Signals and Communication Technology

Ganesh R. Naik
Wenwu Wang *Editors*

Blind Source Separation

Advances in Theory, Algorithms and Applications

Springer

Signals and Communication Technology

For further volumes:
http://www.springer.com/series/4748

Ganesh R. Naik · Wenwu Wang
Editors

Blind Source Separation

Advances in Theory, Algorithms and Applications

Springer

Editors
Ganesh R. Naik
University of Technology
Sydney
Australia

Wenwu Wang
University of Surrey
Guildford
UK

ISSN 1860-4862 ISSN 1860-4870 (electronic)
ISBN 978-3-642-55015-7 ISBN 978-3-642-55016-4 (eBook)
DOI 10.1007/978-3-642-55016-4
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014940320

© Springer-Verlag Berlin Heidelberg 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Blind source separation (BSS) methods have received extensive attention over the past two decades; thanks to its wide applicability in a number of areas such as biomedical engineering, audio signal processing, and telecommunications. The problem of source separation is an inductive inference problem, as only limited information, e.g., the sensor observations, is available to infer the most probable source estimates. The aim of BSS is to process these observations (acquired by sensors or sensor arrays) in such a way that the original unknown source signals are extracted by, e.g., an adaptive system, or separated simultaneously using, e.g., a block (or batch)-based algorithm, without knowing or with very limited information about the characteristics of the transmission channels through which the sources propagate to the sensors. Independent component analysis (ICA) is one of the early and most widely used techniques for BSS by revealing the hidden factors that underlie the sets of measurements or the observed signals. Recently, a number of new techniques have been emerging in BSS, such as latent variable analysis, non-negative matrix/tensor factorization (NMF/NTF), sparse component analysis, dictionary learning, independent vector analysis, factor analysis, matrix completion, compressed sensing, empirical mode decomposition, and complex-valued adaptive methods. At the same time, the applications of BSS continue to grow and prosper in a number of areas, such as audio, speech, music, image, biomedical, communication, and financial data analysis and processing.

This book aims to disseminate timely to the scientific community the new developments in BSS spanning from theoretical frameworks, algorithmic developments, to a variety of applications. The book covers some emerging techniques in BSS, especially those developed recently, offering academic researchers and practitioners a comprehensive update about the new development in this field. The book provides a forum for researchers to exchange their ideas, and to foster a better understanding of the state of the art of the subject. We envisage that the publication of this book will motivate new ideas and more cutting-edge research activities in this area.

This book is intended for computer science and electronics engineers (researchers and graduate students) who wish to get novel research ideas and some training in BSS, ICA, machine learning, artificial intelligence, and signal processing applications. Furthermore, the research results previously scattered in many scientific articles worldwide are methodically collected and presented in the

book in a unified form. As a result of its twofold character, the book is likely to be of interest to researchers, engineers, and graduates who wish to learn the core principles, methods, algorithms, and applications of BSS. Furthermore, the book may also be of broader interest to researchers working in other areas of science and engineering, due to the multidisciplinary nature of this book.

The book is organized into two parts. Part I is devoted to recent developments in theories, algorithms, and extensions of BSS. In this part, we have collected nine chapters with several novel contributions, namely, the idea of quantum ICA by Yannick Deville and Alain Deville, the singularity-aware dictionary learning approach for BSS by Xiaochen Zhao, Guangyu Zhou, Wenwu Wang, and Wei Dai, the theoretical results on the performance of complex ICA by Benedikt Loesch and Bin Yang, sub-band based BSS by Bo Peng and Wei Liu, independent vector analysis for frequency domain BSS by Yanfeng Liang, Syed Mohsen Naqvi, Wenwu Wang, and Jonathon A. Chambers, sparse component analysis by Yannick Deville, underdetermined source separation by Nikolaos Mitianoudis, NMF based source separation by Bin Gao and Wai Lok Woo, and a BSS related topic of source localisation and tracking by Md Mashud Hyder and Kaushik Mahata. Part II focuses on the various applications of BSS and its links to other relevant areas, such as computational auditory scene analysis (CASA). We have gathered 10 chapters in this part. They are respectively, blind speech extraction algorithms by Hiroshi Saruwatari and Ryoichi Miyazaki, combining superdirective beamforming and BSS for speech separation by Lin Wang, Heping Ding, and Fuliang Yin, ideal ratio mask for CASA by Christopher Hummersone, Toby Stokes, and Tim Brookes, monaural speech enhancement by Masoud Geravanchizadeh and Reza Ahmadnia, background/foreground separation by Zafar Rafii, Antoine Liutkus, and Bryan Pardo, NMF-based sparse coding for cochlear implants by Hongmei Hu, Guoping Li, Mark E. Lutman, and Stefan Bleeck, brain signal analysis using ICA by Ruben Martin-Clemente, BSS for the analysis of large-scale omic datasets by Andrew E. Teschendorff, Emilie Renard, and Pierre A. Absil, ICA for complex domain source separation of communication signals by Ajay K. Kattepur and Farook Sattar, and semi-blind source separation algorithms from non-invasive electrophysiology to neuro-imaging by Camillo Porcaro and Franca Tecchio.

We would like to thank the authors for their excellent submissions (chapters) to this book, and their significant contributions to the review process, which have helped to ensure the high quality of this publication. Without their contributions, it would have not been possible for the book to come successfully into existence.

January 2014

Ganesh R. Naik
Wenwu Wang

Contents

Part I
Theory, Algorithms, and Extensions

Chapter 1
Quantum-Source Independent Component Analysis and Related Statistical Blind Qubit Uncoupling Methods

Yannick Deville and Alain Deville

Abstract Quantum Information Processing (QIP) is an emerging field which yields new capabilities beyond classical, i.e., non-quantum, information processing. QIP methods manipulate quantum bit (qubit) states instead of classical bit values. Undesired coupling between these individual quantum states is expected, in the same way as classical systems involve undesired signal coupling. Methods for recovering individual quantum states from their coupled version are therefore required. To solve this problem, we recently introduced the field of Quantum Source Separation (QSS). We showed how to convert qubit states with cylindrical-symmetry Heisenberg coupling into classical-form data, mixed according to a specific nonlinear model, which was not previously studied in the literature. We therefore started to develop methods for unmixing such data. While we restricted ourselves to nonblind QSS methods and a basic blind approach in those previous works, we here proceed much further for the more difficult, i.e., blind, case: we introduce the concept of Quantum-Source Independent Component Analysis (QSICA), and we develop related QSS methods using various statistical signal processing tools, namely mutual information, likelihood and moments. The performance of the proposed approaches is validated by means of numerical tests. This especially shows the attractiveness of our method focused on second-order moments.

Y. Deville (✉)
UPS-CNRS-OMP, IRAP (Institut de Recherche en Astrophysique et Planétologie),
Université de Toulouse, 14 avenue Edouard Belin, 31400 Toulouse, France
e-mail: yannick.deville@irap.omp.eu

A. Deville
IM2NP, Aix-Marseille Univ, Campus Scientifique Saint-Jérôme, 13997 Marseille, France
e-mail: alain.deville@univ-amu.fr

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_1,
© Springer-Verlag Berlin Heidelberg 2014

1.1 Introduction

Source Separation (SS), also called Signal Separation, is a generic Information Processing (IP) problem. It consists in recovering a set of unknown source "signals" (time series, images...) from a set of observations (i.e. measured signals), which are mixtures of these source signals. In particular, the *Blind* Source Separation (BSS) configuration corresponds to the case when the parameter values of the considered mixing model are unknown. On the contrary, in the nonblind case, these values are either known a priori or first estimated from known observations and *known source signals* [18, 21]. The field of BSS emerged in the 1980s and then quickly expanded, as e.g., detailed in the books [11, 18, 33]. Until recently, all these investigations were performed in a "classical", i.e., nonquantum, framework.

Independently from (B)SS, another field within the overall IP domain rapidly developed during the last decades, namely Quantum Information Processing (QIP) [3, 23, 38, 47, 53]. QIP is closely related to Quantum Physics (QP). It uses abstract representations of systems whose behavior is requested to obey the laws of QP. This already made it possible to develop new and powerful IP methods, to be contrasted with classical methods such as the above-mentioned (B)SS approaches. These new methods manipulate the states of so-called quantum bits, or qubits. Their effective implementation then requires one to develop corresponding practical quantum systems, which is only an emerging topic today [38]: in the introductory paper [53], the author "venture[s] to say that the creation of a practical quantum computer may be possible within the next few decades".

We recently bridged the gap between classical (B)SS and QIP/QP, by introducing a new field, namely Quantum Source (or Signal) Separation (QSS), first proposed in our paper [17] and then especially detailed in [21]. The QSS problem consists in restoring the information contained in individual *quantum* source signals, i.e., source qubit states, only starting from the mixtures (in SS terms [21]) of these source qubit states which result from their undesired coupling. This gives rise to three possible approaches:

1. In the classical-processing approach to QSS [17, 21], one first converts the mixed quantum data into classical-form data (whose properties still reflect their quantum origin) by means of measurements, and then processes the measured data with classical (i.e, again, nonquantum) methods. We showed that original separation methods must be developed in this case, because the specific nonlinear mixing model which results from the considered type of qubit coupling was not previously addressed in the classical (B)SS literature. Without having to wait for the development of practical quantum circuits, this classical-processing version of our approach already applies to possible experiments requiring methods for retrieving information about individual quantum states from measurements performed after undesired coupling between these states, e.g., when dealing with quantum phenomena involving electron or nuclear spins 1/2.
2. Quantum-processing QSS methods [21] keep the quantum form of the available mixed data and process them by means of quantum circuits in order to

retrieve the quantum sources.[1] A potential application of this version of our QSS methods concerns the core of future quantum computers, where both the data to be processed and the processing means will have a quantum form. Quantum-processing QSS may then be used as a preprocessing stage, to remove undesired alterations (e.g., due to Heisenberg coupling between physical qubits, as in this chapter) of the data to be provided to the input of the main processing stage, which then applies the final quantum algorithm to these preprocessed data. This two-stage system architecture is already used in the classical context, where BSS is applied as a preprocessing stage to extract some or all source signals in various application fields, for instance:

a. In some audio systems, the final aim is the automatic recognition of speech by a processing unit, e.g., in order to then control actuators (for instance, a car driver can thus control various car functions by speech). However, when a speech signal is recorded by a set of microphones situated in a noisy environment, each recorded signal is a mixture of speech and of various noise signals. Providing these plain recordings to an automatic speech recognition (ASR) system yields degraded recognition performance. A solution to this problem consists in preprocessing these recordings by means of a BSS system, so as to extract the speech signal, and then providing the denoised speech output of this BSS system to the ASR system (see e.g. [25]).
b. Similarly, when using radio-frequency signals to transmit digital data, reception antennas may simultaneously receive several mixed data streams. BSS is then applied to first unmix these signals. Each extracted signal may then be separately used as required in the considered application. For instance, in the radio-frequency identification (RFID) system described in [16], the main processing stage then consists in decoding each data stream, so as to identify the person or object that emitted these data and to accordingly control the actuators of the considered application, e.g., to allow or not the RFID-tag bearer to access a restricted area.
c. Finally, in the biomedical field, a wide range of signals such as electrocardiograms (ECGs) or electroencephalograms (EEGs) are processed by human experts or computers in order to analyze various health disorders. This "main

