requires taking into account the structure of the mixing process. In a real-life application, however, this process is unknown, but some assumptions may be made about the source statistics. In instantaneous mixing, the source signals are assumed to arrive at the sensors at the same time. This has been considered for separation of narrowband signals with sampling frequency within few hundred Hertz.

However, in real RS recordings, the RS and HS signals arrive at the sensor through multiple paths and therefore with different time lags. Furthermore, due to the broadband nature of the RS, convolutive mixing is suggested in this paper to model the real RS recordings where observations can be considered as the combinations of the unknown filtered versions of the source signals. Under the assumption of anechoic recording condition, the mixing process can be formulated as:

$$x(k) = \sum_{j=1}^{N} w_j b_j(k - l_j) + v(k) \tag{1}$$

where $b_j(k)$, $x(k)$ and $v(k)$ denote respectively the j^{th} source signal, the observed signal and the noise captured by the sensor at time instance k. The attenuation w_j and the delay l_j of the j^{th} source to the sensor would be determined by the physical position of the source relative to the sensor.

4 NON-NEGATIVE MATRIX FACTORIZATION (NMF)

The non-negative matrix factorization technique introduced by (Lee and Seung, 1999), is able to produce useful representations of real world data and can be applied to the problem of single channel source separation. The non-negative constraints usually required for these class of algorithms are relaxed by making use of standard ICA algorithms to zero-mean the observed data. However, particular emphasis should be given to the independence and sparsity of the observed data.

Based on the observed single channel data **x**, the NMF decomposes it into two basis matrices **A** and **S**. This results in reduced representation of the original data where each feature is a linear combination of the original attribute set. The NMF has low computational complexity and unlike time-frequency techniques, it is able to deal with both dense and sparse data sets.

The NMF algorithm may be described in the following steps:

1. Initialize the elements of **A** and **S** to random non-negative values. Normalize each column of **A** to unit 2-norm.
2. Update the matrix **A** by either least squares or Kullback-Leibler Divergence(KLD) as shown:

$$\mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\mathbf{x}\mathbf{S}^T}{\mathbf{A}\mathbf{S}\mathbf{S}^T} \tag{2}$$

$$\mathbf{A} \leftarrow \mathbf{A} \cdot \frac{\frac{\mathbf{x}}{\mathbf{AS}}\mathbf{S}^T}{1 \cdot \mathbf{S}} \tag{3}$$

where '$\cdot$' is the element-wise multiplication operator and '$-$' is the element-wise division operator. **A** values below an assigned threshold ε are approximated to be zero. Normalize each column of **A** to unit norm.

3. Update matrix **S** similarly as in step (2).

$$\mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T\mathbf{x}}{\mathbf{A}^T\mathbf{A}\mathbf{S}} \tag{4}$$

$$\mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T\frac{\mathbf{x}}{\mathbf{AS}}}{\mathbf{A} \cdot 1} \tag{5}$$

4. Iterate steps (2) and (3) till convergence is achieved.

The technique proposed by (Schmidt and Mørup, 2006) is based on 2D deconvolution and non-negative matrix factorization (NMF). In order to successfully separate convolutive mixtures, the NMF model is extended to the 2-dimensional case incorporating the time τ and pitch ϕ of the signal.

$$\mathbf{x} = \sum_{\tau}\sum_{\phi} \overset{\downarrow\phi}{\mathbf{A}^{\tau}} \overset{\rightarrow\tau}{\mathbf{S}^{\phi}} \tag{6}$$

where $\downarrow \phi$ represents the downward shift operator which moves each element of matrix ϕ rows down and $\rightarrow \tau$ denotes the right shift operator which moves each element in the matrix τ columns to the right. The least squares and KLD approach for updating **A** and **S** are then applied to separate the convolutive mixtures.

5 ANALYSIS AND RESULTS

5.1 Analysis on Simulated Data

5.1.1 Model Verification

In order to test the proposed convolutive mixing model, single channel source separation was performed on a mixture of tracheal/lung sounds and heart sounds. Clean tracheal sound, lung sound, and heart sound recordings from (Lehrer, 2002)(Wilkins et al., 2004) are used as the source signals. The instantaneous and convolutive mixing process was performed with specification from (MIT, 1999). Separation was then performed using the LI-TIFROM and NMF techniques. While LI-TIFROM can separate only instantaneously mixed sources, NMF can separate both instantaneous and convolutive mixtures.

As seen from Fig. 1, the separation performance of both algorithms is good in the case of instantaneous mixtures. However, the LI-TIFROM technique is poor in the case of convolutive mixtures. This result, when extended to separation of respiratory and heart sounds, provides interesting insights into the modelling process. Since the LI-TIFROM technique cannot separate even sparse mixtures of the recorded signals, the mixing model must be convolutive. Even though there are single source zones in the time-frequency plane implying sparsity criterion, the LI-TIFROM technique is unable to separate the real RS recordings. So, the real recorded RS mixtures can be modelled as a convolutive BSS problem.

