UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Faculty of Electronic and Computer Engineering

BLIND SOURCE SEPARATION USING TWO-DIMENSIONAL NONNEGATIVE MATRIX FACTORIZATION IN BIOMEDICAL FIELD

Toh Cheng Chuan

Master of Science in Electronic Engineering

2018

BLIND SOURCE SEPARATION USING TWO-DIMENSIONAL NONNEGATIVE MATRIX FACTORIZATION IN BIOMEDICAL FIELD

TOH CHENG CHUAN

A thesis submitted
in fulfillment of the requirements for the degree of
Master of Science in Electronic Engineering

Faculty of Electronic and Computer Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2018

DECLARATION

I declare that this thesis entitled "Blind Source Separation using Two-Dimensional Nonnegative Matrix Factorization in Biomedical Field" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in the candidature of any other degree.

Signature	:	....
Name	:	Toh Cheng Chuan
Date	:	....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Science in Electronic Engineering.

Signature	:	....
Supervisor Name	:	Associate Professor Dr. Abdul Majid bin Darsono
Date	:	....

DEDICATION

I dedicate this thesis to my loving family, my lover, and my beloved friends for their constant support and unconditional love.

ABSTRACT

Blind Source Separation (BSS) refers to the statistical technique of separating a mixture of underlying source signals. BSS denotes as a phenomena and separation on mixed heart-lung sound is one of its example. The challenge of this research is to separate the separate lung sound and heart sound from mixed heart-lung sound. A clear lung sound for diagnosis purpose able to be obtained after separating the mixed heart-lung sound. In biomedical field, lung information is precious due to it has been provided for respiratory diagnosis. However, the interference of heart sound towards lung sound will generate ambiguity and it will lead to drop down the accuracy of diagnosis. Thus, a clean lung sound is needed to increases the accuracy of diagnosis. One of the ways for non-invasive respiratory diagnosis for obtaining lung information is through extracting lung sound from mixed heart-lung sound by using Two-Dimensional Nonnegative Matrix Factorization (NMF2D) algorithm. This method is based on cocktail party effect in which it refers to human brain able to selectively listen to target among a cacophony of conversations and background noise and this considered as a difficult task to machine. Therefore, duplication on cocktail party effect into machine is used to separate the mixed heart-lung sound. This research presents a novel approach NMF2D algorithm in which a suitable model for signal mixture that accommodated the reverberations and nonlinearity of the signals. The objectives of this research are focusing on investigating the useful signal analysis algorithms, defining a new technique of signal separability, designing and developing novel methods for BSS. In order to process estimation results, cost function such as β-divergence and α-divergence is integrated with NMF2D. Provisions of experiment are convolutive mixed signal is sampled and real recording using under single channel, Time-Frequency (TF) domain is computed by using Short Time Fourier Transform (STFT) respectively. Performance evaluation is done in term of Signal-to-Distortion Ratio (SDR). Theoretically, β and α is parameters that used to vary the NMF2D algorithm in order to yield high SDR value. Experimentally, for the simulation results, the highest SDR value for β-divergence NMF2D is SDR = 16.69dB at β = 0.8 and n = 100. For α-divergence NMF2D, the highest SDR value is SDR = 17.85dB at α = 1.5 and n = 100. Additional of sparseness constraints toward β-divergence NMF2D and α-divergence NMF2D lead to even higher SDR value. There are SDR = 17.06dB for sparse β-divergence NMF2D at λ = 2.5 and SDR = 17.99dB for sparse α-divergence NMF2D at λ = 5. This represents sparseness constraints yield to decrease ambiguity and provide uniqueness to the model. In comparison in between β-divergence, α-divergence, sparse β-divergence and sparse α-divergence NMF2D, it found that SDR value of sparse α-divergence NMF2D is the best decomposition method among all divergences. This can be concluded that sparse α-divergence NMF2D is more applicable in separating real data recording.