[1] In the field of (B)SS, the term "source" sometimes refers to a physical object which provides (e.g., emits) a signal, but it is more often used as an abbreviation for "source signal", since (B)SS is more concerned with the processing of these signals than with the objects which provide them. This appears in the name of SS itself: performing "SS" of course does not mean that one physically extracts source objects one from another or from their overall set, but that one extracts the signals associated with such objects (by using the measured mixtures of these signals): to be precise, the field of SS is not "SS" but "source signal separation". For the sake of simplicity, we also often use the term "source" as an abbreviation for "source signal" in the field of QSS. For instance, in the sentence containing this footnote, "retriev[ing] the quantum sources" means "retrieving the quantum source signals", i.e. "retrieving the signals associated with quantum sources", where these quantum sources (i.e., these objects) consist of physical implementations of qubits. Similarly, the "source vector" considered further in this chapter is the vector composed of the values of source signals.

task" is often difficult because each signal in the recorded set is a mixture of various contributions, and the information of interest thus cannot be easily extracted from any such mixed signal. Again, a solution to this problem consists in preprocessing the original recordings by means of BSS methods, so as to extract each signal component of interest separately on each output of this BSS system. For instance, this approach has been used in [14] to preprocess multichannel ECG recordings which are mixtures of large-magnitude mother's heartbeats, low-magnitude fetus's heartbeats, and noise components. This made it possible to extract fetus's heartbeats, which are hardly visible in the original recordings.

3. Hybrid QSS methods [22] combine the above two approaches, by first partly processing the quantum mixtures with quantum circuits, then converting the resulting (simpler) quantum data into classical-form ones by means of measurements, and eventually processing the measured data with classical methods.

In this chapter, we only consider the first approach to QSS, based on classical-processing methods, which are the only easily implementable ones nowadays, due to the above-mentioned current state of quantum system development. As in the classical SS framework, these QSS methods give rise to two configurations, the nonblind and blind ones. The simpler configuration, i.e., the nonblind one, was addressed in our journal paper [21] and is not considered hereafter. In [21], we also briefly studied the more complex configuration, i.e., the blind one, which requires one to estimate the value(s) of the mixing model parameter(s). However, we only described a basic classical-processing method for performing this estimation. That method is based on the first-order moment of a measured signal and has the drawback of requiring some marginal source statistics to be known. Therefore, we here develop and compare various much more powerful methods for solving the considered Blind Quantum Source Separation (BQSS) problem, which involves an original nonlinear mixing model.

In the classical framework, several classes of methods have been proposed for solving the BSS problem for a given mixing model (the linear instantaneous model was mainly considered). The most popular of them is based on statistical signal processing and consists of Independent Component Analysis (ICA) and related methods (see especially [10, 11, 18, 33]). In a similar way, we here develop a class of methods which cover various statistical processing tools for solving the considered BQSS problem. As a first type of solutions, we introduce what we will call " Quantum-Source Independent Component Analysis (QSICA) methods", since they have the following two features. First, they are intended for data which initially have a quantum form (these data are here converted into classical-form data and then processed by classical means). Second, they are based on the application of the ICA principle to BSS in its strictest sense, i.e., they assume the source signals to be mutually statistically independent and they restore them by forcing the output signals of the separating system to become mutually statistically independent. Still considering the same data, we then develop other BQSS methods which are directly related to QSICA, i.e., which again assume the sources to be statistically independent but which use this property in other ways.

More precisely, the remainder of this chapter is organized as follows. As stated above, the fundamental concept used in QIP and QSS is the qubit. Therefore, we first define this concept in Sect. 1.2. Then, in Sect. 1.3, we summarize the mixing model that we developed in [21] for the data derived from two coupled qubits by means of quantum-to-classical conversion. The next sections are dedicated to this original nonlinear mixing model for classical-form data. We first present a separating system suited to this model in Sect. 1.4. In the subsequent sections, we propose various methods for estimating the unknown parameter of this system. The first methods, fully based on QSICA and involving mutual information, are presented in Sect. 1.5. Then, Sect. 1.6 describes an alternative approach based on the maximum likelihood principle and shows its close relationship with the methods of Sect. 1.5. Both the above types of methods exploit the whole statistics of the considered signals. Simpler BQSS methods focused on some moments or cumulants of these signals may also be derived, as was previously done in classical BSS. Several such methods are studied in Sect. 1.7 , where we put the emphasis on methods based on first-order or second-order statistics of specific signals. The numerical performance of the main methods described in this chapter is reported in Sect. 1.8 and conclusions are drawn from all this investigation in Sect. 1.9.

One should once and for all note that, whereas we are here concerned with configurations where one aims at extracting information about quantum states after *undesired* coupling (following Heisenberg's model), on the contrary a two-qubit gate using liquid NMR *takes advantage* [52] of the scalar coupling. Besides, as detailed in [21], QSS, and especially BQSS, are quite different from quantum state tomography and quantum process tomography [38], which were e.g. used in [55] for two-qubit systems. These two types of tomographic techniques cannot achieve BQSS [21].

1.2 Definition of a Single Qubit

As stated above, qubits are used instead of classical bits for performing computations in the field of QIP [38]. Whereas a classical bit can only take two values, usually denoted 0 and 1, a qubit with index i has a quantum state expressed, for a pure state, as

$$|\psi_i\rangle = \alpha_i|+\rangle + \beta_i|-\rangle \tag{1.1}$$

in the basis defined by the two orthonormal vectors that we hereafter[2] denote $|+\rangle$ and $|-\rangle$, where α_i and β_i are two complex-valued coefficients constrained to meet the condition

$$|\alpha_i|^2 + |\beta_i|^2 = 1 \tag{1.2}$$

[2] These vectors $|+\rangle$ and $|-\rangle$ are often respectively denoted as $|0\rangle$ and $|1\rangle$ (see e.g., [38]). We had to use the notations $|+\rangle$ and $|-\rangle$ in [21], to avoid confusion, and we keep them here.

which expresses that the state $|\psi_i\rangle$ is normalized. From a QP point of view, this abstract mathematical model especially concerns electron or nuclear spins 1/2, which are quantum (i.e., non-classical) objects. The component of such a spin, with index i, along a given arbitrary axis Oz defines a two-dimensional linear operator s_{iz}. The two eigenvalues of this operator are equal to $+\frac{1}{2}$ and $-\frac{1}{2}$ in normalized units, and the corresponding eigenvectors are therefore denoted $|+\rangle$ and $|-\rangle$. The value obtained when measuring this spin component can only be $+\frac{1}{2}$ or $-\frac{1}{2}$. Moreover, let us assume this spin is in the state $|\psi_i\rangle$ defined by (1.1) when performing such a measurement. Then, the probability that the measured value is equal to $+\frac{1}{2}$ (respectively $-\frac{1}{2}$) is equal to $|\alpha_i|^2$ (respectively $|\beta_i|^2$), i.e., to the squared modulus of the coefficient in (1.1) of the associated eigenvector $|+\rangle$ (respectively $|-\rangle$).

The above discussion concerns the state of the considered spin at a given time. In addition, this state evolves with time. The spin is here supposed to be placed in a static magnetic field and thus coupled to it. The time interval when it is considered is assumed to be short enough for the coupling between the spin and its environment to be negligible. In these conditions, the spin has a Hamiltonian. Therefore, if the spin state $|\psi_i(t_0)\rangle$ at time t_0 is defined by (1.1), it then evolves according to Schrödinger's equation and its value at any time t is

$$|\psi_i(t)\rangle = \alpha_i e^{-\mathrm{i}\omega_p(t-t_0)}|+\rangle + \beta_i e^{-\mathrm{i}\omega_m(t-t_0)}|-\rangle \tag{1.3}$$

where the imaginary unit i, present, e.g., in $e^{-\mathrm{i}\omega_p(t-t_0)}$, should not be confused with the qubit *index* i, and the real (angular) frequencies ω_p and ω_m depend on the considered physical setup.

1.3 Coupling/Mixing Model for Two Qubits

The above description directly applies to several qubits if they are not "coupled", i.e., if they do not interact with one another. One may however expect that undesired coupling between individual quantum states will have to be considered in the QIP/QP area, in the same way as signal coupling often is undesired in current *classical* signal processing systems. Coupling in quantum physical setups, e.g., occurs when two electron spins interact through exchange. In [21], we considered a two-qubit system composed of two distinguishable spins coupled according to the version of the Heisenberg model which has a cylindrical-symmetry axis, denoted Oz and collinear to the applied magnetic field. We analyzed in detail the global state resulting from that coupling and the associated measured values. Here again, the measured value of the component of each spin along axis Oz can only be $+\frac{1}{2}$ or $-\frac{1}{2}$. Therefore, when measuring the components of both spins, the obtained couple of values is equal to one of the four possible values $(+\frac{1}{2}, +\frac{1}{2})$, $(+\frac{1}{2}, -\frac{1}{2})$, $(-\frac{1}{2}, +\frac{1}{2})$ and $(-\frac{1}{2}, -\frac{1}{2})$. The probabilities of these four values are respectively denoted p_1, p_2, p_3 and p_4 hereafter. These probabilities are related as follows to the state of the overall system composed

of these two spins. This state may be expressed as a linear combination of the vectors of the four-dimensional basis $\{|++\rangle, |+-\rangle, |-+\rangle, |--\rangle\}$ which corresponds to the operators s_{1z} and s_{2z} respectively associated with the components of spin 1 and spin 2 along the symmetry axis Oz. As in Sect. 1.2, each of the probabilities p_1 to p_4 is here equal to the squared modulus of the coefficient of the corresponding basis vector in the expression of the overall system state. In [21], we provided a detailed derivation of the expressions of these probabilities in the following configuration. The two spins are separately initialized (i.e., prepared) at time t_0, with states defined by (1.1) where $i = 1$ for spin 1 and $i = 2$ for spin 2. The overall system state then evolves with time and the spin states thus get "mixed" (in the SS sense) with one another as follows. The time evolution of the overall system state is defined by phase rotations, as in (1.3), and this here involves four frequencies. These frequencies depend on Heisenberg coupling, which is especially characterized by the so-called principal value J_{xy} of the exchange tensor (see [21] for more details). We derived the expressions of the above probabilities p_1 to p_4 at an arbitrary time $t > t_0$, with respect to the polar representation of the qubit parameters α_i and β_i, which reads

$$\alpha_i = r_i e^{\mathrm{i}\theta_i} \qquad \beta_i = q_i e^{\mathrm{i}\phi_i} \qquad i \in \{1, 2\} \tag{1.4}$$

with $0 \leq r_i \leq 1$ and

$$q_i = \sqrt{1 - r_i^2} \tag{1.5}$$

due to (1.2). The above probabilities may then be expressed as follows:

$$p_1 = r_1^2 r_2^2 \tag{1.6}$$

$$p_2 = r_1^2(1 - r_2^2)(1 - v^2) + (1 - r_1^2)r_2^2 v^2 - 2r_1 r_2 \sqrt{1 - r_1^2}\sqrt{1 - r_2^2}\sqrt{1 - v^2} v \sin \Delta_I \tag{1.7}$$

$$p_4 = (1 - r_1^2)(1 - r_2^2) \tag{1.8}$$

with

$$\Delta_I = (\phi_2 - \phi_1) - (\theta_2 - \theta_1) \tag{1.9}$$

$$\Delta_E = -\frac{J_{xy}(t - t_0)}{\hbar} \tag{1.10}$$

$$v = \mathrm{sgn}(\cos \Delta_E) \sin \Delta_E \tag{1.11}$$

where $\hbar$ is the reduced Planck constant. Probability p_3 is not considered, since always

$$p_1 + p_2 + p_3 + p_4 = 1. \tag{1.12}$$

Equations (1.6)–(1.8) yield a QSS problem because, using the SS terminology, they show that some "observations" are "mixtures" of the quantities which define quantum "sources". This "mixing model" (1.6)–(1.8) involves the following items.

