SIGNAL PROCESSING TECHNIQUES FOR EXTRACTING SIGNALS WITH PERIODIC STRUCTURE: APPLICATIONS TO BIOMEDICAL SIGNALS

Foad Ghaderi

Ph.D. 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Signal Processing Techniques for Extracting Signals with Periodic Structure: Applications to Biomedical Signals

---

Thesis submitted to the University of Cardiff in candidature of the degree of Doctor of Philosophy

Foad Ghaderi

Center of Digital Signal Processing
Cardiff University
October 2010

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any other higher degree.

Signed:..Ghosh.....(Candidate) Date:..23/11/2010.....

Statement 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of .....(insert as appropriate PhD, MPhil, EngD)

Signed:..Ghosh.....(Candidate) Date:..23/11/2010.....

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed:..Ghosh.....(Candidate) Date:..23/11/2010.....

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying, inter-library loan and for the title and summary to be made available to outside organisations.

Signed:...Ghosh.....(Candidate) Date:..23/11/2010...

Abstract

In this dissertation some advanced methods for extracting sources from single and multichannel data are developed and utilized in biomedical applications. It is assumed that the sources of interest have periodic structure and therefore, the periodicity is exploited in various forms. The proposed methods can even be used for the cases where the signals have hidden periodicities, i.e., the periodic behaviour is not detectable from their time representation or even Fourier transform of the signal.

For the case of single channel recordings a method based on singular spectrum analysis (SSA) of the signal is proposed. The proposed method is utilized in localizing heart sounds in respiratory signals, which is an essential pre-processing step in most of the heart sound cancellation methods. Artificially mixed and real respiratory signals are used for evaluating the method. It is shown that the performance of the proposed method is superior to those of the other methods in terms of false detection. Moreover, the execution time is significantly lower than that of the method ranked second in performance.

For multichannel data, the problem is tackled using two approaches. First, it is assumed that the sources are periodic and the statistical characteristics of periodic sources are exploited in developing a method to effectively choose the appropriate delays in which the diagonalization takes place. In the second approach it is assumed that the sources of interest are cyclostationary. Necessary and sufficient conditions for extractability of the sources are mathematically proved and the extraction algorithms are proposed.

Ballistocardiogram (BCG) artifact is considered as the sum of a number of independent cyclostationary components having the same cycle frequency. The proposed method, called cyclostationary source extraction (CSE), is able to extract these components without much destructive effect on the background electroencephalogram (EEG).

It is shown that the proposed method outperforms other methods particularly in preserving the remaining signals. The CSE is utilized to remove the BCG artifact from real EEG data recorded inside the magnetic resonance (MR) scanner, i.e., visual evoked potential (VEP). The results are compared to the results of benchmark BCG artifact removal techniques. It is shown that VEPs recorded inside the scanner and processed using the proposed method are more correlated with the VEPs recorded outside the scanner. Moreover, there is no need for electrocardiogram (ECG) data in this method as the cycle frequency of the BCG artifact is directly computed from the contaminated EEG signals.

To my beloved wife *Saideh* for all her support, patience and encouragement.

Acknowledgements

I would like to thankfully acknowledge my supervisor Dr. Saeid Sanei for all his support during my research work at Cardiff University. His enthusiastic supervision, invaluable input, and personal guidance have provided a wonderful basis for the present thesis.

I acknowledge Prof. John McWhirter for encouragement and all valuable discussions and suggestions about my work. I appreciate his support.

I would also like to thank Dr. Kianoush Nazarpour at Newcastle University for fruitful discussions and valuable ideas on the proposed BCG artifact removal method and Dr. Andrew Bagshaw at Birmingham University for provision of the EEG-fMRI data.

I am also grateful to all my friends for providing a motivating research environment within the Centre of Digital Signal Processing, Cardiff University.

Last but not the least, I am indebted to my family. It would have been impossible for me to finish this work without the continuous support and encouragement I received from my beloved wife, Saideh. She has lost a lot due to my research abroad. I thankfully acknowledge all her understanding and love to me and Sepehr. I must also express my gratitude to my parents, brothers, and sister for their unconditional love and prayers for my success. More than all, thank God for providing me the opportunity to step in the excellent world of science and granting me the capability to proceed successfully.

Table of Contents

Publications

The publication list below account partially for the originality of the work presented herein.

Journal publications

[1] F. Ghaderi, H. R. Mohseni, and S. Sanei. Heart sound localization in respiratory signals using singular spectrum analysis. *IEEE Transaction on Biomedical Engineering*, (second revision).

[2] F. Ghaderi, K. Nazarpour, J. McWhirter, and S. Sanei. Removal of ballistocardiogram artifacts using the cyclostationary source extraction method. *IEEE Transactions on Biomedical Engineering*, in press.

Conference papers

[1] F. Ghaderi, B. Makkiabadi, J. G. McWhirter, and S. Sanei. Blind source extraction of cyclostationary sources with common cyclic frequencies. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-10.*, pages 4146–4149, Mar. 2010.

[2] F. Ghaderi, H. R. Mohseni, J. G. McWhirter, and S. Sanei. Blind source extraction of periodic signals. *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-09.*, pages 377–380, Apr. 2009.

[3] F. Ghaderi, H. R. Mohseni, and S. Sanei. A fast second order blind identification method for separation of periodic signals. *18th European Signal Processing Conference, EUSIPCO*, Aug. 2010.

[4] F. Ghaderi, K. Nazarpour, J. G. McWhirter, and S. Sanei. Removal of ballistocardiogram artifacts exploiting second order cyclostationarity. *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, Aug. 2010.

[5] F. Ghaderi, S. Sanei, and M. G. Jafari. A unitary diagonalization method for blind cyclostationary source separation. *ICA Research Network International Workshop, ICArn*, Sep. 2008.

[6] F. Ghaderi, S. Sanei, B. Makkiabadi, Vahid Abolghasemi, and J. G. McWhirter. Heart and lung sound separation using periodic source extraction method. *16th International Conference on Digital Signal Processing, DSP*, pages 1–6, July 2009.

Other collaborative publications

[1] B. M. Abadi, A. Sarrafzadeh, F. Ghaderi, and S. Sanei. Semi-blind channel estimation in mimo commumication by tensor factorization. *IEEE/SP 15th Workshop on Statistical Signal Processing, SSP*, pages 313–316, Aug. 2009.

[2] V. Abolghasemi, S. Sanei, S. Ferdowsi, F. Ghaderi, and A. Belcher. Segmented compressive sensing. *IEEE/SP 15th Workshop on Statistical Signal Processing, SSP*, pages 630–633, Aug. 2009.

[3] B. Makkiabadi, F. Ghaderi, and S. Sanei. A new tensor factorization approach for convolutive blind source separation in time domain. *Presented at 18th European Signal Processing Conference (EUSIPCO)*, 2010.

[4] H. R. Mohseni, F. Ghaderi, E. E. Wilding, and S. Sanei. A beamforming particle filter for eeg dipole source localization. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-09*, pages 337–340, Apr. 2009.

[5] H. R. Mohseni, F. Ghaderi, E. L. Wilding, and S. Sanei. Variational bayes for spatiotemporal identification of event-related potential subcomponents. *IEEE Transactions on Biomedical Engineering*, 57(10):2413–2428, Oct. 2010.

List of Figures

List of Tables

List of abbreviations

AAS	Average Artifact Subtraction
AMUSE	Algorithm for Multiple Unknown Signals Extraction
AV	Atrioventricular
BCG	Ballistocardiogram
BOLD	Blood-Oxygen-Level Dependent
BSE	Blind Source Extraction
BSS	Blind Source Separation
CPU	Central processing unit
CSE	Cyclostationary Source Extraction
CWT	multi-resolution decomposition using Continuous Wavelet Transform
dB	decibel
ECG	Electrocardiography/Electrocardiogram
EEG	Electroencephalography/Electroencephalogram
EKG	Electrocardiography/Electrocardiogram
EMF	Electromotive Forces
ENT	Entropy based method
EOG	Electrooculogram
FIR	Finite Impulse Response
fMRI	functional Magnetic Resonance Imaging
GEVD	Generalized Eigenvalue Decomposition
HS	Heart Sound
IC	Independent Component
ICA	Independent Component Analysis
INPS	Improvement in Normalized Power Spectrum ratio
ISTFT	Inverse Short Time Frequency Transform

List of abbreviations, continued

JADE	Joint Approximate Diagonalization of Eigen matrices
MHD	Magnetohydrodynamic
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRL	Mean Rejection Level
OBS	Optimal Basis Set
PC	Personal Computer
PCA	Principal Component Analysis
PCG	Phonocardiogram
PI	Performance Index
PSD	Power Spectral Density
SNR	Signal to Noise Ratio
SOBI	Second Order Blind Identification
SoI	Source of Interest
SSA	Singular Spectrum Analysis
STFT	Short Time Frequency Transform
SVD	Singular Value Decomposition
VEP	Visual Evoked Potentials
πCA	Periodic Component Analysis

List of symbols

a	Scalar
$\mathbf{a}$	Column vector
$\mathbf{A} = [a_{pq}]$	Matrix
a_{pq}	pqth element of matrix $\mathbf{A}$
diag($\mathbf{A}$)	Vector holding the diagonal elements of $\mathbf{A}$
off{$\mathbf{A}$}	Sum of the squared off diagonal elements of the matrix $\mathbf{A}$
E{.}	Expected value
$r_{pq}^{\beta}(\mathbf{x}, \tau)$	Cyclic cross correlation of $x_p(t)$ and $x_q(t)$ at cycle frequency β and time delay τ
$r_{pq}(\mathbf{x}, \tau)$	Cross correlation of $x_p(t)$ and $x_q(t)$ at time delay τ
$\mathbf{R}_{\mathbf{x}}(t, \tau)$	Correlation function of the vector $\mathbf{x}(t)$ at time delay τ
$\mathbf{R}_{\mathbf{x}}^{\beta}(t, \tau)$	Cyclic correlation function of the vector $\mathbf{x}$ at time delay τ and cycle frequency β
j	$\sqrt{-1}$
$[.]^T$	Transpose of a vector or matrix
$[.]^*$	Complex conjugate of a complex number
$[.]^H$	Complex conjugate transpose of a vector or matrix
$\langle . \rangle$	Time averaging operator
$\lfloor . \rfloor$	Floor operator
$*$	Convolution
$\otimes$	Matrix convolution
$[.]^{\sharp}$	Pseudo-inverse
$\Re(.)$	Real part of a complex variable
$\Im(.)$	Imaginary part of a complex variable

List of symbols, continued

$\mathbf{I}_m$	m-dimensional identity matrix
$\mathbb{R}$	Real space
$\mathbb{C}$	Complex space

Chapter 1

INTRODUCTION

1.1 Overview

In many of the conventional statistical signal processing methods the signals are treated as they are statistically stationary. That means it is assumed that the parameters of the underlying physical mechanisms which generate the signals do not vary with time. But for most of the signals from man-made systems such as those encountered in mechanical systems and communication, radar and sonar systems some parameters vary periodically with time. Because of presence of rhythmic, seasonal, or other cyclic behaviour, periodic structure also arises in some of the signals which have natural origins, as in data encountered in economics, astronomy, climatology, biomedical and atmospheric science [41]. Although in some of these signals the periodicities are not detectable as periodic functions of time, the statistical characteristics of the signals may vary periodically with time [43]. This class of signals are called cyclostationary signals.

Mostly, the periodicities are ignored by signal processing algorithms. However, in many cases much information can be gained to improve the performance of the signal processing techniques by recognizing and exploiting the underlying periodicities [41].

The objective of this thesis is to develop suitable multi and single channel signal processing techniques for extracting signals with periodic structures. The proposed methods have been utilized to mitigate the physiological artifacts originated from heart activity. Most of biomedical signals are somehow contaminated by artifacts synchronized with heartbeats. Therefore, artifact processing is an essential task in analysis of biomedical

signals.

In this chapter the structure and periodic behavior of heart are studied. The theories behind the techniques for analyzing and processing cyclostationary and periodic signals are explained in the next chapters. Cogency of artifact removal in biomedical signal processing is explained in Section 1.2. A list of artifacts which are generated from heart activity is presented in Section 1.3. Heart structure and different recording approaches for measuring heart activity are detailed in Sections 1.4 and 1.5, respectively. Objectives of this research are explained in Section 1.6. Finally, Section 1.7 presents the organization of this dissertation.

1.2 Artifact removal

Removing noise and artifacts from a signal is probably the first and the most common step in signal processing. Distinction between the signal and the noise or artifacts is a heuristic decision made by the user. In some cases, the distinction between the noise and the signal is obvious and in other applications some prior information is required to separate them. Once distinction between the signal and artifacts is identified, the next goal is to remove the artifacts with the least possible distortion of the signals of interest [22]. In biomedical signal recordings, a wide variety of artifacts can appear, some of which can be easily identified by simple classical algorithms, while identification and removal of the others may be extremely difficult.

A common approach for categorizing artifacts is by referring to the root of the artifact, i.e., physiological or technical. It is also worth mentioning that multiple types of artifacts can exist in a single recording at the same time. In practice, the influence of artifacts of technical origin can be reduced to a large extent by using appropriate devices, controlling the experimental environment, and paying extra attention to the attachment of electrodes to the body surface. However, it is impossible to avoid the influence of physiological artifacts [93]. Signal processing techniques are required to identify and remove this kind of artifacts before any further processing.

1.3 Heart originated artefacts

Cardiac pulsation and blood flow in the body are the major sources of a class of physiological artifacts in most of biomedical recordings. These artifacts are directly related to the heart activity and depending on the acquisition technology interfere with the underlying signals in different forms. Some examples of this class of artifacts are as follows. Electroencephalogram (EEG) and electromyogram (EMG) recordings are contaminated by electrocardiogram (ECG) signals. Blood oxygen level dependent (BOLD) signals of brain can contain some periodic artifacts which are originated from blood flow in the brain. Lung sounds in auscultation signals are always obscured by heart sounds as the result of continuous activity of the heart.

The repetitive and regularly occurring waveform pattern of the heart activity characterizes the normal heartbeats. Fortunately, it is always possible to track heart activity with existing techniques to find signatures of the artifact in the recorded data and signals of interest.

1.4 Heart structure

The heart is a muscular organ responsible for pumping blood throughout the body in all animals. The blood is distributed all around the body by repeated, rhythmic contractions of the heart. The heart can be seen as two separate pumps: one of which is the right heart responsible for pumping the blood through the lungs and the other is the left heart which pumps the blood to the other organs in the body. Each of these is a two-chamber pump composed of an atrium and a ventricle [54]; see Fig. 1.1.

The human heart is composed of three types of muscle: atrial, ventricular, and specialized excitatory and conductive muscle fibers. The heart's continuous contract and relax cycles are coordinated by a local periodic electrical stimulation. The cardiac events that occur from beginning of one heartbeat to beginning of the next are called the cardiac cycle. Each cardiac cycle consists of a period of relaxation called diastole, during which the heart fills with blood, followed by a period of contraction called systole [54].

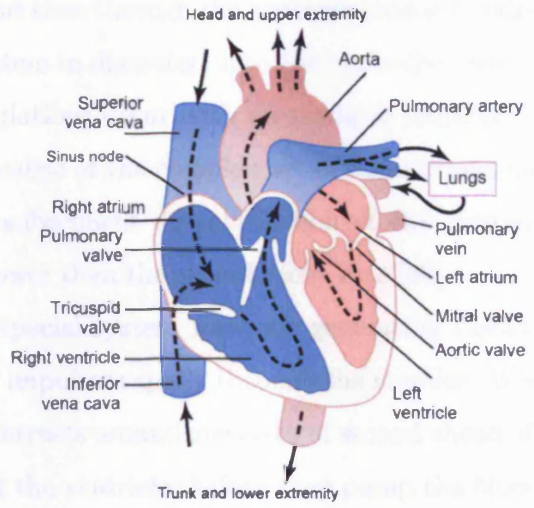


Figure 1.1: Structure of the heart and blood flow inside this organ; taken from [54].

Small valves prevent backflow of blood from the ventricles into the atrium during systole. These valves are anchored to the wall of the ventricle by chordae tendineae. The mitral valve lies between the left atrium and the left ventricle. The tricuspid valve is on the right side of the heart, between the right atrium and the right ventricle. As the mitral and the tricuspid valves separate the atria and the ventricles of the heart, they are called the atrioventricular (AV) valves [54].

There are two other valves at the base of both the pulmonary artery and the aorta (the two arteries taking blood out of the ventricles). These valves permit blood to be pumped into the arteries, and similar to AV valves prevent backflow of blood. The aortic valve is located between the left ventricle and the aorta and the pulmonary valve lies between the right ventricle and the pulmonary artery. These valves are called Semilunar valves and do not have chordae tendineae [54].

Heart cycles are initiated by spontaneous generation of an action potential in the sinus node located in the superior lateral wall of the right atrium. Each heart muscle cell has an electrical charge across its membrane. De-polarisation of the cell (reducing this charge towards zero) causes the cell to contract [30].

During each heartbeat a healthy heart has an orderly progression of a wave of de-polarisation that is triggered by the cells in the sinus node, and travels rapidly from there

through both atria and then through the atrioventricular bundle (a bundle of conductive fibers several millimeters in diameter) into the ventricles [30].

The regular stimulations occur with an intrinsic frequency of 100 to 120 beats per minute. However, because of the complex set of chemical exchanges between the initial stimulation and the subsequent de-polarisation of the surrounding cardiac tissue the actual heart rate is lower than the stimulations rate [30].

The heart has a special system for both generating rhythmical electrical impulses and conducting these impulses rapidly through the muscles. When this system functions normally, the atria contracts around one sixth of second ahead of ventricular contraction, which allows filling of the ventricles before they pump the blood through the lungs and peripheral circulation [54].

1.5 Heart signal acquisition

Different techniques have been developed to measure heart activity. Based on the features and the quality, hardware and software availability and the requirements one of these techniques are selected by the researchers or the clinicians to assess the heart activities. Auscultation, electrocardiography, and Carotid arteriogram are the most used techniques.

1.5.1 Auscultation

Auscultation is the act of listening, usually with a stethoscope, to the internal sounds of the body. This technique is used usually to listen to three main organs/organ systems: heart, lungs, and the gastrointestinal system. When auscultating the heart, physicians listen for abnormal sounds coinciding with heartbeats. When listening to the lung, breathing sounds such as wheezes and crackles are identified. The gastrointestinal system is auscultated to note the presence of bowel sounds [107]. Auscultation is performed as one of the initial and very important steps in the diagnosis of heart or lung abnormalities.

Heart sounds are caused by flow of blood into and out of heart through valves and also heart tissue movements [49]. By placing a stethoscope over the chest, close to the heart location, four basic heart sounds can be identified which are referred to as S1, S2, S3,

and S4. The first and second heart sounds (S1 and S2) are the most fundamental heart sounds. S1 is caused by closure of the mitral and tricuspid valves at the beginning of ventricular contraction. During this contraction cycle the blood is pumped from heart to body. S2 is caused by closure of the aortic and pulmonic valves at the beginning of ventricular relaxation; see Fig. 1.1. The third heart sound (S3), when audible, occurs early in ventricular filling and the fourth heart sound (S4), when audible, is caused by vibration of the ventricular wall during atrial contraction. Both S3 and S4 do not have significant amplitude and mostly are not audible in healthy subjects. These components of heart sound are ignored in most of heart sound processing applications. Main frequency components of the heart sound are concentrated in the range of 20-150 Hz.

One approach for visualizing heart sounds is to record the signals using a microphone and to display the electrical signals graphically. This is called a phonocardiogram (PCG), in which the x and y axis represent time and voltage, respectively.

1.5.2 Electrocardiography

Electrocardiography (ECG or EKG) is an interpretation of the electrical activity of heart over time captured non-invasively from the chest wall using skin electrodes.

The ECG works by detecting and amplifying the tiny voltage changes on the skin caused when the heart muscles de-polarize during each heart beat. The voltage changes are displayed on screen or on paper as a wavy line which indicates the overall rhythm of the heart. ECG signals are used to identify possible malfunctions or weaknesses in different parts of the heart muscle.

In practice more than two electrodes are used to record ECG signals and they can be combined into a number of pairs. The output voltage of each pair is called a lead. Usually, an ECG is taken while the patient is in rest.

ECG is known as the best measure for diagnosis of heart abnormal rhythms [21]. It may also be used to assess the success of drug administration or coronary revascularisation such as coronary intervention or bypass surgery.

A typical ECG waveform is comprised of a sequence of almost similar structure. The

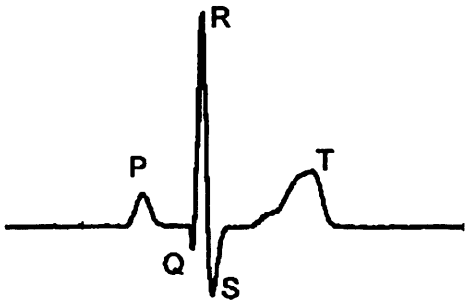


Figure 1.2: A typical ECG waveform for one heart beat.

first positive peak is called P-wave, followed by three peaks which are known as 'QRS complex' and then a trailing T-wave (see Fig. 1.2). These waves are defined as follows:

P-wave: It is a low voltage fluctuation which is caused by de-polarization of the atria prior to contraction.

QRS complex: The highest peaks of the ECG signal caused by the ventricular de-polarization.

T-wave: Caused by ventricular re-polarisation.

1.5.3 Carotid arteriogram

Carotid arteriogram is an x-ray examination to see the blood vessels in neck. These vessels supply blood to the brain. This technique involves inserting a fine plastic tube, called a catheter, into an artery through a needle placed in the groin. The tip of the catheter will need to be guided to its proper position by the radiologist (or x-ray doctor) whilst watching on a TV monitor [80].

1.6 Objectives of this thesis

Heart originated artifacts are referred to as quasi-periodic signals contaminating different biomedical recordings and the objective is to extract the artifacts from the recorded signals. Multi channel signal processing methods in the context of blind source separation (BSS) are powerful candidates for applications for which more than one recorded channel is available. Time series analysis methods are proposed for single channel signal processing.

The objectives of this research are as follows:

- Reviewing state of the art research on cyclostationarity and cyclostationary source separation methods.
- Developing effective and efficient algorithms for extracting signals with periodic structure.
- Reviewing recent achievements in heart and lung sound separation and proposing new approaches in order to enhance the performance of the separation methods.
- Removing periodic artifacts from EEG data recorded inside magnetic resonance imaging scanner.

Recording devices, standards, and procedures are not of any concern in this dissertation and therefore, are not covered.

1.7 Organization of the thesis

Concepts of blind source separation and well established instantaneous BSS methods are reviewed in Chapter 2. A more detailed discussion about these methods specifically focussing on signals with periodic structures is also provided. In this part, the cyclostationary sources, which are known to have periodic higher order statistics, are reviewed. Separation techniques based on cyclostationarity are powerful tools for extracting the signals with hidden periodicities.

In Chapter 3 an efficient algorithm for separating periodic sources is presented. The algorithm is proposed based on the assumption that the sources are periodic and the periodicities are known *a priori*. The algorithm is used for separation of artificial mixtures. Human voice and music as quasi-periodic signals are used to evaluate the separation performance of the algorithm.

Heart and lung sound separation is a challenging problem in the field of biomedical signal processing. Localization of heart sound components is an essential pre-processing step in some removal methods. In Chapter 4 a novel method for localizing S1 and S2 components is proposed. This method is based on singular spectrum analysis of

the respiratory signals recorded over the chest. Comparison between the results of the proposed method and those of two localization methods are reported.

Two methods for extracting cyclostationary sources are proposed in Chapter 5. In the first method it is assumed that the sources of interest have distinct periodic structure, while in the second method the sources can have identical periodic structures. Necessary and sufficient conditions for extracting the sources are considered and mathematical proofs are provided. Preliminary experiments on synthetic data are conducted and the results are reported.

Complementary experiments using the method proposed in Chapter 5 are carried out for synthetic data and the results are presented in Chapter 6. Quantitative indices are defined for more accurate evaluation of the proposed method in comparison with other methods. Ballistocardiogram artifact is one of the major artifacts of EEG recorded jointly with fMRI. This artifact is generated as the result of interactions between the patient's body, inter-electrode loops, and the magnetic field inside the MR machine. Details of the proposed approaches for identifying periodic behaviour of the sources, deflating the extracted sources, and managing the heart rate variability are presented.

Finally, in Chapter 7 the work presented herein is summarized and some future research directions are proposed.

Chapter 2

EXTRACTING PERIODIC SIGNALS

2.1 Overview

Blind source separation (BSS) is currently one of the most attractive areas of research in statistical signals processing and unsupervised machine learning due to its potential applications in various areas such as digital communications, biomedical signal processing, and financial time series analysis [62, 90]. The term 'source separation' refers to the fact that the objective is recovering unknown underlying sources which are mixed through an unknown environment from a set of existing observations. The main feature of BSS lies in the word 'blind' which means that the separation process is done without any training data. However, weak assumptions about the sources or the unknown environment are permitted in the BSS context.

One of these assumptions is statistical independence of the sources, which is the foundation for most of the BSS methods. The term for the operation of this family of algorithms is independent component analysis (ICA). The objective in ICA methods, as powerful statistical tools, is therefore to transform the data into a set of signals that are mutually statistically independent.

2.2 Blind source separation

The BSS problem is to recover the original vector of sources $\mathbf{s}(t)$ from a set of observations $\mathbf{x}(t)$. It is assumed that $\mathbf{x}(t)$ is a mixture of the underlying sources and the mixing

medium is unknown. In general, blind source separation is formulated as follows.

Assume that the outputs of a MIMO (multi input, multi output) linear dynamical system are observed via n sensors:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} \tag{2.2.1}$$

The aim is to design a stable inverse system that can estimate the inputs to the MIMO system just based on the observations and *a priori* knowledge about the system. As it is shown in Fig. 2.1, m input signals

$$\mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_m(t) \end{bmatrix} \tag{2.2.2}$$

are mixed in an unknown dynamic system. n sensors measure these mixtures, while additive noise is added to the measurements. The conventional method used in BSS is estimating the unknown mixing matrix by processing the outputs of the dynamical system. In this case based on some assumptions that come from *a priori* knowledge about the system and using suitable algebraic or optimization methods, the source signals can be estimated implicitly. BSS is referred to as *blind* because recovery of the sources is carried out without knowledge about the sources or characteristics of the mixing channels, except for some minor assumptions about the sources or the mixing system.

2.2.1 Instantaneous BSS

In many BSS problems, the sources are instantaneously mixed by a linear system of zero memory. In mathematical terms, it is assumed that m unknown source signals $s_q(k)$, $(q \in \{1, 2, ...m\})$ are mixed linearly to generate a set of n mixtures $x_p(k)$, $(p \in$

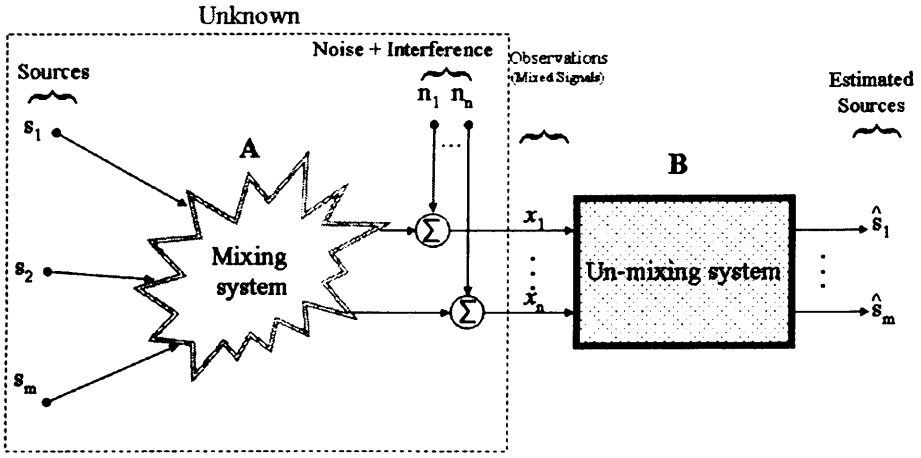


Figure 2.1: General model illustrating blind source separation.

$\{1, 2, ...n\}$). Usually it is assumed that $n \geqslant m$. This can be written as

$$x_p(k) = \sum_{q=1}^{m} a_{pq} s_q(k) + n_p(k) \tag{2.2.3}$$

or in the matrix form

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{n}(k) \tag{2.2.4}$$

where $\mathbf{n}(k) = [n_1(k), n_2(k), ..., n_n(k)]^T$ is a vector of additive noise, $\mathbf{A}$ is the unknown full column rank mixing matrix, and T denotes the transpose of a matrix. In general, just the observations are available and the number of sources is unknown [27].

In this way, a typical BSS solution estimates a linear full rank separating (unmixing) matrix $\mathbf{B}$ such that the output signal vector $\hat{\mathbf{s}}(k) = [\hat{s}_1(k), \hat{s}_2(k), ..., \hat{s}_m(k)]^T$, defined by $\hat{\mathbf{s}}(k) = \mathbf{B}\mathbf{x}(k)$ is an estimate of the input vector $\mathbf{s}(k)$ [28].

2.2.2 Convolutive BSS

In a broad category of BSS applications the mixing media are more complicated than instantaneous systems. If the mixtures are weighted and delayed, and each source contributes to the mixtures with multiple weights and delays corresponding to multiple paths, then, the system is referred to as convolutive. Examples of such systems can be seen in acoustics, biomedical engineering, and telecommunications [84].

In a convolutive model the following equation governs the relationship between the pth observation, the original source signals, and additive noise $n_p(k)$:

$$x_p(k) = \sum_{q=1}^{m} \sum_{\tau=0}^{T-1} a_{pq\tau} s_q(k-\tau) + n_p(k), \qquad (p = 1, 2, ..., n) \tag{2.2.5}$$

where τ represents the time delay corresponding to the path from the pth source to the qth sensor, and $a_{pq\tau}$ is the corresponding mixing matrix filter coefficient.

In practice, these coefficients may also change in time, but for simplicity the mixing model is often assumed stationary. In theory, the filters may be of infinite length, but in practice it is assumed that $T < \infty$. In matrix form, the convolutive model of (2.2.5) is written as

$$\mathbf{x}(k) = \sum_{\tau=0}^{T-1} \mathbf{A}_\tau \mathbf{s}(k-\tau) + \mathbf{n}(k) \tag{2.2.6}$$

where $\mathbf{A}_\tau$ is an $n \times m$ matrix which contains the τth filter coefficients. In the z-domain the convolutive mixture of (2.2.6) can be written as:

$$\mathbf{X}(z) = \mathbf{A}(z)\mathbf{S}(z) + \mathbf{N}(z) \tag{2.2.7}$$

where $\mathbf{A}(z)$ is a matrix with an FIR polynomial in each entry [84].

2.2.3 Anechoic BSS

In the third class of BSS applications the mixtures are delayed and weighted, however, unlike the convolutive BSS the weights do not change for different time delays. The mixing model of an anechoic system is formulated as follows:

$$x_p(k) = \sum_{q=1}^{m} a_{pq} s_q(k - \tau_{pq}) + n_p(k), \qquad (p = 1, 2, ..., n) \tag{2.2.8}$$

Although some anechoic blind separation methods have been proposed [98, 109], instantaneous and convolutive source separation methods are the main trends within the BSS community. In theory, convolutive BSS algorithms should perform much better in an anechoic scenario.

2.2.4 Independent component analysis

Independent Component Analysis (ICA) is the most widely used technique in BSS. The following assumptions are made to ensure that the sources can be estimated.

1. The components are assumed to be statistically independent.

This assumption is enough for the model to be estimated. In most of the ICA methods, nothing more than this assumption is needed. By definition, random variables, v_1, v_2, ..., v_n are said to be independent if no information about values of v_i can be extracted from the value of v_j where $i \neq j$. Mathematically speaking, v_1, v_2, ..., v_n are statistically independent if and only if their joint probability density function (pdf) is factorizable as

$$p(v_1, v_2, ..., v_n) = p_1(v_1) \times p_2(v_2) \times ... \times p_n(v_n) \tag{2.2.9}$$

where $p(v_1, v_2, ..., v_n)$ and $p_i(v_i)$ denote the joint pdf of the random variables, and the marginal pdf of the random variable v_i, respectively.

2. The components must have non-Gaussian distributions.

In methods based on higher order statistics no more than one source can be Gaussian since the moments and cumulants of more than 2nd order are zero for Gaussian signals while they are needed for estimating the mixing matrix and the source signals [62]. In some of the ICA methods the assumption of non-Gaussianity is replaced by some assumptions on time structure of the signals.

2.2.5 Ambiguity of the problem

In BSS it is not possible to uniquely estimate the source signals without some *a priori* knowledge. In other words, the source signals are estimated up to the following indeterminacies:

- The amplitude of the original sources can not be determined.

Since both $\mathbf{A}$ and $\mathbf{s}(k)$ are unknown, any scalar multiplier α_p of source s_p can be canceled by dividing the corresponding column $\mathbf{a}_p$ by the same multiplier;

$$\mathbf{x} = \sum_p \left(\frac{1}{\alpha_p}\mathbf{a}_p\right)(\alpha_p s_p) \tag{2.2.10}$$

Therefore, to overcome this problem in ICA methods the variances of the source signals are assumed to be one, that is $\forall p \quad E\{s_p^2\} = 1$. It is notable that still there is an ambiguity in the sign of signals, i.e., a multiplier of -1 can be changed between a source and the corresponding column vector of the mixing matrix. Fortunately this is not so important, since in most of the real world applications the waveform contains most of the required information.

- The order of independent components can not be determined.

Again due to the lack of information about $\mathbf{A}$ and $\mathbf{s}$, the order of the sources can be changed and so by changing the order of the corresponding column vector in the mixing matrix, no change will be observed in the sensor measurements. In other words, a permutation matrix and its inverse can be substituted in the mixing model to give $\mathbf{x}(k) = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}(k)$. The elements of $\mathbf{P}\mathbf{s}(k)$ are the reordered form of the original signals. The matrix $\mathbf{A}\mathbf{P}^{-1}$ is a new unknown mixing matrix to be estimated [62].

These indeterminacies are usually expressed as scaling and permutation of estimated source signals which are not so important in most of the real world applications. In majority of signal processing applications, it is desirable to have only the waveforms of the original sources and useful information can be exploited form the waveforms. In these cases estimation of the exact amplitude, order of the signals, or even time delays are not very crucial [28]. It is also notable that there is no guarantee that the waveforms of estimations are exactly the same as the source signals. Those are just estimations of the source waveforms and thus their accuracy varies for different sources and different separation methods.

2.3 Protagonists in BSS

Several algorithms have been developed so far in the BSS context each relying on different assumptions and exploit different characteristics of the signals. Here, we review Infomax (derived from information maximization), fast fixed-point algorithm for independent component analysis (known as FastICA), second order blind identification

(SOBI), and joint approximate diagonalization of eigen matrices (JADE). These methods are respectively based on maximization of the mutual information, maximization of the negentropy of every estimated signal, diagonalization of a set of time delayed covariance matrices, and minimization of the sum of the squared cross-cumulants of the estimates. These methods will be used in the next subsections to evaluate the performance of the proposed methods.

2.3.1 Infomax

This algorithm maximizes the output entropy or information flow of a neural network with nonlinear outputs. Assume $\mathbf{x}$ is the input to the neural network and $\mathbf{y}$ is the output vector. The following equation describes the relation between the inputs and the outputs of neurons:

$$y_i = \phi_i(\mathbf{w}_i^T\mathbf{x}) + n_i \tag{2.3.1}$$

where ϕ_i are some nonlinear scalar functions, $\mathbf{w}_i$ are the weight vectors of the neurons, and $\mathbf{n} = [n_1, n_2, ...n_n]^T$ is the additive Gaussian white noise vector. The entropy of the output is expressed by

$$H(\mathbf{y}) = H(\phi_1(\mathbf{w}_1^T\mathbf{x}), \ldots, \phi_n(\mathbf{w}_n^T\mathbf{x})) \tag{2.3.2}$$

For a typical invertible transformation of the random vector $\mathbf{x}$, $\mathbf{y} = \mathbf{f}(\mathbf{x})$, the relationship between the entropies of $\mathbf{y}$ and $\mathbf{x}$ can be expressed as

$$H(\mathbf{y}) = H(\mathbf{x}) + E\{\log|\det J\mathbf{f}(\mathbf{x})|\} \tag{2.3.3}$$

where $J\mathbf{f}(.)$ is the Jacobian matrix of the function $\mathbf{f}(.)$ [62].

Using (2.3.3) and assuming that $\mathbf{y} = \mathbf{f}(\mathbf{x}) = [\phi_1(\mathbf{w}_1^T\mathbf{x}), \ldots, \phi_n(\mathbf{w}_n^T\mathbf{x})]$ denotes the nonlinear function defined by the neural network, the transformation of the entropy in (2.3.2) can be obtained as

$$H(\mathbf{y}) = H(\mathbf{x}) + E\{\log|\det\frac{\partial\mathbf{f}}{\partial\mathbf{W}}(\mathbf{x})|\} \tag{2.3.4}$$

The second term in the right hand side of (2.3.4) can be easily derived and simplified as follows:

$$E\{\log|\det\frac{\partial \mathbf{f}}{\partial \mathbf{W}}(\mathbf{x})|\} = \sum_i E\{\log\phi_i'(\mathbf{w}_i^T\mathbf{x})\} + \log|\det\mathbf{W}| \tag{2.3.5}$$

If the nonlinear functions ϕ_i are chosen as the culumative distribution functions corresponding to the density p_i of the ith source, i.e. $\phi_i' = p_i$, the output entropy is equal to the likelihood. In other words, Infomax is equivalent to maximum likelihood estimation and therefore, all the methods used to maximize the likelihood can be used here to maximize the entropy of the neural network output.