5.1.2 Separation Performance

The separation performance of NMF which was convolutively mixed based on the specification in (MIT, 1999) was tested based on the signal to interference ratio (SIR) improvement. This is given by:

$$SIR = 10\log_{10}\frac{\|\mathbf{s}_{target}\|^2}{\|e_{interf}\|^2} \tag{7}$$

where $\mathbf{s}_{target}$ is the target signal and e_{interf} is an allowed deformation of the sources which accounts for the interferences of the unwanted sources.

In order to test the separation performance of the NMF algorithm, the scenario presented in Fig. 2 was used. For each symmetric location of the sensor in the (x, y) plane, the received power of the single channel mixture was captured. This was then separated using the NMF algorithm and the average SIR improvement was measured. As shown in Fig. 2, two cases are analyzed including placing the sensor away from both the sources (Case A) and in between the sources

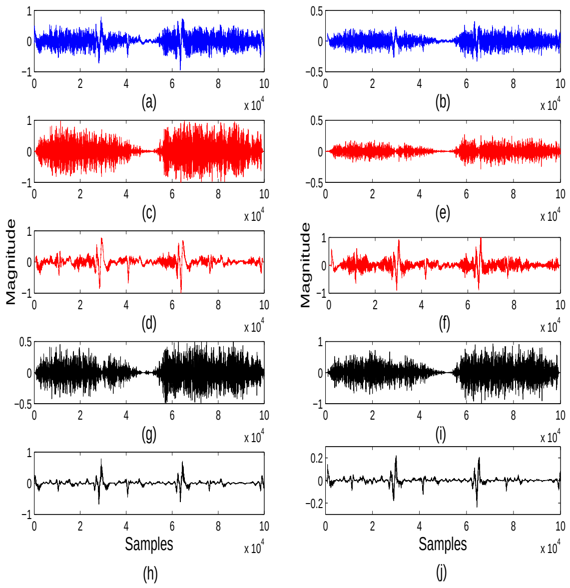

Figure 1: (a) Single-channel instantaneous mixture; (b) Single-channel convolutive mixture; LI-TIFROM separated signals for (c),(d) instantaneous mixture and (e),(f)convolutive mixture; NMF separated signals for (g),(h) instantaneous mixture and (i),(j) convolutive mixture.

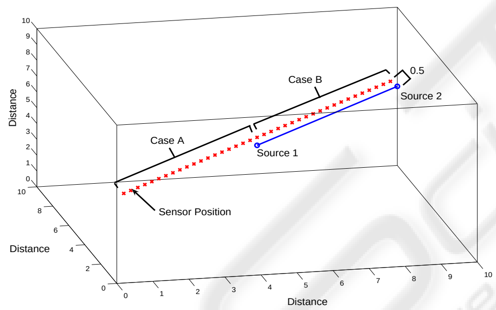

Figure 2: Scenario used for modelling single channel source separation.

(Case B). The tracheal and lung sounds have been analyzed as separate sources with interference from heart sounds in each scenario. As seen in the Figs. 3 and 4, the SIR improvements are superior when the sensor position is far away from both the sources. Similarly, the SIR improvements are superior midway between both the sources. When compared to the received power in each case, a power level of less than -5 dB indicates optimal separation performance. The corresponding received power can be used as a look-up graph for the selection of optimum sensor location during the actual recording of RS.

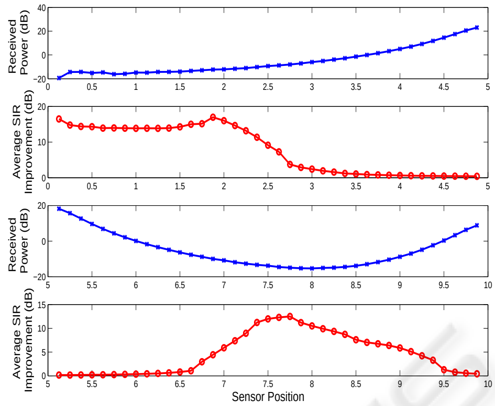

Figure 3: Received power and SIR improvement for a mixture of tracheal and heart sounds. The top two figures refer to sensor positions in case A, while the bottom two figures refer to case B.

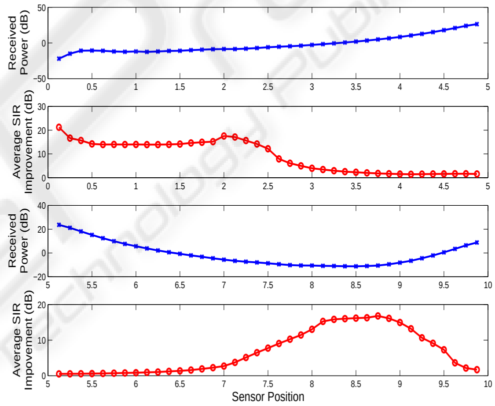

Figure 4: Received power and SIR improvement for a mixture of lung and heart sounds. The top two figures refer to sensor positions in case A, while the bottom two figures refer to case B.