The observations are the probabilities p_1, p_2 and p_4 measured for each choice of the initial states (1.1) of the qubits. More precisely, these probabilities are not known exactly but estimated in practice. The procedure that we developed to this end in [21] operates as follows for each choice of the initial states (1.1) of the qubits. We repeatedly perform two operations: (i) we first initialize these qubits according to (1.1) and (ii) after a fixed time interval when coupling occurs, we measure the two spin components along Oz associated with the system composed of these two coupled qubits. The relative frequencies of occurence of all four possible couples of values of spin components (i.e. $(+\frac{1}{2}, +\frac{1}{2})$ to $(-\frac{1}{2}, -\frac{1}{2})$) then yield estimates of the corresponding probabilities. At this stage, we ignore the resulting estimation errors and therefore consider the exact mixing model (1.6)–(1.8). Using standard SS notations, the observation vector is therefore $x = [x_1, x_2, x_3]^{\mathrm{T}}$, where $^{\mathrm{T}}$ stands for transpose and[3]

$$x_1 = p_1, \qquad x_2 = p_2, \qquad x_3 = p_4. \tag{1.13}$$

Equations (1.6)–(1.8) show that the source vector to be retrieved from these observations turns out to be $s = [s_1, s_2, s_3]^{\mathrm{T}}$ with $s_1 = r_1, s_2 = r_2$ and $s_3 = \Delta_I$. The parameters q_i are then derived from (1.5). The four phase parameters in (1.4) cannot be individually extracted from their combination Δ_I (only two phases have a physical meaning [22]). To avoid ambiguities, one may therefore fix three of the phase parameters θ_1, ϕ_1, θ_2, and ϕ_2 (e.g., to 0) and only use the fourth parameter to store information. The transform from the sources to the observations defined by the nonlinear mixing model (1.6)–(1.8) involves a single "mixing parameter", namely v. As shown by (1.11), this parameter always meets the condition $0 \leq v^2 \leq 1$. In most configurations, the values of the coupling parameter J_{xy} and therefore of v (see (1.10)–(1.11)) are unknown (the sign of J_{xy} is however known). This corresponds to the *blind* version of this QSS problem. In this configuration, retrieving the sources first requires one to estimate the unknown mixing parameter v, which is the main topic of this chapter.

1.4 Separating System

In [21], we showed that the mixing model (1.6)–(1.8) is invertible (with respect to the considered domain of source values), for any fixed v such that $0 < v^2 < 1$, provided the source values meet the following conditions:

[3] It should be noted that the observed signals involved in this QSS problem have a specific nature, as compared to standard nonquantum BSS problems. In the latter problems, each value of an observed signal is usually the value of a measured physical quantity, such as the value of a voltage measured at a given time. On the contrary, as shown by (1.13), each value of an observed signal is here the value of a *probability* (which is estimated in practice). The overall signal composed of all successive values of a given observation (e.g., all values of x_1) therefore consists of a set of values of probabilities (e.g., all values of p_1), which depend on the values of the states used for initializing the qubits.

$$0 < r_1 < \tfrac{1}{2} < r_2 < 1 \tag{1.14}$$

$$-\tfrac{\pi}{2} \leq \Delta_I \leq \tfrac{\pi}{2}. \tag{1.15}$$

In these conditions, we now have to define a separating system, which aims at deriving an output vector $y = [y_1, y_2, y_3]^{\mathrm{T}}$ by combining the observations so that this output vector is equal to (an estimate of) the source vector $s = [s_1, s_2, s_3]^{\mathrm{T}}$ corresponding to these observations. This separating system therefore ideally aims at achieving the inverse of the mixing function (1.6)–(1.8). However, it uses a tunable parameter $\overline{v}$ instead of the actual value of the parameter v of the mixing model, since this value of v is unknown in the considered *blind* configuration. The idea is then to constrain the function achieved by the separating system to belong to a given class of functions which depend on the parameter $\overline{v}$, and to create BQSS methods for setting $\overline{v}$ to an estimate of v since, (only) for this value of $\overline{v}$, the output vector y becomes equal to the source vector s which yields the considered observed vector. The class of functions used for the separating system is derived as follows. One first determines the inverse of the mixing function, i.e., the inverse image $s = [s_1, s_2, s_3]^{\mathrm{T}}$ of a given observation vector $x = [x_1, x_2, x_3]^{\mathrm{T}}$, by solving (1.6)–(1.8). The resulting expression of s is provided in [21]. In this vector $s = [s_1, s_2, s_3]^{\mathrm{T}}$, one then replaces v by $\overline{v}$, and s_1, s_2, s_3 respectively by y_1, y_2, y_3. The resulting expression defines the output vector $y = [y_1, y_2, y_3]^{\mathrm{T}}$ of the separating system. Using the above-mentioned expression of s from [21], the elements of y read

$$y_1 = \sqrt{\frac{1}{2}\left[(1 + p_1 - p_4) - \sqrt{(1 + p_1 - p_4)^2 - 4p_1}\right]} \tag{1.16}$$

$$y_2 = \sqrt{\frac{1}{2}\left[(1 + p_1 - p_4) + \sqrt{(1 + p_1 - p_4)^2 - 4p_1}\right]} \tag{1.17}$$

$$y_3 = \mathrm{Arcsin}\left[\frac{y_1^2(1 - y_2^2)(1 - \overline{v}^2) + (1 - y_1^2)y_2^2\overline{v}^2 - p_2}{2y_1y_2\sqrt{1 - y_1^2}\sqrt{1 - y_2^2}\sqrt{1 - \overline{v}^2}\overline{v}}\right] \tag{1.18}$$

So, to summarize:

- Equations (1.16)–(1.18) define the intput/output relationship of our separating system, from the observations to the estimated sources.
- $\overline{v}$ is the tunable parameter of this separating system and should be set to an estimate of v. The next sections of this chapter describe the BQSS methods that we propose to this end.
- When $\overline{v} = v$, the outputs y_1, y_2 and y_3 respectively restore the sources $s_1 = r_1, s_2 = r_2$ and $s_3 = \Delta_I$.

1.5 QSICA Methods Based on Mutual Information

1.5.1 Proposed Approach

In the classical framework, ICA has been considered in two ways. It has first been used as a transform applied to given observations $x_i(t)$ (without referring to any sources), in order to obtain output signals $y_i(t)$ which are mutually statistically independent (or, at least, as independent as possible). Most often, this multidimensional transform has the following properties: (i) it is memoryless, i.e., its output at a given time t only depends on its input at the same time, and (ii) it is restricted to linear combinations of the observations.

The second situation is found when, in addition, the observations x_i are known to be a transformed version of a set of random source signals s_i. This "mixing transform" is often constrained to belong to a given functional class and to have unknown parameter values, while the source signals are unknown but constrained to be mutually statistically independent. ICA then again consists of a (separating) transform applied to the observations x_i, in order to obtain output signals y_i which are statistically independent. The functional form selected for the separating transform is matched to the mixing functional form, i.e., it is chosen so that, for certain values of the parameters of the separating transform (which depend on the parameter values of the mixing transform), the outputs y_i of the separating transform are equal to the source signals s_i, up to some acceptable residual transforms (permutation, scaling factors, ...). The latter transforms are the so-called "indeterminacies" of the considered "global model." This global model goes from the sources s_i to the outputs y_i, i.e. it combines the studied mixing and separating transforms.

A given global model is then said to be "ICA separable" [20] (for given marginal source statistics) if by using the above ICA principle, i.e., by considering mutually statistically independent sources and by adapting the separating transform only so that its outputs become independent, it is guaranteed that these outputs become equal to the sources, up to the above indeterminacies. If the considered model is ICA separable, then ICA can be used as one of the possible tools for performing BSS. On the contrary, applying ICA to a model which is not ICA separable may yield outputs which are still source mixtures, which is not acceptable for BSS. In the classical framework, many investigations have been devoted to linear instantaneous (i.e. memoryless) mixing and separating models, and it has been shown that this configuration is ICA separable for almost all marginal source statistics [10]. Nonlinear mixtures have been addressed in much less detail, because analyzing their ICA separability and developing associated ICA methods is then a much tougher problem. It has been shown that, if no constraints are imposed upon the nature of the mixing model, it yields an ICA-nonseparable configuration [11, 49].

In this section, we extend the above approach to the considered initially quantum data, which yield the above-defined mixing and separating models. We eventually aim

at developing corresponding QSICA methods for performing BQSS. To guarantee the complete relevance of these methods, we first address the ICA separability of the considered *nonlinear* global model.

1.5.2 ICA Separability of Studied Global Model

The global model studied in this chapter is called "the Heisenberg global model" hereafter. It is obtained by combining the mixing model (1.6)–(1.8) and the separating model (1.16)–(1.18). Some calculations show that this yields

$$y_1 = s_1 \tag{1.19}$$

$$y_2 = s_2 \tag{1.20}$$

$$y_3 = \text{Arcsin}\left[\frac{(\overline{v}^2 - v^2)(s_2^2 - s_1^2)}{2s_1 s_2\sqrt{1-s_1^2}\sqrt{1-s_2^2}\sqrt{1-\overline{v}^2}\overline{v}} + \frac{\sqrt{1-v^2}v}{\sqrt{1-\overline{v}^2}\overline{v}}\sin s_3\right], \tag{1.21}$$

again when ignoring the deviations which are due to the fact that p_1, p_2 and p_4 are estimated in practice.