Gradient, natural gradient, and fast fixed-point algorithms have been proposed to find the maximum point of the likelihood function [62]. Here, we discuss the Bell-sejnowsky algorithm as the simplest algorithm obtained by gradient method [15].

Using the stochastic gradient of the log-likelihood expression in (2.3.5) the update relation of the neural network weight vector is as follows:

$$\Delta\mathbf{W} \propto [(\mathbf{W}^T)^{-1} - \varphi(\mathbf{y})\mathbf{x}^T] \tag{2.3.6}$$

where $\varphi(\mathbf{y})$ is a nonlinear function represented by a column vector whose i-th component is

$$\varphi_i(y_i) = -\frac{\partial p(y_i)/\partial y_i}{p(y_i)} \tag{2.3.7}$$

The approximated probability density function (pdf) of the i-th source signal is shown by $p(y_i)$. In practice, $\varphi(y_i) = -2\tanh(y_i)$ and $\varphi(y_i) = \tanh(y_i) - y_i$ are used for super-Gaussian and sub-Gaussian signals, respectively.

2.3.2 FastICA

This method is one of the well-known instantaneous ICA methods that maximizes the statistical independence of the estimated sources by maximizing the non-Gaussianity. FastICA is inspired from the central limit theorem which states that the distribution of sum of independent random variables tends to have Gaussian distribution [62]. Based on this theorem it is assumed that the distribution of the mixtures are closer to Gaussian

distribution than that of the original sources. Therefore, statistical independence and non-Gaussianity are known to be equivalent in the context. However, this would result in the main limitation of the method which implies that at most one source can possess a Gaussian distribution.

Negentropy is a non-negative function which quantifies how much a random variable deviates from Gaussianity. This function can be formulated as [62]:

$$N(y) = H(x_{gauss}) - H(x) \tag{2.3.8}$$

where $H(.)$ denotes the entropy of the enclosed term, and x_{gauss} is a Gaussian random variable of the same variance as x. Due to properties of the entropy function it is concluded that negentropy is always non-negative. Due to the computational complexity of calculating negentropy, Hyvrinen *et al.* proposed to use the following approximation instead [62]:

$$N(y) \propto [E\{g(y)\} - E\{g(x_v)\}]^2 \tag{2.3.9}$$

where v is a Gaussian variable of zero mean and unit variance and $g(u)$ can be any non-quadratic function. In particular, choosing a g that does not grow too fast can provide more robustness. The following choices of g have been approved to be very useful [62]:

$$g_1(u) = \frac{1}{a_1} \log\cosh(a_1 u) \tag{2.3.10}$$

$$g_2(u) = -\exp(-u^2/2) \tag{2.3.11}$$

where $1 \leq a_1 \leq 2$ is a suitable constant, often taken equal to one.

Differentiating (2.3.9) with respect to the separating vector $\mathbf{w}_i$ corresponding to the ith source yields:

$$\nabla \mathbf{w}_i = \alpha E\{\mathbf{z} g'(\mathbf{w}_i^T \mathbf{z})\} \tag{2.3.12}$$

where $\alpha = E\{g(\mathbf{w}_i^T \mathbf{z})\} - E\{g(x_v)\}$ and $\mathbf{z}$ is the whitened mixtures. The following fixed point iteration is then suggested intuitively:

$$\mathbf{w}_{i,k+1} \longleftarrow E\{\mathbf{z} g'(\mathbf{w}_{i,k}^T \mathbf{z})\} \tag{2.3.13}$$

As the convergence of the fixed point iteration of (2.3.13) is not satisfactory, a Lagrangian approach is employed in [62] to yield a convergent fixed point iteration as follows:

$$\mathbf{w}_i \longleftarrow E\{\mathbf{z}g'(\mathbf{w}_i^T\mathbf{z}) - E\{g''(\mathbf{w}_i^T\mathbf{z})\mathbf{w}_i\} \tag{2.3.14}$$

This method has been widely used in biomedical signal processing [57, 67, 102], which stems from the fact that most natural signals are non-Gaussian.

2.3.3 SOBI

If the underlying sources in an ICA problem are not random variables, i.e. they are time series with particular time structure, additional statistics can be used to improve the estimation of the model. As for example, the cross-covariance or cross-cumulant of the sources are used in some methods. In this case the non-Gaussianity assumption can be relaxed in some algorithms.

The covariance matrix of an observation vector at delay zero, i.e. $C_0^x = E\{\mathbf{x}(t)\mathbf{x}(t)^T\}$ does not contain enough parameters to allow estimation of the mixing matrix. That means simply trying to whiten the observation vector will not yield independent components. This is why in the basic ICA methods one has to exploit the non-Gaussian structure of the components.

The key point here is that the time delayed covariance matrix $C_\tau^x = E\{\mathbf{x}(t)\mathbf{x}(t-\tau)^T\}$ can be used to obtain more information about the underlying sources. In the simplest case, as it is proposed in the algorithm for multiple unknown signals extraction (AMUSE) algorithm [97] only one time delay is used. It is assumed that $\mathbf{z}(t)$ is the zero-mean white version of the observation vector, $\mathbf{x}(t)$. In the next step, the eigenvalue decomposition of $\bar{\mathbf{C}}_\tau^z = \frac{1}{2}[C_\tau^z + C_\tau^{z^T}]$ for some τ is calculated. The rows of the separating matrix are given by the eigenvectors. A similar algorithm has been proposed by Molgedey *et al.* in [76]. Although these algorithms are very simple and fast, they only work when the eigenvectors of the matrix $\bar{\mathbf{C}}_\tau^z$ are all distinct, which is not always possible.

An extension of the AMUSE algorithm that reduces the dependency on the appropriate time delay and hence improves the performance is to consider the covariance matrices at several time delays. This method, called SOBI, simultaneously diagonalizes

the covariance matrices calculated at different time delays [16]. Although practically it is not possible to perfectly diagonalize all the matrices, the objective is to minimize the value of the following cost function:

$$\mathcal{J}(\mathbf{W}) = \sum_{\tau \in \mathcal{T}} \text{off}(\mathbf{W}\mathbf{C}_\tau^z \mathbf{W}^T) \tag{2.3.15}$$

where $\mathcal{T}$ is the set of chosen lags τ and off(.) is the sum of squared off diagonal elements. The idea is based on the Essential Uniqueness of Joint Diagonalization theorem [16]:

Theorem 1. Let $\mathcal{M} = \{\mathbf{M}_1, ..., \mathbf{M}_K\}$ be a set of matrices where, for $1 \leq k \leq K$, matrix $\mathbf{M}_K$ is in the form $\mathbf{M}_K = \mathbf{U}\mathbf{D}_k\mathbf{U}^H$ with $\mathbf{U}$ a unitary matrix and $\mathbf{D}_k = diag[d_1(k), ..., d_n(k)]$. Any joint diagonalizer of $\mathcal{M}$ is essentially equal to $\mathbf{U}$ if and only if

$$\forall\ 1 \leq i \neq j \leq n \quad \exists k, 1 \leq k \leq K \quad d_i(k) \neq d_j(k). \tag{2.3.16}$$

Therefore, instead of exactly diagonalizing a single covariance matrix, the approximate joint diagonalization allows the information extracted from a set of covariance matrices to be integrated in a single unitary matrix. Generalization of the Jacobi technique to the case of multiple matrices is straightforward as in [16]. This technique consists of computing the unitary diagonalizer as a product of Givens rotations calculated at each iteration of the algorithm.

2.3.4 JADE

The other approach in ICA consists of using higher order cumulant tensors. Tensors are generalization of the matrices and hence the cumulant tensors are generalization of the covariance matrices. An idea similar to whitening the data by using eigenvalue decomposition of the covariance matrices is used which results in another class of ICA methods [62].

The fourth order cumulant tensor (sometimes called cumulant tensor for simplicity)

can be considered as a four dimensional matrix. The entries of the cumulant tensor are given by the fourth order cross cumulants of the data, i.e, $\text{cum}(x_i, x_j, x_k, x_l)$, where $1 \leq i, j, k, l \leq n$. The important property of fourth order statistics which is used in ICA methods is that if the sources are independent, all the cumulants with at least two different indices are zero.

The cumulant tensor defines a linear transformation $\mathbf{F} = [f_{ij}]$ in the space of $n \times n$ matrices. The i,jth element of the matrix given by the transformation is defined as follows:

$$f_{ij}(\mathbf{M}) = \sum_{kl} m_{kl} \text{ cum}(x_i, x_j, x_k, x_l) \tag{2.3.17}$$

where m_{kl} is the k,lth element of the matrix $\mathbf{M}$ and cum() is the fourth-order cumulant tensor. An eigenmatrix of the cumulant tensor is, by definition, a matrix $\mathbf{M}$ such that

$$\mathbf{F}(\mathbf{M}) = \lambda \mathbf{M} \tag{2.3.18}$$

that is $f_{ij}(\mathbf{M}) = \lambda m_{ij}$, where λ is a scalar eigenvalue. The transformation has n^2 and $\frac{n*(n+1)}{2}$ real eigenvalues for complex and real variables respectively [23].

This concept is used in ICA context by assuming that $\mathbf{z}$ is the whitened data, i.e.

$$\mathbf{z} = \mathbf{VAs} = \mathbf{W}^T\mathbf{s} \tag{2.3.19}$$

where $\mathbf{W}^T$ is the whitened mixing matrix. If we assume that $\mathbf{w}_m$ is the mth row of $\mathbf{W}$, it can be proved that

$$f_{ij}(\mathbf{w}_m\mathbf{w}_m^T) = w_{mi}w_{mj} \text{ kurt}(s_m) \tag{2.3.20}$$

where kurt() refers to kurtosis. That means every matrix of the form $\mathbf{M} = \mathbf{w}_m\mathbf{w}_m^T$ is an eigenmatrix of the tensor and the corresponding eigenvalues are given by the kurtoses of the independent components. Moreover, it can be proven that all other eigenvalues of the tensor are zero. If the eigenmatrices of the cumulant tensor are known, the independent

components can be identified easily. If the eigenvalues of the tensor, i.e. the kurtoses of the independent components, are distinct, every eigenmatrix will provide one of the columns of the whitened mixing matrix. However, in practice the eigenvalues may not be distinct and therefore, the eigenmatrices are the linear combinations of the matrices $\mathbf{w}_m\mathbf{w}_m^T$.

Different methods have been proposed for eigenvalue decomposition of cumulant tensors, e.g., power methods. JADE is one of the well established methods to solve the problem of degenerate eigenvalues of the cumulant tensor. The eigenvalue decomposition can be viewed as diagonalization, therefore, assuming that the ICA model holds the problem is approached by assuming that the matrix $\mathbf{W}$ is the separating matrix. In this case, $\mathbf{W}$ diagonalizes $\mathbf{F(M)}$, i.e, $\mathbf{WF(M)W}^T$ is diagonal for any $\mathbf{M}$. This is because matrix $\mathbf{F}$ is a linear combination of the eigenmatrix terms $\mathbf{w}_m\mathbf{w}_m^T$.

Thus, we have to choose a set of different matrices $\mathbf{M}_i$ and diagonalize the matrices $\mathbf{WF(M}_i\mathbf{)W}^T$ as much as possible. Obviously, because of sampling errors and uncertainties in the model it is not possible to exactly diagonalize the set of matrices. The best choice of the matrices $\mathbf{M}_i$ is the eigenmatrices of the tensor matrix of the whitened data. The first n significant eigenpairs are selected.

The diagonality of the matrices $\mathbf{Q} = \mathbf{WF(M}_i\mathbf{)W}^T$ can be measured using the following cost function, which is equal to off(.) operator:

$$\mathcal{J}_1(\mathbf{W}) = \sum_{k \neq l} q_{kl}^2 \tag{2.3.21}$$

Extended Jacobi technique for simultaneous diagonalization method [25] is used to jointly diagonalize the matrices $\mathbf{WF(M}_i\mathbf{)W}^T$ [24].

Although JADE provides competitive performance for low dimensional data, it can not perform well for high dimensional spaces [62].

2.4 Periodic component analysis

Periodic component analysis method (πCA), has been originally developed to enhance the periodic signature of the speaker's pitch in speech processing applications. In this method, it is assumed that the speech source is periodic at some pitch period $\tau = 1/f_0$ where f_0 is the fundamental frequency [91]. A linear combination $\hat{s}(t) = \mathbf{w}\mathbf{x}(t) = \sum_i w_i x_i(t)$ is sought such that the following cost function which is a measure of periodicity has minimum value at some τ:

$$\epsilon(\mathbf{w}, \tau) = \frac{\sum_t |\hat{s}(t+\tau) - \hat{s}(t)|^2}{\sum_t |\hat{s}(t)|^2} \tag{2.4.1}$$

This cost function measures the normalized prediction error. To compensate the phase changes across the channels Hilbert transform is applied to $x_i(t)$ resulting in an analytic function $\tilde{x}_i(t)$. The Fourier series of these signals are related by:

$$x_i(t) = \sum_k \alpha_k \cos(\omega_k t + \phi_k) \quad \Longleftrightarrow \quad \tilde{x}_i(t) = \sum_k \alpha_k e^{j(\omega_k t + \phi_k)} \tag{2.4.2}$$

Substituting $x_i(t)$ with $\tilde{x}_i(t)$ in (2.4.1) and simplifying the relations the cost function generalizes in a straightforward way to:

$$\epsilon(\mathbf{w}, \tau) = \frac{\sum_{ij} w_i^\star w_j a_{ij}(\tau)}{\sum_{ij} w_i^\star w_j b_{ij}} \tag{2.4.3}$$

where $\star$ represents the complex conjugate of a variable, and $\mathbf{A}(\tau) = [a_{ij}(\tau)]$ and $\mathbf{B} = [b_{ij}]$ are Hermitian matrices with complex elements

$$a_{ij}(\tau) = \sum_t [\tilde{x}_i^\star(t)\tilde{x}_j(t) + \tilde{x}_i^\star(t+\tau)\tilde{x}_j(t+\tau) - \tilde{x}_i^\star(t)\tilde{x}_j(t+\tau) - \tilde{x}_i^\star(t+\tau)\tilde{x}_j(t)]$$

and

$$b_{ij} = \sum_t \tilde{x}_i^\star(t)\tilde{x}_j(t)$$

Using the Rayleigh-Ritz theorem of linear algebra the weight vector minimizing (2.4.3) is given by the eigenvector corresponding to the smallest eigenvalue of the matrix $\mathbf{B}^{-1}\mathbf{A}(\tau)$. The pitch periods are estimated by first diagonalizing the matrices $\mathbf{B}^{-1}\mathbf{A}(\tau)$ on sliding windows of $\tilde{x}_i$ over a range of pitch periods $\tau \in [\tau_{min}, \tau_max]$ where τ_{min} and τ_{max} are selected based on *a priori* knowledge. The pitch period is estimated by the value of τ that minimizes the cost function for each sliding window.

In [89] generalized eigenvalue decomposition (GEVD) is used to approach the separation of periodic sources in a different way. The problem of GEVD of a matrix pair (A,B) (for all $n \times n$ symmetric matrices A and B) consists of finding matrices U and D such that:

$$\mathbf{U}^T\mathbf{A}\mathbf{U} = \mathbf{D} \quad and \quad \mathbf{U}^T\mathbf{B}\mathbf{U} = \mathbf{I} \tag{2.4.4}$$

where D is the diagonal generalized eigenvalue matrix corresponding to the eigenvector matrix U. If A and B are symmetric and positive-definite matrices, eigenvalues of D are real and positive.

The problem of periodic component analysis is then restated by modifying the cost function (2.4.1) as:

$$\epsilon(\mathbf{w}, \tau) = \frac{\mathbf{w}^T\mathbf{A}_x(\tau)\mathbf{w}}{\mathbf{w}^T\mathbf{C}_x(0)\mathbf{w}} = 2[1 - \frac{\mathbf{w}^T\mathbf{C}_x(\tau)\mathbf{w}}{\mathbf{w}^T\mathbf{C}_x(0)\mathbf{w}}] \tag{2.4.5}$$

where $\mathbf{A}_x(\tau) = E_t\{[\mathbf{x}(t+\tau) - \mathbf{x}(t)][\mathbf{x}(t+\tau) - \mathbf{x}(t)]^T\} = 2\mathbf{C}_x(0) - 2\mathbf{C}_x(\tau)$ and $\mathbf{C}_x(\tau)$ is the covariance matrix of x(t) at time delay τ.

The weight vector w minimizing (2.4.5) is given by the eigenvector corresponding to the smallest generalized eigenvalue of the pair ($\mathbf{A}_x(\tau)$, $\mathbf{C}_x(0)$), or equivalently, the largest generalized eigenvalue of ($\mathbf{C}_x(\tau)$, $\mathbf{C}_x(0)$).

Using (2.4.4) and assuming D as the diagonal generalized eigenvalue matrix corresponding to the eigenvector matrix U, with real eigenvalues sorted in descending order

on its diagonal, the transformation $\mathbf{U}^T\mathbf{x}(t)$ gives the most periodic components in descending order of periodicity.

Fundamental frequency estimation

The simplest method to analyze periodicity, as done in [91], is to compute the autocorrelation of the signal. For nonstationary signals such as speech the autocorrelation can be calculated over a sliding window. In this case the peaks in the autocorrelation function provide estimates of the fundamental frequency.

For those cases that there might be some variations in the periodicity of the signals, e.g., ECG recordings, a constant τ can not provide enough information about the periodicity of the signal. In [89] a method is presented to utilize the time varying period of the ECG data. By detecting the R-peaks of the ECG, a linear phase $\phi(t)$ ranging from $-\pi$ to π is assigned to each ECG sample, with the R-peak being fixed at $\phi(t) = 0$. The linear phase $\phi(t)$ provides a means for phase-wrapping the RR-interval onto the $[-\pi, \pi]$ interval. Therefore, the ECG may be converted to a polar representation in which the ECG components in different beats, such as the P, Q, R, S, and T waves, are more or less phase-aligned with each other, especially over the QRS segment.

2.5 Blind source extraction using oblique projectors

In [99] and [100] blind source extraction (BSE) is performed using approximate joint diagonalization (AJD). It is assumed that the sources are spatially uncorrelated and wide sense stationary. The time lagged autocorrelation matrix is defined as:

$$\mathbf{R}(\tau_k) = E\{\mathbf{x}(t)\mathbf{x}^T(t-\tau_k)\}, \quad k = 1, 2, ..., K \tag{2.5.1}$$

where $\mathbf{x}(t)$ is the observation vector and K is the index of maximum time lag.

We know that the vector $x_p(t)$ (for all $1 < p < n$) is a linear combination of the

columns of matrix $\mathbf{A}$ (the mixing matrix). Therefore, to extract the pth source one can project $\mathbf{x}(t)$ onto a subspace in $\mathbb{R}^n$ orthogonal to all columns of $\mathbf{A}$ except $\mathbf{a}_p$, i.e., $\{\mathbf{a}_1, ..., \mathbf{a}_{p-1}, \mathbf{a}_{p+1}, ..., \mathbf{a}_n\}$. Suppose $\mathbf{t} \equiv \mathbf{a}_p$ and $\mathbf{v}$ is a vector orthogonal to the space spanned by columns of $\mathbf{A}$ excluding $\mathbf{a}_p$, i.e.

$$\mathbf{v} \perp \{\mathbf{a}_1, ..., \mathbf{a}_{p-1}, \mathbf{a}_{p+1}, ..., \mathbf{a}_n\}.$$

Using oblique projection notation [14] the following relation is obtained:

$$y(t)\mathbf{t} = \mathbf{E}_{t|\mathbf{v}^\perp}\mathbf{x}(t) \tag{2.5.2}$$

where $y(t)$ is an estimation of the source of interest, $\mathbf{v}^\perp$ is a subspace in $\mathbb{R}^n$ perpendicular to $\mathbf{v}$ and $\mathbf{E}_{t|\mathbf{v}^\perp} = (\mathbf{t}\mathbf{v}^T)/(\mathbf{v}^T\mathbf{t})$ is the oblique projector onto direction $\mathbf{t}$ along $\mathbf{v}^\perp$. By omitting the scalar $1/(\mathbf{v}^T\mathbf{t})$ and dropping $\mathbf{t}$ from both sides of (2.5.2) we will have:

$$y(t) = \mathbf{v}^T\mathbf{x}(t) \tag{2.5.3}$$

In blind source extraction based on second order statistics, both vectors $\mathbf{t}$ and $\mathbf{v}$ are unknown and they both have to be estimated. The following criterion is exploited to find the unknown parameters:

$$[\hat{\mathbf{t}}, \hat{\mathbf{v}}, \hat{\mathbf{d}}] = arg \min_{\mathbf{t},\mathbf{v},\mathbf{d}} \mathcal{J}(\mathbf{t}, \mathbf{v}, \mathbf{d}) \tag{2.5.4}$$

where the cost function $\mathcal{J}(\mathbf{t}, \mathbf{v}, \mathbf{d})$, is defined as:

$$\mathcal{J}(\mathbf{t}, \mathbf{v}, \mathbf{d}) = \sum_{k=1}^{K} \|\mathbf{R}(\tau_k)\mathbf{v} - d_k\mathbf{t}\|^2 \tag{2.5.5}$$

and $\mathbf{d} = [d_1, d_2, ..., d_K]^H$ is the vector of unknown scalars.

The cost function in (2.5.5) utilizes the fact that for blind source extraction $\mathbf{R}(\tau_k)\mathbf{v}$

should be collinear with $\mathbf{t}$ incorporating the coefficients d_k which provide $\mathbf{t}$ with proper scaling. It is clear that immediate global minimum point of the cost function (2.5.5) is $\mathbf{t} = \mathbf{v} = \mathbf{d} = \mathbf{0}$. This solution has been avoided by imposing the condition $\|\mathbf{t}\| = \|\mathbf{d}\| = 1$.

Minimization of the cost function (2.5.5) with respect to $\mathbf{v}$ leads to the identification of vector $\mathbf{v}$ in equation (2.5.3) which can thereby be used to extract one of the sources.

The sequential approximate diagonalisation algorithm (SDA) proposed in [71], is used to optimize the cost function (2.5.5) by adjusting the parameters sequentially. After extracting one source a deflation process is carried out to remove the extracted source from the mixed signals [28]:

$$\mathbf{x}_{i+1}(t) = \mathbf{Z}^T\mathbf{x}(t), \quad \mathbf{x}_1(t) = \mathbf{x}(t) \tag{2.5.6}$$

where $\mathbf{x}(t)$ is the original observation signal and

$$\mathbf{Z} = \mathbf{I} - \frac{\mathbf{R}_i(0)\mathbf{w}\mathbf{w}^T}{\sigma_y^2} \tag{2.5.7}$$

and $\mathbf{R}_i(0) = \boldsymbol{E}\{\mathbf{x}_i(t)\mathbf{x}_i^T(t)\}$, $\mathbf{I}$ is an $n \times n$ identity matrix, and $\sigma_y^2 = \boldsymbol{E}\{y^2\}$. The autocorrelation matrix has to be updated by the following relation:

$$\mathbf{R}_{i+1}(0) = \mathbf{Z}^T\mathbf{R}_i(0)\mathbf{Z} \tag{2.5.8}$$

This algorithm can extract the source signals one by one. While we are interested in extracting just a source of interest (SoI) this algorithm may not help unless some *a priori* information about SoI is known.

If the fundamental period of the SoI is, τ samples, then its autocorrelation matrix will have the same value at time lags corresponding to integer multiples of τ.

Therefore, the autocorrelation matrices $\mathbf{R}(\tau_k) = E\{\mathbf{x}(t)\mathbf{x}^T(t - \tau_k)\}$ can be jointly diagonalized at time lags τ, 2τ, ..., $K\tau$ along with the constraint $d_1 = d_2 = ... = d_K$.

However, if the source of interest has a varying period this extraction algorithm will result in erroneous results [100].

2.6 Cyclostationary BSS

By definition, stationary signals represent physical phenomena that maintain a constant statistical behaviour in time. This type of signals are conceptually well established and plenty of techniques and tools are developed to process and analyze them.

However, this property is hardly met by some systems such as biomedical and mechanical systems which undergo a nonstationary operation. Even under constant conditions, a succession of phenomena can happen within continuous cycles which causes the system to release energy on a rhythmic basis. Such phenomena typically produce transient signatures in the signals, which are likely to carry some critical information about the operating condition of the system. It is emphasized in [11] that nonstationarity is intimately related to the concept of information. As an example speech or music signals can carry information because they consist of nonstationary sequences.

In practice it is very difficult to trace the nonstationarities included in the signals acquired from real world applications. Therefore, it is usually assumed that the signals are stationary and the stationary signal processing methods are utilized to analyze and process them.

There are many processes in nature that originate from periodic phenomena. Although these processes may not have periodic behaviour in time, give rise to random data whose statistical characteristics vary periodically with time and are called cyclostationary processes [43]. Most of the pioneer works on cyclostationary signals have been limited to communications research and applications in which the modulation of signals for transmission purposes results in cyclostationary behaviour. By further attempts such as in [11] and [40] the cyclostationary framework was generalized and it was shown that

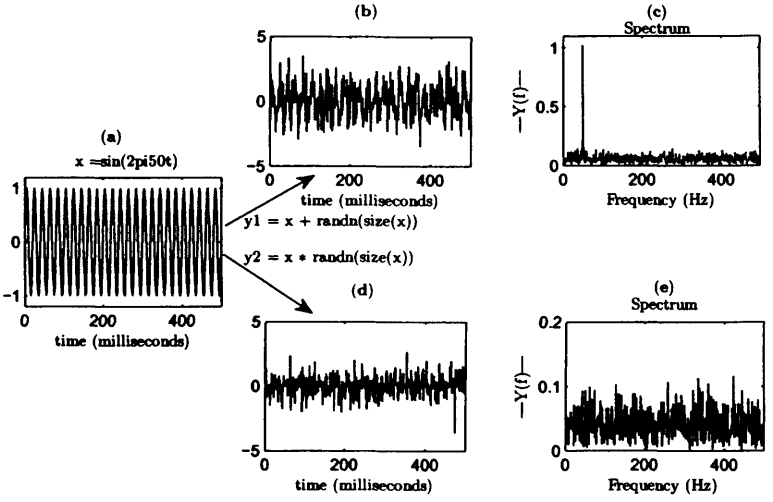


Figure 2.2: Addition and modulation of a sine wave with noise. Using Fourier transform one can identify the periodicity hidden in the signal with additive noise. This technique is not working for the modulated signal.

it is a powerful tool for analyzing any signal originated from phenomena with periodic statistical behaviour. Recently, different methods for cyclostationary signal processing have been developed and used commonly in different fields including telecommunications, radar and sonar applications, mechanics, econometrics and biological and atmospheric science [43].

It is shown in [38] that even successful BSS methods such as JADE [24] that are derived on the assumption that the source signals are zero mean and statistically stationary, often do not provide good results when the sources are cyclostationary. The authors have shown that the estimator that is based on higher-order statistics generates biased estimates when the original data has a cyclostationarity property.

Different approaches for defining the concept of cyclostationarity are provided. In [11], a cyclostationary signal is defined to be the one that exhibits some hidden periodicity in its energy flow. Periodic activity of some parts of the systems may cause periodic modulation of the other signals or noise in the system. As an example, a simple sine

wave of $x(t) = \sin(2\pi 50t)$ is presented in Fig 2.2.(a). The time and frequency representations of $x(t)$ with an additive random Gaussian noise are illustrated in Fig. 2.2.(b) and 2.2.(c), respectively. Although, the periodicity of the signal is not recognizable from the time domain representation of the signal, applying Fourier transform can help to identify the periodicity of the signal as a high peak at 50 Hz in the spectrum of the signal. The signal in Fig 2.2.(b) is said to have first order hidden periodicity [40]. On the other hand, if the sine wave is amplitude modulated with the noise, the resulted signal will be different from the previous case. The time and frequency domain representations of the result are shown in Fig 2.2.(d) and 2.2.(e). In the case of modulated signal Fourier transform can not help in identifying the periodic structure of the underlying sources. Signal $y_2(t)$ has second order hidden periodicity [40].

Gardner has defined the cyclostationary signals in two ways [42]. Based on the first definition a signal is cyclostationary of order n in wide sense if and only if an nth order nonlinear transformations of the signal can be found which generate finite strength additive sine wave components. These components yield spectral lines, e.g. for $n = 2$ the squared signal or the product of the signal with a delayed version of itself generates spectral lines. For stationary signals only one spectral line at frequency zero is generated.

The second definition in [42] is that a signal is cyclostationary of order n in wide sense if and only if the time fluctuations in n spectral bands with center frequencies that sum to certain discrete nonzero values are statistically dependent[1]. For stationary signals, only those bands whose centre frequencies sum to zero have statistical dependence.

It is shown in [42] that the two definitions are completely equivalent. For the simple case of $n = 2$ a signal $x(t)$ is cyclostationary with cycle frequency β if and only if for some τ some of the delayed products of the form $y(t) = x(t)x(t-\tau)$ show a spectral line at frequency β (For complex signals $y(t) = x(t)^* x(t-\tau)$ can be used). On the other hand, $x(t)$ is cyclostationary if and only if the time fluctuations in at least some pairs of

[1]To demonstrate dependency it is enough to show that their nth order moment is nonzero.

spectral bands of $x(t)$, whose two centre frequencies sum (or difference for the complex case) to β, are correlated.

In this subsection we will first introduce cyclostationarity in mathematical terms and then different approaches that exploit this property to separate the sources will be reviewed.

2.6.1 Cyclostationarity theory

Let's define a continuous time real valued stochastic process $\{x(t,\omega), t \in \mathbb{R}, \omega \in \Omega\}$ where Ω is the sample space. We just use abbreviated notation $x(t)$ when it does not create ambiguity. The process $x(t)$ is said to be *Nth-order cyclostationary in the strict sense* if its Nth-order distribution function

$$F_{x(t+\tau_1)\ldots x(t+\tau_{N-1})x(t)}(\xi_1, \ldots, \xi_{N-1}, \xi_N) \triangleq P\{x(t+\tau_1) \leqslant \xi_1, \ldots x(t+\tau_{N-1}) \leqslant \xi_{N-1}, x(t) \leqslant \xi_N\} \tag{2.6.1}$$

is periodic in t with some period T_0:

$$F_{x(t+\tau_1+T_0)\ldots x(t+\tau_{N-1}+T_0)x(t+T_0)}(\xi_1, \ldots, \xi_{N-1}, \xi_N) = F_{x(t+\tau_1)\ldots x(t+\tau_{N-1})x(t)}(\xi_1, \ldots, \xi_{N-1}, \xi_N) \tag{2.6.2}$$

$$\forall t \in \mathbb{R}, \forall (\tau_1, \ldots, \tau_{N-1}) \in \mathbb{R}^{N-1}, \forall (\xi_1, \ldots, \xi_N) \in \mathbb{R}^N$$

The mean of a process is defined by

$$E\{x(t)\} = \int_{-\infty}^{+\infty} x f(x) dx \tag{2.6.3}$$

where $f(x)$ is the probability density function of $x(t)$. If the mean of $x(t)$ varies with a period, say T_0, i.e.,

$$E\{x(t+T_0)\} = E\{x(t)\} \tag{2.6.4}$$

then, the process is referred to as *first order cyclostationary in the wide sense.* The process $x(t)$ is said to be *second order cyclostationary in the wide sense* if its mean (2.6.3) and autocorrelation function

$$\mathcal{R}_x(t,\tau) \triangleq E\{x(t)x(t+\tau)\} \tag{2.6.5}$$

are periodic with some period T_0:

$$E\{x(t+T_0)\} = E\{x(t)\} \tag{2.6.6}$$

$$\mathcal{R}_x(t+T_0,\tau) = \mathcal{R}_x(t,\tau) \tag{2.6.7}$$

for all t and τ [43]. If we assume that the Fourier series expansion of $\mathcal{R}_x(t,\tau)$ converges to $\mathcal{R}_x(t,\tau)$, then, we can write:

$$\mathcal{R}_x(t,\tau) = \sum_{n=-\infty}^{+\infty} \mathcal{R}_x^{n/T_0}(\tau) e^{j2\pi(n/T_0)t}, \tag{2.6.8}$$

where the Fourier coefficients are defined as:

$$\mathcal{R}_x^{n/T_0}(\tau) \triangleq \frac{1}{T_0} \sum_{n=-T_0/2}^{T_0/2} \mathcal{R}_x(t,\tau) e^{-j2\pi(n/T_0)t} dt \tag{2.6.9}$$

and $j = \sqrt{-1}$.

The Fourier coefficients $\mathcal{R}_x^{n/T_0}(\tau)$ are called *cyclic autocorrelation functions* and the frequencies $\{n/T_0\}$ (for $n \in \mathbb{Z}$) are called *cycle frequencies.*

Let us consider this type of process in more detail. A function $y(t)$ is said to be *almost periodic* if it is a limit of a uniformly convergent sequence of trigonometric polynomials

in t:

$$y(t) = \sum_{\beta \in A} y_\beta e^{j2\pi\beta t} \tag{2.6.10}$$

$$y_\beta \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} y(t) e^{-j2\pi\beta t} dt \tag{2.6.11}$$

where A is a countable set and βs are incommensurate.

If the autocorrelation function $\mathcal{R}_x(t,\tau)$ is almost periodic in t (for each τ) a more general class of stochastic processes is obtained. A process $x(t)$ is said to be *Nth order almost cyclostationary in the strict sense* if its Nth order distribution function (2.6.1) is almost periodic in time. Using properties of the autocorrelation function, if the autocorrelation function of a stochastic process $x(t)$ is an almost periodic function of t, then, we call it *almost cyclostationary in the wide sense.* In this case $\mathcal{R}_x(t,\tau)$ is a limit of convergent sequence of trigonometric polynomials in t:

$$\mathcal{R}_x(t,\tau) = \sum_{\beta \in A} \mathcal{R}_x^\beta(\tau) e^{j2\pi\beta t} \tag{2.6.12}$$

where

$$\mathcal{R}_x^\beta(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \mathcal{R}_x(t,\tau) e^{-j2\pi\beta t} dt \tag{2.6.13}$$

is the *cyclic autocorrelation function* at cycle frequency β. It is shown in [40] that if $x(t)$ is an almost cyclostationary process, then this process and its frequency-shifted version $x(t)e^{j2\pi\beta t}$ are correlated for all $\beta \in A$. Consequently if β is not in A, then $x(t)$ and $x(t)e^{j2\pi\beta t}$ are not correlated and thus the cyclic autocorrelation function will be zero for such frequency β. This property as will be seen in the rest of this document can be used for blindly separating and extracting almost cyclostationary signals which in this document are referred to as cyclostationary signals. In [43] it is shown that cyclic autocorrelation function at cycle frequency β ($\beta \in \mathbb{R}$) is:

$$\mathcal{R}_x^{\beta}(\tau) \triangleq \langle x(t)x(t+\tau)e^{-j2\pi\beta t}\rangle \tag{2.6.14}$$

where $\langle . \rangle$ is the time averaging operator. For the case of complex-valued processes, one may define the same model for cyclostationarity. If $x(t)$ is a complex-valued process then, the cyclic autocorrelation function is defined as [43]:

$$\mathcal{R}_x^{\beta}(\tau) \triangleq \langle x(t)x^*(t+\tau)e^{-j2\pi\beta t}\rangle. \tag{2.6.15}$$

Finding hidden periodicities

The periodic behaviour of some periodic signals can be identified simply by inspecting the temporal waveform. However, this is not the case for all the signals, specially if the signal is mainly random in nature whereby even conventional spectral analysis may not help. One solution is to investigate how the signal energy is propagating with time.

Let $x_{\Delta f}(t, f)$ be the filtered version of the signal $x(t)$ through a frequency band of width Δf centred at frequency f. The average energy in this frequency band is

$$P_x(f, \Delta f) = \lim_{T\to\infty} \frac{1}{T} \int_T |x_{\Delta f}(t, f)|^2 dt \tag{2.6.16}$$

For a random signal with no pure component at frequency f the value of energy tends to zero as the bandwidth Δf of the filter tends to zero. A density of energy flow per frequency can be measured, however, by taking the limit of the ratio $P_x(f, \Delta f)/\Delta f$:

$$P_x(f) = \lim_{\Delta f\to 0} \frac{P_x(f, \Delta f)}{\Delta f} = \lim_{\Delta f\to 0} \lim_{T\to\infty} \frac{1}{T.\Delta f} \int_T |x_{\Delta f}(t, f)|^2 dt \tag{2.6.17}$$

$P_x(f)$ is called power spectral density (PSD) and is a measure of how the energy of the signals is distributed in the frequency domain.

The PSD of a random signal which does not contain any periodic waveform will

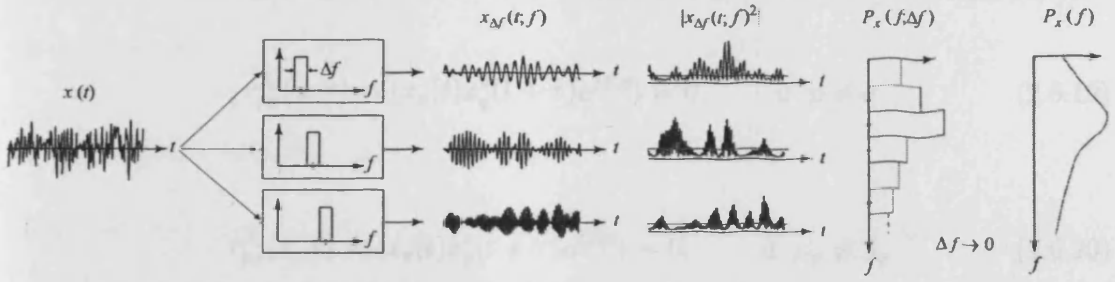


Figure 2.3: Interpretation of the power spectral density as the averaged power measured at the output of a filter bank; taken from [11].

be a (piecewise) continuous function of frequency. On the other hand, presence of a pure periodic component of amplitude a and frequency f will produce a peak at that frequency with magnitude $\lim_{\Delta f \to 0} \frac{a^2}{4\Delta f}$. Therefore, a common practice to investigate whether a signal of interest contains periodic components at a typical frequency f is checking for the presence of peaks in the PSD of the signal.