ABSTRAK

BSS merujuk kepada teknik statistik memisahkan campuran isyarat sumber asas. BSS merupakan satu fenomena dan salah satu contohnya ialah pemisahan campuran antara bunyi dari jantung dan paru-paru. Cabaran kajian ini ialah untuk memisahkan bunyi paru-paru dan jantung dari bunyi jantung-lung yang bercampur. Bunyi paru-paru yang jelas untuk tujuan diagnosis boleh diperolehi selepas memisahkan bunyi jantung-lung yang bercampur. Dalam bidang Bioperubatan, maklumat dari paru-paru adalah berharga kerana ia telah disediakan untuk diagnosis pernafasan. Walau bagaimanapun, gangguan bunyi jantung terhadap bunyi paru-paru akan menghasilkan kekaburan dan ia akan menurunkan ketepatan diagnosis. Oleh itu, bunyi paru-paru yang bersih diperlukan untuk meningkatkan ketepatan diagnosis. Salah satu cara untuk diagnosis pernafasan yang tidak invasif untuk mendapatkan maklumat dari paru-paru adalah dengan mengekstrak bunyi paru-paru daripada campuran bunyi jantung dan paru-paru dengan menggunakan algoritma NMF2D. Kaedah ini adalah berdasarkan kesan parti koktel di mana ia merujuk kepada otak manusia mampu mampu mendengar sasaran antara cacophony perbualan dan bunyi bising dan ini dianggap sebagai satu tugas yang sukar untuk mesin. Oleh yang demikian, pertindihan pada kesan parti koktel kepada mesin untuk memisahkan bunyi jantung-lung yang bercampur. Oleh itu, kajian ini membentangkan algoritma baru pendekatan NMF2D di mana model yang lebih sesuai dibina yang mengambil kira gema dan ketaklelurusan isyarat untuk meniru sistem pendengaran manusia yang mampu memberikan pemisahan. Objektif tesis ini adalah untuk menyiasat algoritma analisis isyarat yang berguna, menentukan teknik baru memisahan isyarat, mereka bentuk dan membangunkan kaedah baru untuk BSS. Dalam melakukan proses keputusan anggaran, fungsi kos seperti β-perbezaan dan α-perbezaan disepadukan dengan NMF2D. Eksperimen dilakukan dengan isyarat campuran convolutif disampel dan rakaman sebenar menggunakan di bawah saluran tunggal, masa frekuensi domain (TF) dikira dengan menggunakan masa pendek jelmaan Fourier (STFT) Penilaian prestasi yang digunakan ialah nisbah isyarat-kepada-penyelewengan (SDR). Secara teorinya, β dan α adalah parameter yang digunakan untuk mengubah algoritma NMF2D untuk menghasilkan nilai SDR yang tinggi. Di dalam ujikaji, untuk keputusan simulasi, SDR nilai yang tertinggi untuk β-perbezaan NMF2D adalah SDR = 16.69dB pada β = 0.8 dan n = 100. Untuk NMF2D α-perbezaan, nilai SDR tertinggi adalah SDR = 17.85dB pada α = 1.5 dan n = 100. *Tambahan lagi, kekangan jarangan ke atas β-perbezaan NMF2D dan α-perbezaan NMF2D menghasilkan nilai SDR lebih tinggi. Iaitu, SDR = 17.06dB untuk NMF2D jarang β-perbezaan pada λ = 2.5 dan SDR = 17.99dB untuk NMF2D jarang α-perbezaan pada λ = 5. Kekangan jarangan ini berjaya mengurangkan kekaburan dan menyediakan keunikan kepada model yang dicadangkan. Perbandingan antara β-perbezaan, α-perbezaan, jarang β-perbezaan dan jarang α-perbezaan NMF2D, ia mendapati bahawa nilai SDR untuk jarang α-perbezaan NMF2D adalah kaedah penguraian terbaik antara semua perbezaan. Ini dapat disimpulkan bahawa jarang α-perbezaan NMF2D adalah lebih sesuai dalam memisahkan data yang sebenar.*

ACKNOWLEDGEMENTS

First of all, I would like to express my greatest gratitude to my parents who nurtured me with love and enrolling our tertiary education in this university, Universiti Teknikal Malaysia Melaka (UTeM). This has granted me opportunity to enrich my knowledge which is instrumental in determining my future career prospects. They will give me everything I need by sacrificing their time, money and heart in an effort to raise me up as a useful person to the society.