In this global model, all output signals are therefore "reference signals", i.e., unmixed signals, except for one of them, y_3, which is a specific nonlinear function of *all* source signals, for a given arbitrary value of the separating parameter $\overline{v}$. This BQSS problem may therefore be considered as a specific nonlinear extension of the (linear) adaptive noise cancelation (ANC) problem, which was especially studied by Widrow et al. and e.g., reported in [56].[4]

In [20], we proved that a wide class of global nonlinear models involving reference signals are ICA separable. More precisely, we investigated memoryless mixing and separating models. We considered the random variables (RVs) defined by all continuous-valued signals at a single time, and we analyzed the case when the source RVs have given marginal statistics and are mutually statistically independent. We showed that, under mild conditions, if the output RVs of the separating system are mutually statistically independent, then they are equal to the source RVs , up to some acceptable indeterminacies which depend on the considered model.[5]

Beyond separability of the above-mentioned wide class of global nonlinear models, we also briefly considered the (memoryless) Heisenberg global model (1.19)–(1.21) in [20], as a spin-off of our general investigation. We showed that this specific model is ICA separable. Whereas our considerations about this Heisenberg

[4] Our configuration is also an extension of ANC in the sense that (i) it involves signals which initially have a quantum form and (ii) reference signals are not directly available as observations here, but only after the adequate fixed processing (1.16)–(1.17) of some observations, which yields the signals defined by (1.19)–(1.20). A reference-based model is thus obtained for the *global* model, not directly for the *mixing* model, unlike in ANC.

[5] For example only one sign indeterminacy for the Heisenberg global model, as detailed hereafter.

global model were limited to a proof of ICA separability in [20], we here aim at proceeding much further in the investigation of this BQSS problem, by deriving practical qubit separation methods based on this separability property. This is the topic of the remainder of this section.

1.5.3 Separation Criterion

Based upon the above results, we here investigate the case when all source signals of our BQSS problem are stochastic, mutually statistically independent and each of them is continuous valued and identically distributed (i.d). The observations and separating system outputs are then also stochastic, continuous-valued and i.d. We consider the RVs defined by all these signals at a single time, and we denote Y_i the RVs thus associated with the outputs of the separating system.

The above-defined ICA separability of the Heisenberg global model is a very attractive property, because it directly yields a means for adapting the parameter $\overline{v}$ of the separating system: this separability property ensures that, by adapting $\overline{v}$ so that the output RVs Y_i become statistically independent, it is guaranteed that they become equal to the source RVs (up to the indeterminacies of the Heisenberg global model). To derive a practical QSICA method from this property, we need a quantity which measures the degree of dependence of the output RVs. A well-known quantity which meets this condition is the mutual information (MI) of these RVs [12, 18, 46]: their MI is zero when they are independent and positive otherwise. The output MI of a separating system has already been used to derive nonquantum BSS methods for linear instantaneous [10, 30, 39, 41, 42], and convolutive [42] mixtures. It has then been applied to some nonlinear mixing and/or separating models: see, e.g., [1, 19, 24, 36, 48] (see also the general analysis in [49]). However, ICA separability was not proved for most of those nonlinear models, unlike in our present investigation.

So, a separation criterion for the Heisenberg global model consists in adapting $\overline{v}$ so as to minimize (and thus cancel) a function, therefore called "the cost function", defined as the MI of the output RVs of our separating system, and denoted $I(Y)$, where $Y = [Y_1, Y_2, Y_3]^T$. The above ICA separability property means that $I(Y)$ has no spurious global minimum points, i.e., that it reaches its global minimum value only when SS is achieved , up to the indeterminacies of this specific model. These indeterminacies are defined as follows. Equations (1.19)–(1.20) show that the first two output signals are always equal to the corresponding source signals. They therefore yield no indeterminacies at all. The indeterminacies for y_3 are then derived from the general analysis provided in [20]. The "interference term" of that investigation is here the first term in the argument of the Arcsin(.) function in (1.21). The general analysis provided in [20] proves that, when the output RVs are independent, this term remains constant when the "interfering sources" (here s_1 and s_2) vary, Eq. (1.21) shows that this here entails

$$\overline{v}^2 = v^2. \tag{1.22}$$

This yields two possible situations. If no prior knowledge about v is available, a QSICA method may provide either $\overline{v} = v$ or $\overline{v} = -v$ (up to estimation errors in practice). Due to (1.21) and (1.15), these two values of $\overline{v}$ respectively yield $y_3 = s_3$ and $y_3 = -s_3$. The third output of our separating system thus yields a sign indeterminacy (which may be avoided e.g., by further restricting s_3 to positive values). One may also face the situation when the sign of v is known, as detailed in [21] (briefly, since the sign of J_{xy} is known, (1.10)–(1.11) show that when $|\Delta_E| \leq \frac{\pi}{2}$ the sign of v is known). In that situation, when a QSICA method yields one of the two values $\overline{v} = \pm v$, one may then select its sign so as to have $\overline{v} = v$. Using the latter value in the separating system, one then gets $y_3 = s_3$, without any indeterminacy.

The cost function thus obtained may be expressed as

$$I(Y) = \left(\sum_{i=1}^{3} h(Y_i) \right) - h(Y). \tag{1.23}$$

In this expression, each term $h(Y_i)$ is the differential entropy of the RV Y_i, which may be expressed as

$$h(Y_i) = -E\{\ln f_{Y_i}(Y_i)\} \tag{1.24}$$

where $f_{Y_i}(.)$ is the probability density function (pdf) of Y_i and $E\{.\}$ stands for expectation. Similarly, $h(Y)$ is the joint differential entropy of all RVs Y_i, which reads

$$h(Y) = -E\{\ln f_Y(Y)\} \tag{1.25}$$

where $f_Y(.)$ is the joint pdf of all RVs Y_i.

Moreover, we here use the following general property. Let us consider an arbitrary N-dimensional random vector X, to which an arbitrary invertible transform ϕ is applied. We thus get the N-dimensional random vector Y defined as

$$Y = \phi(X). \tag{1.26}$$

This transform has the following effect on joint differential entropy [18]:

$$h(Y) = h(X) + E\{\ln |J_\phi(X)|\} \tag{1.27}$$

where $J_\phi(x)$ is the Jacobian of the transform $y = \phi(x)$, i.e., the determinant of the Jacobian matrix of ϕ. Each element with indices (i, j) of this matrix is equal to $\frac{\partial \phi_i(x)}{\partial x_j}$,where $\phi_i = y_i$ is the ith component of the vector function ϕ and x_j is its jth argument.

This property here applies to the output joint differential entropy defined in (1.25), and the transform ϕ here consists of the separating model defined by (1.13) and (1.16)–(1.18). Equations (1.16) and (1.17) show that y_1 and y_2 do not depend on x_2. Therefore, $J_\phi(x)$ here reduces to

$$J_\phi(x) = J_1 J_2 \tag{1.28}$$

where

$$J_1 = \frac{\partial y_3}{\partial x_2} \tag{1.29}$$

$$= -\operatorname{sgn}(\overline{v}) \left\{ 4y_1^2 y_2^2 (1-y_1^2)(1-y_2^2)(1-\overline{v}^2)\overline{v}^2 \right.$$

$$\left. -[y_1^2(1-y_2^2)(1-\overline{v}^2) + (1-y_1^2)y_2^2\overline{v}^2 - x_2]^2 \right\}^{-\frac{1}{2}} \tag{1.30}$$

$$J_2 = \frac{\partial y_1}{\partial x_3}\frac{\partial y_2}{\partial x_1} - \frac{\partial y_1}{\partial x_1}\frac{\partial y_2}{\partial x_3} \tag{1.31}$$

$$= \frac{1}{4y_1 y_2 \sqrt{(1+x_1-x_3)^2 - 4x_1}}. \tag{1.32}$$

Combining (1.23) and (1.27), the considered cost function becomes

$$I(Y) = \left(\sum_{i=1}^{3} h(Y_i) \right) - h(X) - E\{\ln |J_\phi(X)|\}. \tag{1.33}$$

Its term $h(X)$ does not depend on the separating system parameter $\overline{v}$ to be optimized, but only on the fixed available observations. Besides, (1.16) and (1.17) show that the outputs y_1 and y_2, and therefore the differential entropies $h(Y_1)$ and $h(Y_2)$ also do not depend on $\overline{v}$. Therefore, minimizing $I(Y)$ with respect to $\overline{v}$ is equivalent to minimizing the following cost function:

$$C_2(Y) = h(Y_3) - E\{\ln |J_\phi(X)|\}. \tag{1.34}$$

So, the separation *criterion* used in the proposed MI-based QSICA methods consists in looking for a value $\overline{v}_{MI}$ of $\overline{v}$ which yields the global minimum value of the cost function $I(Y)$ defined by (1.33), i.e.

$$\overline{v}_{MI} = \arg\min_{\overline{v}}(I(Y)), \tag{1.35}$$

or, equivalently, in looking for the minimum of $C_2(Y)$. Once this criterion has been fixed, various practical versions of this type of methods may be derived, depending on which *algorithm* is used to optimize these cost functions. We now describe several such optimization algorithms.

1.5.4 Gradient-Based Separation Algorithms

Various standard optimization algorithms, such as steepest descent/ascent (also called gradient descent/ascent) or Newton's method, make use of the gradient of the consid-

ered cost function, with respect to the set of parameters to be optimized. Therefore, we first determine the expression of the gradient of our cost functions, which here reduces to their derivative with respect to the single adaptive parameter $\overline{v}$. Based on the above comments about the independence of some differential entropies with respect to $\overline{v}$, Eqs. (1.33) and (1.34) first yield

$$\frac{\mathrm{d}I(Y)}{\mathrm{d}\overline{v}} = \frac{\mathrm{d}C_2(Y)}{\mathrm{d}\overline{v}} \tag{1.36}$$

$$= \frac{\mathrm{d}h(Y_3)}{\mathrm{d}\overline{v}} - \frac{\mathrm{d}E\{\ln|J_\phi(X)|\}}{\mathrm{d}\overline{v}}. \tag{1.37}$$

We then determine the two derivatives involved in (1.37).