This concept is illustrated in Fig 2.3.

Cyclic correlation matrix

In BSS problems we often have a number of sources that are mixed thorough an unknown medium. Therefore, the aim is estimation of the unmixing matrix and the original sources. In cyclostationary blind source separation, we model the sources as cyclostationary stochastic processes and thus we can use the properties of cyclic correlation functions. Suppose we have a vector $\mathbf{x}(t)$ that is defined in $\mathbb{R}^n$. We define *Cyclic cross-correlation function* of signals $x_p(t)$ and $x_q(t)$ at frequency β to be:

$$r_{pq}^{\beta}(\mathbf{x},\tau) = \langle x_p(t)x_q^{\star}(t+\tau)e^{j\beta t}\rangle \triangleq \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1} x_p(t)x_q^{\star}(t+\tau)e^{j\beta t} \tag{2.6.18}$$

If $p = q$ and β is the cycle frequency of source x_p, $r_{pq}^{\beta}(\mathbf{x},\tau)$ is referred to as cyclic autocorrelation and is denoted by $\rho_p(\mathbf{x},\tau)$. If $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_n(t)]^T$ is a vector

of n independent signals, cyclic cross-correlation function has the following properties:

$$r_{pq}^{\beta_p}(\mathbf{x},\tau) = \langle x_p(t)x_q^{\star}(t+\tau)e^{j\beta_p t}\rangle = 0, \qquad \text{if } p \neq q \tag{2.6.19}$$

$$r_{pp}^{\beta_q}(\mathbf{x},\tau) = \langle x_p(t)x_p^{\star}(t+\tau)e^{j\beta_q t}\rangle = 0, \qquad \text{if } \beta_p \neq \beta_q \tag{2.6.20}$$

$$r_{pp}^{\beta_p}(\mathbf{x},0) = \rho_p(\mathbf{x},0) = \langle x_p(t)x_p^{\star}(t)e^{j\beta_p t}\rangle \neq 0, \qquad \forall\ p \tag{2.6.21}$$

Accordingly, cyclic correlation matrix of the $n \times 1$ complex signal vector $\mathbf{x}(t)$ is defined as:

$$\mathbf{R}_x^{\beta_p}(t,\tau) \triangleq \langle \mathbf{x}(t)\mathbf{x}^H(t+\tau)e^{j\beta_p t}\rangle \tag{2.6.22}$$

where superscript H denotes complex conjugate transpose of a matrix. For the sake of simplicity, the time index t is ignored when appropriate and $\mathbf{R}_x^{\beta_p}(\tau)$ is used instead. This matrix is used widely in cyclostationary BSS. If we assume that $\rho_p(\mathbf{x},0) = 1$, based on the properties (2.6.19)-(2.6.21), it is clear that cyclic correlation matrix of n independent sources $\mathbf{s}(t)$, satisfies the following relation [64]:

$$\mathbf{R}_s^{\beta_p}(0) = \mathbf{I}_p^{'} \tag{2.6.23}$$

where

$$[\mathbf{I}_p^{'}]_{l,g} = \begin{cases} 1 & \text{if } l \in \{1,2,...,m\}, g = l = p \\ 0 & \text{otherwise} \end{cases} \tag{2.6.24}$$

2.6.2 A review of cyclostationary BSS methods

Different methods have been proposed for separating (or extracting) cyclostationary sources from multi channel data. In most of them it is assumed that the cycle frequencies

are known *a priori*. In this section some of the most well-known cyclostationary BSS methods are reviewed.

Natural gradient algorithm

Based on the concept of blind source separation which is a two-stage process, including prewhitening and rotation, a cyclostationary natural gradient algorithm is constructed in [64]. It is assumed that $\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{n}(k)$ is the observation vector and $\mathbf{y}(k) = \mathbf{W}^{(k)}\mathbf{x}(k)$ is the output vector, where $\mathbf{W}^{(k)}$ is the estimation of the mixing matrix at kth iteration. Considering the cyclic frequency of all the sources a new cyclic correlation matrix function is defined as:

$$\begin{aligned}\tilde{\mathbf{R}}_s^{\beta}(k,0) &= E\{\mathbf{s}(k)\mathbf{s}^H(k)e^{j\beta_1 t}\rangle + ... + \langle \mathbf{s}^H(k)\mathbf{s}(k)e^{j\beta_m k}\rangle \\ &= \sum_{i=1}^{m}\langle \mathbf{s}^H(k)\mathbf{s}(k)e^{j\beta_i k}\rangle \\ &= \sum_{i=1}^{m}\mathbf{I}_i' = \mathbf{I}_m\end{aligned} \tag{2.6.25}$$

where $\mathbf{I}_i'$ is defined as (2.6.24) and $\mathbf{I}_m$ is the m-dimensional identity matrix. The mixture and output cyclic correlation matrices at time k and delay $\tau = 0$ are

$$\mathbf{R}_x^{\beta_p}(k,0) = E\{\mathbf{x}(k)\mathbf{x}^H(k)e^{j\beta_p k}\} \tag{2.6.26}$$

$$\mathbf{R}_y^{\beta_p}(k,0) = E\{\mathbf{y}(k)\mathbf{y}^H(k)e^{j\beta_p k}\} \tag{2.6.27}$$

We can rewrite (2.6.26) and (2.6.27) as:

$$\begin{aligned}\mathbf{R}_x^{\beta_p}(k,0) &= \mathbf{A}E\{\mathbf{s}(k)\mathbf{s}^H(k)e^{j\beta_p k}\}\mathbf{A}^H \\ &= \mathbf{A}\mathbf{R}_s^{\beta_p}(k,0)\mathbf{A}^H\end{aligned} \tag{2.6.28}$$

$$\begin{aligned}\mathbf{R}_y^{\beta_p}(k,0) &= \mathbf{W}^{(k)} E\{\mathbf{x}(k)\mathbf{x}^H(k)e^{j\beta_p k}\}\mathbf{W}^{(k)^H} \\ &= \mathbf{W}^{(k)}\mathbf{R}_x^{\beta_p}(k,0)\mathbf{W}^{(k)^H}\end{aligned} \tag{2.6.29}$$

The main idea behind the natural gradient cyclostationary blind source separation method is to decorrelate the estimated sources in the cyclostationary sense. To follow the original approach of natural gradient algorithm it is assumed that all the source signals and the mixing matrix have real values [64]. Let

$$\tilde{\mathbf{R}}_y^{\beta_p}(k,0) = \sum_{p=1}^{m} E\{\mathbf{y}(k)\mathbf{y}^H(k)e^{j\beta_p k}\} = \sum_{p=1}^{m} R_y^{\beta_p}(k,0) \tag{2.6.30}$$

Then, in the limit, as $k \to \infty$, each of the output cyclic correlation matrices $\mathbf{R}_y^{\beta_p}(k,0)$ converges to a matrix with only one non-zero entry placed at the pth position along the main diagonal. Consequently, the sum of all correlation matrices converges to the identity matrix:

$$\lim_{K\to\infty} \tilde{\mathbf{R}}_y^{\beta}(k,0) = \mathbf{I}_m \tag{2.6.31}$$

Using (2.6.31), we can rewrite (2.6.29) as:

$$\begin{aligned}\tilde{\mathbf{R}}_y^{\beta}(k,0) &= \sum_{p=1}^{m} \mathbf{W}^{(k)}\mathbf{R}_x^{\beta_p}(k,0)\mathbf{W}^{(k)^T} \\ &= \mathbf{W}^{(k)}\left(\sum_{p=1}^{m}\mathbf{R}_x^{\beta_p}(k,0)\right)\mathbf{W}^{(k)^T} \\ &= \mathbf{W}^{(k)}\tilde{\mathbf{R}}_x^{\beta_p}(k,0)\mathbf{W}^{(k)^H}\end{aligned} \tag{2.6.32}$$

Therefore, the following cost function is defined that represents the sum of Kullback-Leibler divergence between the output joint probability density function and the product

of the marginal probability distribution functions, and the distance between the two normal distributions with covariance matrices $\tilde{\mathbf{R}}_y^{\beta}$ and $\mathbf{I}_m$:

$$\begin{aligned}\mathcal{KL}(\mathbf{W}^{(k)}) = &- \log\det(\mathbf{W}^{(k)}) - \sum_{p=1}^{m} \log q_p(y_p(k)) \\ &+ \frac{1}{2}Tr(\tilde{\mathbf{R}}_y^{\beta}(k,0)) - \frac{1}{2}\log\det(\tilde{\mathbf{R}}_y^{\beta}(k,0)) - \frac{m}{2}\end{aligned} \tag{2.6.33}$$

where q_p is the true or estimated probability distribution function of source $s_p(t)$ and is used as a reference. Applying the stochastic gradient descent, the gradient of (2.6.33) is:

$$\begin{aligned}\frac{\partial\mathcal{KL}(\mathbf{W}^{(k)})}{\partial\mathbf{W}^{(k)}} =&\mathbf{f}(\mathbf{y}(k))\mathbf{x}^T(k) - (\mathbf{W}^{(k)^{-1}})^T \\ &+ \mathbf{W}^{(k)}\tilde{\mathbf{R}}_x^{\beta}(k,0) - (\mathbf{W}^{(k)^{-1}})^T\end{aligned} \tag{2.6.34}$$

where $\mathbf{f}(\mathbf{y}(k))$ is an odd nonlinear function of the output vector $\mathbf{y}(t)$ [9]. In order to obtain natural gradient of the cost function (2.6.34) is post-multiplied by $\mathbf{W}^{(k)^T}\mathbf{W}^{(k)}$, which results is:

$$\frac{\partial\mathcal{KL}(\mathbf{W}^{(k)})}{\partial\mathbf{W}^{(k)}}\mathbf{W}^{(k)^T}\mathbf{W}^{(k)} = \left[\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) - \mathbf{I} + \tilde{\mathbf{R}}_y^{\beta}(k,0) - \mathbf{I}\right]\mathbf{W}^{(k)} \tag{2.6.35}$$

Thus, the update equation can be expressed by:

$$\begin{aligned}\mathbf{W}^{(k+1)} &= \mathbf{W}^{(k)} - \hat{\mu}\frac{\partial\mathcal{KL}(\mathbf{W}^{(k)})}{\partial\mathbf{W}^{(k)}}\mathbf{W}^{(k)^T}\mathbf{W}^{(k)} \\ &= \mathbf{W}^{(k)} + \hat{\mu}\left[\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) + \mathbf{I} - \tilde{\mathbf{R}}_y^{\beta}(k,0)\right]\mathbf{W}^{(k)}\end{aligned} \tag{2.6.36}$$

where $\hat{\mu}$ is the step size parameter. In [64] (2.6.36) is compared with Natural Gradient update equation

$$\mathbf{W}^{(k+1)} = \mathbf{W}^{(k)} + \mu\left[\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)\right]\mathbf{W}^{(k)}. \tag{2.6.37}$$

Equation (2.6.36) contains second and higher order conventional deconvolution terms as

in (2.6.37) plus second order cyclic deconvolution terms.

The output cyclic correlation matrices $\mathbf{R}_y^{\beta_p}(k,0)$, $p \in \{1,2,...,m\}$ are estimated at each iteration using an exponentially weighted average of the instantaneous statistics

$$\hat{\mathbf{R}}_y^{\beta_p}(k+1,0) = (1-\lambda)\hat{\mathbf{R}}_y^{\beta_p}(k,0) + \lambda[\mathbf{y}(k)\mathbf{y}^H(k)e^{j\beta_p k}] \tag{2.6.38}$$

where λ is the Lagrangian multiplier.

BSS using cyclostationary statistics

In [1] it is assumed that m sources impinge on an array of n sensors, where $n \geq m$ as the general formulation of BSS:

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{n}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{2.6.39}$$

As it is mentioned in (2.6.19)-(2.6.21) a vector of zero mean mutually independent cyclostationary sources $\mathbf{s}(t) = [s_1(t),...,s_m(t)]^T$, satisfies the following conditions:

$$r_{pq}^{\beta_p}(\mathbf{s},\tau) = \langle s_p(t)s_q^*(t+\tau)e^{j\beta_p t}\rangle = 0, \qquad if p \neq q \tag{2.6.40}$$

$$r_{pq}^{\beta_q}(\mathbf{s},\tau) = \langle s_p(t)s_p^*(t+\tau)e^{j\beta_q t}\rangle = 0, \qquad if \beta_p \neq \beta_q \tag{2.6.41}$$

$$r_{pp}^{\beta_p}(\mathbf{s},\tau) = \rho_p(\mathbf{s},\tau) = \langle s_p(t)s_p^*(t)e^{j\beta_p t}\rangle > 0, \qquad \forall p \tag{2.6.42}$$

The additive noise vector $\mathbf{n}(t)$ is modeled as a stationary random process, so that

$$\langle n_p(t)n_p^*(t+\tau)e^{j\beta_p t}\rangle = 0, \qquad \forall p,\tau \tag{2.6.43}$$

The output cyclic correlation matrix is defined to be

$$\mathbf{R}_x^{(\beta_p)}(\tau) \triangleq \langle \mathbf{x}(t)\mathbf{x}^H(t+\tau)e^{j\beta_p t}\rangle = \sum_{\{q|\beta_q=\beta_p\}} \rho_q(\mathbf{s},\tau)\mathbf{a}_q\mathbf{a}_q^H \tag{2.6.44}$$

where the sum is defined over all sources with frequency β_p. If all the sources have distinct cycle frequencies the output cyclic correlation matrix will be:

$$\mathbf{R}_x^{\beta_p}(\tau) = \rho_p(\mathbf{s}, \tau)\mathbf{a}_p\mathbf{a}_p^H \tag{2.6.45}$$

The indeterminacy in estimating the separating matrix is used to assume that the source signals have unit-norm zero-lag cyclic autocorrelation coefficients:

$$\rho_p(\mathbf{s}, 0) = \langle s_p(t)s_p^*(t)e^{j\beta_p t}\rangle = 1 \tag{2.6.46}$$

In [1] the necessary and sufficient conditions for cyclostationary BSS have been defined using only cyclic correlation matrices. If there exist multiple sources with a common cycle frequency, separation is possible if the following condition is met [1].

Identifiability condition: For any $K > 0$ blind source separation can be achieved using the output cyclic correlation matrices $\{\mathbf{R}_x^{\beta_p}(\tau)|p = 1, ..., m; \tau \in \{0, \tau_1, ..., \tau_k\}\}$ if and only if there do not exist any two distinct source signals $s_p(t)$ and $s_q(t)$ whose cycle frequencies are the same ($\beta_p = \beta_q$) and whose cyclic autocorrelation vectors $\boldsymbol{\rho}_p(\mathbf{s})$ and $\boldsymbol{\rho}_q(\mathbf{s})$ are linearly dependent, where $\boldsymbol{\rho}_l(\mathbf{s}) = [\rho_l(\mathbf{s}, 0), \rho_l(\mathbf{s}, \tau_1), ..., \rho_l(\mathbf{s}, \tau_K)]$, $1 \leq l \leq m$.

Based on the identifiability condition the following theorems are presented that discuss the necessary and sufficient conditions for separability of a vector of observations [1].

Theorem 2. Assume that the cycle frequencies of the source signals are distinct. For any matrix $\mathbf{B}$, define $\mathbf{z}(t)$ to be the $m \times 1$ vector given by $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$. Then, $\mathbf{B}$ is a separating matrix if and only if

$$r_{pq}^{\beta_p}(\mathbf{z}, 0) = 0 \quad \textit{and} \quad r_{pp}^{\beta_p}(\mathbf{z}, 0) = 1 \qquad \forall p, q \quad 1 \leq p \neq q \leq m. \tag{2.6.47}$$

Proof. Define $\mathbf{C} = \mathbf{BA} = [c_{pq}]_{1 \leq p,q \leq m}$ so that $\mathbf{z}(t) = \mathbf{C}\mathbf{s}(t) + \mathbf{B}\mathbf{n}(t)$. From the mutual

independence of the sources and the stationarity of the noise vector $\mathbf{n}(t)$ we can conclude that $r_{pq}^{\beta_p}(\mathbf{z},\tau) = c_{pp}c_{qp}^{\star}\rho_p(\mathbf{s},\tau)$. $\mathbf{B}$ is a separating matrix if and only if $\mathbf{C}$ is a unitary diagonal matrix, thus it is sufficient to show that (2.6.47) is equivalent to $\mathbf{C}$ being unitary and diagonal. This readily follows from the fact that $\rho_p(\mathbf{s},0) = 1$ for all p. In particular, if $\mathbf{C}$ is unitary and diagonal, then $c_{pp}c_{qp}^{\star} = 0$ for all $p \neq q$, and $c_{pp}c_{qp}^{\star} = |c_{pp}|^2 = 1$ if $p = q$. This implies (2.6.47).

Conversely, if (2.6.47) is true, then $|c_{pp}|^2 = 1$ and $c_{pp}c_{qp}^{\star} = 0$ for all $p \neq q$. This means that $\mathbf{C}$ is unitary and diagonal. □

In Theorem 2 only cyclic correlation functions with $\tau = 0$ were considered. Theorem 3 as follows is a generalization of Theorem 2:

Theorem 3. Assume that the identifiability condition is satisfied, that is if $\beta_p = \beta_q$ then, $\boldsymbol{\rho}_p$ and $\boldsymbol{\rho}_q$ are linearly independent. Then, $\mathbf{B}$ is a separating matrix if and only if

$$r_{pq}^{\beta_p}(\mathbf{z},k) = 0 \quad \textit{and} \quad r_{pp}^{\beta_p}(\mathbf{z},0) = 1 \qquad \forall p,q \quad 1 \leq p \neq q \leq m \quad \textit{and} \quad k \in \{0;\tau_1,...,\tau_K\} \tag{2.6.48}$$

Proof. The proof is extension of proof of Theorem 2. Recall that $\mathbf{C}$=$\mathbf{BA}$. Write it as $\mathbf{C} = [\mathbf{c}_1,...,\mathbf{c}_m]$ where $\mathbf{c}_l$ is the lth column of $\mathbf{C}$. Consider an arbitrary source p and let $p_1,...,p_I$ be all the sources (including p) with the same cycle frequency as source p. Define $\mathbf{s}_p(t) = [s_{p_1}(t), s_{p_2}(t), ...s_{p_I}(t)]^T$ as the corresponding column vectors of $\mathbf{C}$ i.e., $\mathbf{C}_p = [s_{p_1}, s_{p_2}, ...s_{p_I}]^T$, and

$$\mathbf{S}_p(\tau) = \langle \mathbf{s}_p(t)\mathbf{s}_p^H(t+\tau)e^{j\beta_p t}\rangle = diag(\rho_{p_1}(\tau),...,\rho_{p_I}(\tau))$$
$$\mathbf{R}_z^{\beta_p}(\tau) = \langle \mathbf{z}(t)\mathbf{z}^H(t+\tau)e^{j\beta_p t}\rangle = \mathbf{C}_p\mathbf{S}_p(\tau)\mathbf{C}_p^H$$

where the last equality comes from (2.6.44). Applying matrix notations to (2.6.48)

it leads to

$$\sum_{\beta_p} \mathbf{R}_z^{\beta_p}(0) = \mathbf{C}\mathbf{C}^H = \mathbf{I}.$$

That means $\mathbf{C}$ is unitary (and thus, in particular, $\mathbf{C}_p$ is full column rank), and $\mathbf{R}_z^{\beta_p}(k) = \mathbf{C}_p\mathbf{S}_p(k)\mathbf{C}_p^H$ is diagonal for $k = \tau_1, ..., \tau_K$. We can then conclude that $\mathbf{C}_p = \mathbf{P}\mathbf{\Lambda}$, where $\mathbf{P}$ is a permutation matrix and $\mathbf{\Lambda} = [\mathbf{\Lambda}_p^T, 0^T]^T$, where $\mathbf{\Lambda}_p^T$ is an $I \times I$ unitary diagonal matrix by using Theorem 2 of [16]. □

Based on these theorems two algorithms have been developed in [1]. Since Theorem 2 is just an special case of Theorem 3, its implementation is an special case of that of Theorem 3 and therefore, only the general case is discussed here. The following cost function is defined based on the concepts of Theorem 3:

$$\begin{aligned} G(\mathbf{z}) \triangleq & \sum_{k=\tau_0}^{\tau_K} \sum_{1 \le p < q \le m} [|r_{pq}^{\beta_p}(\mathbf{z}, k) + r_{qp}^{\beta_p}(\mathbf{z}, k)|^2 + |r_{pq}^{\beta_p}(\mathbf{z}, k) - r_{qp}^{\beta_p}(\mathbf{z}, k)|^2] \\ & + \sum_{p=1}^{m} |r_{pp}^{\beta_p}(\mathbf{z}, 0) - 1|^2 \end{aligned} \tag{2.6.49}$$

where $\tau_0 = 0$. According to the separation criterion of Theorem 3 it is clear that

$$G(\mathbf{z}(t)) = 0 \Rightarrow \mathbf{B} \ \textit{is a separating matrix} \tag{2.6.50}$$

where $\mathbf{z}(t) = \mathbf{B}\mathbf{x}(t)$. The solution is an iterative algorithm based on natural gradient algorithm [8], [10].

In each iteration i, the un-mixing matrix $\mathbf{B}$ and the estimations $\mathbf{z}(t)$ are obtained by the following equations:

$$\mathbf{B}^{(i+1)} = (\mathbf{I} + \epsilon^{(i)})\mathbf{B}^{(i)} \tag{2.6.51}$$

$$\mathbf{z}^{(i+1)}(t) = (\mathbf{I} + \epsilon^{(i)})\mathbf{z}^{(i)}(t) \tag{2.6.52}$$

where the matrix $\epsilon^{(i)} = [\epsilon_{pq}^{(i)}]_{1 \le p,q \le m}$ is obtained from a local linearization of $G(\mathbf{Bx}(t))$ calculated as follows.

At the ith iteration, $r_{pq}^{\beta_p^{(i)}}(\mathbf{z}, \tau)$ is approximated by:

$$r_{pq}^{\beta_p^{(i)}}(\mathbf{z}, k) \approx \frac{1}{T-k} \sum_{t=1}^{T-k} z_p^{(i)}(t) z_q^{(i)*}(t+k) e^{j\beta_p t} \tag{2.6.53}$$

where T is the number of observations. If $\beta_p = \beta_q$, using (2.6.52) it can be concluded that

$$\begin{aligned} r_{pq}^{\beta_p^{(i+1)}}(\mathbf{z}, k) &\triangleq \langle z_p^{(i+1)}(t) z_q^{(i+1)*}(t+k) e^{j\beta_p t} \rangle \\ &= \left\langle \left(z_p^{(i)}(t) + \sum_{h=1}^{m} \epsilon_{ph}^{(i)} z_h^{(i)}(t) \right) \cdot \left(z_q^{(i)*}(t+k) + \sum_{l=1}^{m} \epsilon_{pl}^{(i)*} z_l^{(i)*}(t+k) \right) e^{j\beta_p t} \right\rangle \end{aligned}$$

If we assume that $\mathbf{B}^i$ is close to a separating matrix, then the following terms are negligible: $|\epsilon_{pq}^{(i)}| \ll 1$, $|\langle z_p^{(i)}(t) z_l^{(i)*}(t+k) e^{j\beta_p t} \rangle| \ll 1$ for $l \neq p$, and $|\langle z_h^{(i)}(t) z_q^{(i)*}(t+k) e^{j\beta_p t} \rangle| \ll 1$ for $h \neq q$. A first-order approximation of $r_{pq}^{\beta_p^{(i+1)}}(\mathbf{z}, k)$ is thus given by

$$r_{pq}^{\beta_p^{(i+1)}}(\mathbf{z}, k) \approx r_{pq}^{\beta_p^{(i)}}(\mathbf{z}, k) + \epsilon_{qp}^{(i)*} r_{pp}^{\beta_p^{(i)}}(\mathbf{z}, k) + \epsilon_{pq}^{(i)} r_{qq}^{\beta_p^{(i)}}(\mathbf{z}, k) \tag{2.6.54}$$

When $p \neq q$, $\epsilon_{pq}^{(i)}$ is chosen to be the solution of the following least squares (LS) minimization problem obtained by substituting (2.6.54) into (2.6.49)

$$min \left\| [\mathbf{r}_{qq}^{\beta_p^{(i)}}(\mathbf{z}), \mathbf{r}_{pp}^{\beta_p^{(i)}}(\mathbf{z})] \mathbf{E}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) + [\frac{1}{2}(\mathbf{r}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) + \mathbf{r}_{qp}^{\beta_p^{(i)}}(\mathbf{z})), \frac{1}{2j}(\mathbf{r}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) - \mathbf{r}_{qp}^{\beta_p^{(i)}}(\mathbf{z}))] \right\|^2$$

where

$$\mathbf{E}_{pq}^{(i)} \triangleq \begin{bmatrix} \Re(\epsilon_{pq}^{(i)}) & \Im(\epsilon_{pq}^{(i)}) \\ \Re(\epsilon_{qp}^{(i)}) & -\Im(\epsilon_{qp}^{(i)}) \end{bmatrix} \tag{2.6.55}$$

$$\mathbf{r}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) = \left[r_{pq}^{\beta_p^{(i)}}(\mathbf{z},0), r_{pq}^{\beta_p^{(i)}}(\mathbf{z},\tau_1), ..., r_{pq}^{\beta_p^{(i)}}(\mathbf{z},\tau_K)\right]^T \tag{2.6.56}$$

here $\Re(x)$ and $\Im(x)$ denote the real and imaginary parts of the complex scalar x. Solving the least square minimization problem for ϵ, we obtain:

$$\begin{aligned} \mathbf{E}_{pq}^{(i)} &= -\left[\mathbf{r}_{qq}^{\beta_p^{(i)}}(\mathbf{z}), \mathbf{r}_{pp}^{\beta_p^{(i)}}(\mathbf{z})\right]^{\sharp} \left[\frac{1}{2}(\mathbf{r}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) + \mathbf{r}_{qp}^{\beta_p^{(i)}}(\mathbf{z})), \frac{1}{2j}(\mathbf{r}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) - \mathbf{r}_{qp}^{\beta_p^{(i)}}(\mathbf{z}))\right] \\ &= \left([\mathbf{r}_{qq}^{\beta_p^{(i)}}(\mathbf{z}), \mathbf{r}_{pp}^{\beta_p^{(i)}}(\mathbf{z})]^H [\mathbf{r}_{qq}^{\beta_p^{(i)}}(\mathbf{z}), \mathbf{r}_{pp}^{\beta_p^{(i)}}]\right)^{-1} [\mathbf{r}_{qq}^{\beta_p^{(i)}}(\mathbf{z}), \mathbf{r}_{pp}^{\beta_p^{(i)}}(\mathbf{z})]^H \\ &\quad \cdot \left[\frac{1}{2}(\mathbf{r}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) + \mathbf{r}_{qp}^{\beta_p^{(i)}}(\mathbf{z})), \frac{1}{2j}(\mathbf{r}_{pq}^{\beta_p^{(i)}}(\mathbf{z}) - \mathbf{r}_{qp}^{\beta_p^{(i)}}(\mathbf{z}))\right] \end{aligned} \tag{2.6.57}$$

where $\sharp$ denotes pseudo-inverse of a matrix. For diagonal elements, i.e. when $p = q$ (2.6.57) is simplified to:

$$\epsilon_{pp}^{(i)} = \frac{1 - r_{pp}^{\beta_p^{(i)}}(\mathbf{z},0)}{2r_{pp}^{\beta_p^{(i)}}(\mathbf{z},0)}. \tag{2.6.58}$$

The iterative algorithm for Theorem 2 is similar to the above algorithm with the simplification that the off-diagonal elements are calculated by:

$$\epsilon_{pq}^{(i)} = -\left(\frac{r_{qp}^{\beta_p^{(i)}}(\mathbf{z},0)}{r_{qq}^{\beta_p^{(i)}}(\mathbf{z},0)}\right)^{\star}. \tag{2.6.59}$$

Sequential BSS for periodic sources

In a separate work, the separation is performed by sequentially converging to a solution which diagonalizes the output covariance matrix constructed at a lag corresponding to the smallest fundamental period [65]. It is assumed that m zero mean periodic sources $\mathbf{s}(k)$, are mixed and their instantaneous mixtures $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ are measured. The pth

original source is temporally periodic in time with fundamental period T_p

$$s_p(k) = s_p(k + nT_p) \tag{2.6.60}$$

and it is assumed that T_p is the smallest period such that $T_p < T_q$, $\forall q \in \{1, 2, ..., p - 1, p + 1, ..., m\}$.

The following theorem is proved in [65] which deals with the distinction between the eigenvalues of covariance matrix and will be used in the identification criterion.

Theorem 4. The eigenvalues of the source covariance matrix $\mathbf{R}_s(T_p)$ are nonzero and distinct, provided the following:

1) there exist m unit variance, discrete-time source signals with distinct periods;

2) $\rho_{s_q}(T_p) \neq 0$, $\forall q \in \{1, 2, ..., p - 1, p + 1, ..., m\}$

3) $\rho_{s_q}(T_p) \neq \rho_{s_l}(T_p)$, $\forall q, l \in \{1, 2, ..., p - 1, p + 1, ..., m\}$

where $\mathbf{R}_s(\tau) = \langle \mathbf{s}(k)\mathbf{s}^T(k + T_p) \rangle$ and $\rho_{s_i}(\tau) = \langle s_i(k)s_i(k + \tau) \rangle$.

Identification criterion: Suppose decorrelation of observation signals is carried out at $\tau = T_p$, an instantaneous output covariance matrix can be formed from:

$$\mathbf{R}_y(k, T_p) = \mathbf{P}(k)\mathbf{\Lambda}_s(T_p)\mathbf{P}^T(k + T_p) \tag{2.6.61}$$

where $\mathbf{\Lambda}_s(T_p) = \mathbf{R}_s(T_p)$, $\mathbf{P}(k) = \mathbf{W}^{(k)}\mathbf{A}$ and $\mathbf{y}(k) = \mathbf{W}^{(k)}\mathbf{x}(k)$. Let $\mathbf{R}_y^o(T_p)$ denote the output covariance matrix at convergence of a sequential algorithm, that is

$$\lim_{k \to \infty} \mathbf{R}_y(k, T_p) = \mathbf{R}_y^o(T_p) \tag{2.6.62}$$

By definition, the whitening operation implies that

$$\mathbf{R}_y^o(T_p) = \mathbf{P}^o\mathbf{\Lambda}_s(T_p)\mathbf{P}^{oT} = \hat{\mathbf{D}}(T_p) \tag{2.6.63}$$

where $\hat{\mathbf{D}}(T_p)$ is a diagonal matrix and, since the system (2.6.63) is in steady-state, $\mathbf{P}^o = \lim_{k\to\infty} \langle \mathbf{P}(k) \rangle$. The columns of $\mathbf{P}^o$ are the eigenvectors of $\mathbf{R}_y^o(T_p)$. This leads to the following theorem:

Theorem 5. Provided that Theorem 4 holds, the eigenvalues of $\mathbf{R}_y^o(T_p)$ are distinct and nonzero.

Proof. Pre- and post-multiplying (2.6.63) by $\mathbf{P}^{oT}$ and $\mathbf{P}^o$ respectively, we have

$$\mathbf{P}^{oT}\mathbf{R}_y^o(T_p)\mathbf{P}^o = \mathbf{P}^{oT}\mathbf{P}^o\Lambda_s(T_p)\mathbf{P}^{oT}\mathbf{P}^o = \Lambda_s(T_p) \tag{2.6.64}$$

where $\Lambda_s(T_p)$ is the diagonal matrix of eigenvalues of $\mathbf{R}_s(T_p)$. Clearly, from (2.6.64), it also contains the eigenvalues of $\mathbf{R}_y^o(T_p)$, and therefore, provided that Theorem 4 holds, its diagonal entries are distinct and nonzero. □

In the next theorem it is shown that whitening the output covariance matrix at a lag $\tau = T_p$, simultaneously whitens all the covariance matrices at $\tau = nT_p$.

Theorem 6. Diagonalization of $\mathbf{R}_y^o(T_p)$ effectively amounts to simultaneously diagonalising the covariance matrix at every delay $\tau = T_p, n \in \mathbb{Z}$

Proof of this theorem can be found in [65]. This theorem indicates that a separating matrix $\mathbf{W}^o$ exists that diagonalizes all the matrices $\mathbf{R}_y^o(nT_p)$, if its eigenvalues are distinct and so it results in blind separation of periodic sources.

Minimizing the following information theoretic cost function is the main idea used in a sequential algorithm for separation of periodic signals:

$$\mathcal{J}(\mathbf{W}^{(k)}) = -\frac{1}{2}\log(\det(\mathbf{W}^{(k)}\mathbf{W}^{(k)^H})) + \frac{1}{2}\sum_{i=1}^{m}\langle |y_i(k)y_i^\star(k+\tau)| \rangle \tag{2.6.65}$$

Role of the first term in the right-hand side of (2.6.65) is to avoid a zero solution for $\mathbf{W}^{(k)}$. The second term is used to ensure that at the chosen lag τ the output signals have minimum sum absolute auto-correlation values. Differentiating (2.6.65) results in

the following gradient function:

$$\nabla \mathcal{J}(\mathbf{W}^{(k)}) = -\left[\mathbf{I} - \frac{1}{2}(\mathbf{R}_y^H(k,T_p) + \mathbf{R}_y(k,T_p))\right]\mathbf{W}^{(k)}. \tag{2.6.66}$$

thus the following learning rule is provided:

$$\mathbf{W}^{(k+1)} = \mathbf{W}^{(k)} + \eta\left[\mathbf{I} - \frac{1}{2}(\mathbf{R}_y^H(k,T_p) + \mathbf{R}_y(k,T_p))\right]\mathbf{W}^{(k)}. \tag{2.6.67}$$

Convolutive cyclostationary BSS

As explained before, in many real world applications such as those in biomedical engineering, telecommunications, speech processing, econometrics, and atmospheric science, the sources are not mixed instantaneously [84]. In these applications the observations are mixed convolutively and conventional blind source separation methods can no longer be applied to separate the underlying sources.

As convolutive BSS is not the main concern of this thesis, only two methods are reviewed here briefly. In the first selected research work, ICA is used to separate heart and lung sound signals [87] and in the second one a method for extracting cyclostationary sources from convolutive mixtures is proposed [105].

In [87] heart and lung sounds separation is performed using ICA. It is assumed that the observations are convolutive mixtures of the source signals. Therefore, $x_p(t)$ is expressed as:

$$x_p(t) = \sum_{q=1}^{m}(\sum_{\tau} a_{pq}(\tau)s_q(t-\tau)) = \sum_{q=1}^{m} a_{pq}(t) * s_q(t) \tag{2.6.68}$$

where $*$ denotes convolution and a_{pq} represents the transfer function of the transmission path from source q to the sensor p. For simplicity it is assumed that the number of sensors is the same as the number of sources, i.e., $m = n$. The convolutive model of

(2.6.68) can be written in the matrix form as follows

$$\mathbf{x}(t) = \mathbf{A}(t) \otimes \mathbf{s}(t) \tag{2.6.69}$$

where

$$\mathbf{A}(t) = \begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t) & a_{n,2}(t) & \cdots & a_{n,n}(t) \end{bmatrix}$$

and $\otimes$ represents matrix convolution.

Equation (2.6.69) is then represented in time-frequency domain using short time Fourier transform (STFT) as

$$\hat{X}(\omega, t_s) = \hat{A}(\omega)\hat{S}(\omega, t_s), \qquad t_s = 0, \Delta T, 2\Delta T, ... \tag{2.6.70}$$

where $\hat{A}(\omega)$ is the Fourier transform of $\mathbf{A}(t)$, $\hat{S}(\omega, t_s)$ is the STFT of the source signals and ΔT represents shifting time. Short time Fourier transform of a signals is commonly called the *spectrogram* of the signal. Now the mixing model is changed such that the mixing matrix is simply multiplied to the source matrix. Thus, one of the conventional source separation methods can be used. In [87] JADE is used. The separation method is run for different frequencies and so different components of the sources in different frequencies are separated.

Permutation problem is solved using the fact that the components of a nonstationary signal at different frequencies are correlated [77]. Combining proper independent components for each frequency band, the spectrograms of the estimated signals are produced. Using the Inverse Short Time Frequency Transform (ISTFT) the separated signals are

transformed to time domain to reconstruct the original sources of interest.

Recently a work has been reported in [105] that uses cyclo-stationarity of the source signals and extends the natural gradient method used in [64] to separate convolutive mixtures of cyclo-stationary signals. It is assumed that the sources and observations are related through the following equation:

$$\mathbf{x}(k) = \mathbf{H}(z)\mathbf{s}(k) + \mathbf{v}(k) \tag{2.6.71}$$

where $\mathbf{H}(z)$ is the z-transform of the mixing matrix with entries $H_{pq}(z) = \sum_{i=0}^{P-1} h_{pqi} z^{-i}$, P is the filter length, and z^{-1} is the time shift operator, i.e., $z^{-1}s_q(k) = s_q(k-1)$.