Secondly, I appreciate the chance that was granted by UTeM to allow me proceed in postgraduate by research. Through this research, I have equipped a lot of skills and techniques that related to my scope of the research in the aspect of hardware and software. Besides that, I also gained a lot of knowledge that regarding to my project which is signal audio processing.

Finally, I would like to thank my dear research supervisor, Associate Professor Dr. Abdul Majid bin Darsono who guided me throughout this research with effort and patience. Associate Professor Dr. Abdul Majid bin Darsono gave me sufficient time and freedom in completing this research. Advices were sought after his and through with his guidance, I am able to complete the report in time.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

representation of ELS, (c) time-frequency representation domain of EHS, (d) time-frequency representation domain of ELS

LIST OF ABBREVIATIONS

BSS	-	Blind Source Separation
CASA	-	Computer Auditory Scene Analysis
dB	-	decibel
DKLI	-	Dual Kullback-Leibler I
EHS	-	estimated heart sound
ELS	-	estimated lung sound
EM	-	Expectation-Maximization
FDE	-	Fall Detection System
fMRI	-	functional Magnetic Resonance Imaging
ICA	-	Independent Component Analysis
IS	-	Itakura-Saito
ISNMF	-	Improved Sparse Nonnegative Matrix Factorization
KL	-	Kullback-Leibler
KLI	-	Kullback-Leibler I
LS	-	Least Square
MAP	-	Maximum A Posteriori
MCBSS	-	Multi Channel Blind Source Separation
MCNMF	-	Multi Channel Nonnegative Matrix Factorization
ML	-	Maximum Likelihood

MNMF	-	Multilayer Nonnegative Matrix Factorization
MU	-	Multiplicative Update
NMF	-	Nonnegative Matrix Factorization
NMF2D	-	Two-Dimensional Nonnegative Matrix Factorization
NMPCF	-	Nonnegative Matrix Partial Co-Factorization
OHS	-	original heart sound
OLS	-	original lung sound
PCA	-	Principal Component Analysis
PCs	-	Principle Components
SAR	-	source-to-artifact ratio
SCBSS	-	Single Channel Blind Source Separation
SCSS	-	Single Channel Source Separation
SDR	-	signal-to-distortion ratio
SFTF	-	Short time Fourier Transform
SH	-	Squared Hellinger
SIR	-	source-to-interference ratio
SiSEC	-	Signal Separation Evaluation Campaign
SMR	-	signal-to-music ratio
SNMF	-	Segmental Nonnegative Matrix Factorization
SVD	-	Singular Value Decomposition
TF	-	time-frequency
α-divergence	-	alpha divergence
β-divergence	-	beta divergence

LIST OF SYMBOLS

$\approx$	-	Approximation
λ	-	Sparseness constraints
η	-	Step size
dB	-	Decibel
Hz	-	Frequency unit
V	-	Original signal
Λ	-	Estimated signal
W	-	spectral basis
H	-	temporal code

LIST OF PUBLICATIONS

Journals:

1. Darsono, A. M., Toh, C. C., Md Saat, M. S., Isa, A. A. M., Manap, N. A., and Ibrahim, M. M., 2017. β-Divergence Nonnegative Matrix Factorization on Biomedical Blind Source Separation. *Journal of Telecommunication, Electronic and Computer Engineering*, 9(2), pp. 1-4. *(Scopus)*
2. Darsono, A. M., Toh, C. C., Saat, S., Manap, N. A., Ibrahim, M. M., and Ahamd, M. I., 2017. β-Divergence Two-Dimensional Nonnegative Matrix Factorization with Sparseness Constraints for Biomedical Signal Separation. *Journal of Telecommunication, Electronic and Computer Engineering*, 9(2-6), pp. 7-10. *(Scopus)*
3. Chuan, T. C., Darsono, A. M., Md Saat, M. S., Isa, A. A. M., and Hashim, N. M. Z., 2016. Blind Source Separation on Biomedical Field by using Nonnegative Matrix Factorization. *ARPN Journal of Engineering and Applied Sciences*, 11(13), pp.8200–8206.
4. Darsono, A. M., Toh, C. C., Saat, M. S., and Isa, A. A. M., 2016. A Study of Blind Source Separation using Nonnegative Matrix Factorization. *ARPN Journal of Engineering and Applied Sciences*, 11(18), pp.10702–10708.