1.5.4.1 Gradient of Differential Entropy

We here aim at determining the gradient of differential entropy, i.e., $\frac{\mathrm{d}h(Y_3)}{\mathrm{d}\overline{v}}$, which appears in (1.37). To this end, we apply a property which holds for a general nonlinear separating system. This property is established in [48], and a related proof using some more accurate notations is also provided in Sect. 14.10 of [18] for linear instantaneous mixtures. The general framework that we first consider involves a stochastic, continuous-valued, i.d. output signal $y_i(t)$ of a separating system, defined at time t as

$$y_i(t) = w_i(x(t)) \tag{1.38}$$

where $x(t)$ is a stochastic, continuous-valued, i.d. vector of observed values at time t and $w_i(.)$ is an arbitrary, memoryless, differentiable, possibly nonlinear function, which depends on a set of adaptive parameters of the considered separating system. Each of these parameters is denoted as c_j hereafter.[6] The RV Y_i defined by this output at any given time t may be expressed with respect to the random vector X, composed of the RVs X_j defined by the observations at that time, according to

$$Y_i = w_i(X). \tag{1.39}$$

Considering the differential entropy $h(Y_i)$, it may be shown that its derivative with respect to a parameter c_j reads

$$\frac{\mathrm{d}h(Y_i)}{\mathrm{d}c_j} = E\left\{\psi_{Y_i}(w_i(X))\frac{\mathrm{d}w_i(X)}{\mathrm{d}c_j}\right\} \tag{1.40}$$

$$= E\left\{\psi_{Y_i}(Y_i)\frac{\mathrm{d}Y_i}{\mathrm{d}c_j}\right\} \tag{1.41}$$

[6] The index i of these coefficients associated with $w_i(.)$ is omitted for readability, i.e., j is used as the overall single index of all coefficients of $w_i(.)$.

where $\psi_{Y_i}(.)$ is the score function of Y_i defined as

$$\psi_{Y_i}(u) = \frac{\partial[-\ln f_{Y_i}(u)]}{\partial u}. \tag{1.42}$$

We now apply the general result (1.41) to the term $\frac{\mathrm{d}h(Y_3)}{\mathrm{d}\overline{v}}$ of (1.37), i.e., to $c_j = \overline{v}$ and to the function $w_3(x) = y_3$ which is defined by (1.13) and (1.18). Equation (1.41) shows that we still just have to calculate the corresponding derivative $\frac{\mathrm{d}w_3(x)}{\mathrm{d}c_j} = \frac{\mathrm{d}y_3}{\mathrm{d}\overline{v}}$. This yields

$$\begin{aligned}\frac{\mathrm{d}y_3}{\mathrm{d}\overline{v}} = &\Big\{4y_1^2y_2^2(1-y_1^2)(1-y_2^2)(1-\overline{v}^2)\overline{v}^2 \\ &- [y_1^2(1-y_2^2)(1-\overline{v}^2) + (1-y_1^2)y_2^2\overline{v}^2 - x_2]^2\Big\}^{-\frac{1}{2}} \\ &\times \Big\{2(y_2^2-y_1^2)(1-\overline{v}^2)\overline{v}^2 - [y_1^2(1-y_2^2)(1-\overline{v}^2) + (1-y_1^2)y_2^2\overline{v}^2 - x_2](1-2\overline{v}^2)\Big\} \\ &\times [(1-\overline{v}^2)|\overline{v}|]^{-1}. \end{aligned} \tag{1.43}$$

1.5.4.2 Gradient Associated with Jacobian

Using (1.28)–(1.32) and the resulting property that J_2 does not depend on $\overline{v}$, some calculations yield

$$\frac{\mathrm{d}\ln|J_\phi(x)|}{\mathrm{d}\overline{v}} = \frac{1}{|J_1|}\frac{\mathrm{d}|J_1|}{\mathrm{d}\overline{v}} \tag{1.44}$$

$$\begin{aligned} &= -2\overline{v}\Big\{4y_1^2y_2^2(1-y_1^2)(1-y_2^2)(1-\overline{v}^2)\overline{v}^2 \\ &\quad -[y_1^2(1-y_2^2)(1-\overline{v}^2) + (1-y_1^2)y_2^2\overline{v}^2 - x_2]^2\Big\}^{-1} \\ &\quad \times \Big\{2y_1^2y_2^2(1-y_1^2)(1-y_2^2)(1-2\overline{v}^2) \\ &\quad -[y_1^2(1-y_2^2)(1-\overline{v}^2) + (1-y_1^2)y_2^2\overline{v}^2 - x_2](y_2^2-y_1^2)\Big\} \end{aligned} \tag{1.45}$$

for $\overline{v} \neq 0$.

1.5.4.3 Overall Gradient of Cost Functions

Gathering all above results, the gradient (1.37) becomes

$$\frac{\mathrm{d}I(Y)}{\mathrm{d}\overline{v}} = \frac{\mathrm{d}C_2(Y)}{\mathrm{d}\overline{v}} \tag{1.46}$$

$$= E\left\{\psi_{Y_3}(Y_3)\frac{\mathrm{d}Y_3}{\mathrm{d}\overline{v}}\right\} - E\left\{\frac{\mathrm{d}\ln|J_\phi(X)|}{\mathrm{d}\overline{v}}\right\}. \tag{1.47}$$

In this expression, ψ_{Y_3} is the score function of Y_3, which is estimated in practice, as in other BSS methods based on mutual information [24, 39, 48]. Besides, the explicit expression of $\frac{\mathrm{d}Y_3}{\mathrm{d}\overline{v}}$ in (1.47) is obtained by replacing all considered signals by the corresponding RVs in (1.43). By also replacing signals by RVs in (1.45) and taking the expectation of the resulting expression, one obtains the explicit form of the term $E\left\{\frac{\mathrm{d}\ln|J_\phi(X)|}{\mathrm{d}\overline{v}}\right\}$ of (1.47). The resulting version of (1.47) reads

$$\begin{aligned}
\frac{\mathrm{d}I(Y)}{\mathrm{d}\overline{v}} = E\Big\{&\psi_{Y_3}(Y_3)\\
&\times\Big(4Y_1^2Y_2^2(1-Y_1^2)(1-Y_2^2)(1-\overline{v}^2)\overline{v}^2\\
&\quad-[Y_1^2(1-Y_2^2)(1-\overline{v}^2)+(1-Y_1^2)Y_2^2\overline{v}^2-X_2]^2\Big)^{-\frac{1}{2}}\\
&\times\Big(2(Y_2^2-Y_1^2)(1-\overline{v}^2)\overline{v}^2\\
&\quad-[Y_1^2(1-Y_2^2)(1-\overline{v}^2)+(1-Y_1^2)Y_2^2\overline{v}^2-X_2](1-2\overline{v}^2)\Big)\Big\}\\
&\times[(1-\overline{v}^2)|\overline{v}|]^{-1}\\
&+2\overline{v}E\Big\{\Big(4Y_1^2Y_2^2(1-Y_1^2)(1-Y_2^2)(1-\overline{v}^2)\overline{v}^2\\
&\quad-[Y_1^2(1-Y_2^2)(1-\overline{v}^2)+(1-Y_1^2)Y_2^2\overline{v}^2-X_2]^2\Big)^{-1}\\
&\quad\times\Big(2Y_1^2Y_2^2(1-Y_1^2)(1-Y_2^2)(1-2\overline{v}^2)\\
&\quad-[Y_1^2(1-Y_2^2)(1-\overline{v}^2)+(1-Y_1^2)Y_2^2\overline{v}^2-X_2](Y_2^2-Y_1^2)\Big)\Big\}.
\end{aligned} \tag{1.48}$$

1.5.4.4 Gradient Descent Algorithm

The simplest gradient-based algorithm for minimizing the above cost functions is the gradient descent algorithm, which initializes $\overline{v}$ (typically with a random value) and then iteratively updates it according to the rule

$$\overline{v}(n+1) = \overline{v}(n) - \mu\frac{\mathrm{d}I(Y)}{\mathrm{d}\overline{v}}\bigg|_{\overline{v}=\overline{v}(n)} \tag{1.49}$$

where n is the iteration index and μ is a positive (fixed or varying) adaptation gain, which controls convergence speed and accuracy. Although this algorithm is based on a simple principle, its attractiveness eventually turns out to be limited in the QSICA

problem tackled in this chapter, due to the complexity of the gradient (1.48) that we derived above. Moreover, a general drawback of the gradient descent algorithm is that it may converge toward a *local* minimum of the considered cost function. These limitations motivate us to develop another algorithm, described hereafter.

1.5.5 A Gradientless Separation Algorithm

We here still consider the separation criterion that we derived in Sect. 1.5.3, which consists in looking for a value of $\overline{v}$ which yields the global minimum value of the cost function $I(Y)$. We now take advantage of the fact that this function only depends on a *single* tunable parameter of our separating system, namely $\overline{v}$, and that this parameter is to be varied in a *bounded* interval, namely $[-1, 1]$, due to (1.11). Therefore, a straightforward and relatively cheap algorithm for reaching the *global* minimum value of $I(Y)$ is here a sweep-based approach: this consists of increasing $\overline{v}$ with a small discrete step over $[-1, 1]$, in computing the values (estimates in practice) of $I(Y)$ which correspond to each tested value of $\overline{v}$, and in keeping a value of $\overline{v}$ which minimizes $I(Y)$. Due to (1.22), one thus gets either $\overline{v} = v$ or $\overline{v} = -v$ (up to estimation errors). Moreover, in the above-defined situation when the sign of v is known, one can then reassign the sign of $\overline{v}$ in order to obtain $\overline{v} = v$, or one should preferably directly perform the sweep of $\overline{v}$ only over the adequate reduced interval $[-1, 0]$ or $[0, 1]$. *This sweep-based BQSS method using mutual information is called BQSS-MI in the tests reported further in this chapter.*

It should be noted that such sweep-based approaches are also used in other types of multisource problems than BSS, such as array processing. In particular, they have been widely applied to the well-known MUSIC algorithm, where the single parameter varied in the sweep is the tested direction of arrival of a propagating wave [4, 18, 43, 44, 54].

1.6 QSICA-Oriented Methods Based on Likelihood

1.6.1 Separation Criterion

Another approach to the BSS problem which has been used in the classical framework is based on the Maximum Likelihood (ML) principle , which is a very general estimation technique, e.g. described in Chap. 4 of [33]. This approach has first been applied to linear mixtures [7, 27–30, 40, 41] and then extended to nonlinear ones [9, 19, 31, 57]. It is closely related to ICA in the sense that it assumes statistically independent *source* signals, although its separation criterion is not initially based

on enforcing the independence of the *outputs* of the separating system.[7] Whereas this ML principle has only been applied to BSS in the *classical* framework in the literature, we here show how to extend it to the BQSS problem, so as to estimate the parameter v of our mixing model. This estimated value is then used as the parameter $\overline{v}$ of the separating system of Sect. 1.4, in order to restore the source signals.