It is assumed that $\mathbf{y}(k)$ is the source estimation, thus

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{W}(z,k)\mathbf{x}(k) \\ &= \mathbf{W}(z,k)\mathbf{H}(z)\mathbf{s}(k) = \mathbf{C}(z,k)\mathbf{s}(k) \end{aligned} \tag{2.6.72}$$

where entries of $\mathbf{y}(k)$ are $y_p(k) = \sum_{q=1}^{m} W_{pq}(z,k)x_q(k)$ and $\mathbf{W}(z,k)$ is the unmixing matrix with entries $W_{pq}(z,k) = \sum_{i=0}^{L-1} w_{pqi}(k)z^{-i}$, where L is the filter length. It is assumed that $\mathbf{H}(z)$ and $\mathbf{W}(z,k)$ are both stable with no zero eigenvalues on the unit circle $|z| = 1$

The goal is to adjust $W_{pq}(z,k)$ such that $\lim_{k\to\infty} \mathbf{C}(z,k) = \mathbf{P\Lambda D}(z)$ where $\mathbf{P}$ is a permutation matrix, $\mathbf{\Lambda}$ is a non-singular diagonal scaling matrix, $\mathbf{D}(z)$ is a diagonal matrix whose ppth entry is $\sum_i c_{pi} z^{-i}$, c_{pi} is a complex scalar weighting, and i is an integer delay value. It is also assumed that the source signals are cyclostationary and satisfy (2.6.23).

From algebraic point of view, convolutive and instantaneous BSS have equivalent mathematical models except for description of the mixing media.

To simplify the complexity of convolutive BSS, Z-transform is used. Similar to [64],

the following cost function for convolutive mixtures of cyclostationary sources is defined:

$$\begin{aligned}\mathcal{J}(\mathbf{W}(z,k)) = &-\sum_{i=1}^{m}\log(p_i(y_i(k)) - \frac{1}{2\pi j}\oint \log|\det \mathbf{W}(z,k)|z^{-1}dz \\ &+\frac{1}{2}\left\{Tr(\tilde{\mathbf{R}}_y^{\beta}(k,0)) - \log\det(\tilde{\mathbf{R}}_y^{\beta}(k,0)) - N\right\}\end{aligned} \tag{2.6.73}$$

where $\tilde{\mathbf{R}}_y^{\beta}(k,0)$ is defined as in (2.6.25). The minimum point of (2.6.73) is estimated using natural gradient algorithm following the same approach of [64], which was discussed before.

2.7 Concluding remarks

The main focus of this chapter was blind source separation/extraction of sources with periodic structure. General framework for BSS problem and mathematical formulation were described. Based on the characteristics of the mixing media different models for the system and mixtures can be generated and therefore, different approaches for separation the sources of interest have been developed. Theoretical details of four well-established instantaneous separation methods were reviewed and πCA as one of the ICA methods for separating periodic sources was described.

A review of cyclostationarity as a concept for analyzing the signals with hidden periodicity was provided. This concept has attracted researchers from different areas and powerful techniques have been developed for processing and analyzing this type of signals. Recent research works on blind separation and extraction of cyclostationary sources were described as well.

In the next chapters periodicity of the sources will be exploited to develop suitable separation methods and the results will be compared with those of benchmark methods, when applicable.

Chapter 3

A FAST BLIND SEPARATION METHOD FOR PERIODIC SOURCES

3.1 Overview

In this chapter a fast algorithm for blind identification of periodic sources is presented. In the well-known SOBI method, the information is extracted from instantaneous mixtures by simultaneously diagonalizing several time-delayed covariance matrices, however, the delays are chosen arbitrarily. This imposes computational cost which is linearly related to the number of covariance matrices. Statistical characteristics of periodic sources are exploited in the new algorithm to develop a method to effectively choose the appropriate delays in which the diagonalization should happen. Detailed theory together with the corresponding algorithm will be presented. Software simulations verify the superior performance of the algorithm in the face of different noise and frequency variation levels over alternative methods.

The outline of the chapter is as follows. Next section covers the related work. Problem formulation is detailed in section 3.3. The proposed algorithm and the simulated experiments are presented in sections 3.4 and 3.5, respectively. Section 3.6 contains concluding remarks.

3.2 Related works

Knowledge about periodicity of the signals has been exploited by researchers in different science and engineering fields to separate the sources that originate from periodic phenomena. In [89] and [91] a method based on GEVD is used to diagonalize the covariance matrices of the observation vector at zero lag and a lag equal to the period of the source of interest. The method called periodic component analysis (πCA) maximizes a cost function which is a measure of periodicity of the estimated source. In the case of varying periods, the observations have to be adjusted to have exact periods [89]. The performance of this method depends on proper detection of the cycles of the periodic source signal(s). More details about the πCA method are provided in section 2.4.

Second order statistics are widely used in source separation context. Using the concept developed in Essential uniqueness theorem in [16], an average eigen structure of the data is obtained by simultaneous diagonalization of a set of covariance matrices each calculated at a different delay of the pre-whitened data. It has been shown that the sources can be estimated using the joint diagonalizer of the covariance matrices [16, 25]. This method is called second order blind identification (see section 2.3.3). Whitening a nonzero delay covariance matrix is suggested in [26] to reduce white noise effects in the non-stationary data. In order to reduce the effects of spatially colored noise on the separation performance, the whitening is performed on a positive definite matrix in [17]. This matrix is a linear combination of covariance matrices at different delays. To minimize the effects of spatially colored noise on separation performance, a bank of subband filters is proposed in [47]. The method is based on reducing the covariance matrix of noise subband from the covariance matrix of the observations.

Despite the good performance of the methods in [16, 17, 26, 47], there is no guideline regarding the selection of appropriate delays in order to achieve the best performance and the least computational cost in separation. Moreover, it is not known how many delayed covariance matrices are required such that the condition of essential uniqueness

theorem [16] is met. In the implementation of the algorithm provided by the authors, the first $min(100, N/3)$ delayed covariance matrices are used as default, where N is the total number of samples. Using this number of covariance matrices leads to an acceptable average eigenvalue decomposition. However, the computational cost of jointly diagonalizing c matrices is proportional to c [16, 66] and therefore, if one can reduce the number of covariance matrices the separation algorithm will converge faster.

Under the periodicity assumption of the sources, a method for selecting appropriate delays is presented in this chapter. It is shown that for p periodic signals of different periodicities using just p delayed covariance matrices is enough to obtain a high quality estimation. This method is also robust to noise and performs well in those cases where the main frequency of the sources vary with time.

3.3 Problem formulation

As in general instantaneous BSS problem assume m unknown mutually statistically independent sources are mixed through an unknown medium and measured at n ($n \geq m$) sensors (see section 2.2). Also, let the mixing medium be modeled by matrix $\mathbf{A}$. Such a system, therefore, can be formulated in a vector form as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) = \mathbf{y}(t) + \mathbf{n}(t) \tag{3.3.1}$$

where $\mathbf{s}(t) = [s_1(t)...s_m(t)]^T$ is an $m \times 1$ source vector, $\mathbf{n}(t) = [n_1(t)...n_n(t)]^T$ is an $n \times 1$ stationary zero mean white noise vector independent of the source signals, $\mathbf{x}(t) = [x_1(t)...x_n(t)]^T$ is an $n \times 1$ measurement vector, and $\mathbf{A}$ is an $n \times m$ unknown full column rank mixing matrix. The additive noise is stationary, temporally white, zero mean random process, and independent of the source signals.

Here, it is assumed that the source signals are periodic with distinct fundamental frequencies. Furthermore, to simplify the notation and with no loss of generality we

assume that m=n.

The covariance matrix of vector $\mathbf{v}(t)$ at time t and delay τ is defined as

$$\mathbf{R}_v(t,\tau) = \langle \mathbf{v}(t)\mathbf{v}^H(t+\tau)\rangle = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{N} \mathbf{v}(t)\mathbf{v}^H(t+\tau) \tag{3.3.2}$$

where $\langle . \rangle$ is the expected value of the enclosed term, N is the total number of samples and superscript H denotes complex conjugate transpose of a matrix. We refer to the ijth element of $\mathbf{R}_v(t,\tau)$ as $r_v^{ij}(t,\tau)$.

In order to overcome the scaling problem, without loss of generality, we assume that the sources are unit norm signals, which means

$$\mathbf{R}_s(t,0) = \langle \mathbf{s}(t)\mathbf{s}^H(t)\rangle = \mathbf{I} \tag{3.3.3}$$

where $\mathbf{I}$ is an $n \times n$ identity matrix . From the above assumption we can easily conclude the following relations for ijth element of the covariance matrix (3.3.3):

$$|r_s^{ii}(t,\tau)| \leq |r_s^{ii}(t,0)| \quad \forall t,\tau; \quad \forall i\ 1 \leq i \leq n \tag{3.3.4}$$

$$|r_s^{ij}(t,\tau)| = 0 \quad \forall t,\tau; \quad \forall i,j\ 1 \leq i \neq j \leq n \tag{3.3.5}$$

To estimate the original sources,

the signal part of the observation vector is first pre-whitened to obtain $\mathbf{W}\mathbf{y}(t) = \mathbf{W}\mathbf{A}\mathbf{s}(t)$. $\mathbf{C} = \mathbf{W}\mathbf{A}$ is a unitary matrix because $\mathbf{R}_y(t,0) = \mathbf{W}\mathbf{A}\langle \mathbf{s}(t)\mathbf{s}^H(t)\rangle \mathbf{A}^H\mathbf{W}^H = \mathbf{C}\mathbf{C}^H = \mathbf{I}$. The fundamental idea behind the method is to find a matrix $\mathbf{B}$ which estimates the source signals by a rotation applied to the whitened data, i.e., $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$. In other words, by finding the appropriate matrix $\mathbf{B}$ the sources can be identified as $\hat{\mathbf{s}}(t) = \mathbf{B}^H\mathbf{z}(t)$.

The covariance matrix of the whitened data at lag τ is:

$$\mathbf{R}_z(t,\tau) = \langle \mathbf{z}(t)\mathbf{z}^H(t+\tau)\rangle = \mathbf{C}\langle \mathbf{s}(t)\mathbf{s}^H(t+\tau)\rangle\mathbf{C}^H \tag{3.3.6}$$

which is obviously a normal matrix. The objective is to find a matrix $\mathbf{B}$ which is equal to $\mathbf{C}$ or essentially equal to $\mathbf{C}$. In the case of successful estimation of $\mathbf{B}$, the estimation of mixing matrix, $\hat{\mathbf{A}}$, is found by $\hat{\mathbf{A}} = \mathbf{W}^\sharp\mathbf{B}$ where $\mathbf{W}^\sharp$ denotes Moore-Penrose pseudo-inverse of $\mathbf{W}$.

It is known from linear algebra that all normal square matrices are diagonalizable by some unitary matrices (spectral theorem in [96]) which may lead to separation. The unitary diagonalizer matrix of a whitened covariance matrix at some lag τ is the separating matrix if the covariance matrix has distinct eigenvalues. However, without having any prior knowledge it is difficult to find a time delay in which the covariance matrix is full rank. In order to reduce the probability that an unfortunate choice of time lag τ results in unidentifiability of $\mathbf{C}$ from $\mathbf{R}_z(t,\tau)$ the joint diagonalization of several covariance matrices is proposed in [16]. The consequent problem in joint diagonalization to be addressed is the uniqueness of the unitary diagonalizer matrix. Here, the periodicity of the sources is used to obtain the unique unitary diagonalizer, which is the separating matrix.

We know that the source signal $s_i(t)$ is periodic for all $1 \leq i \leq n$. This requires that for every source s_i, we have:

$$r_s^{ii}(t, kT_i) = r_s^{ii}(t,0) \quad \forall t, i \quad 1 \leq i \leq n \tag{3.3.7}$$

where T_i is the period of source s_i and k is an arbitrary integer. $r_s^{ii}(t,0)$ is the maximum possible value for the covariance of the ith source. It means that except for the integer multiples of T_i the value of this function is less than $r_s^{ii}(t,0)$ for all the other time delays. Since the original sources are unit norm, the covariance matrix of $\mathbf{s}$ in delay kT_i obeys

the following structure:

$$\begin{gathered}\mathbf{R}_s(t, kT_i) = diag(r_1, ..., r_{i-1}, 1, r_{i+1}, ..., r_n) \\ |r_l| < 1,\ l \neq i,\ kT_i \neq T_j,\ 1 \leq i, j, l \leq n,\ k \in \mathbb{N}\end{gathered} \tag{3.3.8}$$

Assume that a unitary matrix $\mathbf{B}$ diagonalizes the covariance matrix $\mathbf{R}_z(t, \tau)$ at lag τ such that $\mathbf{B}^H\mathbf{R}_z(t, \tau)\mathbf{B} = \mathbf{B}^H\mathbf{C}\mathbf{R}_s(t, \tau)\mathbf{C}^H\mathbf{B} = \mathbf{\Lambda}$. Both $\mathbf{B}$ and $\mathbf{C}$ are unitary matrices, so $\mathbf{D} = \mathbf{B}^H\mathbf{C}$ is also a unitary matrix and the diagonal matrix $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_n)$ is the eigenvalue matrix of $\mathbf{R}_s(t, \tau)$ which is equal to $\mathbf{R}_s(t, \tau)$. Therefore, for each time delay T_i, the covariance matrix $\mathbf{R}_z(t, T_i)$ is diagonalizable and only one of its eigenvalues is equal to 1. This means for each T_i we have:

$$\lambda_i = 1 \neq \lambda_j \quad \forall j \ \ 1 \leq j \leq n,\ j \neq i \tag{3.3.9}$$

This fact is used in the following theorem to guarantee uniqueness of the unitary diagonalizer.

Theorem 1. Assume that $\mathbf{z}(t)$ is a white mixture of periodic sources with distinct periods and the covariance matrices of the source vector $\mathbf{s}(t)$ satisfy (3.3.8). If a unitary matrix $\mathbf{B}$ simultaneously diagonalizes the set of covariance matrices $\mathcal{R} = \{\mathbf{R}_z(t, T_i) \ \ \forall i \ 1 \leq i \leq n\}$, (i.e. for all i $\mathbf{R}_z(t, T_i) = \mathbf{B}\mathbf{D}_i\mathbf{B}^H$, where $D_i = diag(d_1(i), d_2(i), ..., d_n(i))$) then, any joint diagonalizer of elements of $\mathcal{R}$ is essentially equal to $\mathbf{B}$.

Proof. To prove the sufficiency of the theorem, it is enough to show that any linear combination of the columns of $\mathbf{B}$ cannot be a common eigenvector of the members of $\mathcal{R}$. Let's assume $\mathbf{e}$ is a common eigenvector, i.e.,

$$\mathbf{R}_z(t, T_j)\mathbf{e} = \lambda_j\mathbf{e} = \sum_{i=1}^{n} \lambda_j\alpha_i\mathbf{b}_i \qquad \text{for all} \quad 1 \leq j \leq n \tag{3.3.10}$$

where $\mathbf{b}_i$ is the ith column of $\mathbf{B}$ and λ_j is an eigenvalue of $\mathbf{R}_z(t,T_j)$. We arbitrarily assume that $\alpha_p \neq 0$, $1 \leq p \leq n$. Then, index j can be found such that $d_p(j) = \mathbf{b}_p^H \mathbf{R}_z(t,T_j)\mathbf{b}_p = 1$. Using the definition of $\mathbf{e}$ and the assumption of the theorem, $\mathbf{R}_z(t,T_j)\mathbf{e}$ can also be restated as:

$$\begin{aligned}\mathbf{R}_z(t,T_j)\mathbf{e} &= \mathbf{R}_z(t,T_j)\sum_{i=1}^{n}\alpha_i\mathbf{b}_i \\ &= \sum_{i=1}^{n}\alpha_i\mathbf{R}_z(t,T_j)\mathbf{b}_i \\ &= \sum_{i=1}^{n}\alpha_i d_i(j)\mathbf{b}_i\end{aligned} \tag{3.3.11}$$

From (3.3.10) and (3.3.11) one can conclude that for all i, $\alpha_i(\lambda_j - d_i(j)) = 0$. We know that $\alpha_p \neq 0$, therefore, $\lambda_j = d_p(j) = 1$. Because the sources are periodic, using (3.3.8) it can also be concluded that $d_i(j) \neq d_p(j) = 1$ for all $i \neq p$ and hence $\alpha_i = 0$.

For the necessity condition assume that for two arbitrary indices (p,q) $d_p(j) = d_q(j)$ for all j. It's clear that any linear combination of the columns of $\mathbf{B}$ is a common eigenvector of the members of $\mathcal{R}$.

□

Although the above analysis is based on the assumption that the periods of the signals are exactly known, the analysis is still true for some time intervals close to the exact periods. In other words, when there is uncertainty about the fundamental periods of the sources or fundamental periods vary with time, the method can still be used successfully. In (3.3.7) and (3.3.8) we showed that for each periodic signal there is a set of delays in which the source covariance has a maximum value. An approximation of the source frequencies may be obtained by different time and frequency domain methods. It is very likely that the maximum covariance value can be found in a delay close to the estimated period. Hence, to best cover the estimation indeterminacy or the frequency variations a window centered at the delay corresponding to the estimated frequency is

used. A suitable window length L (which depends on the nature of the sources) meets the condition of Theorem 1 and so can lead to separation of the source signals. It is also noteworthy that choosing the appropriate window length is not our major concern here.

3.4 Separation algorithm

Based on Theorem 1 our objective is to find a unitary matrix $\mathbf{B}$ which jointly diagonalizes the set of selected delayed covariance matrices. In other words, the desired $\mathbf{B}$ is the one which minimizes the squared off-diagonal elements of the set of all $\mathbf{B}^H\mathbf{R}_z(t,T_i)\mathbf{B}$ for all periods T_i. The implementation of the proposed method is presented in Algorithm 1.

Algorithm 1

1) Estimate periods of the sources
2) Pre-whiten the data by $\mathbf{W}$ as $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$. $\mathbf{W} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T$ where $\mathbf{\Lambda}$ is the eigenvalue matrix of $\mathbf{x}(t)$ and $\mathbf{E}$ is the corresponding eigenvector matrix.
3) Calculate $\mathcal{R} = \bigcup_{i=1}^{n} \mathcal{R}_i$ where $\mathcal{R}_i = \{\mathbf{R}_z(t,T_i), \mathbf{R}_z(t,T_i \pm 1), ..., \mathbf{R}_z(t,T_i \pm \lfloor L/2 \rfloor)\}^*$
4) Find $\mathbf{B}$, the joint diagonalizer of the set of covariance matrices R.
5) The estimated sources are formed by $\mathbf{B}$ and $\mathbf{W}$ as $\hat{\mathbf{s}}(t) = \mathbf{B}^H\mathbf{W}\mathbf{x}(t)$.

where $\lfloor . \rfloor$ is the floor operator, i.e. the largest integer not greater than the operand.
* The set of covariance matrices $\mathcal{R}_i$ can be formed for any delay kT_i instead of T_i as long as $kT_i \neq T_j$, $j \neq i$ and $k \in \mathbb{N}$.

Remarks:

- Computational cost of the second order blind identification algorithm is

$$crm^2/2 + 4m^3/3 + (c-1)m^3/2 + Itn^2[4n(c-1) + 17(c-1) + 4n + 75]/2 \quad (3.4.1)$$

  where c is the number of covariance matrices, It is the number of sweeps required by the joint diagonalization algorithm, and r is the length of the data [66].
- Periodic component analysis [91] is a special case of the presented method in which the diagonalization is performed in only two lags (i.e. zero and the one corresponding to the frequency of periodic signal). However, the accuracy of this

method is subject to the frequency variations. The presented method in [89] for adjusting the period is useful as long as the cycles of the periodic signals can be accurately recognized by some means.

- The proposed method can be considered as a special case of the well-known SOBI method. SOBI is a widely used method and has an established performance, however, a large number of matrices is usually used in this method. As it is shown in Section 3.5 selecting a small number of covariance matrices does not provide a correct separation by SOBI and a large number of matrices require higher computational cost.

3.5 Experiments

To evaluate the performance of the proposed method, different experiments were designed for both synthetic and real world data. The first data set used here includes four periodic sources, which are composed of sine waves with normalized frequencies of 0.023, 0.01, 0.037, and 0.017 Hz and few harmonics. To evaluate the performance of the method where the main frequencies of the signals vary in time, a random coefficient is also applied to the frequencies in each cycle. Fig. 3.1 illustrates 500 samples of pure periodic sources along with their distorted versions. The main frequencies of the distorted sources have a random variation of up to 10%. The second data set is a mixture of human voice and music signals. Experiments 1-4 are performed for the synthetic data and in the last experiment the proposed method is applied to real world data.

In the ideal case of the separation, $\hat{\mathbf{A}}$ should be equal to $\mathbf{A}$, or in other words $\hat{\mathbf{A}}^{\sharp}\mathbf{A} = \mathbf{I}$. Therefore, the sum of the squared off-diagonal elements of $\hat{\mathbf{A}}^{\sharp}\mathbf{A}$ which is called mean rejection level (MRL) is used here as a quantitative measure to evaluate the algorithm. The lower the value of MRL is, the better performance from the algorithm is expected [16].

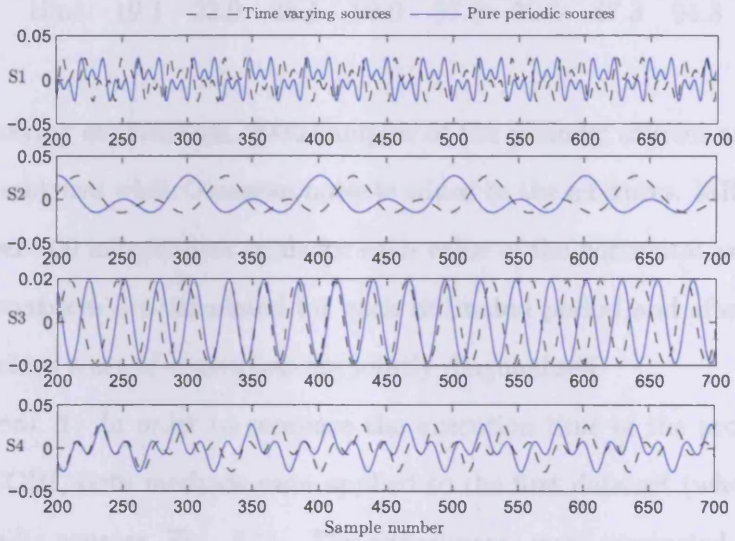


Figure 3.1: Four periodic sources used in the experiments. For some experiments the frequencies of the sources are changed by time. The black dashed lines show the distorted sources when change of up to 10% in frequencies is permitted in each cycle.

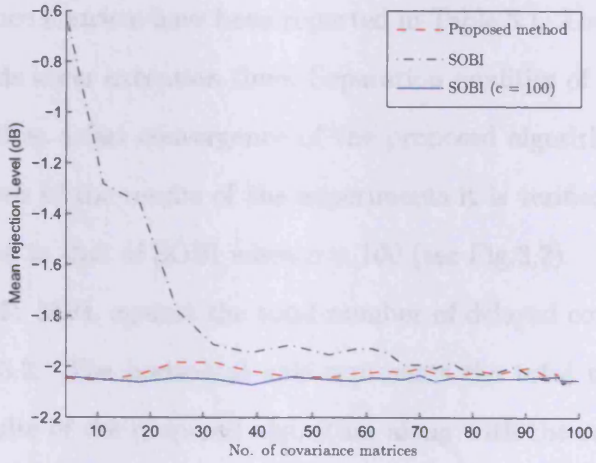


Figure 3.2: Mean rejection level vs. the number of covariance matrices.

Table 3.1: Simulation time vs. number of covariance matrices for the proposed method. The average time for SOBI with $c = 100$ is 94.1ms. (All times are in milliseconds.)

	No. of covariance matrices (c)							
	11	18	39	53	64	79	94	100
time	19.1	23.0	38.1	50.0	57.0	70.5	87.3	94.3

In the following experiments, 2000 samples of the periodic sources are mixed using linear mixing matrices while Gaussian noise is added to the mixtures. MRL (when used) is averaged over 100 independent trials for each value of the horizontal axis of the plots. L covariance matrices are calculated for each estimated period and after omitting the overlapping delays a set of c matrices are jointly diagonalized.

Experiment 1: In order to compare the execution time of the proposed method with that of SOBI, both methods were applied to the first data set (where the sources are pure periodic sources, Fig. 3.1). The experiments were conducted on a PC with 3.2 GHz Pentium IV CPU and 1.5GB of RAM. The average execution time of 100 independent trials of the original SOBI algorithm and the proposed method for different number of covariance matrices have been reported in Table 3.1. Lower number of covariance matrices yields lower execution time. Separation qualities of the methods are also compared. As well as a fast convergence of the proposed algorithm, by analyzing the mean rejection levels of the results of the experiments it is verified that the separation quality is very close to that of SOBI when $c = 100$ (see Fig.3.2).

Experiment 2: MRL against the total number of delayed covariance matrices (c) is shown in Fig. 3.2. The horizontal axis represents the total number of covariance matrices. The results of the proposed algorithm along with the results of SOBI, when the same number of covariance matrices are used, are presented in this figure. The result of SOBI with 100 covariance matrices is also presented as a reference. In this experiment the signal-to-noise ratio (SNR) defined as SNR $= -10\log_{10}\sigma^2$ is set to -1dB, where σ^2 is the noise variance. For almost all values of c the proposed method performs very close

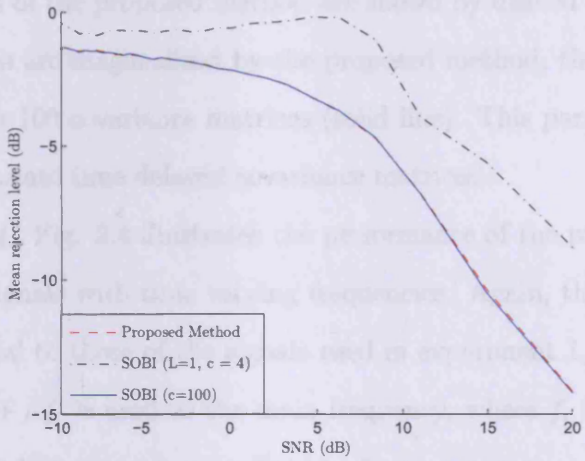


Figure 3.3: Mean rejection level vs. SNR. The proposed method and SOBI (c=100) perform similarly.

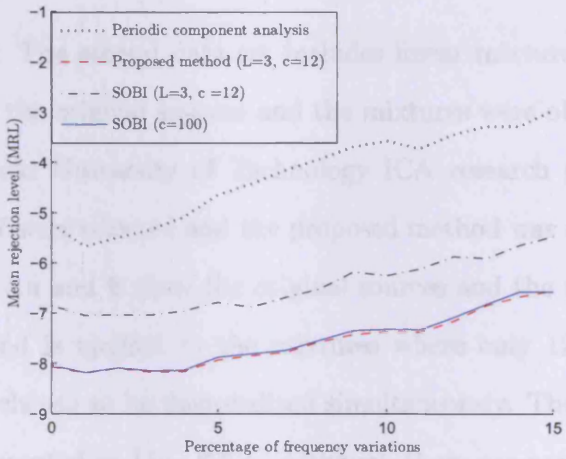


Figure 3.4: Mean rejection level vs. variations in frequency.

to SOBI with 100 covariance matrices.

Experiment 3: In this experiment the effects of SNR on the performance of the methods were investigated and the results were depicted in Fig. 3.3. Different levels of noise were added to the linear mixtures of the same sources used in previous experiments. The results of the proposed method are shown by dashed line. Although only 4 covariance matrices are diagonalized by the proposed method, the performance is very close to SOBI with 100 covariance matrices (solid line). This performance is the result of choosing appropriate time delayed covariance matrices.

Experiment 4: Fig. 3.4 illustrates the performance of the proposed algorithm for a set of periodic signals with time varying frequencies. Again, the main frequencies of the signals are equal to those of the signals used in experiment 1, but for each cycle of the ith source, $f_i + \beta f_i$ is used as the main frequency, where f_i is the main frequency and $-0.10 < \beta < 0.10$ is a random coefficient. As expected, the proposed method with c=12 and SOBI with c=100 provide better performance compared to πCA and SOBI with c=12.

Experiment 5: The second data set includes linear mixture of human voice and music signals. Both the original sources and the mixtures were obtained from the ICA demo page at Helsinki University of Technology ICA research group website[1]. The sources 2, 4, 5, and 7 were selected and the proposed method was applied to their linear mixtures. Figures 3.5.a and b show the original sources and the mixtures respectively. The proposed method is applied to the mixtures where only 12 covariance matrices ($L = 3,\ c = 12$) are chosen to be diagonalized simultaneously. The result of application of the method is presented in Fig. 3.5.c. Although there are scaling and permutation ambiguities, the estimated sources are very similar to the original ones. $\hat{Src}_2$ and $\hat{Src}_4$ are estimates of Src_1 and Src_4, respectively and $\hat{Src}_1$ and $\hat{Src}_3$ estimate Src_2 and Src_3 with a negative multiplier.

[1]The demo page is accessible from: `http://www.cis.hut.fi/projects/ica/cocktail/cocktail_en.cgi`

The period of the original sources has to be known *a priori*. Here, the appropriate time delays are chosen using the PSD of the sources. In Fig. 3.6 portions of the PSD of the sources are presented.

High peaks at PSD of signals show the periodic structure of the signal. Several peaks are detectable in this figure indicating the periodic behaviour of the sound signals. For this case 34.9, 97.5, 134.6, and 164.4 hertz are selected empirically.

3.6 Conclusions

In this chapter an efficient method for selecting the optimal delays for second order blind identification of periodic signals has been presented. The cost of computations for simultaneous diagonalization of covariance matrices in the second order blind identification method is a linear function of the number of covariance matrices. In the proposed method, using a considerably smaller set of covariance matrices results in a fast and still high separation quality. Different experiments show that, with lower computational cost, the results of the proposed method are very close to the asymptotic results of SOBI.

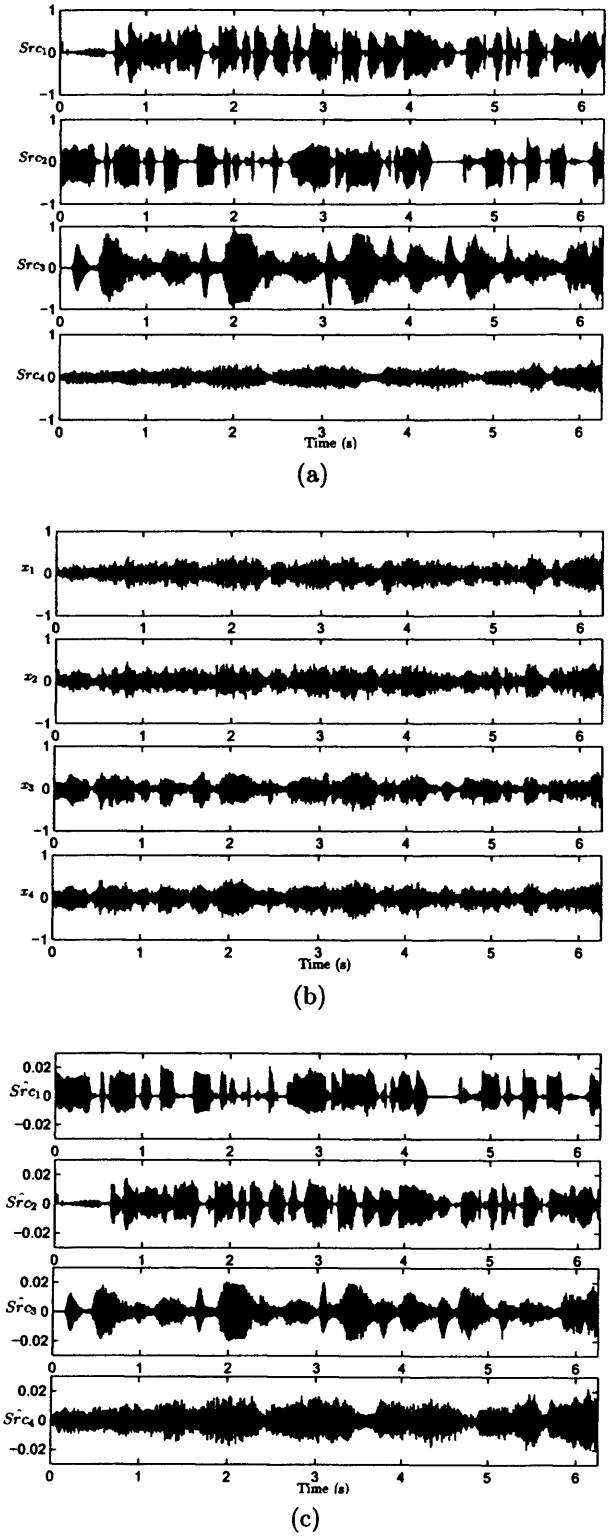


Figure 3.5: A set of voice and music signals; (a) original sources, (b) linear mixtures, and (c) output of the proposed method.

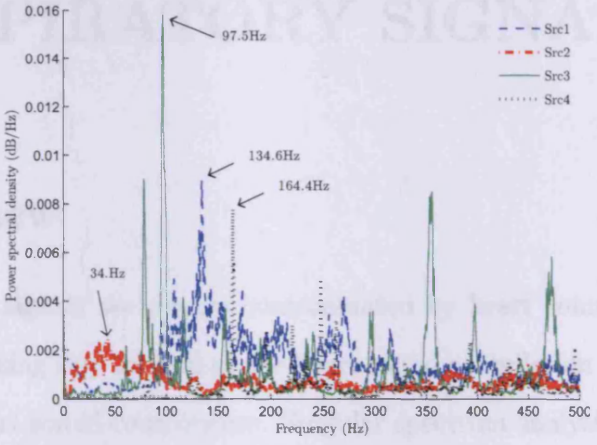


Figure 3.6: Part of power spectral densities of the original sources in Fig. 3.5.a. Arrows point to some of the frequencies corresponding to the appropriate time delays to be used in the proposed algorithm.

Chapter 4

LOCALIZING HEART SOUNDS IN RESPIRATORY SIGNALS

4.1 Overview

Respiratory sound signals are always contaminated by heart sound interference. An essential pre-processing step in most of the heart sound cancellation methods is localization of primary heart sound components. Singular spectrum analysis (SSA), a powerful time series analysis technique in separating periodic structures from non-periodic structures and noise, is used in this chapter to find a subspace that contains information about the underlying heart sound. Artificially mixed and real respiratory signals are used for evaluating the performance of the method. Selecting an appropriate length for the SSA window results in a good decomposition quality and low computational cost for the algorithm. The results of the proposed method are compared with the results of well-established methods which use wavelet transform and entropy of the signal to detect the heart sound components. The proposed method outperforms the wavelet based method [88] in terms of false detection and also correlation with the underlying heart sound. The performance of the proposed method is generally superior to that of the

entropy based method [108]. Moreover, the execution time of the former is significantly lower than that of the latter.

This chapter is organized as follows. The related work is reviewed in the next section. The theoretical background and the localization method are presented in sections 4.3 and 4.4, respectively. Experimental results are detailed in section 4.5. A summary and the concluding remarks are provided in the last section.

4.2 Related work

Auscultation has been the first and the easiest method used for years to diagnose lung abnormalities. Recordings from the chest wall are mixtures of lung sound, heart sound (HS), environmental sounds, effects of thoracic tissues, and measurement noise [49]. Although modern stethoscopes can assist in hearing the sounds more clearly, heart sounds still interfere with the respiratory sounds.

Lung sound signals exhibit wideband power spectrum, however, most of the energy is concentrated in frequencies below 200 Hz [94]. On the other hand, HS has four basic time components. The first two components (called S1 and S2) are the most fundamental ones. The other components (called S3 and S4) do not have significant amplitudes and are mostly inaudible in healthy subjects. The main frequency components of the HS are found in the range of 20-150 Hz. Peak frequencies of heart sounds are shown to be lower than those of lung sounds [13].

Removing HS signals from respiratory signals has been studied in many research works so far. The easiest way to cancel heart sounds is to highpass filter the respiratory signals. However, due to temporal and spectral overlaps of the heart and lung sounds, part of the useful signal information may be lost. Different methods based on adaptive filtering [55, 63, 101], Wavelet denoising [56, 59], time-frequency filtering [88], and modulation filtering [37] have been proposed to overcome this problem. In [46] and [87]

blind source separation methods have been used to separate heart and lung sounds from multi-channel recordings. Due to the complex nature of the mixing system common BSS methods, however, do not result in accurate separation in this problem.

An important pre-processing step in some of the HS removal methods is to find the locations of fundamental HS components. Knowing the approximate location of S1 and S2, different methods such as adaptive filtering [39, 69] and nonlinear prediction [3] can be exploited to remove the heart sounds.

In [63] a modified ECG signal was used to provide a reference for the HS. Using the amplitude of the respiratory signal, the distance between the S1 and S2 peaks, and a noise-reduction percentage criterion, a HS localizer algorithm has been proposed in [55]. Both of these methods rely on a fixed time distance between S1 and S2, which is not always a valid assumption.

In [88] the time-frequency spectrogram of the respiratory sound, calculated using continuous wavelet transform, was used to localize the HS components. The recurrence time statistic is used in [3] to detect the location of heart sounds. The idea is based on determining the dynamics of a system by its possible states using a multivariate vector space (state space) and describing the transitions between the states by vectors. The entropies of the respiratory sound segments including HS components are greater than those of the other segments. Based on this observation a localization method is proposed in [108]. Despite the good performance of this method, calculating the entropies for all segments is a time consuming task. In [4] the first HS component is localized using wavelet denoising and ECG-gated ensemble averaging. The focus of that work is on exact localization of S1 and does not deal with S2 component.