CHAPTER 1

INTRODUCTION

1.1 Research Background

In the past few decades, signal processing engineers and scientists have been devoted in audio processing especially in Blind Source Separation (BSS) (J. Zhang et al., 2007; Naik and Wang, 2014; Parra and Sajda, 2003) through different approaches. It becomes a useful technique when separating or differentiating the mixed audios, sounds or voices. The example of mixtures is mixture of song, music composition, medical instrument like heart sound and lung sound (P. L. P. Li et al., 2006), audio separation system (Hu and Wang, 2004), music separation (Radfar and Dansereau, 2007; Pedersen et al., 2008) and etc. BSS denotes to a set of source signals from a set of mixed signals can be separated without the assistance of information about the source signals or the mixing process.

There is one famous phenomenon which is related to BSS that has been discovered and studied for years by many researchers, which is the cocktail party effect or known as cocktail party problem. Acoustically, the cocktail party effect refers to the ability to one's auditory attention on a particular stimulus while filtering out a range of other stimuli, much the same way that a partygoer can concentrate on a single conversation in an obstreperous place. This specialized listening ability is because of the characteristics of the human speech production system, the auditory system, or high-level perceptual and language processing. It is also known as selective attention. For instance, one's auditory attention on a single talker is increased although he or she surrounded by a cacophony of conversations and background noise (Kim, 2013; McDermott, 2009). Therefore, it can be considered as selectively attend to and recognize one source of auditory input which means of separation on the mixed audio source. However, machine unable to separate the mixtures when the mixtures are recorded from multiple speakers who speak simultaneously. Thus, the process of human auditory system is replicated into machine to separate the voice mixtures.

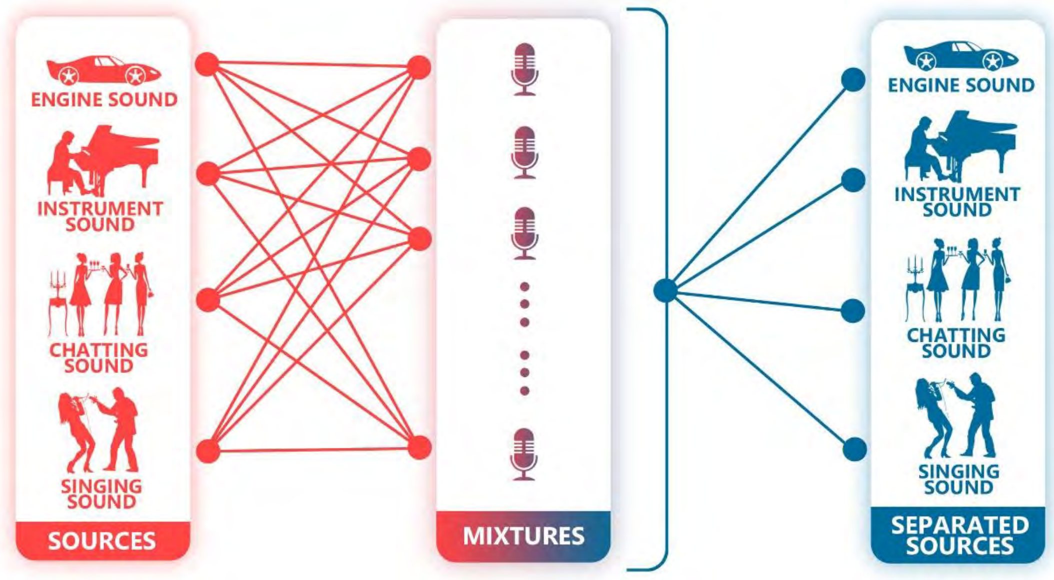

Figure 1.1: The cocktail party problem

Furthermore, the replication of the human auditory system process into machine is also referring to other aspects such as mixture of song with human voice, music composition, mixture of heart sound and lung sound, audio recognition system, data mining, pattern recognition and etc.