The ML procedure detailed in [7, 18][8] applies as follows to our BQSS problem. We first express the mixing model (1.6)–(1.8) in compact form as $x = F(s, v)$, where the nonlinear function $F(.,.)$ has three components $F_1(.,.)$ to $F_3(.,.)$, with $x_i = F_i(s, v)$, $\forall i \in \{1, \ldots, 3\}$, and these $F_i(.,.)$ are respectively defined by (1.6)–(1.8). Starting from the above model of the actual data, we then build the associated model of the ML approach: still considering the functional form $F(.,.)$, we introduce the variables $\tilde{s} = [\tilde{r_1}, \tilde{r_2}, \tilde{\Delta}_I]^{\mathrm{T}}$, $\tilde{v}$ and $\tilde{x} = [\tilde{p_1}, \tilde{p_2}, \tilde{p_4}]^{\mathrm{T}}$, which are respectively associated with s, v, and x, and which are therefore such that

$$\tilde{x} = F(\tilde{s}, \tilde{v}). \tag{1.50}$$

In the statistical ML approach, we consider the random vectors $\tilde{S}$ and $\tilde{X}$, respectively defined by $\tilde{s}$ and $\tilde{x}$ at a single time, which are such that

$$\tilde{X} = F(\tilde{S}, \tilde{v}). \tag{1.51}$$

Using the standard ML approach, we study the case when $\tilde{S}$ has the same joint pdf as the actual source random vector S, i.e., $f_S(.)$ (comments about the case when they are different are available in [7, 11]). The joint pdf of $\tilde{X}$ is denoted as $f_{\tilde{X}}(.)$. It depends on $f_S(.)$, which is fixed but possibly only partly known, and on $\tilde{v}$, which is varied as explained below. The resulting family of pdf $f_{\tilde{X}}(.)$ is used as a parametric model of the pdf $f_X(.)$ of the actual observations. Due to Sect. 1.4, the function $F(., \tilde{v})$ from $\tilde{s}$ to $\tilde{x}$ is invertible, for a given value of $\tilde{v}$. Equation (1.51) then yields

$$f_{\tilde{X}}(\tilde{x}) = \frac{f_S(\tilde{s})}{|J_F(\tilde{s}, \tilde{v})|} \tag{1.52}$$

where $J_F(\tilde{s}, \tilde{v})$ is the Jacobian of the function $F(., \tilde{v})$, defined in the same way as in Sect. 1.5.3. For the function $F(., \tilde{v})$ considered in this chapter, our calculations show that

$$J_F(\tilde{s}, \tilde{v}) = 8\tilde{r_1}^2\tilde{r_2}^2(\tilde{r_1}^2 - \tilde{r_2}^2)\sqrt{1 - \tilde{r_1}^2}\sqrt{1 - \tilde{r_2}^2}\sqrt{1 - \tilde{v}^2}\tilde{v}\cos\tilde{\Delta}_I. \tag{1.53}$$

[7] One may also choose to define the concept of ICA for BSS in a broader sense, i.e., as the estimation of statistically independent source signals from their mixtures, using any suitable approach. The ML-based approach then completely belongs to ICA.

[8] We here aim at avoiding any ambiguity between the actual "fixed data" of the considered problem and the corresponding variables introduced in the ML approach. We therefore use different notations for these corresponding quantities, e.g., v for the fixed (unknown) mixing parameter and $\tilde{v}$ for the corresponding variable of the ML approach. In the framework of BSS, this type of notations was especially introduced in [7].

Taking the logarithm of (1.52), and studying the case when the sources are mutually statistically independent, we obtain

$$\ln f_{\tilde{X}}(\tilde{x}) = \sum_{i=1}^{3} \ln f_{S_i}(\tilde{s}_i) - \ln |J_F(\tilde{s}, \tilde{v})|. \tag{1.54}$$

We then consider the overall set of observed values, which consists of M samples $x(m)$ of the observation vector, for integer values of the time index m ranging from 1 to M. Using the extension of $f_{\tilde{X}}(.)$ to all these times m, the likelihood that observed values are drawn with a particular pdf $f_{\tilde{X}}(.)$ in the considered family is defined as

$$L = f_{\tilde{X}}(x_1(1), x_2(1), x_3(1), \ldots, x_1(M), x_2(M), x_3(M)). \tag{1.55}$$

Using the standard ML approach, we focus on the case when each random signal involved in the considered memoryless models is independent and identically distributed (i.i.d). We then have

$$L = \prod_{m=1}^{M} f_{\tilde{X}}(x_1(m), x_2(m), x_3(m)). \tag{1.56}$$

Defining the (normalized) log-likelihood as

$$\mathscr{L} = \frac{1}{M} \ln L, \tag{1.57}$$

Eq. (1.56) yields

$$\mathscr{L} = \frac{1}{M} \sum_{m=1}^{M} \ln f_{\tilde{X}}(x_1(m), x_2(m), x_3(m)) \tag{1.58}$$

$$= E_t[\ln f_{\tilde{X}}(x_1(t), x_2(t), x_3(t))], \tag{1.59}$$

where $E_t[.]$ is the temporal averaging operator. Equation (1.54) then yields

$$\mathscr{L} = \sum_{i=1}^{3} E_t[\ln f_{S_i}(\tilde{s}_i(t))] - E_t[\ln |J_F(\tilde{s}(t), \tilde{v})|] \tag{1.60}$$

where $\tilde{s}(t)$ is the inverse image of the vector $x(t)$ of observed values at time t for the mapping (1.50), and $\tilde{s}_i(t)$ are the components of $\tilde{s}(t)$. The mapping (1.50) and its Jacobian depend on $\tilde{v}$. The log-likelihood $\mathscr{L}$ therefore depends on $\tilde{v}$. The ML estimator of the mixing parameter v is eventually defined as the value (or one of the values) of $\tilde{v}$ which yields the global maximum value of the likelihood L or, equivalently, of the log-likelihood $\mathscr{L}$, i.e:

$$\hat{v}_{ML} = \arg\max_{\tilde{v}}(\mathscr{L}) \tag{1.61}$$

with $\mathscr{L}$ defined by (1.60). The maximum likelihood approach thus consists in selecting the value of $\tilde{v}$ which maximizes the pdf (1.55) associated with the observations, i.e., which makes the obtained measurements most likely, hence the name of this approach.

1.6.2 Algorithms and Connection with Mutual Information

Various optimization algorithms may then be derived from the above ML separation criterion (1.61), e.g., using the gradient of the cost function $\mathscr{L}$ with respect to $\tilde{v}$. These algorithms may be developed independently from those that we proposed above for the mutual information minimization criterion, but by using similar principles. Moreover, such calculations may be avoided thanks to the explicit connection between the mutual information and log-likelihood cost functions (1.33) and (1.60) that we will now derive. To this end, their respective tunable parameters (both related to the mixing parameter v) are here set to the same value, i.e., $\overline{v} = \tilde{v}$. The tunable function achieved by the separating system,[9] here denoted $\phi(.,\overline{v})$ for the sake of clarity, is then equal to the inverse of the tunable function $F(.,\tilde{v})$ which corresponds to the mixing model. Therefore, when $x = F(\tilde{s},\tilde{v})$, we have

$$y = \phi(x,\overline{v}) = \tilde{s} \tag{1.62}$$

$$J_{\phi}(x,\overline{v}) = J_F(\tilde{s},\tilde{v})^{-1}. \tag{1.63}$$

Besides, in a blind configuration, the pdf of the source signals involved in (1.60) are unknown. In practice, they are therefore replaced by (estimates of) the pdf $f_{Y_i}(.)$ of the outputs of the separating system, since the latter signals provide estimates of the source signals. Also using (1.62)–(1.63), the log-likelihood (1.60) is thus replaced by

$$\mathscr{L}_2 = \sum_{i=1}^{3} E_t[\ln f_{Y_i}(y_i(t))] + E_t[\ln|J_{\phi}(x(t),\overline{v})|]. \tag{1.64}$$

This quantity is a relevant estimate of

$$\mathscr{L}_3 = \sum_{i=1}^{3} E\{\ln f_{Y_i}(Y_i)\} + E\{\ln|J_{\phi}(X,\overline{v})|\} \tag{1.65}$$

because $\mathscr{L}_2$ converges toward $\mathscr{L}_3$ for ergodic signals, when the number of samples involved in the temporal averaging operator $E_t[.]$ tends to infinity. Moreover, comparing (1.65) with (1.33) and (1.24) shows that

[9] See function ϕ defined on p. 13.

$$\mathscr{L}_3 = -I(Y) - h(X). \tag{1.66}$$

Since $h(X)$ does not depend on $\overline{v}$, this shows that optimizing $\overline{v}$ so as to maximize $\mathscr{L}_3$ is fully equivalent to optimizing it so as to minimize $I(Y)$. The ML and MI approaches to this BQSS problem are therefore closely connected. This result, which is well known for linear instantaneous mixtures [8, 30], is thus here extended to the nonlinear mixing model considered in this chapter (and beyond it, since the above calculations straightforwardly extend to quite general nonlinear mixing and separating functions $F(.,.)$ and $\phi(.,.)$). Since the ML approach is thus merged with the MI approach, only the latter approach is considered in the tests reported below.

1.7 QSICA-Oriented Methods Focused on Moments or Cumulants

1.7.1 Methods Focused on Higher Order Statistics

The exact QSICA cost function (1.34) involves the differential entropy, and therefore the pdf, of a separating system output. It therefore requires one to estimate this pdf (or its derivative, used in some optimization algorithms), which is cumbersome. An alternative approach consists in deriving an approximation of this pdf, which yields an associated approximate QSICA criterion. A standard method for defining an approximation of a pdf, and then of the associated differential entropy or shifted[10] negentropy, consists in using the Edgeworth expansion, which is, e.g., detailed in [35]. This approach may be summarized as follows. The considered pdf of an RV U is expressed as the product of a reference pdf, here selected as a Gaussian RV G with the same mean and variance as U, and of a factor expressed as a series (see its explicit expression e.g. in [35]). This then makes it possible to express the shifted negentropy of U, i.e.,

$$\mathscr{J}(U) = h(G) - h(U), \tag{1.67}$$

as a series. Then truncating that series to a given order provides a corresponding approximation of that shifted negentropy, expressed in terms of the higher order cumulants of the considered RV.

[10] The expression "negentropy" is often used by the signal processing community for the quantity $\mathscr{J}(U)$ defined in (1.67). We call this quantity "*shifted* negentropy" because, whereas negentropy literally means "negative of entropy" [5], the shifted negentropy $\mathscr{J}(U)$ of U has the property of *never being* negative [33]. The expression "shifted negentropy" is quite compatible with two other uses of the word negentropy. The first one occurs in the context of living organisms, since Schrödinger first spoke of "negative entropy" in [45], in order to describe the ability of *living organisms* to fight against the tendency to disorder. The second appeared in the field of information theory, when Brillouin explicitly introduced the word negentropy, in [5], when establishing a link between *information processing and the behavior of the physical systems* making this processing.

That approach was used and detailed by Comon in [10], for linear instantaneous mixtures. It is less attractive for our nonlinear mixing model, for several reasons. First, once the resulting approximate cost function has been derived, one should check whether it has spurious global minimum points (as defined in Sect. 1.5.3). This could be done for linear instantaneous mixtures, where the eventual cost function only consists of cumulants. In our case, (1.34) shows that our cost function also involves the complicated Jacobian of the nonlinear transform achieved by the separating system. It is therefore not guaranteed that one can show whether this cost function is free from spurious global minimum points. Besides, such an approach was required for linear instantaneous mixtures for deriving a cost function which is simpler than mutual information (it is based on higher order cumulants) since that BSS problem cannot be solved by only using first-order and second-order statistics. On the contrary, this Edgeworth expansion can be avoided here, because even simpler cost functions can be derived by other means for our specific BQSS problem (i.e., for our mixing model). So, we will now present these methods, focused on first-order and/or second-order statistics of specific signals derived from our configuration.