In this chapter, a new method for HS localization is proposed. The method is based on singular spectrum analysis of the respiratory signal. SSA is a well-established time series analysis method widely used in different applications. In order to select the components that provide more information about the HS components in reconstruction stage,

a set of heuristic criteria are defined. The performance of the method is examined both on synthetically mixed and real respiratory signals by evaluating localization errors, false detection percentage, correlation with the original HS, and execution time. The results of the proposed method are compared with those of other localization methods reported to have the best performance.

4.3 Singular spectrum analysis

SSA is essentially a model-free technique, which decomposes the data into a number of interpretable components including slowly varying trends, oscillatory components, and unstructured noise [52].

Assume s is a non-zero real-valued time series of length r samples, i.e. $\mathbf{s} = [s_0, s_1, ..., s_{r-1}]$, where $r > 2$. Singular spectrum analysis consists of two complementary stages: decomposition and reconstruction, and each of the stages has two consecutive steps.

1) ***Decomposition***: This stage is composed of two steps: embedding and singular value decomposition (SVD).

Embedding: In this step, the time series s is mapped to k multidimensional lagged vectors of length l,

$$\mathbf{x}_i = [s_{i-1}, s_i, ..., s_{i+l-2}]^T, \quad 1 \leq i \leq k \tag{4.3.1}$$

where $k = r - l + 1$, l is the window length ($1 \leq l \leq r$), and superscript T denotes the transpose of a vector. A proper window length totally depends on the application and the prior information about the signals of interest. Theoretically, l should be large but not greater than $r/2$ [48].

The trajectory matrix of the series s is constructed by inserting each $\mathbf{x}_i$ as the ith column of an $l \times k$ matrix, i.e.,

$$\mathbf{X} = [x_{ij}] = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k]$$
$$= \begin{bmatrix} s_0 & s_1 & s_2 & \cdots & s_{k-1} \\ s_1 & s_2 & s_3 & \cdots & s_k \\ s_2 & s_3 & s_4 & \cdots & s_{k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{l-1} & s_l & s_{l+1} & \cdots & s_{r-1} \end{bmatrix} \tag{4.3.2}$$

From (4.3.2) it is evident that the trajectory matrix has equal elements for all the diagonals $i + j = constant$, and therefore, is a Hankel matrix.

Singular value decomposition: Let $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ and assume $\lambda_1, \lambda_2, ..., \lambda_l$ are the eigenvalues of $\mathbf{S}$ in descending order; that is $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_l \geq 0$ and $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_l$ are the corresponding eigenvectors.

Projecting the time series onto the direction of each eigenvector yields the corresponding temporal principal component, i.e., $\mathbf{v}_i = \mathbf{X}^T\mathbf{e}_i/\sqrt{\lambda_i}$. The SVD of the trajectory matrix can be written as

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_d \tag{4.3.3}$$

where $d = arg\max_i\{\lambda_i > 0\}$ and $\mathbf{X}_i = \sqrt{\lambda_i}\mathbf{e}_i\mathbf{v}_i^T$. The collection $(\sqrt{\lambda_i}, \mathbf{e}_i, \mathbf{v}_i)$ is called the ith eigentriple of the SVD (4.3.3). The matrices $\mathbf{X}_i$ have rank 1.

2) *Reconstruction*: This stage has two steps: grouping and diagonal averaging.

Grouping: The grouping step divides the set of indices $\{1, 2, ..., d\}$ to m disjoint subsets $I_1, I_2, ..., I_m$. For every group $I_j = \{i_{j1}, ..., i_{jp}\}$ we have $\mathbf{X}_{I_j} = \mathbf{X}_{i_{j1}} + ... + \mathbf{X}_{i_{jp}}$. Grouping the eigentriples and expanding all matrices $\mathbf{X}_{I_j}$, (4.3.3) can be written as

$$\mathbf{X} = \mathbf{X}_{I_1} + ... + \mathbf{X}_{I_m} \tag{4.3.4}$$

There is no general rule for grouping. For each application the grouping rule totaly

depends on the special requirements of the problem and the type of the contributing signals and noise.

Diagonal averaging: In the final stage of analysis, each group is transformed into a series of length r. For a typical $l \times k$ matrix $\mathbf{Y}$ the qth element of the resulted time series, g_q, is calculated by averaging the matrix elements over the diagonal $i+j = q+2$, where i and j are the row and column indices of $\mathbf{Y}$, respectively [52].

The eigenvectors resulting from the SSA contain information about the frequency content of the data. If there is a periodic component in the data, it will be reflected in the output of the SSA as a pair of (almost) equal eigenvalues [73]. Moreover, the highest peaks in the Fourier transform of the corresponding eigenvectors are related to the frequency of the periodic component [85]. These features are used to construct a set of data driven filters [86].

4.4 Localization method

In principal component analysis (PCA) the principal components (which are the results of singular value decomposition of the covariance matrix) are categorized into signal and noise subspaces using the sorted eigenvalues. In SSA the trajectory matrix $\mathbf{X}$ is the result of embedding the original time series and therefore, in order to obtain the underlying signals, one has to carefully group the eigentriples into a number of disjoint subsets.

In real world applications of SSA usually *a priori* knowledge about the problem is used to define some heuristic criteria to reject the noise subspace and extract the trends or particular periodicities. The auscultation signal contains heart and lung sounds and noise. Applying SSA to this signal will decompose it into a number of principal components. Because of the frequency overlap of the heart and lung sounds, it is not possible to categorize the components into distinct heart and lung subspaces. Therefore, our

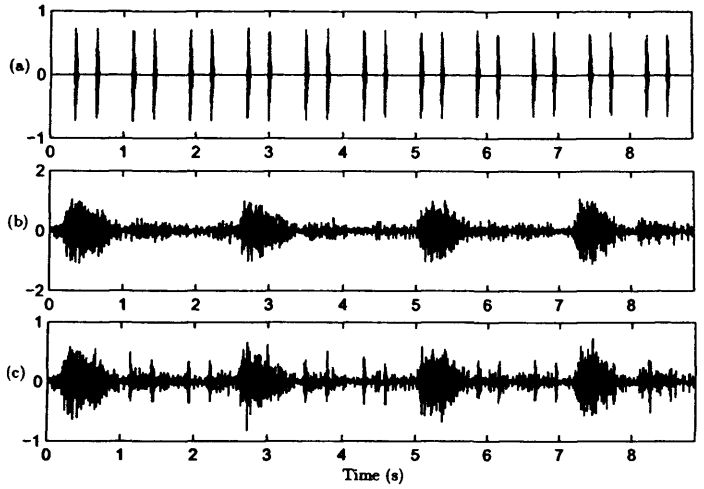


Figure 4.1: Heart and lung sound signals used to evaluate the performance of the heart sound localization method; (a) heart sound, (b) lung sound, and (c) a typical randomly generated convolutive mixture of the sounds.

approach in localization is based on finding a subspace which contains more information about the HS signal.

As each eigenvalue in the decomposition stage is equal to the variance of the signal in the direction of the corresponding principal component, one can conclude that only the largest eigenvalues belong to the signal subspace, and interpret the smallest eigenvalues as noise [73]. Investigating different recorded respiratory signals, the following criterion is defined heuristically as in [73] to reject the components most probably relating to the noise subspace.

Rejection criterion: All the components relating to the eigenvalues λ_i are rejected if $i > \mathcal{L}$, where

$$\mathcal{L} = arg\min_{a}\{\frac{\sum_{j=1}^{a}\lambda_j}{\sum_{j=1}^{l}\lambda_j} > 0.95\} \tag{4.4.1}$$

Assuming that the recording system has acceptable quality and the environmental noise is under control, the dominant part of the signal after rejecting the noise is a mixture of heart and lung sounds. Periodicity of the HS is utilized here to select a group

of principal components which best reconstruct the HS subspace. Thus, we seek for the eigenvalues which exist in pairs. However, the following facts have to be considered when selecting the pairs:

1) Noise can generate equal eigenvalue pairs,

2) The eigenvalues generated by quasi-periodic sources may not be exactly equal [73], [85].

The corresponding eigenvectors must also have almost equal frequency peaks in their spectrum and have similar waveform in time domain [85]. Therefore, in order to obtain only the periodic components, the eigenvalues λ_i and λ_j are selected as a pair if the following conditions are met [73]:

1) i and j are less than $\mathcal{L}$,

2) $|1 - \frac{\lambda_i}{\lambda_j}| < 0.05$,

3) $|1 - \frac{f_i}{f_j}| < 0.03$ and cross-correlation maximum (over different lags) of $\mathbf{e}_i$ and $\mathbf{e}_j$ is greater than 0.8 where f_i has the highest peak value in the Fourier transform of eigenvector $\mathbf{e}_i$.

As discussed in the next section, the first few eigentriple pairs represent a subspace which contains more information about the HS and hence the S1 and S2 peaks are clearly distinguishable.

At the final step, the locations of S1 and S2 components are identified in the reconstructed signal using an adaptive threshold defined as $(\mu + \sigma)$, where μ and σ are respectively mean and standard deviation of the reconstructed signal [108].

A disadvantage of SSA is its computational complexity for calculation of the SVD [5]. However, this shortcoming can be reduced by using modern and parallel algorithms for SVD. Moreover, selecting smaller values for l reduces the dimensions of the problem and therefore reduces the cost. As mentioned before, the length of the window is totaly a problem dependent task.

4.5 Experiments

The objective of our work is to detect the segments of the respiratory signal that include one of the HS components. In order to evaluate the performance of the proposed method different experiments were carried out using synthetically mixed data and real respiratory sound signals. The HS localization methods presented in [88] and [108] were also used for comparison.

HS localization in [88] is performed using multi-resolution decomposition. The wavelet transform of the original lung sound signal was calculated using Daubechies mirror filters of order 4 (db4) with eight levels. Thresholding the spectrogram of the coefficients at the fifth level results in detection of the segments which include HS. In [108], localization is performed by dividing the respiratory data into 20 ms segments with 50% overlap. The Shannon entropy of each segment is calculated as $H(p) = -\sum_{i=1}^{N} p_i \log p_i$ where $\{p_i, i = 1, ...N\}$ are the pdf of the samples and N is the number of samples in each segment. An estimate of the pdf of the samples in each segment is calculated using a Normal kernel estimator. Because of higher entropy of the heart sound included segments, they can be detected using an adaptive threshold. The outputs of the three methods are a set of windows which indicate the estimated location of each HS component. In sequel the multi-resolution decomposition and the entropy based methods are referred to as CWT and ENT, respectively

Synthetic data

The respiratory sounds recorded from the chest wall are not simple instantaneous mixtures of the heart and lung sound signals [35]. Therefore, in our experiments convolutive mixtures of the signals are used. Normal heart and lung sounds[1] were mixed after being down-sampled to 4kHz and filtered by two randomly generated FIR filters of length four. Fig. 4.1 illustrates the original sound signals and a typical convolutive

[1]Available to download from `http://www.dundee.ac.uk/medther/Cardiology/hsmur.html` and `http://www.acoustics.org/press/132nd/2aea4.html`.

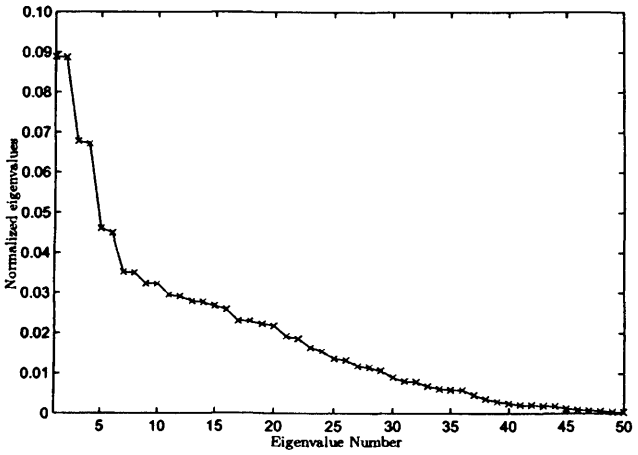


Figure 4.2: The first 50 eigenvalues normalized by cumulative sum of all the eigenvalues. The rest of eigenvalues are almost zero.

mixture.

An appropriate window length, l, is a compromise between computational cost and decomposition quality. After trying different l values, we set $l = 200$ which provides a fair decomposition and does not impose high computational cost. Before applying SSA, the simulated respiratory signal is highpass filtered at 50 Hz. Fig. 4.2 shows the eigenvalue spectra of the first fifty eigentriples. The values are normalized by the cumulative sum of all the eigenvalues. The rest of the eigenvalues (from 51 to 200) are almost zero and therefore not shown in this figure.

Inspecting the eigenvalue spectra provides some insight into the time series. The first six eigenvalues can be categorized as three almost equal pairs. Because of the significant change in the trend of the spectra, we can conclude that the first three pairs belong to a subspace different from that of the others. Therefore, we choose the first six principal components and reconstruct the corresponding time series. The reconstructed and the remaining signals are shown in Fig. 4.3. As expected, the heart and lung sound signals are not separated completely. However, the S1 and S2 peaks are clearly distinguishable in Fig. 4.3.(a) and can be localized using an adaptive threshold.

Although the time series presented in Fig. 4.3.(b) is very similar to the waveform

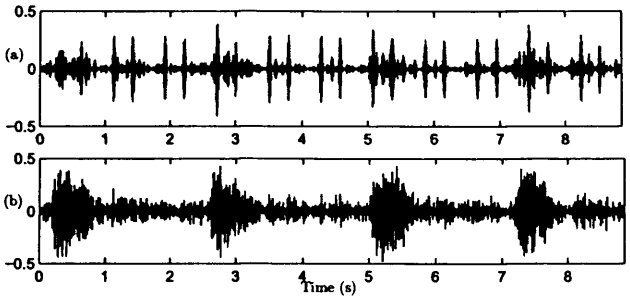


Figure 4.3: Result of SSA applied to the signal of Fig. 4.1.(c); (a) reconstructed signal using the first six eigentriples and (b) the remaining signal.

of the underlying lung sound (see Fig 4.1.(b)), there are some parts of the lung sound mixed with the HS in Fig. 4.3.(a). Investigating the power spectrum of the signal in Fig. 4.3.(b) also confirms existence of dominant frequencies that belong to the HS spectrum.

To evaluate the performance of the localization methods for different breathing flow rates, separate experiments were conducted on low and medium flow rate synthetic signals. The respiratory signals were simulated by normalizing the coefficients of the HS FIR filter to one. Each mixture was empirically interpreted as having low and medium flow rate if the norm of the lung sound filter coefficients were in the range of [0.1, 0.7) and [0.7, 1.6), respectively. 100 experiments with random filter coefficients were carried out for low and medium flow rates independently. The experiments were conducted on a PC with 3.2 GHz Pentium IV Intel CPU and 1.5GB of internal memory.

SSA, CWT [88], and ENT [108] methods were utilized to localize the HS components. Fig. 4.4 illustrates a typical medium breathing flow mixture along with the original HS and the output of different methods. A segment of the Fig. 4.4.(a) including two heart cycles is selected and represented in Fig. 4.4.(b) to show the details of the signal. Fig. 4.4.(d)-(f) illustrate the same segment of the outputs of different methods. The center of each HS component is marked by a vertical dashed line. The duration and center of the resulted windows for each method vary based on the possible overlaps with the lung sounds. If the lung sounds are strong, they may impose errors in both length and center of the window, as can be seen in Fig. 4.4. To evaluate the accuracy of

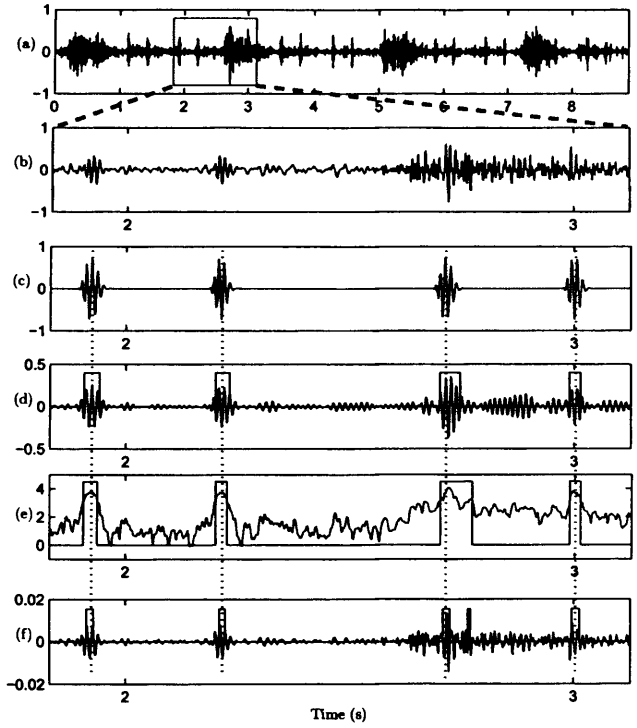


Figure 4.4: Localizing S1 and S2 in the synthetically mixed data; (a) the artificial respiratory signal. A segment of this signal (enclosed in the rectangle) is selected to compare the original heart sound and the output of localization methods, (b) the selected segment of part (a), (c) original heart sound corresponding to the segment, (d), (e), and (f) the results provided by SSA, ENT, and CWT respectively. Vertical lines show the center of the heart sound components derived from part (c).

localization methods, the difference between the center of every heart sound component in Fig. 4.1.(a) and the center of corresponding windows found by different methods are averaged over 100 experiments. Moreover, the number of false positive and negative errors were investigated manually and their averages were reported in Table 4.1. In ideal case the results of different methods have to be highly correlated with the underlying HS. Therefore, the heart sound of Fig.4.1.(a) was used as a reference and its correlations with the outputs of the localization methods were calculated for each random mixture. As the sample correlation coefficients (r_s values) are not distributed normally, under the null hypothesis that the population correlation coefficient (ρ) equals zero, the Fisher's Z transformation [75] is employed to transform the coefficients into normally distributed values, i.e. $z = \frac{\sqrt{n-3}}{2} \ln \frac{1+r_s}{1-r_s}$, where n is the number of sample correlation coefficients. Setting α (the level of significance) to 0.05, the critical values to reject the null hypothesis will be ± 1.96. The averages over all 100 independent experiments are reported in Table 4.1. High correlation values (compared with 1.96) for all the methods lead to rejection of the null hypothesis $\rho = 0$.

From Table 4.1 it can be concluded that the localization errors of low flow signals are almost in the same range for all the methods. For medium breathing flow, this parameter measures lower for CWT. SSA and ENT outperform CWT both in terms of component detection and correlation. Although the average percentage of errors in detecting the components are similar for SSA and ENT and the localization error of CWT is lower, SSA provides lower false negative values which is more desirable for the HS localization problem (as localization is a pre-processing step). Table 4.1 also indicates that the results of SSA are more correlated with the original HS. Moreover, the average execution time for 100 independent trials were 1.21 ± 0.06, 0.09 ± 0.01, and 3.62 ± 0.16 seconds for SSA, CWT, and ENT methods, respectively.

Despite the low computational cost and localization error of CWT, the number of false detections is higher in this method and the correlation of its output with the HS is

Table 4.1: SSA, CWT, and ENT methods were applied to the randomly generated convolutive mixtures of the heart and lung sound of Fig. 4.1.(c). The experiments were carried out for 100 independent trials and the results were averaged. L and M represent low and medium breathing flows, respectively.

		Localization error (samples)	False negative (%)	False positive (%)	Correlation
L.	SSA	4.79 ± 3.61	0.0 ± 0.0	0.03 ± 0.17	8.27 ± 0.73
	CWT	4.26 ± 3.09	0.08 ± 0.33	0.11 ± 0.44	6.18 ± 0.33
	ENT	4.92 ± 3.58	0.01 ± 0.10	0.04 ± 0.19	7.70 ± 0.56
M.	SSA	21.77 ± 33.25	0.27 ± 0.14	0.31 ± 0.54	5.75 ± 0.67
	CWT	19.56 ± 15.24	1.78 ± 0.83	1.16 ± 0.74	4.02 ± 0.26
	ENT	23.11 ± 34.02	0.33 ± 0.2	0.30 ± 0.48	5.13 ± 0.56

lower. On the other hand, ENT provides less false detection rates and higher correlation, however, the computational cost is considerably high. The computational cost and the localization error of SSA are less than those of ENT, and at the same time this method has the least false negative detection and the most correlated output with the underlying HS.

Respiratory data

The proposed HS localization methods were applied to respiratory signals recorded from three healthy subjects (male, aged 28 to 32 years) with no known cardiac or pulmonary problem. A Welch Allyn Tycos stethoscope was used to record the sounds from the chest. The subjects were reposed in a comfortable position and the stethoscope was fixed over the chest in the left midclavicular area, 2nd intercostals space. The subjects were trained and asked to breath at low or medium flow rate during each experiment. For each subject the respiratory data was recorded for 3 minutes, amplified, and digitized at 44 kHz. The respiratory signals were then down-sampled to 4 kHz and highpass filtered at 50Hz. A 5.32 second segment of one of the respiratory signals is illustrated in Fig. 4.5.(a).

The length of SSA window was set to 200. Using the rejection criterion (4.4.1) more than 150 of the eigentriples were rejected. In order to find the component pairs which have almost equal eigenvalues and frequency peaks the eigen spectrum of three subjects

Table 4.2: False component detection for each of the methods on low and medium breathing flow rate respiratory signals. The values are the averages of the results obtained from the respiratory sounds of three different healthy subjects.

		False negative (%)	False positive (%)
Low flow	SSA	0.41 ± 0.09	0.90 ± 0.13
	CWT	1.10 ± 0.45	1.92 ± 0.72
	ENT	0.40 ± 0.12	0.93 ± 0.19
Medium flow	SSA	1.79 ± 0.52	1.86 ± 0.61
	CWT	4.38 ± 0.75	5.23 ± 1.12
	ENT	1.86 ± 0.60	2.07 ± 0.97

at low and medium flow rates were analyzed and the correlation of the eigenvectors were also evaluated. Therefore, the first six eigentriples were selected. The eigenvectors of the selected pairs are illustrated in Fig. 4.6. The high correlation between eigenvector pairs 1-2, 3-4, and 5-6 is clearly distinguishable.

The principal components corresponding to the selected pairs were reconstructed and using the defined adaptive threshold the locations of the heart components were identified. The result is presented in Fig. 4.5.(b). Fig. 4.5.(c) and (d) respectively illustrate the results of the ENT and CWT methods. In order to provide a quantitative performance measure, the number of false negative and positive errors of the proposed method and the two other methods are presented in Table 4.2. The values are averaged for three subjects in low and medium flow rates.

The execution times averaged over three subjects were 8.09 ± 0.31, 0.72 ± 0.12, and 24.13 ± 1.57 seconds for SSA, CWT, and ENT methods, respectively.

The difference between the results of ENT and SSA for low breathing flow rates is negligible. For medium flow rates, SSA generally provides better results than ENT. Moreover, both of the methods provide lower false negative results for both low and medium flow rates than CWT does.

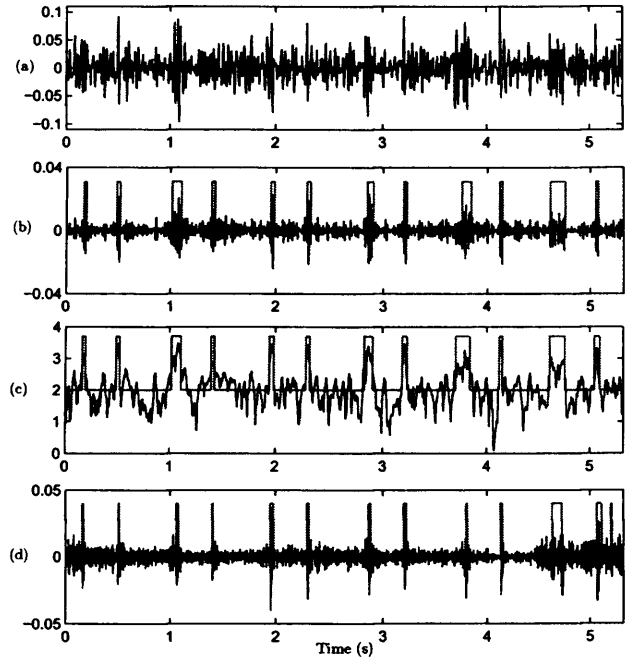


Figure 4.5: A segment of one of the respiratory signals along with the results of different localization methods; (a) the original signal, (b) result of SSA by reconstructing the output using the first six eigentriples, (c) localization using entropy based method (ENT), and (d) result of multi-resolution decomposition (CWT).

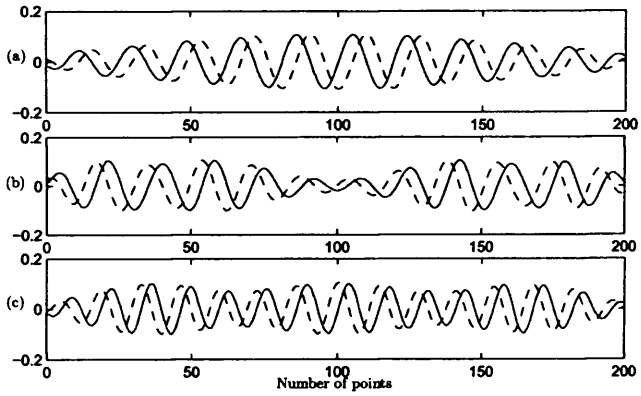


Figure 4.6: The first six eigenvectors. (a) eigenvalue pairs 1, 2, (b) pairs 3, 4, and (c) pairs 5, 6.

4.6 Summary and conclusions

Singular spectrum analysis was used in this chapter to construct a data driven filter to localize the primary HS components in the single channel auscultation signal. The main advantage of using SSA, particularly compared with conventional PCA, is its power in extracting periodic components in the time series. Investigating the characteristics of different principal components and the resulting reconstructed signals, the required criteria for selecting the best subspace were defined.

The error in localization and the correlation between the original heart sound and the outputs of different methods were used to quantitatively evaluate the performance of the methods on synthetic data. False negative and positive detections were used as a measure which can be used for both synthetic and real world signals.

Among different steps of the singular spectrum analysis, SVD is the one with the highest computational cost. Therefore, selecting a suitable window length is of great importance. On the other hand, ENT is a segment based approach. That means, for each segment of the data pdf estimation is required, which imposes a high computational cost. One would suggest to increase the length of the segments in order to reduce the computational cost. This approach cannot help, as it affects the estimated entropies.

In order to quantitatively evaluate the performance of the proposed method normal heart and lung sounds were mixed to generate artificial respiratory signals. The average numbers of false detection by the proposed and the entropy based methods were better than that by the multi-resolution decomposition method. Comparing the results of different methods with the original HS, it is concluded that the result of the proposed method is more correlated with the HS. CWT method resulted in the least localization error and had the lowest computational cost among the three methods. The drawback of this method was its high false detection which is very important when evaluating the performance of a typical HS localization pre-processing method. On the other hand, SSA generally performed better than ENT in terms of false detection, correlation between

HS and the output of the methods, and error in localization while it was much faster.

Real world respiratory signals from three different subjects recorded in low and medium breathing flow rates were also used to evaluate the methods. The results were almost similar to the results when synthetic signals were used, where the lower values of false negative detection were provided by SSA and ENT and the execution time of ENT was the largest among the others.

Chapter 5

CYCLOSTATIONARY SOURCE EXTRACTION

5.1 Overview

There are many processes in nature that originate from periodic phenomena. These processes may generate some random data that represent periodicity in their statistical properties, and are often called cyclostationary processes. Examples of cyclostationary data can be found in telecommunication, radar and sonar applications, mechanics, econometrics, biological science, and meteorology [43], see Chapter 2 for more information about cyclostationarity.

Real world observations are often mixtures of different signals plus noise and researchers in different application areas try to develop methods that can blindly estimate the original sources. In these problems the original sources and the mixing mechanism are generally unknown and only weak assumptions about them are available.

Novel methods for blind source extraction of cyclostationary sources are developed and presented in this chapter. It is assumed that the cycle frequencies of the sources are known *a priori*. Different cases of distinct and common cycle frequencies are considered and necessary and sufficient conditions are introduced. In order to estimate the sources, steepest descent method for complex matrices (which stems from differential geometry) and extended Jacobi technique for diagonalization of complex matrices are used. The

proposed algorithms are applied to simulated data and the effectiveness and performance of the algorithms are verified.

5.2 Related work

Second and higher order statistics are widely used in BSS problems and many algorithms have been developed based on such information. A typical algorithm performs a pre-whitening operation followed by a rotation to estimate the parameters.

SOBI algorithm is a well known method in which several time-delayed covariance matrices of observations are jointly diagonalized [16]. Joint approximate diagonalization method is applied to forth-order cumulant matrices in [24] to estimate a set of narrow band sources from their mixtures. Details of these well-established methods are presented in Chapter 2. In [65] the periodicity of the correlation matrix of cyclostationary signals is used to separate periodic signals. In this attempt, the covariance matrices have been diagonalized in different lags. Authors of [1] have presented two iterative and two algebraic methods to separate and extract the cyclostationary sources.

Convolutive mixtures of cyclostationary signals have been considered as well. In [104] the conventional higher-order statistics and second-order cyclostationarity of the signals are used and it is shown that exploitation of the statistical properties of this kind of signals can enhance the performance of blind cyclostationary source separation.

In this chapter extraction of cyclostationary sources is considered in two different cases: 1) sources with distinct cycle frequencies and 2) sources with common cycle frequencies. In the case of distinct cycle frequencies, an extraction algorithm is proposed which exploits cyclostationarity of the source signals to extract them. Subject to presence of *a priori* knowledge about the cyclic frequencies, it is possible to extract all the cyclostationary sources one by one [45].

In the second case, it is assumed that the cyclostationary sources can have common cyclic frequencies. The necessary and sufficient conditions to blindly extract the

cyclostationary sources are introduced and based on the extended Jacobi method for diagonalization of complex matrices [25] the extraction algorithm is developed.

In this chapter an overview of two diagonalization methods for complex matrices is introduced first. Next, the fundamental theory of blind cyclostationary extraction and the proposed algorithms are provided. Finally, the time varying cycle frequency problem is discussed and the results of the experiments on synthetic data are reported.

5.3 Diagonalization of complex matrices

Before going more into the details of BSE methods let first introduce some techniques which are used in sequel. Regardless of the original signals being complex or real, the definition of the cyclic correlation matrix of equation (2.6.18) transfers the data to complex domain. Therefore, selecting appropriate data manipulation techniques is of great importance.

Matrix diagonalization is a technique used widely in the BSS context. The problem can be formulated as follows:

Assume $\mathbf{A} = [a_{ij}]$ is an $n \times n$ complex matrix. $\mathbf{A}$ is diagonal if the norm of $\mathbf{A}$ is not zero and $\text{off}\{\mathbf{A}\} = 0$, where operator $\text{off}\{.\}$ computes the sum of squared magnitudes of the off-diagonal elements of the enclosed term, i.e, $\text{off}\{\mathbf{A}\} = \sum_{1 \le i \ne j \le n} |a_{ij}|^2$. If $\mathbf{A}$ is inherently a non-diagonal symmetric matrix it is possible to diagonalize it as much as possible by defining the function $f(\mathbf{B}) = \mathbf{B}^H\mathbf{A}\mathbf{B}$ and minimizing the cost function $\mathcal{J}(\mathbf{B}) = \text{off}\{f(\mathbf{B})\}$, where $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ and $\mathcal{J} : \mathbb{C}^{n\times n} \to \mathbb{R}$ $\mathbf{B}$ is a unitary matrix.

In both of the extraction methods considered in this chapter, we have a diagonalization problem. Two well established methods are introduced here and used in developing the extraction methods. Both of these methods result in unitary diagonalization which is a requirement of the proposed extraction methods. Steepest descent algorithm for complex matrices is an optimization method that can be used for diagonalization by

defining the appropriate cost function. Extended Jacobi method is essentially an iterative algorithm for simultaneous diagonalization of several complex matrices.

5.3.1 Steepest Descent algorithm

In order to find the minimum point of $\mathcal{J}(\mathbf{B})$, the steepest descent method for complex matrices [2] can be used. This method stems from differential geometry and optimizes the cost function under unitary constraint. It is shown in [2] that to find the optimum point of a cost function under unitary constraint on the Riemannian space it is enough to evaluate the gradient of the cost function at a point $\mathbf{B}$ and then translate it to identity. In mathematical terms, if $\nabla\mathcal{J}(\mathbf{B})$ is the gradient of the cost function $\mathcal{J}(\mathbf{B})$ on Euclidean space, then the gradient direction on the Riemannian space is:

$$\widehat{\nabla}\mathcal{J}(\mathbf{B}) = \nabla\mathcal{J}(\mathbf{B})\mathbf{B}^H - \mathbf{B}\nabla\mathcal{J}(\mathbf{B})^H. \tag{5.3.1}$$

The Riemannian steepest descent algorithm is therefore as follows [2]:

Algorithm 2

$k \leftarrow 0$, $\mathbf{B}_k \leftarrow \mathbf{I}$
repeat
  Compute the Euclidean gradient of the cost function: $\nabla\mathcal{J}(\mathbf{B}_k)$
  Compute the gradient direction on the Riemannian space: $\widehat{\nabla}\mathcal{J}(\mathbf{B}_k)$
  Compute the rotation matrix $\mathbf{P}_k = \exp(-\mu\widehat{\nabla}\mathcal{J}(\mathbf{B}_k))$
  $\mathbf{B}_{k+1} \leftarrow \mathbf{P}_k\mathbf{B}_k$, $k \leftarrow k+1$
until $\|\nabla\mathcal{J}(\mathbf{B}_k)\| \leq Tr$

where μ is the learning rate, Tr is a pre-defined threshold level, and the exponential of an $n \times n$ complex matrix $\mathbf{M}$ is given by the convergent power series $\exp(\mathbf{M}) = \sum_{m=0}^{\infty}(\mathbf{M}^m/m!)$.

5.3.2 Extended Jacobi method

The other approach to find the minimum point of the cost function $\mathcal{J}(\mathbf{B})$ is the extended Jacobi technique [25] which can be used to diagonalize the complex matrices. In this

method, which is originally developed for simultaneous diagonalization of a set of normal commuting matrices (i.e., $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_K\}$), a unitary matrix $\mathbf{U}$ is sought such that it minimizes the cost function $\sum_{k=1,K} \text{off}\{\mathbf{U}\mathbf{A}_k\mathbf{U}\}$. The complex rotation matrix $\mathbf{V}(i,j,c,s) = v_{pq}$ is defined to be equal to identity matrix except for the following entries:

$$\begin{bmatrix} v_{ii} & v_{ij} \\ v_{ji} & v_{jj} \end{bmatrix} = \begin{bmatrix} c & s^* \\ -s & c^* \end{bmatrix} \quad c, s \in \mathbb{C} \ \ and \ \ |c|^2 + |s|^2 = 1 \tag{5.3.2}$$

In each iteration of the diagonalization algorithm and for each selection of i and j, it is desired to find the complex angles c and s which minimize the cost function

$$f(c,s) = \sum_{k=1}^{K} \text{off}\{\mathbf{V}(i,j,c,s)\mathbf{A}_k\mathbf{V}(i,j,c,s)^H\} \tag{5.3.3}$$

It is shown in [25] that the solution for (5.3.3) is

$$c = \sqrt{\frac{x+r}{2r}} \quad s = \frac{y - jz}{\sqrt{2r(x+r)}} \quad r = \sqrt{x^2+y^2+z^2} \tag{5.3.4}$$

where $[x,y,z]^T$ is any eigenvector associated with the largest eigenvalue of $\mathbf{G}$, $\mathbf{G} = real(\sum_{k=1,K} h^H(\mathbf{A}_k)h(\mathbf{A}_k))$, and $h(\mathbf{A}) = [a_{ii} - a_{jj}, a_{ij} + a_{ji}, j(a_{ji} - a_{ij})]$. In each iteration, i and j can be selected cyclic by row (or by column), i.e., for the first row, select all the columns and then carry on with the other rows and the rest of the columns [51]. Details of the method are presented in Algorithm 3.

Algorithm 3

```
B ← I
while err > eps do
  for p = 1 to n − 1 do
    for q = p+1 to n do
      (c,s) = Jacobi(A, p,q)
      B = BV(p, q, c, s)
    end for
  end for
end while
```

where $\mathcal{A}$ is the input to the algorithm, *err* is sum of the errors, *eps* is the maximum allowed level of error (here *err* is a measure of nondiagonality of the cyclic correlation

matrix), and *Jacobi* is a function which calculates complex angles, c and s, based on the values of $\boldsymbol{A}$, p, and q as inputs.

5.4 Cyclostationary source extraction

Assume m unknown mutually statistically independent sources are mixed instantaneously through an unknown medium and n sensors ($n \geq m$) are used to measure these signals. This system can be formulated in a vector form as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{5.4.1}$$

where $\mathbf{s}(t) = [s_1(t), ..., s_m(t)]^T$ is an $m \times 1$ source vector, $\mathbf{n}(t) = [n_1(t), ..., n_n(t)]^T$ is an $n \times 1$ stationary noise vector, $\mathbf{x}(t) = [x_1(t), ..., x_n(t)]^T$ is an $n \times 1$ observation vector, $\mathbf{A}$ is an $n \times m$ unknown full column rank mixing matrix. Furthermore, to simplify the problem it is assumed that m=n.

It is also assumed that some of the source signals are second order cyclostationary in wide sense.[1] The objective here is to estimate a set of cyclostationary sources assuming that their cycle frequencies are known *a priori*.