This research is focusing on the application of biomedical field which is to separate the mixture of heart sound and lung sound. This is because in human life, respiratory organ plays a very important role within the systema respiratorium. Various kinds of information are used to analyze the respiratory organ condition, one amongst the non-invasive diagnoses relies on the respiratory organ sound information that could be a valuable indicator of metabolism health and sickness. It is conjointly the foremost directly or indirectly accessible information by merely employing a medical instrument. In additional, lung sound recording conjointly finds application to the medical analysis. A clean lung sound recording will increase the accuracy of diagnosis, however, it is hard to be achieved within the real recording scenario. In most of the time, heart sound interferes with lung sound and leads difficult to extract clear lung sound from mixed heart-lung sound signal (Pasterkamp et al., 1997a; Lin and Hasting, 2013). Hence, extraction of a clear lung sound by separating the mixed heart-lung sound signal regarded as an example of BSS issue with respect in biomedical field.

As we know, there are several approaches that have been developed to solve the BSS problem such as the supervised Independent Component Analysis (ICA) (Y. Li et al., 2006; Jang and Lee, 2003; Fevotte and Godsill, 2006) , model-based Single Channel Source Separation (model-based SCSS) (Radfar and Dansereau, 2007), Computer Auditory Scene Analysis (CASA) (Pedersen et al., 2008; P. L. P. Li et al., 2006; Hu and Wang, 2004), Nonnegative Matrix Factorization (NMF) (Ozerov and Fevotte, 2010; Buciu et al., 2008), NMF2D (Gao et al., 2012) and so on. ICA, Supervised ICA and model-based SCSS are principally relying on a priori knowledge obtained during the training phase before estimating the sources. These methods are not robust and expensive. As for CASA method, it is unsupervised methods which aim to resemble the process of human auditory system. In CASA, low level perceptual cues are used to segment a Time-Frequency (TF) representation of a mixture into regions consistent with being generated by a single source. NMF gives part-based decomposition where the spectral bases and temporal code of each source is estimated to reconstruct the estimated sources. The NMF decomposition is unique making it unnecessary in the analysis to enforce the constraints in the form of orthogonality or independence under certain conditions the decomposition.

Novel methods proposed based on NMF which is extended into two-dimensional such as Two-Dimensional Nonnegative Matrix Factorization (NMF2D) deconvolution factor matrices that represent the spectral basis and temporal code. This means the proposed model represents each signal compactly by a single TF profile convolved in both time and frequency by a time-pitch weight matrix. The factorization will consider the convolutive mixing in the decomposition by introducing frequency constrained parameters in the model. The method aims to separate the mixture into its constituent spectral-temporal source components while alleviating the effect of convolutive mixing. In addition, family of β-divergence will be used as a cost function which brings the beneficial property of scale-invariant. Therefore, NMF2D will be the best approach on solving the BSS problem.

In short, NMF able to give an additive parts based representation by factorizing the input spectrogram into a linear combination of basis vectors with non-negativity constraints on both basis and encoding matrices. At the meanwhile, NMF2D separates the instruments by representing each instrument using a single time-frequency profile convolved in both time and frequency. Hence, the difference between NMF and NMF2D methods is that NMF2D performs separation in a straightforward manner without requiring a clustering scheme.

In addition, a statistical techniques will be investigated namely, the Maximum A Posteriori (MAP) estimation (Sparacino et al., 2000) approach which maximizes the joint probability of a mixed signal using Multiplicative Update (MU) rules. To further improve this research work, adaptive sparseness will be incorporate into the cost function to resolve the ambiguity and hence, improve the algorithm performance. The theoretical foundation and mathematical formulation of the propose solutions will be rigorously developed and discuss in details.

There are multiple challenges in BSS which can be summarized as following points, first, modeling the mixture which accommodated the reverberations and nonlinearity of the signals become one of the issue. Second, estimation on the number of sources in the mixture automatically becomes another issue. This research sets out to overcome those problems and to investigate the separation of monaural mixed sources without relying on training