1.7.2 Methods Focused on First-Order Statistics

A very simple class of statistical methods for estimating the mixing parameter v may be derived by considering the first-order moment of p_2, i.e., by taking the expectation of (1.7). We will not provide all its details here, because this yields the (only) class of blind methods that we described in our initial journal paper [21] dealing with QSS. We summarize its principle, however, since we will use its basic version in the tests reported below. This approach sets a constraint on the source statistics, i.e., it requires the value of $E\{\sin \Delta_I\}$ to be known. Moreover, we here only consider its version intended for the case when $E\{\sin \Delta_I\} = 0$, since this simplifies the approach. Besides, all source signals are assumed to be mutually statistically independent, which again connects this approach with QSICA, as was done above for our ML-based approach. One may then easily check that taking the expectation of (1.7) yields a linear equation with respect to v^2. The solutions of this equation (not detailed in [21]) read

$$v = \pm \sqrt{\frac{E\{p_2\} - E\{r_1^2\}(1 - E\{r_2^2\})}{E\{r_2^2\} - E\{r_1^2\}}} \tag{1.68}$$

and the sign of v is known a priori in some configurations, as explained above (for the case when it is unknown, see [21]). All parameters in (1.68) are known,[11] since they are statistics of the observation p_2 and of the sources r_1 and r_2, which are derived by the first two separating system outputs without knowing the value of v (see (1.16)–(1.17) and (1.19)–(1.20)). Using the corresponding sample statistics, (1.68) yields

[11] In practice, they are estimated from a sequence of i.d. (therefore possibly i.i.d) source values.

an estimate of v, which is denoted as $\hat{v}_1$ since it is initially based on the *first-order* moment of observation p_2. Note that it eventually also uses second-order moments of other signals, due to the nonlinear nature of mixing Eq. (1.7). This estimate $\hat{v}_1$ is then used as the value of the parameter $\overline{v}$ of the separating system, in order to estimate all sources.

This BQSS method focused on a first-order moment is called BQSS-m1 in the tests reported below. Its drawback is its constraint on the source statistics $E\{\sin \Delta_I\}$. This motivates us to introduce a new approach hereafter, to avoid this limitation.

1.7.3 Methods Focused on Second-Order Statistics

1.7.3.1 Separation Criterion

The methods that we proposed in Sect. 1.5 are based on ICA separability: they use the mutual information of the output signals y_i of our separating system to measure their dependence, and they therefore take into account the whole pdf. We will now show that our quantum mixing model is also second-order separable, in the sense that its source signals may be estimated by using a criterion only based on the covariance of transformed versions of the above-defined signals y_i. One of these signals, denoted z_3, is derived from the observations and from $\overline{v}$. It is defined as the argument of the Arcsin(.) function which yields y_3 in (1.18), i.e.,

$$z_3 = \frac{y_1^2(1-y_2^2)(1-\overline{v}^2) + (1-y_1^2)y_2^2\overline{v}^2 - p_2}{2y_1y_2\sqrt{1-y_1^2}\sqrt{1-y_2^2}\sqrt{1-\overline{v}^2\overline{v}}}. \tag{1.69}$$

In other words, we have $z_3 = \sin y_3$. Thanks to (1.21), z_3 may also be expressed with respect to the source signals as

$$z_3 = \frac{(\overline{v}^2 - v^2)(s_2^2 - s_1^2)}{2s_1s_2\sqrt{1-s_1^2}\sqrt{1-s_2^2}\sqrt{1-\overline{v}^2\overline{v}}} + \frac{\sqrt{1-v^2}v}{\sqrt{1-\overline{v}^2\overline{v}}}\sin s_3. \tag{1.70}$$

The first term of (1.70) contains a factor which only depends on the source signals s_1 and s_2, i.e.

$$z_{12} = \frac{s_2^2 - s_1^2}{2s_1s_2\sqrt{1-s_1^2}\sqrt{1-s_2^2}}. \tag{1.71}$$

Equations (1.19)–(1.20) then yield

$$z_{12} = \frac{y_2^2 - y_1^2}{2y_1y_2\sqrt{1-y_1^2}\sqrt{1-y_2^2}}. \tag{1.72}$$

Therefore, by using (1.16)–(1.17) and then (1.72), the signal z_{12} may indeed be derived from the observations without having yet determined the adequate value of $\overline{v}$. This signal z_{12} may therefore be used to subsequently select the value of $\overline{v}$. To this end, let us consider the covariance of the RVs Z_{12} and Z_3 defined by z_{12} and z_3 for a single random value of the initial states of the two qubits. This covariance is denoted as $\mathrm{cov}(Z_{12}, Z_3)$, whereas var(.) stands for variance. Again assuming all source signals to be mutually statistically independent, (1.70)–(1.71) can be shown to yield

$$\mathrm{cov}(Z_{12}, Z_3) = \mathrm{var}(Z_{12}) \frac{\overline{v}^2 - v^2}{\sqrt{1 - \overline{v}^2 \overline{v}}}. \tag{1.73}$$

Considering the nondegenerate case when $\mathrm{var}(Z_{12}) \neq 0$, Eq. (1.73) shows that

$$\mathrm{cov}(Z_{12}, Z_3) = 0 \Leftrightarrow \overline{v}^2 = v^2 \Leftrightarrow \overline{v} = \varepsilon v \tag{1.74}$$

with $\varepsilon = \pm 1$. When condition (1.74) is met, (1.15) and (1.21) yield $y_3 = \varepsilon s_3$. Moreover, the sign indeterminacy on $\overline{v}$ and therefore on y_3 may be avoided by using physical knowledge about the sign of v, as explained in Sect. 1.5.3.

The analysis presented so far shows that a second-order criterion which guarantees separation (i.e., without spurious points) for our model consists in selecting $\overline{v}$ so as to cancel $\mathrm{cov}(Z_{12}, Z_3)$. We now proceed to practical algorithms that may be used to blindly tune $\overline{v}$ to such a value.

1.7.3.2 Separation Algorithms

Here again, iterative methods for updating $\overline{v}$ might be developed. However, a better approach can here be derived by considering the expression of $\mathrm{cov}(Z_{12}, Z_3)$ only with respect to known quantities, i.e., $\overline{v}$ and observations and/or outputs of the separating system. Using (1.69) and (1.72), it can then be shown that

$$\mathrm{cov}(Z_{12}, Z_3) = \frac{\mathrm{var}(Z_4)\overline{v}^2 + \mathrm{cov}(Z_4, Z_5)}{4\sqrt{1 - \overline{v}^2 \overline{v}}} \tag{1.75}$$

where Z_4 and Z_5 are the RVs associated, for a single random value of the initial states of the two qubits, with the signals z_4 and z_5 defined as

$$z_4 = 2z_{12} = \frac{y_2^2 - y_1^2}{y_1 y_2 \sqrt{1 - y_1^2}\sqrt{1 - y_2^2}} \tag{1.76}$$

$$z_5 = \frac{y_1^2(1 - y_2^2) - p_2}{y_1 y_2 \sqrt{1 - y_1^2}\sqrt{1 - y_2^2}}. \tag{1.77}$$

Equation (1.75) shows that condition (1.74) is also equivalent to setting $\overline{v}$ to a value $\overline{v}_2$, whose squared value is defined by

$$\overline{v}_2^2 = -\frac{\mathrm{cov}(Z_4, Z_5)}{\mathrm{var}(Z_4)}. \tag{1.78}$$

This value is denoted as $\overline{v}_2$ because it corresponds to our method based on *second-order* moments of the signals z_{12} to z_5. *This method is therefore called BQSS-m2.* Note, however, that these signals are nonlinear functions of the signals y_i (they also involve p_2) and this method is therefore based on a few *generalized* moments of the signals y_i.

Equation (1.78) provides a closed-form solution for selecting $\overline{v}^2$. A practical estimate of this expression may then again be derived from an i.d. sequence of signal values, and the sign indeterminacy is handled as explained above. At this stage, this second-order approach is more attractive than the other main solutions proposed in this chapter, because: (i) as compared to the MI-based approach, it uses simpler statistics and yields a closed-form solution, (ii) as compared to the first-order method, it does not require some statistics of source $s_3 = \Delta_I$ to be known. However, the comparison of all three methods should also take into account the numerical performance of their practical implementations. This is the topic of the next section. Before proceeding to that numerical part of the chapter, the interested reader may refer to the Appendix for some comments about the relationship between the BQSS methods introduced in this section and classical BSS and ANC methods which are also based on cumulants, moments or generalized moments.

1.8 Numerical Results

As stated above, the physical implementation of qubits is only an emerging topic, which is beyond the scope of this chapter. Therefore, we here assess the performance of the above methods by means of tests performed with data derived from our software simulation of the behavior of coupled qubits. This software was described in [21], where the only blind method tested with it was BQSS-m1. It performs each elementary test as follows. It creates a set of observed vectors x corresponding to known source vectors s, mixed according to the considered Heisenberg model with a given value of the mixing parameter v. During the "mixture estimation stage", it first uses part of the above observed vectors (100 or 1,000 of them in our tests) to blindly estimate v or the associated separating parameter $\overline{v}$. During the "source estimation stage", this software is then used to process 100 other observed vectors with the above estimate of the separating parameter, and it thus derives estimates y of the actual sources s from which these observed vectors x were computed. As detailed below, analyzing these estimated sources and comparing them to the actual sources then makes it possible to check that the proposed methods succeed in separating

these sources, and to determine the accuracy of this separation, i.e. the magnitude of the deviation of the estimated sources y with respect to the actual sources[12] s.

In the mixture and source estimation stages of each elementary test, the qubit parameter values r_1, r_2 and Δ_I are randomly selected within the 20–80 % subrange of their 0–100 % allowed range defined by (1.14) and (1.15), with a uniform distribution. These data are thus such that $E\{\sin \Delta_I\} = 0$. This is required in these tests, because we want to include in our comparison the version of our BQSS-m1 method which requires $E\{\sin \Delta_I\} = 0$, in addition to our new BQSS-MI and BQSS-m2 methods. In our tests using the BQSS-MI method, the estimation of output mutual information is performed with the approach of [13] as implemented at [50]. The estimators used in our BQSS-m1 and BQSS-m2 methods are respectively obtained by replacing expectation by sample mean in (1.68) and in the square root of (1.78). As stated above, for each choice of the initial states of the qubits, a set of measurements is used to derive the corresponding observed vector, based on the relative frequencies of occurence of measured values. The number of measurements performed per observed vector is K_m in the mixture estimation stage and K_s in the source estimation stage. These two independent parameters are varied in the tests reported below. The parameter v is set to 0.5 in all these tests, and the sign of v is assumed to be known by the considered BQSS methods.[13]

For each considered set of conditions and each BQSS method, we perform 100 above-defined elementary tests with different sets of source values (so that we perform 100 estimations of the same mixing parameter value).[14] For some of these tests, we may get complex values for the estimated sources during the source estimation stage, for the following reason. As noted above, the observations used to derive the outputs of the separating system are only estimates of the actual probabilities p_1 to p_4. These estimates are more or less accurate, depending on the number K_s of measurements performed to estimate the above probabilities. Inserting these estimates in (the complex-valued extension of) the Arcsin(.) function in (1.18) may yield complex-valued estimated sources. The cases when any output of the separating system is complex-valued may be detected in practice and correspond to non-satisfactory situations. Therefore, the first performance criterion used hereafter is the "success rate" of each method, i.e., the number of elementary tests (among all 100 elementary tests for each considered set of conditions) for which all separating system outputs obtained during the source estimation stage are real-valued. These success rates are shown in Figs. 1.1 and 1.2.