As mentioned earlier in Chapter 2 if $s(t)$ is cyclostationary, $s(t)$ and its frequency-shifted version $s(t)e^{j\beta t}$ are correlated. Using the definition of cyclic cross-correlation function, equation (2.6.18), the following properties are concluded:

$$\begin{aligned} r_{pq}^{\beta_p}(\mathbf{s}, \tau) &= \langle s_p(t) s_q^\star(t+\tau) e^{j\beta_p t} \rangle = 0, \quad \text{if } p \neq q \\ r_{pp}^{\beta_q}(\mathbf{s}, \tau) &= \langle s_p(t) s_p^\star(t+\tau) e^{j\beta_q t} \rangle = 0, \quad \text{if } \beta_p \neq \beta_q \\ r_{pp}^{\beta_p}(\mathbf{s}, 0) &= \langle s_p(t) s_p^\star(t) e^{j\beta_p t} \rangle \neq 0, \qquad \forall p \end{aligned} \tag{5.4.2}$$

where β_i is the cycle frequency of the source $s_i(t)$. Based on the assumption about the stationarity of the noise vector it can also be concluded that $\mathbf{R}_{\mathbf{n}}^{\beta}(\tau) = \langle \mathbf{n}(t)\mathbf{n}^H(t+$

[1]For definition of wide sense cyclostationarity see section 2.6.

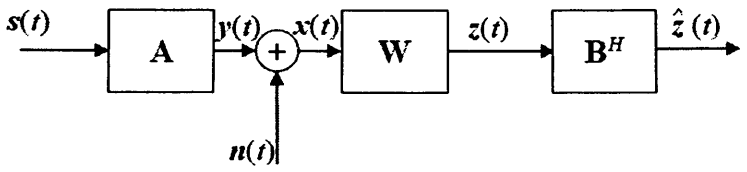


Figure 5.1: The overall block diagram of the BSS problem.

$\tau)e^{j\beta t}\rangle = \mathbf{0}$ $\forall \tau, \beta$, [40]. These properties are used here to blindly extract the cyclostationary sources of interest. The problem is considered in two different cases of sources with distinct and common cycle frequencies.

5.4.1 Distinct cycle frequencies

To estimate the original sources, the observations are first pre-whitened to obtain $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{C}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t)$, where $\mathbf{C} = \mathbf{W}\mathbf{A}$ (see Fig. 5.1). To find an estimation of the source signals we need to rotate the whitened data by a matrix $\mathbf{B}$, which will generate $\hat{\mathbf{z}}(t) = \mathbf{B}^H\mathbf{z}(t)$. In [24], fourth order cumulants are used to find the rotation matrix for non-Gaussian sources. Here we exploit the cyclostationarity property of the sources to estimate and extract the original sources.

In the following theorem the necessary and sufficient conditions for extracting one cyclostationary source from the set of observations are considered.

Theorem 1. Assume $\mathbf{z}(t)$ is a white mixture of cyclostationary sources with distinct cycle frequencies β_i, $i \in \{1, ..., n\}$. For any matrix $\mathbf{B}$, define $\hat{\mathbf{z}}(t) = \mathbf{B}^H\mathbf{z}(t)$. Then $\mathbf{b}_i$ (column i of $\mathbf{B}$) is an extracting vector of the pth source if and only if $|r_{ii}^{\beta_p}(\hat{z}, 0)| = 1$ and $r_{ij}^{\beta_p}(\hat{z}, 0) = 0$ for all $j \in \{1, ..., i-1, i+1, ..., n\}$.

Proof. Define $\mathbf{D} = \mathbf{B}^H\mathbf{C} = [d_{pq}]$, so that $\hat{\mathbf{z}}(t) = \mathbf{B}^H\mathbf{z}(t) = \mathbf{B}^H\mathbf{W}(\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t))$. Due to independence of the source signals, stationarity of the noise signals, and the assumption that the cyclic autocorrelation of each source is equal to one, the elements of the cyclic

correlation matrix can be simplified to

$$r_{pq}^{\beta_p}(\hat{\mathbf{z}},0) = d_{pp}d_{qp}^{\star}\langle \mathbf{s}_p(t)\mathbf{s}_p^H(t)e^{j\beta_p t}\rangle = d_{pp}d_{qp}^{\star}.$$

B is a separating matrix if and only if **D** is a unitary diagonal matrix. Therefore, to prove the theorem it is enough to show that the conditions of the theorem are equivalent to **D** being unitary and diagonal.

If **D** is unitary and diagonal, then

$$d_{pp}d_{qp}^{\star} = \begin{cases} 0 & \text{if } p \neq q \\ |d_{pp}|^2 = 1 & \text{if } p = q \end{cases}.$$

That means the conditions of the theorem are satisfied. Conversely, if the conditions are true then $|d_{pp}|^2 = 1$, and $d_{pp}d_{qp}^{\star} = 0$ for all $p \neq q$, i.e, **D** is a unitary and diagonal matrix. □

Therefore, if a matrix **B** can be found such that the conditions of Theorem 1 are satisfied, the mixed sources can be extracted one by one. To simplify the calculations, it is assumed that the sources are real, and it is easy to extend the results to complex case. Fourier series expansion of the periodic sources exist and for the vth source it can be written as:

$$s_v(t) = \alpha_{v0} + \sum_{l=1}^{L} \alpha_{vl}\cos(2\pi l f_v t + \phi_l) \tag{5.4.3}$$

where f_v is the fundamental frequency of source v, and L is the number of coefficients in the Fourier series expansion. It is possible to assume that the source signals are zero mean, therefore, $\alpha_{v0} = 0$. To calculate $r_{pq}^{\beta}(\hat{\mathbf{z}},0)$ we first evaluate $\hat{z}_p(t)\hat{z}_q(t)$ for all p and

q as

$$\begin{aligned}\hat{z}_p(t)\hat{z}_q(t) &= \sum_{h=1}^{n} d_{ph}s_h(t) \sum_{g=1}^{n} d_{qg}s_g(t) \\ &= \sum_{h=1}^{n}\sum_{g=1}^{n} d_{ph}d_{qg} \sum_{l=1}^{L}\sum_{e=1}^{E} \alpha_{hl}\alpha_{ge} \\ &\quad \times \cos(2\pi l f_h t + \phi_l)\cos(2\pi e f_g t + \phi_e)\end{aligned} \tag{5.4.4}$$

From basic trigonometry we know that

$$\begin{aligned}&\cos(2\pi l f_h t + \phi_l)\cos(2\pi e f_g t + \phi_e) = \\ &\quad \frac{1}{2}[\cos(2\pi(l f_h + e f_g)t + \phi_l + \phi_e) + \\ &\quad \cos(2\pi(l f_h - e f_g)t + \phi_l - \phi_e)]\end{aligned} \tag{5.4.5}$$

Substituting (5.4.4) and (5.4.5) in the cyclic cross correlation function (defined in (2.6.18)) and remembering the Fourier transform properties of cosine function and ignoring the frequency component at $\beta = 0$ it can be concluded that:

- For the vth estimated signal, $\hat{z}_v(t)$, the cyclic auto-correlation function $r^{\beta}_{vv}(\hat{\mathbf{z}}, 0)$ has nonzero values for all $\beta_v \in \mathcal{F}_v$, where $\mathcal{F}_v = \{\beta | \beta = k f_v; \forall k \in \{1, .., 2L\}, \beta \neq 0\}$. β_v is a cyclic frequency of $\hat{z}_v(t)$.

- For any two estimated signals, $\hat{z}_p(t)$ and $\hat{z}_q(t)$, the cyclic cross-correlation function $r^{\beta}_{pq}(\hat{\mathbf{z}}, 0)$ can have nonzero values at cyclic frequencies β, where $\beta \in \{l f_1 \pm e f_1, ..., l f_1 \pm e f_n, ..., l f_n \pm e f_1, ..., l f_n \pm e f_n, \beta \neq 0\}$

In other words, the spectrum of the quadratic function $\hat{z}_p(t)\hat{z}_q(t)$ has nonzero components at some β, where the values of these components are composed of two parts. The first one is resulted from a cosine term with frequency $f = \beta/2$ and the second one comes from the values of cross components which are created by mutually multiplying the

harmonics of the source signals while their harmonic frequencies satisfy:

$$lf_h \pm ef_g = \beta \tag{5.4.6}$$

For the ith source with cycle frequency β_i, provided that the source signals satisfy

$$\forall h, g \neq i \qquad lf_h \pm ef_g \neq \beta_i \tag{5.4.7}$$

if a matrix B is found such that for $\beta_i \in \mathcal{F}_i$

$$\text{off}\{\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_i}(0)\} = 0 \tag{5.4.8}$$

it means that the cosine term with the frequency of $f = \beta_i/2$ exists in just one row of $\hat{\mathbf{z}}(t)$. In other words, $\mathbf{B}$ can extract a cosine term with frequency β_i in the source of interest, i.e., $s_i(t)$. All the terms in the Fourier series expansion of the source $s_i(t)$ are multiplied by the same elements of the matrices $\mathbf{A}$, $\mathbf{W}$, and $\mathbf{B}$ and therefore, outcome of the extraction process is the complete source $s_i(t)$.

To extract the ith source we can relax the condition of Theorem 1 and simply diagonalize the cyclic correlation matrix of the estimation vector.

Although (5.4.7) imposes a restriction, it provides some redundancy in selecting the appropriate cycle frequency. In other words, we don't need to select the fundamental cycle frequency (the one corresponding to the fundamental frequency) as the required cycle frequency of the source. As long as the cycle frequencies satisfy (5.4.7) there is no more restriction in the selection.

Algorithm

The objective is to find a matrix $\mathbf{B}$ for which (5.4.8) is satisfied. In order to find $\mathbf{B}$, Algorithm 2 is used. The following cost function is defined, which measures nondiagonality

of the cyclic correlation matrix of $\hat{\mathbf{z}}(t)$:

$$\mathcal{J}(\mathbf{B}) = \text{off}\{\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)\} = \text{off}\{\mathbf{B}^H \mathbf{R}_{\mathbf{z}}^{\beta_p}(0)\mathbf{B}\}. \tag{5.4.9}$$

For our problem, the Euclidean gradient of $G(\mathbf{B})$ is

$$\begin{aligned} \nabla\mathcal{J}(\mathbf{B}) &= \frac{\partial \mathcal{J}(\mathbf{B})}{\partial \mathbf{B}^H} \\ &= \mathbf{R}_{\mathbf{z}}^{\beta_p}(0)\mathbf{B}\left(\mathbf{B}^H \mathbf{R}_{\mathbf{z}}^{\beta_p}(0)\mathbf{B} - diag(\mathbf{B}^H \mathbf{R}_{\mathbf{z}}^{\beta_p}(0)\mathbf{B})\right). \end{aligned} \tag{5.4.10}$$

Therefore, the minimizer of (5.4.9) can be found using Algorithm 2 and applying $\mathbf{R}_{\mathbf{z}}^{\beta_p}(0)$ to the input.

Deflating sources of interest

Upon convergence of the above algorithm, $\hat{z}_p(t)$ is an estimaton of the pth source. It is notable to mention that due to permutation ambiguity, the source of interest may appear in a different row of $\hat{\mathbf{z}}(t)$, but it has the same properties. To deflate the extracted signal we use the deflation method proposed in [46] which is based on minimizing the cyclic correlation between the extracted source and the remaining mixtures. In this case $\mathbf{z}_{p+1}(t) = \mathbf{z}_p(t) - \tilde{\mathbf{b}}_p \hat{z}_p(t)$ and $\tilde{\mathbf{b}}_p$ is the $n \times 1$ deflating column vector:

$$\tilde{\mathbf{b}}_p = \frac{\langle \mathbf{z}_p(t)\mathbf{z}_p^H(t)e^{j\beta_p t}\rangle \mathbf{b}_p}{\langle \hat{z}_p^2(t)e^{j\beta_p t}\rangle} \tag{5.4.11}$$

where $\mathbf{b}_p$ is a column in $\mathbf{B}$ corresponding to the pth source.

5.4.2 Common cycle frequencies

So far, we assumed that the cyclostationary sources have distinct cycle frequencies. In this subsection we consider a more complicated case of when a number of sources have a common cycle frequency. This case happens when the cyclostationary sources

are originated from the same periodic phenomena inside the system. Such a system is governed by a model which follows equation (5.4.1) while it is assumed that d sources have a common cycle frequency.

In [1] a condition for identifiability of cyclostationary sources with common cycle frequency is presented and proved as follows:

Identifiability condition:

If $\rho_p(\mathbf{s}, \tau)$ is the cyclic autocorrelation function of the pth source signal, the $1 \times (K+1)$ cyclic autocorrelation vector is defined as $\boldsymbol{\rho}_p(\mathbf{s}) = [\rho_p(\mathbf{s}, 0), \rho_p(\mathbf{s}, \tau_1), ..., \rho_p(\mathbf{s}, \tau_K)]$. Separation of the cyclostationary sources can be achieved if and only if two distinct sources $s_p(t)$ and $s_q(t)$ do not exist whose cycle frequencies are the same and whose cyclic autocorrelation vectors $\boldsymbol{\rho}_p(\mathbf{s})$ and $\boldsymbol{\rho}_q(\mathbf{s})$ are linearly dependent.

Proof. without loss of generality it can be assumed that the source vector $\mathbf{s}(t)$ is such that $\beta_1 = \beta_2$ and $\boldsymbol{\rho}_1(\mathbf{s}) = \boldsymbol{\rho}_2(\mathbf{s})$. For any mixing matrix $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_m]$, define another mixing matrix, $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \mathbf{a}_3, ..., \mathbf{a}_m]$, where

$$[\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2] = [\mathbf{a}_1, \mathbf{a}_2] \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Similarly, define another source vector, $\tilde{\mathbf{s}}(t) = [\tilde{s}_1(t), \tilde{s}_2(t), s_3(t), ..., s_m(t)]^T$, where

$$\begin{bmatrix} \tilde{s}_1(t) \\ \tilde{s}_2(t) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix}.$$

It can be easily verified that the output vectors $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)$ and $\tilde{\mathbf{x}}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(t) + \mathbf{n}(t)$ have the same cyclic correlation matrices at time delays $0, \tau_1, ..., \tau_K$ as well as the source vectors $\mathbf{s}(t)$ and $\tilde{\mathbf{s}}(t)$. □

In order to extract the sources of interest, the observations are first pre-whitened by a matrix $\mathbf{W}$ to obtain $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$, where $\mathbf{C} = \mathbf{W}\mathbf{A}$. To find an estimation of

the source signals the whitened data is then rotated by a matrix $\mathbf{B}$, which will generate $\hat{\mathbf{z}}(t) = \mathbf{B}^H\mathbf{z}(t)$. The objective is to find an extracting matrix $\mathbf{B}$ such that d elements of $\hat{\mathbf{z}}(t)$ are estimations of the cyclostationary sources of interest.

Each entry of the cyclic correlation matrix of $\hat{\mathbf{z}}$ at a typical cycle frequency β can be written as:

$$\begin{aligned} r^{\beta}_{pq}(\hat{\mathbf{z}}, \tau) &= \langle \hat{z}_p(t)\hat{z}^*_q(t+\tau)e^{j\beta t}\rangle \\ &= \mathbf{b}_p^H\langle \mathbf{z}(t)\mathbf{z}^*(t+\tau)e^{j\beta t}\rangle \mathbf{b}_q \\ &= \mathbf{b}_p^H\mathbf{C}\langle \mathbf{s}(t)\mathbf{s}^*(t+\tau)e^{j\beta t}\rangle \mathbf{C}^H\mathbf{b}_q \end{aligned} \tag{5.4.12}$$

where $\mathbf{b}_i$ is the ith column of matrix $\mathbf{B}$. If the cycle frequencies of the sources are distinct and known *a priori*, Theorem 1 considers the necessary and sufficient conditions for extracting one cyclostationary source from the set of observations. This theorem is extended here to the more complicated case of the sources with common cycle frequencies.

Assume that d sources have the same cycle frequency β_p and satisfy the identifiability condition. We exploit the scaling ambiguity of ICA methods to assume that the sources are unit norm and hence the cyclic correlation matrix of $\mathbf{s}_p(t) = [s_{p_1}(t), s_{p_2}(t), ..., s_{p_d}(t)]$ at $\tau = 0$ is equal to $\mathbf{I}_d$, where $\mathbf{I}_d$ is a $d \times d$ identity matrix. In other words, cyclic correlation matrix of $\mathbf{s}(t)$ at $\tau = 0$ obeys:

$$\mathbf{R}^{\beta_p}_{\mathbf{s}}(0) = \begin{bmatrix} \mathbf{I}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{5.4.13}$$

This property is used in the following theorem to find the extracting matrix:

Theorem 2. Assume $\mathbf{z}(t)$ is a white instantaneous mixture of independent sources with stationary additive noise and the sources satisfy the identifiability condition. $\mathbf{s}_p(t) = [s_{p_1}(t), s_{p_2}(t), ..., s_{p_d}(t)]$ is a vector of d sources which are all cyclostationary with common cycle frequency β_p. Define $\hat{\mathbf{z}}(t) = \mathbf{B}^H\mathbf{z}(t)$, then, any unitary matrix $\mathbf{B}$ is an extracting

matrix of $\mathbf{s}_p(t)$ if and only if for d columns (rows) of $\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)$ the off-diagonal elements are zero while the diagonal elements are not.

Proof. To simplify the calculations we prove the theorem for the case of $m = n = 3$, while two of the sources have common cycle frequencies. The proof can be generalized to higher dimensions and more numbers of cyclostationary sources.

We first prove sufficiency of the theorem. From (5.4.13) and the assumption that two of the sources have a common cycle frequency, it is known that the cyclic correlation matrix of $\mathbf{s}(t)$ has two nonzero diagonal elements while all other elements are zero . Without loss of generality we can assume that the two first elements of this matrix are 1. Therefore,

$$\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0) = \mathbf{B}^H\mathbf{C}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\mathbf{C}^H\mathbf{B}. \tag{5.4.14}$$

Expanding (5.4.14) results in

$$r_{ij}^{\beta_p}(\hat{z}, 0) = \mathbf{b}_i^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{b}_j + \mathbf{b}_i^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{b}_j. \tag{5.4.15}$$

If $\mathbf{B}$ is an extracting matrix, $\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)$ must be diagonal and since its rank is 2, two of its diagonal elements are nonzero, while all the off-diagonal elements in each column of $\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)$ are zero, i.e., $\forall i \neq j$

$$\begin{aligned} &\mathbf{b}_i^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{b}_j + \mathbf{b}_i^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{b}_j = \\ &\mathbf{b}_i^H(\mathbf{c}_1\mathbf{c}_1^H\mathbf{b}_j + \mathbf{c}_2\mathbf{c}_2^H\mathbf{b}_j) = \mathbf{b}_i^H(\alpha_{1j}\mathbf{c}_1 + \alpha_{2j}\mathbf{c}_2) = 0 \end{aligned} \tag{5.4.16}$$

where $\alpha_{1j} = \mathbf{c}_1^H\mathbf{b}_j$ and $\alpha_{2j} = \mathbf{c}_2^H\mathbf{b}_j$ are two scalar values that form a linear combination of $\mathbf{c}_1$ and $\mathbf{c}_2$. The equality in equation (5.4.16) is met in only one of the following cases:

- $\mathbf{b}_i = 0$. $\mathbf{B}$ is a unitary matrix, so this can never happen.

- $\alpha_{1j}\mathbf{c}_1 + \alpha_{2j}\mathbf{c}_2 = 0$. This is not possible, unless $\alpha_{1j} = 0$ and $\alpha_{2j} = 0$.

- $\mathbf{b}_i \perp \alpha_{1j}\mathbf{c}_1 + \alpha_{2j}\mathbf{c}_2 \ \forall i, 1 < i \leq n = 3$. This means $\mathbf{b}_2$ and $\mathbf{b}_3$ are both orthogonal to the plane made by $\mathbf{c}_1$ and $\mathbf{c}_2$. But as long as α_{1j}, $\alpha_{2j} \neq 0$ this is not possible because $\mathbf{b}_i$s are orthogonal vectors.

- ($\mathbf{b}_i^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{b}_j = 0$ and $\mathbf{b}_i^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{b}_j = 0 \ \forall i, j \ (i \neq j)$). For each term one of the two following alternative conditions must be true:

$$\mathbf{b}_i^H\mathbf{c}_1\mathbf{c}_1^H\mathbf{b}_j = 0 \ \Rightarrow \begin{cases} \mathbf{b}_i^H\mathbf{c}_1 = 0 & \text{ⓐ} \\ or & \\ \mathbf{c}_1^H\mathbf{b}_j = 0 & \text{ⓑ} \end{cases}$$

$$\mathbf{b}_i^H\mathbf{c}_2\mathbf{c}_2^H\mathbf{b}_j = 0 \ \Rightarrow \begin{cases} \mathbf{b}_i^H\mathbf{c}_2 = 0 & \text{ⓒ} \\ or & \\ \mathbf{c}_2^H\mathbf{b}_j = 0 & \text{ⓓ} \end{cases}$$

As $\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)$ is a symmetric matrix, we just evaluate the lower triangular elements of this matrix. For each off-diagonal element of cyclic correlation matrix there are four different combinations of the above conditions and it is easy to see that ⓐ-ⓒ and ⓑ-ⓓ cause the diagonal elements $r_{22}^{\beta_p}(\hat{z}, 0)$ and $r_{33}^{\beta_p}(\hat{z}, 0)$ to be zero simultaneously, so these are not valid combinations.

For $j = 1$ if we assume that the combination ⓐ-ⓓ is true, then, it can be concluded that $\mathbf{b}_2 \perp \mathbf{c}_1$, $\mathbf{b}_3 \perp \mathbf{c}_1$, and $\mathbf{b}_1 \perp \mathbf{c}_2$. That means $\mathbf{b}_1 \parallel \mathbf{c}_1$. As $\mathbf{c}_2^H\mathbf{b}_2 \neq 0$ the necessary condition for $r_{32}^{\beta_p}(\hat{z}, 0) = 0$ is $\mathbf{b}_3^H\mathbf{c}_2 = 0$ or in other words $\mathbf{b}_3 \perp \mathbf{c}_2$. It means $\mathbf{b}_2 \parallel \mathbf{c}_2$ and $\mathbf{b}_3 \parallel \mathbf{c}_3$, and therefore,

$$\begin{bmatrix} \hat{z}_1(t) \\ \hat{z}_2(t) \end{bmatrix} = \begin{bmatrix} \alpha_1 s_1(t) + v_1(t) \\ \alpha_2 s_2(t) + v_2(t) \end{bmatrix}$$

where α_1 and α_2 are scalar values and $v_1(t)$ and $v_2(t)$ are undesirable effects of noise on the estimations that are inevitable in all ICA methods. The same analysis for the combination ⓑ-ⓒ results in

$$\begin{bmatrix} \hat{z}_1(t) \\ \hat{z}_2(t) \end{bmatrix} = \begin{bmatrix} \alpha_1 s_2(t) + v_1(t) \\ \alpha_2 s_1(t) + v_2(t) \end{bmatrix}$$

For necessity condition it is obvious that for any separating matrix $\mathbf{B}$ the cyclic correlation matrix $\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)$ satisfies (5.4.13). □

Therefore, if a unitary matrix $\mathbf{B}$ can be found such that the conditions of the above theorem are satisfied, the mixed cyclostationary sources can be extracted.

Algorithm

Assume $\boldsymbol{\beta}$ is a set of P *a priori* known cycle frequencies of the sources of interest, i.e. $\boldsymbol{\beta} = \{\beta_p | 1 \leq p \leq P\}$. The proposed algorithm is developed based on the assumption that there exist d_p sources having the same cycle frequency β_p.

Our objective here is to find a matrix $\mathbf{B}$ for which $\mathcal{J}(\mathbf{B}) = \text{off}\{\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)\}$ is minimum.

Using the extended Jacobi approach, the proposed source extraction method is implemented as follows:

Algorithm 4 Cyclostationary source extraction (CSE)

1. Calculate the covariance matrix of $\mathbf{x}(t)$; denote the eigenvalues of this matrix by λ_1, λ_2, ... and λ_n and the corresponding eigenvectors by $\mathbf{e}_1$, $\mathbf{e}_2$, ... and $\mathbf{e}_n$.
2. Obtain the whitened data from $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$, where $\mathbf{W} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T$, $\mathbf{\Lambda} = diag\{\lambda_1, \lambda_2, ..., \lambda_n\}$ and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ...\mathbf{e}_n]$.
3. Calculate $\mathbf{R}_{\mathbf{z}}^{\beta_p}(0)$, the cyclic correlation matrix of $\mathbf{z}(t)$ at $\tau = 0$ using (2.6.22).

4. Use the extended Jacobi technique to find the unitary matrix $\mathbf{B}$, the minimizer of cost function $\mathcal{J}(\mathbf{B}) = \text{off}\{\mathbf{B}^H \mathbf{R}_{\mathbf{z}}^{\beta_p}(0)\mathbf{B}\}$.
5. Estimate d_p cyclostationary sources as $\hat{\mathbf{z}}_{d_p}(t) = [\mathbf{b}_{d_1}\ \mathbf{b}_{d_2}\ ...\ \mathbf{b}_{d_p}]^H \mathbf{x}(t)$, where $\mathbf{b}_{d_i}$'s are d_p column vectors of $\mathbf{B}$ corresponding to the largest diagonal values of $\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)$.

Deflating sources of interest

The easiest and fastest way to deflate the sources of interest is to simply switch off all the estimated cyclostationary sources and project the data back to the sensor space. If we define $\bar{\mathbf{z}}(t)$ to be equal to $\hat{\mathbf{z}}(t)$ except for the rows $p_1, p_2, ..., p_d$ which are all set to zero, then $\hat{\mathbf{x}}(t) = (\mathbf{B}^H \mathbf{W})^{\sharp} \bar{\mathbf{z}}(t)$ is the remaining observation vector.

5.5 Time varying cycle frequencies

In the proposed algorithms we assumed that the cycle frequencies of the sources are known and fixed. The knowledge about the cycle frequency of the signals can be obtained using power spectral density of the signals. However, in some of the real world applications such as speech processing, biomedical signal processing, and econometrics, the cycle frequencies may not be fixed and are subject to change with time.

If by any means one can identify the cycle durations of the source of interest in the mixtures, then, it is possible to warp the signals such that the cycles have the same duration. Here based on the assumption that the sources of interest have a peak in each cycle which can be detected easily, a method is proposed to overcome the changes in frequency. This assumption is true for some signals such as heart sound signals or ECG signals in which the high amplitude of the heart signals can be recognized easily.

The main idea in warping (de-warping) is changing the duration of a cycle from $l_{k+1} - l_k$ samples to l_{avg} samples, where l_{avg} is the average duration of the cycles and l_k represents the sample number of the kth peak in the sequence of cycles. The warping

procedure is as follows:

***Procedure* warping**

-Find the peaks of the SoI and recognize each cycle duration*

-l_{avg} = average of cycle durations

-Assign the phase range of $[-\pi, \pi]$ to the range

$[0, l_{avg} - 1]$

-For all cycles, assign the phase range of $[-\pi, \pi]$ to the sequence of samples in the range of $[l_k, l_{k+1} - 1]$

-Warp the signal by mapping every individual cycle

to the cycle with the duration of T samples.

* SoI: Source of interest

5.6 Experiments

Computer simulations were carried out to illustrate the performance of the proposed algorithms and the results were compared with those of the other methods. More experiments are reported in the next chapter.

5.6.1 Distinct cycle frequencies

The objective of the proposed algorithm in subsection 5.4.1 is to estimate the source signals such that there is minimum cyclic correlation between them. Thus, the following performance index is used to measure the cyclic correlation between the estimated sources:

$$PI = \sum_{i=1}^{m} \frac{\text{off}\{\mathbf{R}_{\hat{z}}^{\beta_i}(0)\}}{\sum_{1 \le p,q \le n} |r_{pq}^{\beta_i}(\hat{z}, 0)|^2} \tag{5.6.1}$$

The lower values of this performance index are preferred.

Experiment 1:

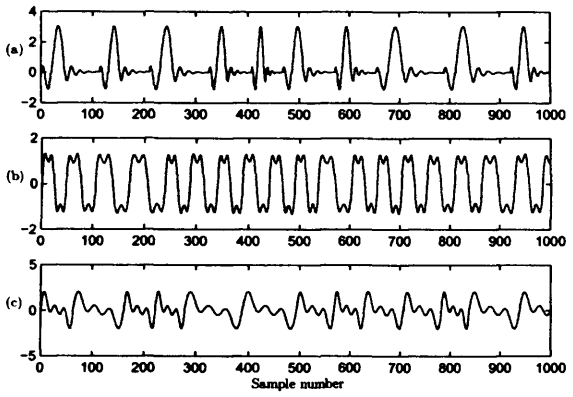


Figure 5.2: Synthetic source signals.

The proposed algorithm was applied to 2000 samples of three periodic signals illustrated in Fig. 5.2 while different levels of noise were added and the performance index was averaged over 100 independent trials for each level of noise. The signal-to-noise ratio is defined as SNR $= -10\log_{10}\sigma^2$, where σ^2 is the noise variance. The warping/de-warping method is also utilized.

The values of the performance indices for SOBI, JADE, and the proposed algorithm are presented in Fig. 5.3. Obviously, performances of all the methods are affected by high levels of noise. Although for high noise levels all the methods perform similarly with high PI values (which means the performance is low), the performance of the proposed method is better for normal and low levels of noise.

Experiment 2:

The proposed algorithm has been used to separate heart and lung sounds from a set of two synthetically generated mixtures. A typical random mixture is illustrated in Fig. 5.4. Because of the changes in the period of the heart sound signal, the warping/de-warping method is utilized in advance to adjust the cycle frequency (β). The output of algorithm is presented in Fig. 5.5. Listening to the signals also confirms separation of the sound signals.

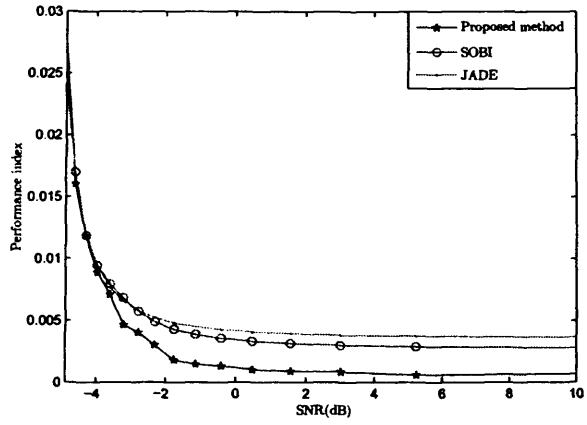


Figure 5.3: Performance comparison of the proposed method, SOBI and JADE algorithms for different levels of noise.

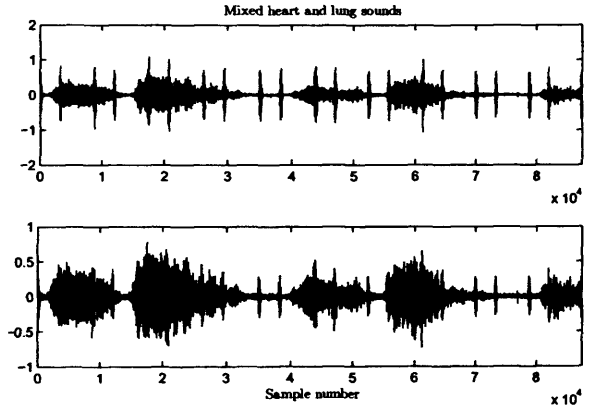


Figure 5.4: Amplitude normalized mixtures of heart and lung sounds.

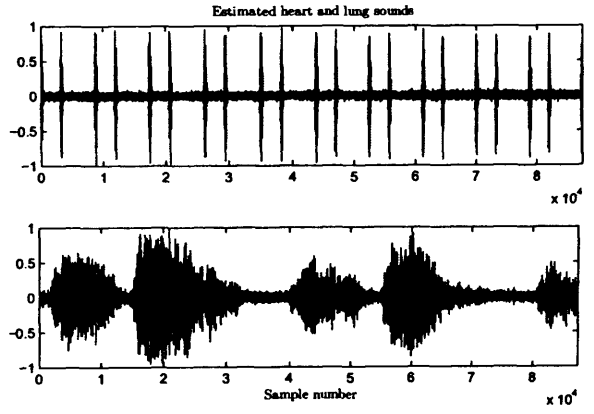


Figure 5.5: Amplitude normalized estimated heart (top) and lung (bottom) sounds.

5.6.2 Common cycle frequencies

Computer simulations were carried out to illustrate the performance of the CSE method (subsection 5.4.2) and the results were compared with those of well known ICA algorithms, i.e., SOBI and JADE. We have also implemented Theorem 4 (ATH4) in [1] which is developed for extracting cyclostationary sources with common cycle frequency. This theorem is as follows:

ATH4 from [1]: Define matrix $\mathbf{B}_i = \mathbf{R}^{-H/2}\mathbf{U}_i\mathbf{V}_i$, where $\mathbf{R}^{-1/2}$ is the pseudo-inverse of an $n \times m$ square root of $\mathbf{R} = \langle \mathbf{x}(t)\mathbf{x}^*(t) \sum_{j=1}^{m} e^{j\beta_j t} \rangle$, $\mathbf{U}_i$ is an $m \times d_i$ matrix whose column vectors form an orthogonal basis of $\ker(\mathbf{R}^{-1/2}\tilde{\mathbf{R}}_i\mathbf{R}^{-H/2})$, $\tilde{\mathbf{R}}_i = \langle \mathbf{x}(t)\mathbf{x}^*(t) \sum_{\beta_j \neq \beta_i}^{m} e^{j\beta_j t} \rangle$, $\mathbf{V}_i$ is a $d_i \times d_i$ unitary matrix that jointly diagonalizes the matrices $\mathbf{M}_i(k)$ for $k = \tau_1, ..., \tau_K$, $\mathbf{M}_i(k) = \mathbf{U}_i^H \mathbf{R}_{-1/2}\mathbf{R}_i(k)\mathbf{R}_{-H/2}\mathbf{U}_i$, and $\mathbf{R}_i(k) = \langle \mathbf{x}(t+k)\mathbf{x}^*(t)e^{j\beta_i t} \rangle$. Then, $\mathbf{B}_i$ separates out the source signals with common cycle frequency β_i, i.e., $\mathbf{z}_i(t) = \mathbf{B}_i^H \mathbf{x}(t)$ is an estimate of $\mathbf{s}_i(t)$.

The ATH4 algorithm can be implemented simply using the above theorem. The following normalized performance index which is a standard measure of the performance of BSS algorithms [29] is also used for evaluating the performance of the methods:

$$\mathbf{PI} = \frac{1}{m}\left[\sum_{i=1}^{m}\left\{\sum_{j=1}^{m}\frac{|p_{ij}|^2}{max_q|p_{iq}|^2} - 1\right\} + \sum_{j=1}^{m}\left\{\sum_{i=1}^{m}\frac{|p_{ij}|^2}{max_q|p_{qj}|^2} - 1\right\}\right]$$

PI is evaluated and averaged for 100 independent trials at different noise levels.

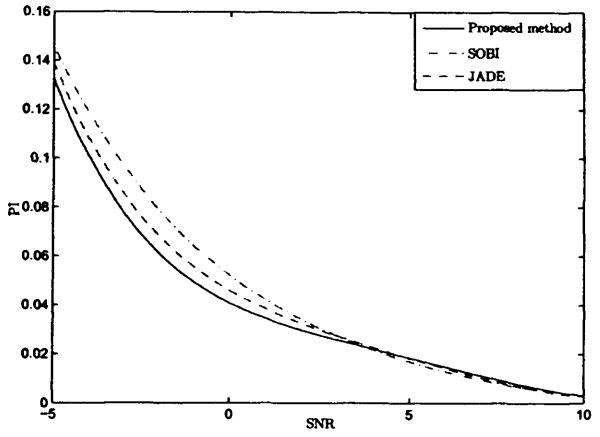


Figure 5.6: Performance index of three algorithms vs. SNR over an average of 100 different experiments on each noise level.

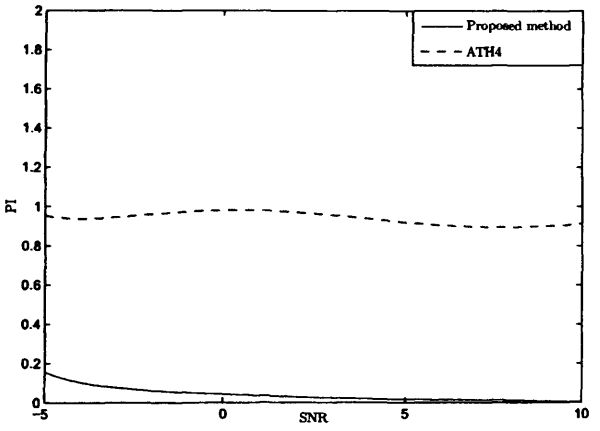


Figure 5.7: Performance index of the proposed algorithm vs. SNR compared with that of ATH4.

Experiment 3:

The two sources are first order autoregressive Gaussian processes with complex coefficients $a_1 = 0.8e^{j0.4}$ and $a_2 = 0.5e^{j0.6}$ which are modulated with two separate sine waves. The cyclic correlation of the sources are $\alpha_1 = 2f_1$ and $\alpha_2 = 2f_2$, where f_1 and f_2 are the carrier frequencies [40]. The sources are mixed by random 2×2 full rank matrices. The signal-to-noise ratio is defined as SNR $= -10\log_{10}\sigma^2$, where σ^2 is the noise variance. Fig. 5.6 demonstrates the PI values for different noise levels resulted from the proposed, SOBI, and JADE algorithms. Although the difference is not significant, the proposed method has lower performance index values.