5.2 Analysis on Real RS Recordings

5.2.1 Data Acquisition

Real RS recordings were done in anechoic room with the subjects in sitting position. Single electret condenser microphone (ECM-77B, Sony Inc., Japan) was inserted into a hemispherical rubber chamber of 2 cm in diameter, and was placed over suprasternal notch. The recording environment and equipments were chosen based on the standard given by (Sovijärvi et al., 2000). The choice of microphone together with the recording condition, the environmental noises were suppressed to the largest extend. Recording software WAVEPAD (V3.05, NCH Swift Sound Software) was used and the respiratory sound recordings have been saved as mono-channel '*.wav' files with sampling frequency F_s=11.025 kHz. Test subjects were asked to breathe normally with no targeted flow. The characteristics due to sex, age, weight were not taken into consideration.

5.2.2 Evaluation

To evaluate the effectiveness of the NMF based technique, the separation performance are tested on single channel separation of heart and breath sounds. As shown in Fig. 5, the separation performance is quite good and is shown to pass the subjective test for separation. This proves that the proposed model and associated parameters are consistent with those required for separating the real recorded data.

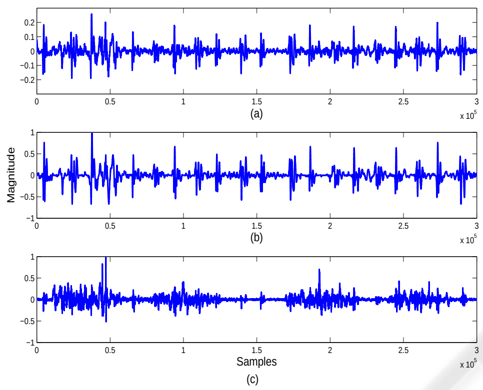

Figure 5: Single channel source separation using NMF. (a) Observed signal; (b) Separated HS; (c) Separated RS.

6 CONCLUSIONS

Non-negative matrix factorization techniques are shown to perform well in case of single channel source separation. The convolutive mixing model for respiratory sounds has been verified based on the separation performance. The NMF technique, when used on respiratory sounds, provides an SIR improvement of over 10 dB for optimal sensor positions. This, on the other hand, suggests an optimal sensor position for sound capturing. Due to the good separation performance, this has potential medical applications for accurate detection of pulmonary and heart diseases based on the separated RS and HS respectively.

REFERENCES

Abrard, F. and Deville, Y. (2005). A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. *Signal Processing*, 85(7):1389–1403.

Bingham, E., Hyvrinen, A., and rinen, A. H. (2000). A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. of Neural Systems*, 10:1–8.

Cardoso, J. F. and Souloumiac, A. (1993). Blind beamforming for non Gaussian signals. *IEE Proceedings*, 140(6):362–370.

Charleston-Villalobos, S., Aljama-Corrales, A. T., and Gonzalez-Camarena, R. (2006). Analysis of simulated heart sounds by intrinsic mode functions. *Proc. of 28th IEEE EMBS Conf.*, pages 2848–2851.

Choi, S., Cichocki, A., Park, H. M., and Lee, S. Y. (2005). Blind source separation and independent component analysis: A review. *Neural Information Processing - Letters and Reviews*, 6(1):1–57.

Cortés, S., Jané, R., Fiz, J. A., and Morera, J. (2005). Monitoring of wheeze duration during spontaneous respiration in asthmatic patients. *Proc. 27th IEEE EMBS Conf.*

Gavriely, N. (1999). Automatic detection and analysis of breath sounds. *Eur. Patent*, (EP 0 951 867 A2).

Jang, G.-J. and Lee, T.-W. (2003). A maximum likelihood approach to single-channel source separation. *J. Machine Learning Res.*, 4:1365–1392.

Kompis, M., Pasterkamp, H., Motai, Y., and Wodicka, G. R. (1998). Spatial representation of thoracic sounds. *Proc. 20th IEEE EMBS Conf.*, pages 1661–1664.

Lee, D. and Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lehrer, S. (2002). *Understanding Lung Sounds*. Philadelphia, PA: Saunders, Audio CD.

MIT (1999). ICA Synthetic Benchmarks.

Schmidt, M. N. and Mørup, M. (2006). Nonnegative matrix factor 2-D deconvolution for blind single channel source separation. In *Int. Conf. on ICA and Signal Separation*.

Sovijärvi, A. R. A., Vanderschoot, J., and Eavis, J. R. (2000). Standardization of computerized respiratory sound analysis. *European Respiratory Review*, 10(77):585–649.

Van der Veen, A. J. and Paulraj, A. (1996). An analytical constant modulus algorithm. *IEEE Trans. Signal Processing*, 44(5):1136–1155.

Wilkins, R. L., Hodgkin, J. E., and Lopez, B. (2004). *Fundamentals of Lung and Heart Sounds*. Mosby, Audio CD.

Zhao, Z. D. and Wang, Y. (2007). Analysis of diastolic murmurs for coronary artery diseasebased on hilbert huang transform. *Proc. of Int. Conf. Machine Learning and Cybernetics*, 6.