Experiment 4:

The same configuration as that of experiment 3 is used except the value of the carrier frequencies, which are modified so that $f_1 = f_2$. Theorem 4 (ATH4) in [1] is used as a benchmark to evaluate the performance of the proposed method. As it is shown in Fig. 5.7, the proposed algorithm performs significantly better than ATH4. In the CSE method, the sources are extracted only based on the information obtained from the zero lag cyclic correlation matrix. This can be a possible result of better performance of the method.

5.7 Concluding remarks

In this chapter two algorithms for blind extraction of cyclostationary sources are introduced. In developing the first algorithm it is assumed that the cyclostationary sources have distinct cycle frequencies which are known *a priori*. If there are more than one cyclostationary source, it is possible to extract all of them by extracting the sources one by one after deflating the extracted sources.

Extraction of the sources with common cycle frequencies is considered in the second algorithm, called CSE. Both algorithms are based on diagonalization of the cyclic correlation matrix at a specific cycle frequency. Computer simulations verify better performance of the proposed algorithms compared with the other ICA methods. In the next chapter the results of applying the cyclostationary source extraction method to a real world problem is presented. More experiments on synthetic mixtures are also conducted and the results are reported.

Chapter 6

BALLISTOCARDIOGRAM ARTIFACT REMOVAL USING CSE

6.1 Overview

Simultaneous recording of EEG and functional magnetic resonance imaging (fMRI) sequences is increasingly used to monitor various mental and physiological brain activities. The interactions between the scanner magnetic field, the patient's body, and the EEG electrodes generate a pulsation artifact called ballistocardiogram (BCG) which is synchronized with the patient's heart beat. In this case, in order to make the EEG signals clinically useful such artifacts have to be highly reduced. Ballistocardiogram artifact is considered here as the sum of a number of independent cyclostationary components having the same cycle frequency. The proposed method, which is based on cyclostationary source extraction (CSE), is able to extract these components without much destructive effect on the background EEG data. It is shown that the proposed method outperforms other methods particularly in preserving the remaining signals. CSE is utilized to remove the BCG artifact from real EEG data recorded inside the magnetic resonance (MR) scanner, i.e., visual evoked potential (VEP). The results are compared to the results of benchmark BCG artifact removal techniques. Analyzing the power spectral density of the cleaned EEG data, it is shown that CSE effectively removes the frequency

components corresponding to the BCG artifact. It is also shown that VEPs recorded inside the scanner and processed using the proposed method are more correlated with the VEPs recorded outside the scanner. Moreover, since the cycle frequency of the BCG artifact is directly computed from the contaminated EEG signals there is no need for electrocardiogram (ECG) data in this method.

In the next section, the literature of BCG artifact removal techniques is reviewed. The extraction algorithm is presented in Section 6.3. The results of application of this method on synthetic data and EEG recordings are respectively reported in Sections 6.4 and 6.5. Section 6.6 presents the concluding remarks.

6.2 Related work

Information from simultaneous recording of EEG and blood-oxygen-level dependent (BOLD) data provides a great opportunity to examine the temporal and spatial activities of the brain. This information is used by neuroscientists to study spontaneous brain activity which, by its nature, is non-reproducible and hence difficult to study with separate EEG and fMRI. One of the expected results of such studies is identification of different areas of the brain involved during EEG events. However, concurrent EEG-fMRI recording poses some difficulties.

EEG data recorded during MR scanning are affected by the interaction between the patient's body, the EEG electrodes, and the magnetic field inside the MR scanner [7]. During imaging, switching magnetic fields induce electromotive forces (EMFs) which obscure the EEG signals with a regular artifact having an amplitude of up to 100 times larger than the EEG amplitude and with a very short time course [7], [78]. This artifact is coined the gradient artifact and is independent of movements of the head and leads. Several methods have been proposed to remove the gradient artifact. In [58] the frequencies related to the power spectrum template of the artifact are filtered out. As the shape of this artifact is relatively invariant over time [33], an average template of

the artifact can be formed and subtracted from the EEGs as in [6, 18, 53, 81].

Other significant source of artifact is the tilting - movement - of the subject's head within the MR scanner, which changes the area of inter electrode loops normal to the magnetic field. This type of artifact can be divided into two groups. The first group, which is caused by deliberate movements of the head, has higher amplitudes and happens in short courses of recording. The second group, called ballistocardiogram, is caused by micro-movements of the head as a result of cardiac pulsation and obscures the underlying EEG mainly at alpha frequency (8-13 Hz) and below, with amplitudes around 150 μV at 1.5 Tesla magnetic field [20]. It is also likely that BCG artifact is the result of blood flow perpendicular to the static magnetic field inside the scanner [7, 106] and the movements of the electrodes and scalp due to expansion and contraction of scalp arteries between systolic and diastolic phases [92].

In an early attempt to remove the BCG artifact, a method based on average subtraction has been proposed in [7]. The QRS complexes of subject's ECG are first detected. Then, a limited number of the EEG signal slices corresponding to the QRS timing are averaged to create a template of the BCG artifact to be reduced from each channel. This method, which is called average artifact subtraction (AAS), is very popular [78]. However, the assumption that all the waveforms are similar during the scans is not always valid [81]. In order to deal with the heart beat timing variations a weighted averaging approach is proposed in a subsequent study [50]. In [53], the problem of variability of the artifact is addressed using a clustering algorithm. For all the methods which are based on averaging technique a reference ECG channel is essential. However, in some cases this channel is not present or the heart beats are not accurately detectable.

A new type of multi-path EEG cap is proposed in [36] that oversamples the electrode space to provide an overcomplete representation of the data. Using the assumption that neural activity is Kirchhoffian and the BCG artifacts are not, the artifacts are removed by solving an overcomplete representation of the single trial EEG data.

Adaptive filtering has also been used for BCG artifact removal [20], [92], and [103]. The reference signal comes from a movement detector, i.e. a piezoelectric sensor, attached to the body of the subject inside the scanner [20] and median filtering is used instead of simple averaging to create the BCG artifact template [92]. The authors in [68] enhanced their work by exploiting both average subtraction and adaptive filtering.

Different ICA based methods have also been used for BCG artifact removal [15, 70, 74, 78]. These methods assume that the brain neural activity including evoked potentials, oscillatory waves, artifacts caused by muscles, and noise are all mixed linearly and are independent or at least can be categorized in groups of independent components. As mentioned earlier, three phenomena with different characteristics generate the BCG artifact. It implies that BCG artifact consists of more than one independent component added linearly to the EEG data [7, 106]. Hence, the artifact can still be separated using ICA methods. The advantage of these methods is that they do not require an ECG channel. More importantly, they do not assume that the BCG artifacts are reproducible. Infomax [15] is used in [95] to extract the BCG artifact sources. In [74], fastICA [61, 62] is utilized to remove imaging, BCG, and ocular artifacts. In a comparative work, the performance of Infomax, fastICA, SOBI [16], and complexity pursuit [60] are evaluated and compared to AAS in [78]. A sequential blind extraction method [71] is used in [70] to extract the BCG artifacts and a simple peak detector is utilized to track the time varying period.

Based on the assumption that each occurrence of the BCG artifact in any EEG channel is independent of the previous observations, principal component analysis (PCA) is employed in the optimal basis set (OBS) method [81]. In the next step, for each EEG channel few of the principal components are chosen as the basis set, which is then fitted (scaled in time and amplitude) and subtracted from each BCG artifact instance. To remove any possible BCG artifact residual it is proposed in [32] and [33] to apply Infomax to the OBS output.

An important issue of concern in BCG artifact removal is selection of the correct number of BCG artifact components. In ICA based methods, an incorrect assumption about the number of BCG artifacts may influence the independence assumption. It is assumed in [78], that the BCG artifacts are caused only by head movements inside the scanner. In this case, it is mathematically and experimentally shown that the number of independent BCG artifact components is three. Their experiments also show that assuming three BCG artifact components provides reliable results. In another attempt, the number of components is not set fixed and three to six independent components are chosen for different subjects by thresholding the correlation of the estimated independent components (ICs) with the ECG channel [74]. The authors in [81] opted a conservative approach and fixed the number of components to three. In [36], only the strongest component (in terms of power) from the ICA decomposition of the EEG data is labeled as BCG artifact.

CSE is an ICA based blind source extraction method for extracting the sources with periodic statistics. Similar to other ICA methods, it is assumed that the original sources and the mixing medium are generally unknown, however, *a priori* knowledge about the periodicities helps to improve the extraction performance [44, 45]. This method, is used to remove the BCG artifacts from the EEG data recorded inside the MR scanner. The period of the second order statistics is obtained directly from the EEG data (availability of the ECG channel, necessary for some of the other removal methods, is not essential here). In order to find the appropriate number of BCG artifact components, the outputs of different methods are analyzed using the defined performance indices. Moreover, it is shown that the proposed method preserves the remaining data better than the other methods.

6.3 Algorithm

Details of cyclostationary source extraction method development were presented in Chapter 5. An overview of the algorithm is provided here. A technique for estimating the cycle frequency of the BCG artifact channel without using the ECG channel is also presented.

6.3.1 Source extraction

Assume $\mathbf{s}(t)$ is a vector of m unknown mutually (statistically) independent sources which are mixed instantaneously through an unknown medium, $\mathbf{A}$, and $\mathbf{x}(t)$ is the vector of n measurements, i.e., $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t)$, where $\mathbf{v}(t)$ is an $n \times 1$ stationary noise vector. Also assume that d sources are cyclostationary with a common cycle frequency β_p. The objective is to estimate the cyclostationary sources.

The steps of the proposed cyclostationary source extraction method are described below (see Chapter 5 for more details).

1. Calculate the covariance matrix of $\mathbf{x}(t)$; denote the eigenvalues of this matrix by λ_1, λ_2, ... and λ_n and the corresponding eigenvectors by $\mathbf{e}_1$, $\mathbf{e}_2$, ... and $\mathbf{e}_n$.
2. Obtain the whitened data from $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$ where $\mathbf{W} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T$, $\mathbf{\Lambda} = diag\ \{\lambda_1, \lambda_2, ..., \lambda_n\}$ and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n]$.
3. Calculate $\mathbf{R}_{\mathbf{z}}^{\beta_p}(0)$, the cyclic correlation matrix of $\mathbf{z}(t)$ at $\tau = 0$, using (2.6.22).
4. Use the extended Jacobi technique (see subsection 5.3.2) to find the unitary matrix $\mathbf{B}$, the minimizer of cost function $\text{off}\{\mathbf{B}^H\mathbf{R}_{\mathbf{z}}^{\beta_p}(0)\mathbf{B}\}$.
5. Estimate d_p cyclostationary sources as $\hat{\mathbf{z}}_{d_p}(t) = [\mathbf{b}_{d_1}\ \mathbf{b}_{d_2}\ ...\ \mathbf{b}_{d_p}]^H\mathbf{x}(t)$, where $\mathbf{b}_{d_i}$s are d_p column vectors of $\mathbf{B}$ corresponding to the largest diagonal values of $\mathbf{R}_{\hat{\mathbf{z}}}^{\beta_p}(0)$.

6.3.2 Cycle frequency estimation

As the cycle frequency of the cyclostationary signals appear as high peaks in their PSDs [43], the cycle frequencies in real world scenarios can be estimated using the PSD of

the mixtures as long as the SNR and energy of the sources of interest are not too low. Therefore, a simple way to find the cycle frequency is investigating the PSD of those channels which have more periodic behavior (or the average of the PSDs of all or some of the channels). PSDs are obtained by calculating FFT of the squared version of the signals. This approach takes advantage of the cyclostationarity and works for those signals that have hidden periodicity, i.e., any periodic behavior which may be hard to be distinguished from visual inspection of the temporal waveform [11].

6.3.3 Deflating sources of interest

After estimating the extracting matrix, the output channels which show the highest cyclostationarity at cycle frequency β_p can be determined visually. Alternatively, as an automated solution, one can choose those sources which have the highest correlation with a reference signal [95]. It is shown in [32] that the ECG channel contains some features that are not present in the EEG (and vice versa). Therefore, this method is not suitable enough for BCG artifact removal. Here, we identify the sources of interest by examining the PSD of the output signals and selecting those having the highest PSD peaks at β_p and its two harmonics[1] by smoothing the PSD of each output channel and subtracting the smoothed PSD from the original PSD. The outputs are sorted according to the sum of the values of the modified PSD at the main cycle frequency and the harmonics. The first d sources are the sources of interest.

To deflate the estimated sources it is enough to switch off all the d estimated sources of interest and project the remaining data back to the sensor space. If we define $\bar{\mathbf{z}}(t)$ to be equal to $\hat{\mathbf{z}}(t)$ except for the rows $p_1, p_2, ..., p_d$, which are all set to zero, then $\hat{\mathbf{x}}(t) = (\mathbf{B}^H\mathbf{W})^\sharp\bar{\mathbf{z}}(t)$ is the remaining observation vector, where $\sharp$ denotes pseudo-inverse of the matrix.

[1]Investigating the PSD of the BCG artifact we realized that the significant part of the periodic components is concentrated in the first three harmonics.

6.4 Experiments

To evaluate the performance of the proposed algorithm it is compared with different ICA methods using a set of artificially mixed synthetic signals. The CSE method is then applied to the real world EEG data and the results are compared to the outputs of the traditional BCG removal methods, i.e., AAS and OBS. Several performance indices are utilized to assess the quality of the extracted sources and the deflation process.

6.4.1 Performance evaluation

The objective of all extraction algorithms is to find a vector $\mathbf{b}_i$ for each source of interest such that $\hat{z}_i(t) = \mathbf{b}_i\mathbf{W}\mathbf{x}(t)$ is an estimation of the source. Ideally, $\mathbf{f}_i = \mathbf{b}_i\mathbf{W}\mathbf{A} = c_k\mathbf{e}_k$ is a column vector which has just one nonzero element in say kth row and thus $\hat{z}_i(t) = c_k s_k(t)$, where c_k is an arbitrary nonzero scalar and $\mathbf{e}_k$ is the kth unit base vector along the kth coordinate direction.

In order to evaluate the performance of a typical algorithm which extracts d sources the following performance index is defined [28], [72]:

$$PI_1 = \frac{1}{d}\sum_{i=1}^{d} 10\log_{10}\left(\frac{1}{m}\times\left(\frac{\sum_{j=1}^{m} f_{i,j}^2}{max\{f_{i,1}^2, f_{i,2}^2, \ldots, f_{i,m}^2\}} - 1\right)\right) \tag{6.4.1}$$

where $\mathbf{f}_i = [f_{i,1}\ f_{i,2}\ \ldots f_{i,m}]^T$. More negative values of this index show better performance of the extraction algorithm.

The second index used in the sequel is the averaged correlation of the remaining data at time delays close to the period of the deflated data. If the remaining data are represented by $\hat{\mathbf{x}}(t) = [\hat{x}_1(t)\ \hat{x}_2(t)\ \ldots\ \hat{x}_m(t)]^T$ and the cycle frequency of the deflated data is β_p, PI_2 is defined as:

$$PI_2 = \frac{1}{m}\sum_{i=1}^{m} 10\log_{10}\sum_{\tau\in W}\frac{|\sum_t(\hat{x}_i(t-\tau)-\hat{\mu}_i)(\hat{x}_i(t)-\hat{\mu}_i)|}{\sum_t(\hat{x}_i(t)-\hat{\mu}_i)^2} \tag{6.4.2}$$

where W is an l second length time window centered at $1/\beta_p$ seconds and $\hat{\mu}_i$ is the mean

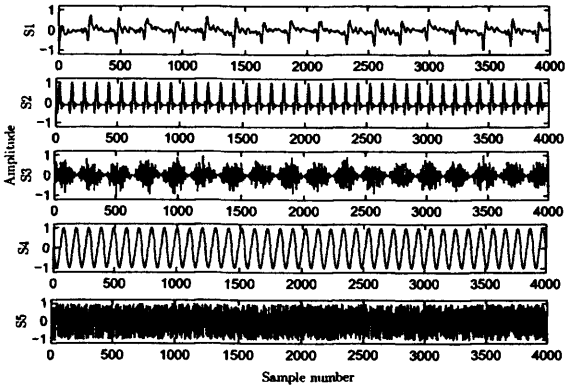


Figure 6.1: Five signals are synthetically mixed to evaluate the performance of the proposed method. S1 and S3 have a common cycle frequency.

of $\hat{x}_i(t)$ [78]. PI_2 is a measure which shows the amount of deflated sources still present in the remaining data. Again, more negative values of PI_2 are more desirable.

Improvement in normalized power spectrum ratio (INPS) [78, 95] is used to measure how the deflation process succeeds in clearing the undesired sources from the data. If $\phi_{x_i}(f)$ denotes power spectral density of $x(i)$, INPS is defined as:

$$PI_3 = \frac{1}{m}\sum_{i=1}^{m} 10\log_{10}\frac{\sum_{f\in F}\phi_{\hat{x}_i}(f)}{\sum_{f\in F}\phi_{x_i}(f)} \tag{6.4.3}$$

where F is a set of intervals around β_p and its harmonics. More negative values of PI_3 represent better performance.

The indices defined so far are useful measures to evaluate the extraction and deflation quality. Additionally, we need to know whether the remaining data are distorted by the deflation process. To do so, we define $\bar{\mathbf{s}}(t)$ to be equal to $\mathbf{s}(t)$ except for the sources of interest which are all set to zero. In definition of the last performance index $\bar{\mathbf{x}}(t) = \mathbf{A}\bar{\mathbf{s}}(t)$ is used as a reference to evaluate the performance of the deflation process:

$$PI_4 = \frac{1}{m}\sum_{i=1}^{m} 10\log_{10}\left|\frac{<\bar{x}_i(t),\hat{x}_i(t)>}{\sqrt{<\bar{x}_i(t),\bar{x}_i(t)><\hat{x}_i(t),\hat{x}_i(t)>}} - 1\right|. \tag{6.4.4}$$

This index measures the similarity between $\bar{\mathbf{x}}$ and $\hat{\mathbf{x}}$. The more negative the value of PI_4, the better the performance.

SOBI, Infomax, and fastICA are selected as benchmarks for comparison when applied to synthetic data. SOBI is a widely used blind source separation method which estimates the sources by simultaneously diagonalizing a set of time delayed covariance matrices. In other words, SOBI defines an average eigen-structure of the data [16]. Infomax is based on maximization of the estimated entropy, i.e. maximizing the mutual information between the observations and the estimations [15]. The main idea in fastICA is to maximize the negentropy of every estimated signal [62].

Average artifact subtraction and optimal basis set methods are fully automated methods commonly used as benchmarks in BCG artifact removal [78], [33]. Therefore, for real world EEG data, CSE results are compared with the results of these methods. In this case, the performance evaluation is restricted to PI_2 and PI_3, as the original data are unknown.

6.4.2 Synthetic data

A set of five independent sources including a BCG artifact signal extracted from an EEG recording (for details of the recording see the next section) is selected as in Fig. 6.1 in which S1 represents the BCG artifact signal, S2 is a periodic signal, S3 is a Gaussian random noise modulated with a sine wave of half the frequency of the BCG artifact channel in S1, S4 is a sine wave, and S5 is a uniform random noise. The first two sources are super-Gaussian, S4 and S5 are sub-Gaussian, and S3 is close to Gaussian. To evaluate the performance of the method the sources are mixed through random 5×5 full column rank matrices and different separation methods are applied to extract the sources that have common cycle frequency, i.e. the BCG artifact channel and the modulated Gaussian noise.

Fig. 6.2 provides an illustrative comparison between the PSD of the mixtures, $\mathbf{x}(t)$, the desired output, $\bar{\mathbf{x}}(t)$, the result of CSE, $\hat{\mathbf{x}}(t)$, and the output of the other ICA-based methods. Visual inspection of the PSDs in Fig. 6.2.a and Fig. 6.2.b suggests that all

Table 6.1: The averages and their corresponding standard deviations of the results of 1000 independent trials of different ICA methods over the linear mixtures of the sources in Fig. 6.1. The length of time window in calculating PI_2 is 0.4s, and the length of frequency intervals for each harmonic of cycle frequency of the sources of interest is 0.5 Hz.

Method	PI_1	PI_2	PI_3	PI_4
SOBI	-67.63 ± 2.39	-8.50 ± 0.75	-1.57 ± 0.73	-19.30 ± 1.19
Infomax	-72.62 ± 5.70	-9.43 ± 0.84	-1.63 ± 0.63	-22.97 ± 1.97
FastICA	-68.69 ± 2.55	-9.42 ± 0.85	-1.67 ± 0.58	-23.65 ± 1.76
CSE	-99.15 ± 3.74	-9.45 ± 0.85	-1.74 ± 0.57	-29.80 ± 1.50

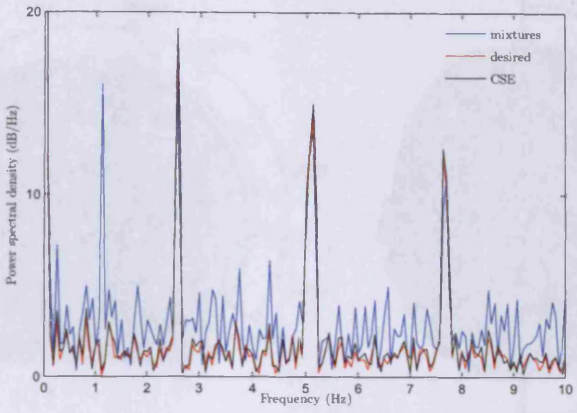
the methods have successfully extracted the cyclostationary sources.

Table 6.1 shows the average values for each performance index after 1000 independent trials and their corresponding standard deviations. For every trial a new random mixing matrix is generated. Since the period of S1 is 1.12s, the length of W in calculating PI_2 is set to 0.4s and the frequency intervals around the main cycle frequency and its two harmonics in calculating PI_3 are 0.5Hz. The values of PI_1, which measure the accuracy of the separation algorithm in terms of sparsity of the product of the extracting vectors and the mixing matrix, show that the CSE method provides the smallest value. Although the methods do not show any significant difference in the values of PI_2 and PI_3, CSE still provides lower values for both indices. The values of PI_4 verify that the CSE method preserves the remaining signals better than the other methods.

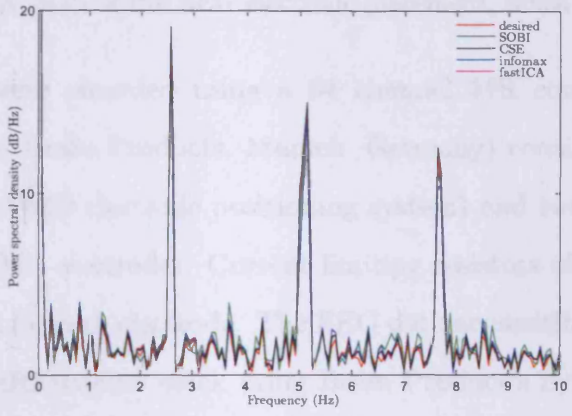
PI_2 and PI_3 measure the correlation and the power of the remaining data in the neighborhoods of the period and cycle frequency of the sources of interest. On the other hand, PI_1 and PI_4 consider the overall behavior of the removal method rather than the performance just within certain time or frequency ranges. The values of performance indices in Table 6.1 show that the CSE method effectively extracts the sources of interest and preserves the remaining data.

6.5 EEG-fMRI data

Data from five of the fourteen subjects participated in the study of [83] were randomly selected. These subjects were all recruited from the University of Birmingham and paid



(a)



(b)

Figure 6.2: A comparison between the average of the power spectral density of the mixtures, the desired outputs and outputs of the ICA methods applied to the mixtures, (a) mixtures, $\mathbf{x}(t)$, desired output, $\bar{\mathbf{x}}(t)$, and CSE output, $\hat{\mathbf{x}}(t)$, and (b) desired and all ICA outputs.

for their participation. All observers had normal or corrected to normal vision, no history of neurological disorders, and gave written informed consent. The study was approved by the local ethics committee.

The experiment was conducted in the University of Birmingham Imaging Centre using a 3 T Philips Achieva MRI scanner. Gradient echo-pulse sequence was acquired from 20 slices (2.5×2.5×3 mm resolution, TR 1500 ms, TE 35 ms, SENSE factor 2, flip angle 80, with equidistant temporal slice spacing to facilitate synchronization of the EEG clock).

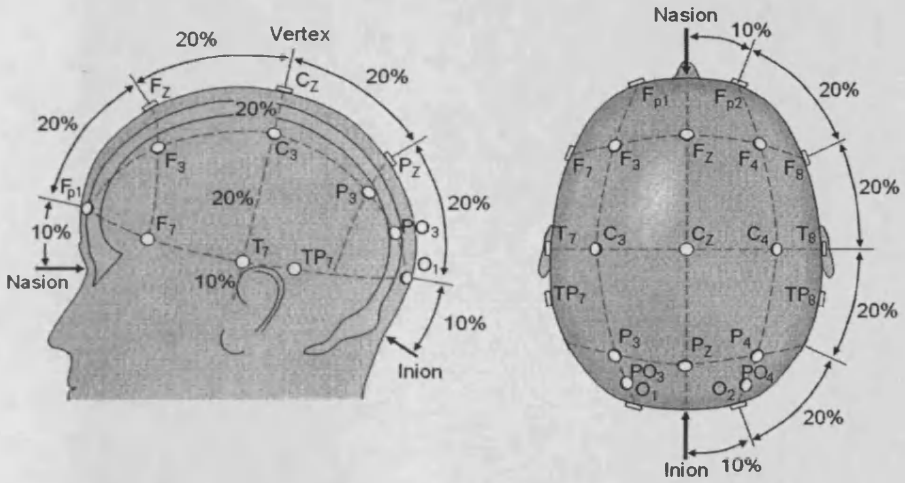


Figure 6.3: Map of the EEG electrode placement, taken from [90].

The EEG data were recorded using a 64 channel MR compatible EEG system (BrainAmp MR Plus, Brain Products, Munich, Germany) consisting of 62 scalp electrodes (following the 10-20 electrode positioning system) and two additional ECG and electrooculogram (EOG) electrodes. Current limiting resistors of 5 kΩ at the amplifier input were considered for each electrode. The EEG data acquisition setup clock was synchronized with the MRI scanner clock using Brain Product's SyncBox. The sampling rate was set to 5 kHz.

As stimuli, left hemi-field reversing checkerboards were presented at a spatial frequency of 2 cycles per degree of visual angle at two different contrasts, i.e., high and low. Stimuli were presented together with a central fixation cross. We were interested in the visual evoked potentials (VEPs) elicited by these stimuli. Further details of recordings can be found elsewhere [83].

Raw EEG data were partitioned into data acquisition sessions and exported to $*.dat$ format using Brain Vision Analyzer (Brain Products, Munich, Germany). Subsequently, the gradient artifacts were removed using the Brain Vision Analyzer built-in functions. After segmenting the data corresponding to high and low contrast stimuli, EEGs were low-pass filtered at 25 Hz and down-sampled to 256 Hz. To reduce the computational

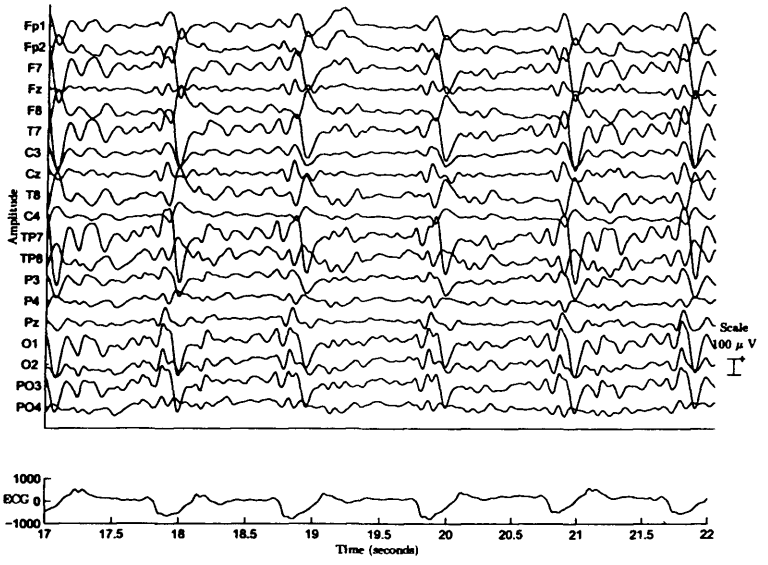


Figure 6.4: Nineteen EEG channels recorded inside MR scanner after cleaning the gradient artifact. ECG channel is also recorded which is used in some BCG artifact removal methods as a reference of cardiac pulsation.

cost of different BCG artifact removal methods, we also spatially down-sampled the data to twenty channels including Fpl, Fp2, F7, Fz, F8, T7, C3, Cz, T8, C4, Tp7, Tp8, P3, P4, Pz, O1, O2, PO3, PO4 and ECG, having electrodes distributed across all areas of the scalp.[2] The measurements were then exported to MATLAB for further analysis.[3] A map of the placement of the electrodes is shown in Fig. 6.3 and a five second segment of the recorded data from one of the subjects is depicted in Fig. 6.4.

6.5.1 Results

Three methods, i.e. AAS, OBS and CSE, were applied to the EEG data in order to remove the BCG artifact. We used the FMRIB plug-in provided in EEGLAB [34] which includes implementations of AAS and OBS and is freely available.

To use CSE, we first checked the validity of the basic assumptions. Since they

[2]EEG-fMRI data analysis is mainly performed off-line and hence the computational cost of different BCG artifact removal methods was not of our major concern here.

[3]In removing the BCG artifact using the CSE algorithm, the eye-blink artifact is also cleaned as a by-product. Although other advanced methods such as [79] can be utilized.

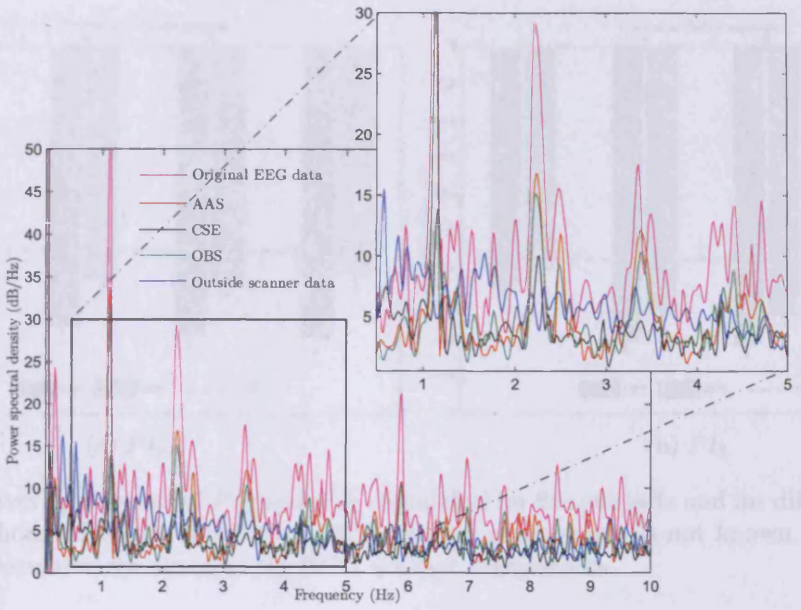
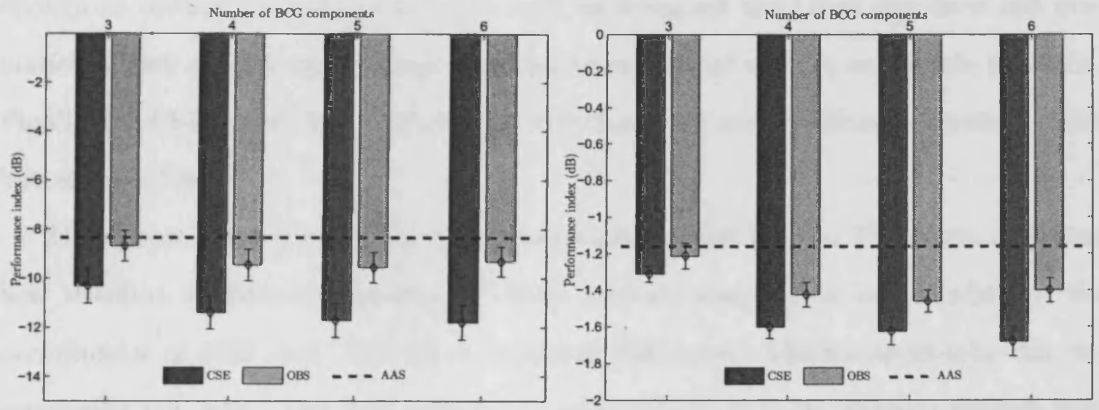


Figure 6.5: Averaged power spectral densities of the outputs of AAS, OBS, and CSE methods. The PSDs of nineteen EEG channels are averaged for each method. Averaged PSD of the same subject's EEG data recorded inside the scanner and the EEGs recorded outside the scanner are also provided for comparison.

originate from the sources of different nature, it is valid to assume that EEG and BCG artifact signals are independent. As BCG artifact components are synchronized with the heart beats, they have the same cycle frequency. Moreover, since the BCG artifacts are independent, the identifiability condition is satisfied for this problem. Therefore, all the prerequisites to apply the CSE method to this problem are met.

In developing CSE, prior knowledge about the cycle frequency of the sources of interest is necessary. The concurrently recorded ECG channel is a good source of information about the BCG artifact cycle frequency, however in those cases where this channel is corrupted or not recorded, the required information can still be extracted from other channels. Due to the high amplitude of BCG artifacts, one can measure the cycle frequency from those channels that bear BCG artifacts with the highest amplitudes. Fig. 6.5 includes the averaged PSDs for one of the subjects before and after artifact removal. The averages are calculated from the PSDs of all EEG channels and show clear peaks



(a) PI_2 (b) PI_3

Figure 6.6: Averaged values of PI_2 and PI_3 calculated for five subjects and for different artifact removal methods when the number of BCG artifact components is not known. Performance indices are calculated for three to six BCG artifact components.

in the original data at the main frequency and harmonics of the subject's heart beat frequency.

Since the heart rate is not strictly fixed, we may encounter some variations in the cycle frequency which influence the performance of the extraction algorithm. One way to overcome this problem is to warp the signals after detecting the peaks, in order to force them to have a fixed cycle frequency, and dewarp the estimated and/or remaining data after deflation, as done earlier in [46]. This method is only applicable in such cases that the peaks are accurately detectable. This, however, imposes a significant computational cost for warping/dewarping procedures. Alternatively, applying CSE to smaller segments of data can also make the extraction procedure more robust. For a pure periodic signal, each cycle frequency generates a high peak in the PSD of the signal. In quasi-periodic signals the cycle frequencies have wider peaks. If the width of the peaks is small, equations (5.4.2) and (5.4.13) are still satisfied and hence, the proposed algorithm is able to extract the sources with common cycle frequency β_p.

Here, the length of each EEG segment is selected to be 10s. This length is short

enough to assume that in the experimental environment the heart rate does not dramatically vary and the cycle frequencies can be calculated with an acceptable precision. Finally, the CSE output is bandpass filtered between 0.1 and 25 Hz using a second order butterworth filter.

The correct number of BCG artifact components is not known. Therefore, we tested how selection of different numbers of BCG artifact components would influence the performance of CSE and OBS. These methods follow two different approaches for decomposing the data. The first method decomposes the data to spatially independent components, whereas in the latter it is assumed that every occurrence of the artifact in each channel is independent of the previous ones and hence a temporal decomposition is performed. However, as the origin of the components in both methods is the same, the numbers of the components are comparable. Starting with three, the performance of the removal methods against different assumptions about the number of BCG artifact components is evaluated in terms of PI_2 and PI_3. The evaluation has been performed for up to 6 components and averaged over five subjects. The results are presented in Fig. 6.6.

The values of PI_2 and PI_3 in Fig. 6.6 imply that CSE and OBS outperform the AAS method. It can also be noticed that generally CSE performs better than OBS in terms of PI_2 and PI_3. The trend of changes in the performance index values suggests the appropriate number of BCG artifact channels. For the OBS method there is a shift in the value of PI_2 and PI_3 between three and four components, but CSE has a continuous decreasing trend for both PIs. From step 5 of the CSE algorithm it can be concluded that the sources are estimated based on their cyclic autocorrelation values at β_p, which implies that as long as the diagonal values of $\mathbf{R}_{\hat{z}}^{\beta_p}(0)$ are not zero, we can expect more components to be extracted. That is why both the indices have decreasing values for CSE.

The appropriate number of BCG artifact components can be the one for which the

performance indices have their minimum values. This number is four for OBS and six for CSE. However, the values returned by PI_2 and PI_3 only show how the periodic sources are removed. None of these performance indices evaluate the possible distortion of the remaining data after deflating more components. Therefore, as there is no significant change in the values of the indices for more than four BCG artifact components, we empirically select this number.

The effects of the removal methods on the signals in frequency domain is presented in Fig. 6.5. The averages of power spectral densities of nineteen EEG channels for the original data of one of the subjects, and outputs of AAS, OBS, and CSE, and the outside scanner data are illustrated. The peaks at the frequencies around the cardiac frequency and its harmonics are dramatically mitigated by CSE. Although OBS has reduced the power of signal at the peak points, the peaks are still visible.

Visual evoked potentials are obtained after baseline correction of the averaged data. The topographic map for the first positive peak at 100ms after the stimulus (P100 peak) of one of the subjects averaged over 90 trials for the EEG data recorded outside the scanner is shown in Fig. 6.7.(a)-(c). It is expected that the highest values can be seen in the electrodes located on the right occipital region. Fig. 6.7.(d)-(f) and Fig. 6.7.(g)-(i) show the topographic maps of the same subject for which the data were recorded inside the scanner and BCG artifacts removed using the CSE and OBS methods, respectively. Correlation of the peak values of the output of the removal methods and the outside scanner data over all the channels is evaluated in order to compare the similarity. However, as the sample correlation coefficients (r_s values) are not distributed normally, under the null hypothesis that the population correlation coefficient (ρ) is equal to zero, Fisher's Z transformation is employed to transform the coefficients into normally distributed values, i.e. $z = \frac{\sqrt{n-3}}{2} \ln \frac{1+r_s}{1-r_s}$ where n is the number of sample correlation coefficients [75]. Setting α, the level of significance, to 0.05, the critical values to reject the null hypothesis will be ± 1.96. The correlation coefficients of the outside scanner

Table 6.2: Averaged correlations between averaged VEPs obtained from EEG data recorded outside scanner and those obtained from the processed EEG data recorded inside the scanner. For each channel, Fisher's Z transform is applied to the correlation coefficient. Averages of correlations over nineteen EEG channels are presented for high and low contrast stimuli separately. In the last two rows the average and standard deviation over all the subjects are displayed.

		AAS	OBS	CSE
Subject 1	High	5.97	6.26	6.42
	Low	5.45	5.70	5.97
Subject 2	High	4.16	4.98	4.45
	Low	3.97	4.55	4.25
Subject 3	High	6.12	6.58	6.92
	Low	5.20	5.70	6.26
Subject 4	High	5.83	6.26	7.10
	Low	5.57	6.42	6.26
Subject 5	High	5.97	6.12	6.58
	Low	5.57	5.45	5.83
Average	High	5.61 ± 0.81	6.04 ± 0.61	6.29 ± 1.06
	Low	5.15 ± 0.68	5.56 ± 0.67	5.72 ± 0.83

peaks and the results of the OBS and CSE methods over nineteen EEG channels are 2.17 and 2.90, respectively.

The waveforms of the obtained VEP from PO4 electrode are illustrated in Fig. 6.8. The VEPs are processed by AAS, OBS, and CSE. Visually, the three methods perform similarly on extracting the VEPs. In order to quantitatively evaluate the performance of the methods, correlation coefficients derived by comparing the outside scanner VEPs and those recorded inside scanner and restored by the removal methods are averaged over nineteen channels for five different subjects and presented in Table 6.2. The correlation coefficients are averaged over all the channels after being normalized by Fisher's Z transform. High correlation values in Table 6.2 (compared with 1.96) lead to rejection of the null hypothesis $\rho = 0$. CSE and OBS provide better correlations for high contrast stimulus VEPs. Although the standard deviations of the CSE results are bigger than those of the other methods, on average, the CSE results are more correlated with the outside scanner data.

To distinguish the signal and noise contributions in the obtained VEPs, the SNR is evaluated by dividing the amplitude of the P100 peak by the standard deviation of the EEG in the 200ms pre-stimulus interval [33]. The SNR values for PO4 channel for

high contrast stimuli VEP are 5.91, 10.13, 8.42, 8.81, and 9.06 for the original (BCG artifact contaminated) data, outside scanner data, and the output of AAS, OBS and CSE methods, respectively. The same measures for low contrast stimuli VEP are 3.46, 5.83, 3.96 4.03, and 4.59, respectively. Although the P100 amplitudes of the data restored by CSE have lower values compared to the AAS and OBS peak values, the higher SNRs of CSE indicate a better noise reduction of the proposed method.

6.6 Summary and conclusions

A cyclostationary source extraction method is developed in this chapter. This method exploits periodicity in statistics of the sources and extracts those sources which have higher power spectral density levels in the cycle frequency of interest. The mathematical proof is provided and the performance of the method for synthetically mixed signals is compared with those of standard ICA methods. The extraction performance is comparable with that of the other methods, while CSE preserves the remaining signals better than the others.

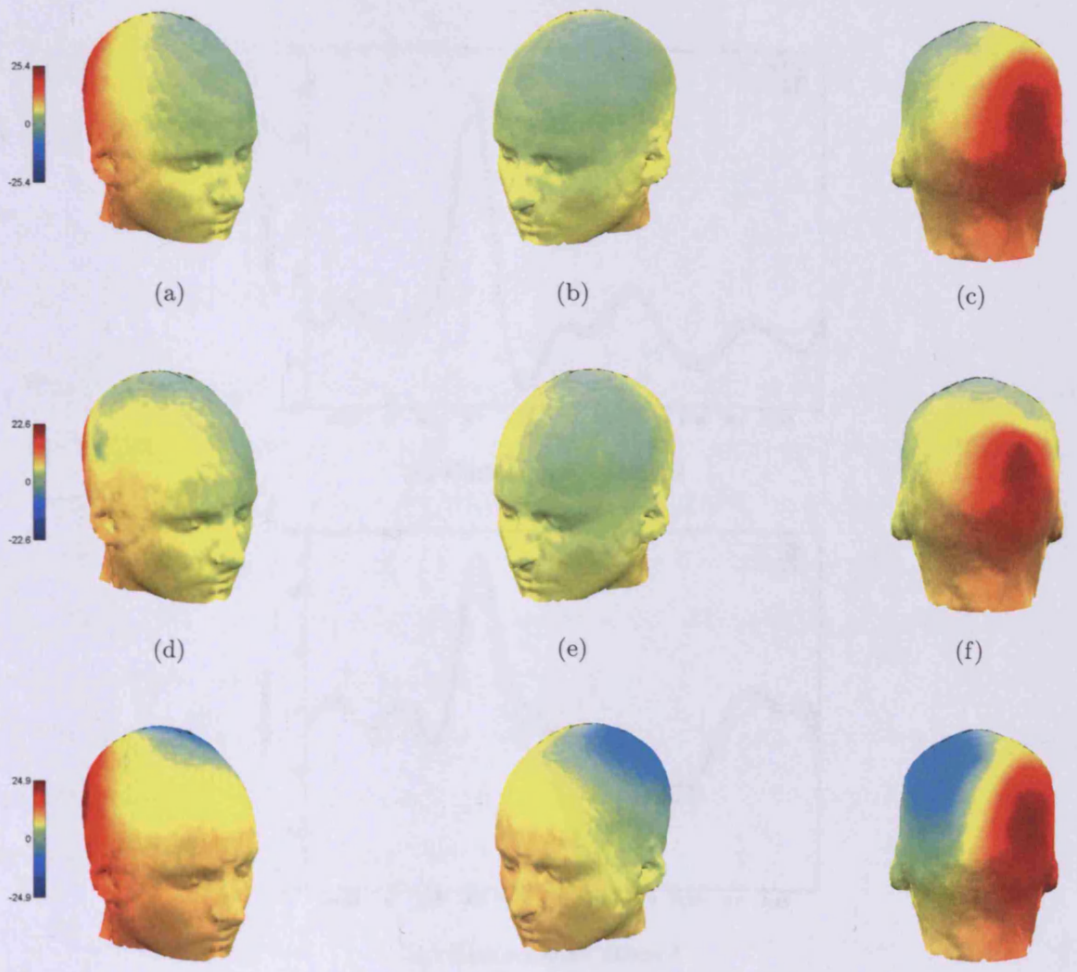
Unlike the standard removal methods which are based on removing a template built from the average of previous samples and time locked to QRS peaks, the proposed method does not need the ECG channel. The cycle frequency of the sources can be calculated using the highest peaks in the average of the smoothed PSD of different channels.

To identify the BCG artifact components, the output channels of the algorithm are sorted based on their PSD values at the main cycle frequency and its two harmonics. Based on the values of the performance indices it is shown that there is no significant improvement in the performance if more than four output channels are deflated. After removing the BCG artifact signals, the data are projected back to the sensor space using the estimated mixing matrix.

Higher performance of the CSE in removing the BCG artifact artifacts is reflected by

two performance indices which measure how well the artifacts are removed from the EEG signals. The power spectral density of the EEG signals restored by different methods is also analyzed. Comparing the VEPs obtained from different removal methods, it is shown that the results of the proposed method are more correlated with the VEPs obtained outside the scanner. The proposed method also results in higher SNR values.

The CSE method can be used for extracting various types of sources or artifacts which originate from periodic phenomena even when the periodicity can not be detected visually in the observations. For those cases that the frequency of the periodic signals vary with time, two solutions are provided which make the CSE a robust method for extracting quasiperiodic sources.

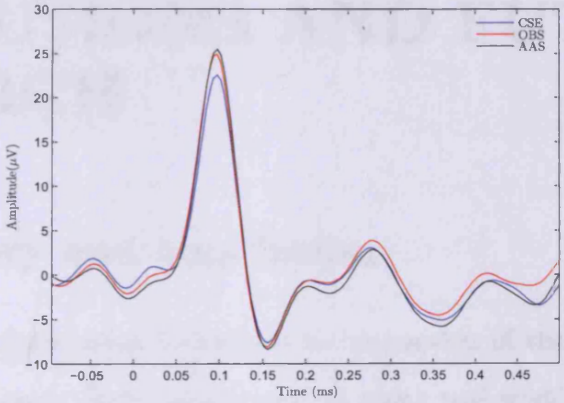


(a) (b) (c)

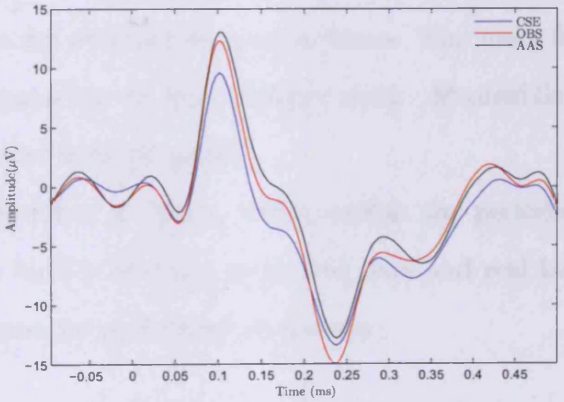
(d) (e) (f)

(g) Right view (h) Left view (i) Back view

Figure 6.7: Topographic maps corresponding to visual evoked potentials of one subject, (a)-(c) recorded outside the scanner, (d)-(i) recorded inside MR scanner. The BCG artifacts are removed using CSE in (d)-(f). Signals are restored using OBS in (g)-(i).



(a) High contrast stimuli



(b) Low contrast stimuli

Figure 6.8: Visual evoked potential on PO4 electrode obtained from EEG recorded inside the MR scanner and restored with CSE and two standard BCG artifact removal methods, AAS and OBS.

Chapter 7

CONCLUSIONS AND FUTURE RESEARCH

7.1 Summary and conclusion

In this dissertation, some novel techniques for extraction of the signals with periodic structures were proposed. Such signals exist in many real world signal processing applications, however, in most applications they are hidden and some advanced signal processing techniques are required to process them. The main focus of this thesis has been blind source separation of multichannel data. Meanwhile, a method for single channel analysis has also been proposed.

The proposed extraction methods, which exploit the periodic structure of the signal, are evaluated on both artificially generated data and real biomedical signals. The methods can be summarized and listed as follows:

1. An efficient method to select the time delays to be used in second order blind identification of periodic signals.

2. A localization method derived from singular spectrum analysis to localize heart sound components in time domain representation of respiratory signals.

3. Blind extraction methods for cyclostationary sources. These approaches are considered for cyclostationary sources with distinct and common cycle frequencies.

The general BSS concept was introduced in Chapter 2, where the mathematical formulations were provided. More details about a number of well known BSS methods were described. Approaches for blind separation of periodic sources including periodic component analysis and cyclostationary source separation methods were reviewed. Also, a comprehensive review of the cyclostationary source separation methods for both instantaneous and convolutive mixtures was provided.

In some of the BSS problems a number of matrices are jointly diagonalized. For example, the covariance matrices are jointly diagonalized in the SOBI method to obtain average eigenvalue decomposition. In Chapter 3, a method for selecting appropriate time delays to calculate the covariance matrices was presented. The results were compared with the results of πCA and SOBI, where different numbers of randomly chosen covariance matrices were used in the separation algorithm. Using the values of the defined performance index, it was shown that although the number of covariance matrices for the proposed method was too low, the performance was almost equal to that of SOBI.

Heart sound localization is an important pre-processing step in most of heart sound removal methods. In Chapter 4 singular spectrum analysis was used to localize the fundamental heart sound components in single channel mixtures of heart and lung sounds. Inspecting the principal components resulted from decomposition of the signals, a criterion for selecting a subspace, which contains more information about S1 and S2 components, was introduced. The proposed localization method was applied to synthetic signals and also real respiratory signals from three subjects in different breathing flow rates and the results were compared with those of two other localization methods.

The proposed methods for extracting cyclostationary sources were expressed in chapters 5 and 6. In a typical system, there might be more than one phenomenon generating cyclostationary sources. If these origins have different periodic behaviours, the cycle frequencies will be distinct. On the other hand, if different independent origins act with identical periodicities or an origin can cause some independent sources, the sources will

have common cycle frequencies. Two methods for these cases were proposed in Chapter 5. Using the first method it is possible to extract the cyclostationary sources subject to having some prior knowledge about the cycle frequency of the source of interest. When more than one cyclostationary source exist, it is enough to extract the sources and deflate them one by one. Second method, called CSE, is based on the assumption that some independent sources have a common cycle frequency.

Four performance indices were defined in Chapter 6 to evaluate the performance of the extraction methods. The CSE method was used to extract the cyclostationary sources from artificial random mixtures of a number of signals. The results were then compared with the results of fastICA, Infomax, and SOBI. It was shown that the CSE method is able to preserve the remaining signals (after deflating the sources of interest) better than the other methods.

Ballistocardiogram artifact is known as one of the main artifacts contaminating EEG data recorded simultaneously with fMRI sequences. This artifact is caused by the subject's body inside the MR scanner as the result of cardiac pulsation. It is usually assumed that BCG artifact is composed of a number of independent components [78]. The CSE method was used to remove BCG artifact and the results were compared with those of AAS and OBS as the benchmarks for this type of artifact.

As there is not that much information available about the correct number of independent BCG artifact channels, the number of BCG artifact channels to be extracted was selected using the performance index values. For AAS, OBS, and CSE methods the values of the PIs were evaluated with different assumptions about the number of independent BCG artifact components. Analyzing the results from five different subjects it was concluded that the artifacts contribute to four independent channels. The main advantage of CSE over the other methods for BCG artifact removal is that there is no need for any QRS detection system needed for the other methods. QRS detection is usually performed by analyzing the ECG channel.

7.2 Future work

Removing ballistocardiogram artifact using the cyclostationary source extraction method is a contribution which will put a step forward in usage of the cyclostationarity concept and corresponding algorithms in biomedical engineering. So far, cyclostationarity has been used mostly in telecommunications [43] and recently in mechanics [11]. Hopefully, this work and the associated publications will attract attention of researchers from other fields of research (especially biomedical signal processing) to the high applied and theoretical potential of the cyclostationary signal processing.

A prerequisite for successful extraction of cyclostationary sources is estimation of the cycle frequency. This can be one of the reasons that the techniques which exploit the cyclostationarity are not widely used. If some information about the sources of interest is available, estimating the cycle frequencies will be possible by investigating the power spectral densities. However, if the energies of cyclostationary sources are low, accurate estimation of cycle frequencies can be difficult. This can be included in any future research on cyclostationary BSS.

Exploiting the periodicity in the time representation or the statistics of a signal provides a great opportunity for analysis and processing of the signal. Cyclostationary signal processing provides useful techniques for different applications dealing with signals with periodic structures. In this dissertation, the CSE method was utilized in a particular biomedical application, i.e., cancelling the artifacts caused by the effects of MR scanner on EEG data. However, there are many more applications which can be dealt with using this method. Some of the candidates for such applications include localization and magnitude estimation of cardiac related artifacts in BOLD data [31] and magnetohydrodynamic (MHD) removal from multichannel ECG recordings during magnetic resonance imaging [82, 19]. Another possible application of this method is condition monitoring of rotating machinery. Cyclostationary techniques have been used in this field and the research is still on going [11, 12].

Singular spectrum analysis is a powerful technique for analyzing time series. This method has been successfully utilized in different applications. However, there is still no general rule for selecting the embedding window length. This can be an attractive research path for future work.

Subspace selection is also a challenging issue in using SSA for different applications. A variety of measures can be utilized to compare the statistical or spectral properties of the candidate subspaces with what we expect from the sources of interest. For example, one can measure the second order statistics of different subspaces at some particular time delays and select among them. Kurtosis can be another suitable measure when there is significant difference between Gaussianity of expected subspaces. Some possible candidate applications of this technique in biomedical signal processing include epilepsy prediction from EEG signals, localizing or removing heart generated artifacts from different biomedical recordings, and extracting murmurs from heart sound signals.

References

[1] K. Abed-Meraim, Y Xiang, J. H. Manton, and Y. Hua, *Blind source-separation using second-order cyclostationary statistics*, IEEE Transactions on Signal Processing **49** (2001), no. 4, 694–701.

[2] T. E. Abrudan, J. Eriksson, and V. Koivunen, *Steepest descent algorithms for optimization under unitary matrix constraint*, IEEE Transactions on Signal Processing **56** (2008), no. 3, 1134–1147.

[3] C. Ahlstrom, O. Liljefeldt, P. Hult, and P. Ask, *Heart sound cancellation from lung sound recordings using recurrence time statistics and nonlinear prediction*, IEEE Signal Processing Letters **12** (2005), no. 12, 812–815.

[4] C. Ahlstrom, T. Lnne, P. Ask, and A. Johansson, *A method for accurate localization of the first heart sound and possible applications*, Physiological Measurements **29** (2008), no. 3, 417–428.

[5] T. Alexandrov, *A method of trend extraction using singular spectrum analysis*, REVSTAT - Statistical Journal **7** (2009), no. 1, 1–22.

[6] P. J. Allen, O. Josephs, and R. Turner, *A method for removing imaging artifact from continuous EEG recorded during functional MRI*, NeuroImage **12** (2000), no. 2, 230–239.

[7] P. J. Allen, G. Polizzi, K. Krakow, D. R. Fish, and L. Lemieux, *Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction*, NeuroImage **8** (1998), no. 3, 229–239.

[8] S. Amari, *Natural gradient works efficiently in learning*, Neural Computation **10** (1998), no. 2, 251–276.

[9] S. Amari and A. Cichocki, *Adaptive blind signal processing-neural network approaches*, Proceedings of the IEEE **86** (1998), no. 10, 2026–2048.

[10] S. Amari and S. C. Douglas, *Why natural gradient?*, International Conference on Acoustics, Speech, and Signal Processing, ICASSP-98. **2** (1998), 1213–1216 vol.2.

[11] J. Antoni, *Cyclostationarity by examples*, Mechanical Systems and Signal Processing **23** (2009), no. 4, 987–1036.

[12] J. Antoni, J. Daniere, and F. Guillet, *Effective vibration analysis of IC engines using cyclostationarity. Part I-A methodology for condition monitoring*, Journal of Sound and Vibration **257** (2002), no. 5, 815–837.

[13] P. J. Arnott, Pfeiffer G. W., and Tavel M. E., *Spectral analysis of heart sounds: relationships between some physical characteristics and frequency spectra of first and second heart sounds in normals and hypertensives*, Journal of Biomedical Engineering **6** (1984), no. 2, 121–128.

[14] R. T. Behrens and L. L. Scharf, *Signal processing applications of oblique projection operators*, IEEE Transactions on Signal Processing **42** (1994), no. 6, 1413–1424.

[15] A. J. Bell and T. J. Sejnowski, *An information-maximization approach to blind separation and blind deconvolution*, Neural Computation **7** (1995), no. 6, 1129–1159.

[16] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, *A blind source separation technique using second-order statistics*, IEEE Transactions on Signal Processing **45** (1997), no. 2, 434–444.

[17] A. Belouchrani and A. Cichocki, *Robust whitening procedure in blind source separation context*, Electronics Letters **36** (2000), no. 24, 2050–2051.

[18] C. Benar, Y. Aghakhani, Y. Wang, A. Izenberg, A. Al-Asmi, F. Dubeau, and J. Gotman, *Quality of EEG in simultaneous EEG-fMRI for epilepsy*, Clinical Neurophysiology **114** (2003), no. 3, 569–580.

[19] B. Bhatt and M. R. Reddy, *ICA based flow artifact removal from ECG during MRI*, International Conference on Advances in Computing, Control, Telecommunication Technologies (2009), 241–243.

[20] G. Bonmassar, P. L. Purdon, I. P. Jskelinen, K. Chiappa, V. Solo, E. N. Brown, and J. W. Belliveau, *Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI*, NeuroImage **16** (2002), no. 4, 1127–1141.

[21] E. Braunwald, *Heart disease: A textbook of cardiovascular medicine, fifth edition*, W.B. Saunders Co., Philadelphia, 1997.

[22] E. N. Bruce, *Biomedical signal processing and signal modeling*, Wiley-Interscience, November 2000.

[23] J.-F. Cardoso, *Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem*, International Conference on Acoustics, Speech, and Signal Processing, ICASSP-90., 3-6 1990, pp. 2655 –2658.

[24] J.-F. Cardoso and A. Souloumiac, *Blind beamforming for non gaussian signals*, IEE Proceedings-F **140** (1993), 362–370.

[25] J.-F. Cardoso and A. Souloumiac, *Jacobi angles for simultaneous diagonalization*, SIAM Journal on Matrix Analysis and Applications **17** (1996), no. 1, 161–164.

[26] S. Choi and A. Cichocki, *Blind separation of nonstationary sources in noisy mixtures*, Electronics Letters **36** (2000), no. 9, 848–849.

[27] S. Choi, A. Cichocki, H. M. Park, and S. Y. Lee, *Blind source separation and independent component analysis: A review*, Neural Information Processing - Letters and Reviews **6** (2005), no. 1, 1–57.

[28] A. Cichocki and S. Amari, *Adaptive blind signal and image processing: learning algorithms and applications*, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[29] A. Cichocki, B. Orsier, A. Back, and S.-I. Amari, *On-line adaptive algorithms in non-stationary environments using a modified conjugate gradient approach*, (1997), 316–325.

[30] G. D. Clifford, *Signal processing methods for heart rate variability*, Ph.D. thesis, Department of Engineering Science, University of Oxford, 2002.

[31] M. S. Dagli, J. E. Ingeholm, and J. V. Haxby, *Localization of Cardiac-Induced Signal Change in fMRI*, NeuroImage **9** (1999), no. 4, 407–415.

[32] S. Debener, K. J. Mullinger, R. K. Niazy, and R. W. Bowtell, *Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 T static magnetic field strength*, International Journal of Psychophysiology **67** (2008), no. 3, 189–199.

[33] S. Debener, A. Strobel, B. Sorger, J. Peters, C. Kranczioch, A. K. Engel, and R. Goebel, *Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: Removal of the ballistocardiogram artefact*, NeuroImage **34** (2007), no. 2, 587–597.

[34] A. Delorme and S. Makeig, *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis*, Journal of neuroscience methods **134** (2004), 9–21.

[35] L. G. Durand and P. Pibarot, *Digital signal processing of the phonocardiogram: review of the most recent advancements*, Critical Reviews in Biomedical Engineering **23(3-4)** (1995), 163–219.

[36] M. Dyrholm, R. Goldman, P. Sajda, and T.R. Brown, *Removal of BCG artifacts using a non-Kirchhoffian overcomplete representation*, IEEE Transactions on Biomedical Engineering **56** (2009), no. 2, 200–204.

[37] T. H. Falk and Wai-Yip C., *Modulation filtering for heart and lung sound separation from breath sound recordings*, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008. (2008), 1859–1862.

[38] A. Ferreol and P. Chevalier, *On the behavior of current second and higher order blind source separation methods for cyclostationary sources*, IEEE Transactions on Signal Processing **48** (2000), no. 6, 1712–1725.

[39] D. Flores-Tapia, Z. M. K. Moussavi, and G. Thomas, *Heart sound cancellation based on multiscale products and linear prediction*, IEEE Transaction on Biomedical Engineering **54** (2007), no. 2, 234–243.

[40] W. A. Gardner, *Exploitation of spectral redundancy in cyclostationary signals*, IEEE Signal Processing Magazine **8** (1991), no. 2, 14–36.

[41] ———, *Cyclostationarity in communications and signal processing*, IEEE Press, New York, NY, USA, 1993.

[42] W. A. Gardner, *An introduction to cyclostationary signals*, IEEE Press, New York, June 1994.

[43] W. A. Gardner, A. Napolitano, and L. Paura, *Cyclostationarity: Half a century of research*, Signal Process. **86** (2006), no. 4, 639–697.

[44] F. Ghaderi, B. Makkiabadi, J. G. McWhirter, and S. Sanei, *Blind source extraction of cyclostationary sources with common cyclic frequencies*, International Conference on Acoustics, Speech, and Signal Processing, ICASSP-10. (2010), 4146–4149.

[45] F. Ghaderi, H. R. Mohseni, J. G. McWhirter, and S. Sanei, *Blind source extraction of periodic signals*, International Conference on Acoustics, Speech, and Signal Processing, ICASSP-09. (2009), 377–380.

[46] F. Ghaderi, S. Sanei, B. Makkiabadi, Vahid Abolghasemi, and J. G. McWhirter, *Heart and lung sound separation using periodic source extraction method*, 16th International Conference on Digital Signal Processing, DSP (2009), 1–6.

[47] R. R. Gharieb and A. Cichocki, *Second-order statistics based blind source separation using a bank of subband filters*, Digital Signal Processing **13** (2003), no. 2, 252–274.

[48] M. Ghodsi, H. Hassani, S. Sanei, and Y. Hicks, *The use of noise information for detection of temporomandibular disorder*, Biomedical Signal Processing and Control **4** (2009), no. 2, 79–85.

[49] J. Gnitecki and Z. M. K. Moussavi, *Separating heart sounds from lung sounds - accurate diagnosis of respiratory disease depends on understanding noises*, IEEE Engineering in Medicine and Biology Magazine **26** (2007), no. 1, 20–29.

[50] R. I. Goldman, J. M. Stern, J. Engel, and M. S. Cohen, *Acquiring simultaneous EEG and functional MRI*, Clinical Neurophysiology **111** (2000), no. 11, 1974–1980.

[51] G. H. Golub and C. F. Van Loan, *Matrix computations*, second ed., Baltimore, MD, 1989.

[52] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of time series structure: SSA and related techniques*, Chapman and Hall/CRC, 2001.

[53] S. I. Goncalves, P. J. W. Pouwels, J. P. A. Kuijer, R. M. Heethaar, and J. C. de Munck, *Artifact removal in co-registered EEG/fMRI by selective average subtraction*, Clinical Neurophysiology **118** (2007), no. 11, 2437–2450.

[54] A. C. Guyton and J. E. Hall (eds.), *Textbook of medical physiology*, eleventh ed., W. B. Saunders Co., 2006.

[55] L. J. Hadjileontiadis and S. M. Panas, *Adaptive reduction of heart sounds from lung sounds using fourth-order statistics*, IEEE Transactions on Biomedical Engineering **44** (1997), no. 7, 642–648.

[56] ———, *Separation of discontinuous adventitious sounds from vesicular sounds using a wavelet-based filter*, IEEE Transactions on Biomedical Engineering **44** (1997), no. 12, 1269–1281.

[57] C. W. Hesse and C. J. James, *The FastICA algorithm with spatial constraints*, IEEE Signal Processing Letters **12** (2005), no. 11, 792–795.

[58] A. Hoffmann, L. Jger, K.J. Werhahn, M. Jaschke, S. Noachtar, and M. Reiser, *Electroencephalography during functional echo-planar imaging: detection of epileptic spikes using post-processing methods*, Magnetic Resonance in Medicine **44** (2000), no. 5, 791–798.

[59] I. Hossain and Z. Moussavi, *An overview of heart-noise reduction of lung sound using wavelet transform based filter*, Proc. of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society **1** (2003), 458–461 Vol.1.

[60] A. Hyvärinen, *Complexity pursuit: separating interesting components from time series*, Neural Computation **13** (2001), no. 4, 883–898.

[61] A. Hyvärinen and E. Oja, *A fast fixed-point algorithm for independent component analysis*, Neural Computation **9** (1997), 1483–1492.

[62] H. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, Wiley-Interscience, 2001.

[63] V. K. Iyer, P. A. Ramamoorthy, H. Fan, and Y. Ploysongsang, *Reduction of heart sounds from lung sounds by adaptive filtering*, IEEE Transactions on Biomedical Engineering **BME-33** (1986), no. 12, 1141–1148.

[64] M. G. Jafari, S. R. Alty, and J. A. Chambers, *New natural gradient algorithm for cyclostationary sources*, IEE Proceedings Vision, Image and Signal Processing **151** (2004), no. 1, 62–68.

[65] M. G. Jafari, W. Wang, J. A. Chambers, T. Hoya, and A. Cichocki, *Sequential blind source separation based exclusively on second-order statistics developed for a class of periodic signals.*, IEEE Transactions on Signal Processing **54** (2006), no. 3, 1028–1040.

[66] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, *ICA: a potential tool for BCI systems*, IEEE Signal Processing Magazine **25** (2008), no. 1, 57–68.

[67] B. Kamousi, Z. Liu, and B. He, *Classification of motor imagery tasks for brain-computer interface applications by means of two equivalent dipoles analysis*, IEEE Transactions on Neural Systems and Rehabilitation Engineering **13** (2005), no. 2, 166–171.

[68] K. H. Kim, H. W. Yoon, and H. W. Park, *Improved ballistocardiac artifact removal from the electroencephalogram recorded in fMRI*, Journal of Neuroscience Methods **135** (2004), no. 1-2, 193–203.

[69] M. Kompis and E. Russi, *Adaptive heart-noise reduction of lung sounds recorded by a single microphone*, Proc of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society **2** (1992), 691–692.

[70] A. Kosma, K. Nazarpour, and S. Sanei, *Removal of ballistocardiogram artifact from electroencephalograms exploiting heart rate variability*, 16th International Conference on Digital Signal Processing (2009).

[71] X. Li and X. Zhang, *Sequential blind extraction adopting second-order statistics*, IEEE Signal Processing Letters **14** (2007), no. 1, 58–61.

[72] W. Liu and D. P. Mandic, *A normalised kurtosis-based algorithm for blind source extraction from noisy measurements*, Signal Processing **86** (2006), no. 7, 1580–1585.

[73] J. Mamou and E. J. Feleppa, *Singular spectrum analysis applied to ultrasonic detection and imaging of brachytherapy seeds*, J. Acoustical Society of America **121** (2007), no. 3, 1790–1801.

[74] D. Mantini, M. G. Perrucci, S. Cugini, A. Ferretti, G. L. Romani, and C. Del Gratta, *Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis*, NeuroImage **34** (2007), no. 2, 598–607.

[75] I. Miller, J. E. Freund, and R. Johnson, *Probability and statistics for engineers*, fourth ed., Englewood Cliffs NJ: Prentice-Hall Inc., 1990.

[76] L. Molgedey and H. G. Schuster, *Separation of a mixture of independent signals using time delayed correlations*, Physical Review Letters **72** (1994), 3634–3637.

[77] N. Murata, S. Ikeda, and A. Ziehe, *An approach to blind source separation based on temporal structure of speech signals*, Neurocomputing **41** (2001), no. 1-4, 1–24.

[78] W. Nakamura, K. Anami, T. Mori, O. Saitoh, A. Cichocki, and S. Amari, *Removal of ballistocardiogram artifacts from simultaneously recorded EEG and fMRI data using independent component analysis*, IEEE Trans. Biomed. Eng. **53** (2006), no. 7, 1294–1308.

[79] K. Nazarpour, Y. Wongsawat, S. Sanei, J.A. Chambers, and S. Oraintara, *Removal of the eye-blink artifacts from EEGs via STF-TS modeling and robust minimum variance beamforming*, IEEE Transactions on Biomedical Engineering **55** (2008), no. 9, 2221–2231.

[80] NHS, *Carotid arteriogram*, [Online; accessed 12-Oct-2010].

[81] R. K. Niazy, C. F. Beckmann, G. D. Iannetti, J. M. Brady, and S. M. Smith, *Removal of FMRI environment artifacts from EEG data using optimal basis sets*, NeuroImage **28** (2005), no. 3, 720–737.

[82] F. Odille, C. Pasquier, R. Abacherli, P.-A. Vuissoz, G. P. Zientara, and J. Felblinger, *Noise cancellation signal processing method and computer system for improved real-time electrocardiogram artifact correction during MRI data acquisition*, IEEE Transactions on Biomedical Engineering **54** (2007), no. 4, 630–640.

[83] D. Ostwald, C. Porcaro, and A. P. Bagshaw, *An information theoretic approach to EEG-fMRI integration of visually evoked responses*, NeuroImage **49** (2010), no. 1, 498–516.

[84] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, *A survey of convolutive blind source separation methods*, Springer Handbook of Speech Processing (2007).

[85] W. C. A. Pereira, S. L. Bridal, A. Coron, and P. Laugier, *Singular spectrum analysis applied to backscattered ultrasound signals from in vitro human cancellous bone specimens*, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control **51** (2004), no. 3, 302–312.

[86] A. Pietilä, M. El-Segaier, R. Vigário, and E. Pesonen, *Blind source separation of cardiac murmurs from heart recordings*, Springer Berlin / Heidelberg, 2006.

[87] M. T. Pourazad, Z. Moussavi, F. Farahmand, and R. K. Ward, *Heart sounds separation from lung sounds using independent component analysis*, 27th Annual International Conference of the Engineering in Medicine and Biology Society, EMBS 2005. (2005), 2736–2739.

[88] M. T. Pourazad, Z. Moussavi, and G. Thomas, *Heart sound cancellation from lung sound recordings using time-frequency filtering*, Medical and Biological Engineering and Computing **44** (2006), no. 3, 216–225.

[89] R. Sameni, C. Jutten, and M.B. Shamsollahi, *Multichannel electrocardiogram decomposition using periodic component analysis*, IEEE Transactions on Biomedical Engineering **55** (2008), no. 8, 1935–1940.

[90] S. Sanei and J. A. Chambers, *EEG Signal Processing*, Wiley, John & Sons, Incorporated, 2007.

[91] L. K. Saul and J. B. Allen, *Periodic component analysis: An eigenvalue method for representing periodic structure in speech*, Advances in Neural Information Processing Systems (2001), 807–813.

[92] J. Sijbers, J. V. Audekerke, M. Verhoye, A. V. der Linden, and D. V. Dyck, *Reduction of ECG and gradient related artifacts in simultaneously recorded human EEG/MRI data*, Magnetic Resonance Imaging **18** (2000), no. 7, 881–886.

[93] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*, Elsevier Academic Press, June 2005.

[94] A. Sovijarvi, L. Malmberg, G. Charbonneau, F. Vanderschoot, J. Dalmasso, C. Sacco, M. Rossi, and J. Earis, *Characteristics of breath sounds and adventitious respiratory sounds*, Euroean Respiratory Review (2000), 591–596.

[95] G. Srivastava, S. Crottaz-Herbette, K.M. Lau, G.H. Glover, and V. Menon, *ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner*, NeuroImage **24** (2005), no. 1, 50–60.

[96] G. Strang, *Linear algebra and its applications*, 3rd ed., Harcourt Brace Jovanovich, 1988.

[97] L. Tong, R. W. Liu, V. C. Soon, and Y. F. Huang, *Indeterminacy and identifiability of blind identification*, IEEE Transactions on Circuits and Systems **38** (1991), no. 5, 499–509.

[98] K. Torkkola, *Blind separation of delayed sources based on information maximization*, vol. 6, May 1996, pp. 3509–3512.

[99] T. Tsalaile, S.M. Naqvi, K. Nazarpour, S. Sanei, and J.A. Chambers, *Blind source extraction of heart sound signals from lung sound recordings exploiting periodicity of the heart sound*, International Conference on Acoustics, Speech, and Signal Processing, ICASSP-08. (2008), 461–464.

[100] T. Tsalaile, R. Sameni, S. Saeid, C. Jutten, and J. Chambers, *Sequential blind source extraction for quasi-periodic signals with time-varying period*, IEEE Transactions on Biomedical Engineering **56** (2009), no. 3, 646 –655.

[101] T. Tsalaile and S. Sanei, *Separation of heart sound signal from lung sound signal by adaptive line enhancement*, 15th European Signal Processing Conference (EUSIPCO) (2007), 1231–1235.

[102] R. Vigario, J. Sarela, V. Jousmiki, M. Hamalainen, and E. Oja, *Independent component approach to the analysis of EEG and MEG recordings*, IEEE Transactions on Biomedical Engineering **47** (2000), no. 5, 589–593.

[103] X. Wan, K. Iwata, J. Riera, T. Ozaki, M. Kitamura, and R. Kawashima, *Artifact reduction for EEG/fMRI recording: Nonlinear reductionof ballistocardiogram artifacts*, Clinical Neurophysiology **117** (2006), no. 3, 668–680.

[104] W. Wang, M. G. Jafari, S. Sanei, and J. A. Chambers, *Blind separation of convolutive mixtures of cyclostationary sources using an extended natural gradient method*, Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on **2** (2003), 93–96.

[105] W. Wang, M.G. Jafari, S. Sanei, and J.A. Chambers, *Blind separation of convolutive mixtures of cyclostationary signals*, International Journal of Adaptive Control and Signal Processing **18** (2004), no. 3, 279–298.

[106] R. E. Wendt, R. Rokey, G. W. Vick, and D. L. Johnston, *Electrocardiographic gating and monitoring in NMR imaging*, Magnetic Resonance Imaging **6** (1988), no. 1, 89–95.

[107] Wikipedia, *Auscultation — Wikipedia, the free encyclopedia*, 2010, [Online; accessed 10-Oct-2010].

[108] A. Yadollahi and Z. Moussavi, *A robust method for heart sounds localization using lung sounds entropy*, IEEE Transactions on Biomedical Engineering **53** (2006), no. 3, 497–502.

[109] A. Yeredor, *Blind source separation with pure delay mixtures*, The 3rd International Workshop on Independent Component Analysis and Blind Source Separation (2001).