---

[12] This performance assessment procedure can only be used when *developing* and testing the considered BQSS methods, with actual source values s which are *known* (but which are not used in the BQSS methods themselves). On the contrary, in the actual setup which is to be eventually used, the actual sources are *unknown*, and one precisely aims at estimating them ! They cannot therefore be compared to their estimated values.

[13] The above conditions for each elementary test are the same as in [21].

[14] We therefore here perform more exhaustive tests than in [21], where only one elementary test was performed for each set of conditions (and we avoided the complex-valued outputs mentioned below).

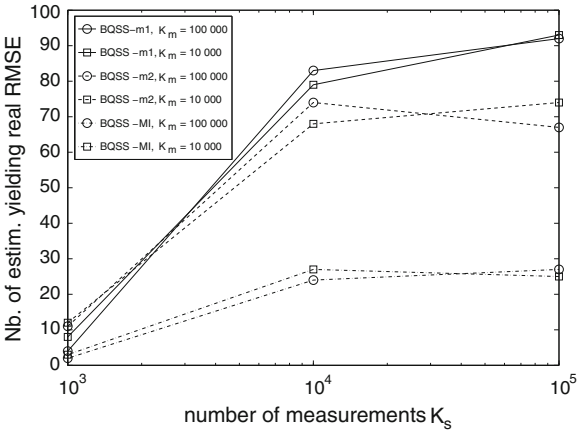


Fig. 1.1 Success rate (in %), i.e., number of elementary tests (among 100 tests per BQSS method) which yield real-valued estimated sources. Tested methods, from high success rate (*top curves*) to low success rates (*bottom curves*): restore sources using a value of $\overline{v}$ blindly estimated from 100 observed vectors by means of BQSS-m1 (*solid lines*), BQSS-m2 (*dashed lines*) or BQSS-MI (*dash-dotted lines*) methods. The success rate is plotted vs number of measurements K_s used in the source estimation stage. Each plot corresponds to a specific number of measurements K_m used in the mixture estimation stage

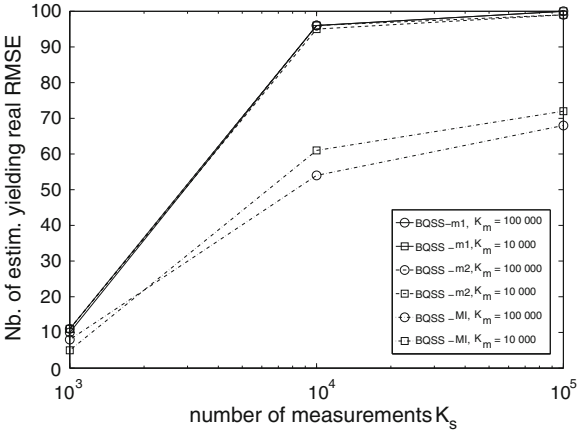


Fig. 1.2 Same as Fig. 1.1, but estimating $\overline{v}$ with 1,000 observed vectors

Only the elementary tests which yield real-valued outputs are then considered and the second performance criterion computed over them is the root-mean square error (RMSE)[15] achieved when estimating the third source, Δ_I. The first two sources, r_1

[15] We here use the standard definition of the RMSE, which was detailed in [21] for an elementary test, and which is straightforwardly extended to the set of elementary tests which yield real-valued separating system outputs.

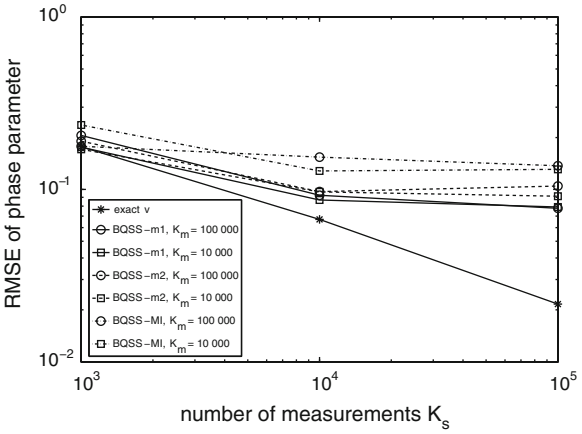


Fig. 1.3 RMSE of source Δ_1. Tested methods, from low RMSE (*bottom curves*) to high RMSE (*top curves*): restore sources with $\overline{v}$ set to the exact value of v (*solid line and stars*) or with a value of $\overline{v}$ blindly estimated from 100 observed vectors by means of BQSS-m1 (*solid lines and circles or squares*), BQSS-m2 (*dashed lines*) or BQSS-MI (*dash-dotted lines*) methods. RMSE is plotted versus number of measurements K_s used in the source estimation stage. Each plot for the BQSS methods corresponds to a specific number of measurements K_m used in the mixture estimation stage

and r_2, are not considered hereafter because their estimates do not depend on the estimated value of the mixing or separating parameter, as shown by (1.16)–(1.17), so that the corresponding RMSE is the same for all three tested methods, and this RMSE was already provided for BQSS-m1 in [21] (for an elementary test). The results obtained for the RMSE of Δ_1 are shown in Figs. 1.3 and 1.4. These RMSE should be analyzed by also taking into account the corresponding success rates in Fig. 1.1 and 1.2: one may get similar RMSE for different BQSS methods because, when their success rates are lower than 100 %, these RMSE are only computed over subsets of tests, i.e., for the "good situations" (real-valued output signals) which have not been filtered out when computing success rates. In such cases, one should first compare success rates: a method with (almost) 100 % success rate is sought, since this means that it is able to extract the source signals from all considered observed values, with an accuracy which is then investigated by checking its RMSE.

All above-defined figures yield the following comments. Figs. 1.1 and 1.2 show that when increasing the number K_s of measurements per observed vector in the source estimation stage, better performance is generally obtained, which was expected because the observed values (i.e., frequency-based estimates of probabilities) thus get closer to their theoretical values. The number K_m of measurements per observed vector in the mixture estimation stage has a limited influence over its considered range.[16] The most standard BQSS method, i.e., the one based on output

[16] Low values of K_m are not considered here, because higher numbers of measurements are more easily accepted in the single initial chararacterization of the system (mixture estimation stage) than

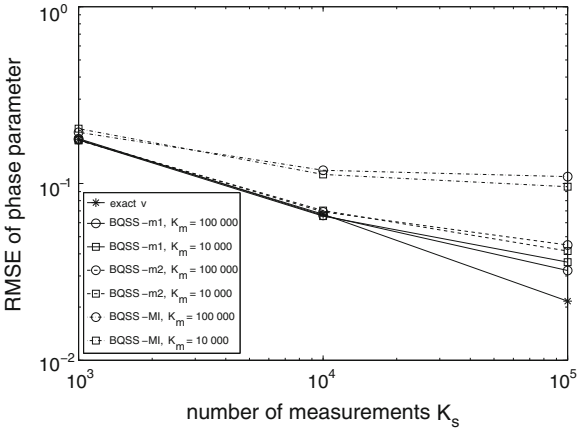


Fig. 1.4 Same as Fig. 1.3, but estimating $\overline{v}$ with 1,000 observed vectors in BQSS methods

mutual information as in classical BSS, turns out to yield rather low success rates, i.e., low performance, even for the highest considered values of the numbers K_m and K_s of measurements. This may especially be due to the fact that this information-based method requires a large number of observed vectors in order to accurately estimate output signal statistics, whereas we wish to restrict ourselves to a limited number of such observed vectors (100 or 1,000), in order to limit the amount of data that must be measured to apply our methods (remember that each observed vector here requires $K_m = 10^4$ or 10^5 measurements in the mixture estimation stage). This low performance would be an issue if we had to restrict ourselves to this information-based BSS criterion.

Fortunately, for the considered mixing model, we could develop simpler BQSS methods focused on moments, which yield much higher performance than the BQSS-MI approach when applied to the limited numbers of observed vectors that we used in the mixture estimation stage. In particular, when estimating the mixing parameter with 1,000 observed vectors and using $K_m = 10^5$ and $K_s = 10^5$, the success rate tends to 100 % and the RMSE gets close to 3×10^{-2} for these BQSS-m1 and BQSS-m2 methods. Allowing larger data sets would yield even better performance. As a bound, Figs. 1.3 and 1.4 also show the RMSE achieved in the source estimation stage with K_s measurements, when $\overline{v}$ is set to the exact value of the mixing parameter v.

Finally, the performance of the BQSS-m1 method compares as follows with that of BQSS-m2. When using a rather low number (100) of observed vectors in the mixture estimation stage, BQSS-m1 is attractive because it provides higher performance, but it should be remembered that (this version of) this method has a drawback: it is only

(Footnote 16 continued)
in its subsequent permanent use (source estimation stage), and because these higher numbers of measurements are preferred, in order to better estimate the mixing parameter and thus to achieve better performance.

applicable to sources which are such that $E\{\sin \Delta_I\} = 0$. For a medium number (1,000) of observed vectors in the mixture estimation stage, the performance of BQSS-m2 is nearly as high as that of BQSS-m1, and BQSS-m2 has the advantage of avoiding the constraint $E\{\sin \Delta_I\} = 0$. Depending on the considered situation, the preferred approach is therefore BQSS-m1 or BQSS-m2. However, the situations involving a medium number of observed vectors are of higher importance (in order to achieve better success rates), and the most prominent method is therefore BQSS-m2.

1.9 Conclusions and Future Work

In our initial paper [21], we introduced the concept of Quantum Source Separation (QSS) and we mainly investigated its non-blind version. In this chapter, we proceeded much further, by considering the blind configuration, with quantum-to-classical data conversion which yields an original nonlinear mixing model. We introduced the field of Quantum-Source Independent Component Analysis (QSICA) and we developed QSS methods based on this QSICA concept by using various statistical signal processing tools, namely mutual information, likelihood and moments. Besides, we showed that cumulant-based methods derived from Edgeworth expansion are not suited to the mixing model considered in this chapter. The performance of the proposed approaches was validated by means of numerical tests. This especially showed the attractiveness of our BQSS-m2 method, which is focused on the use of second-order moments, and which yields a closed-form solution and good numerical performance for a limited amount of measured data, without setting restrictions on the marginal statistics of the source signals. This investigation completes a major step in the emerging field of QSS, but we foresee other major developments of this field, e.g., considering other classes of QSS methods or other quantum coupling models. We will report such developments in future papers.

Appendix

In Sects. 1.5 and 1.6, we developed BQSS methods by explicitly starting from principles which were previously used in classical BSS, namely mutual information minimization and likelihood maximization. On the contrary, in Sects. 1.7.2 and 1.7.3, we derived BQSS methods focused on moments without having to refer to corresponding classical BSS methods. Such relationships however do exist. We therefore describe them hereafter, to avoid any ambiguity about the claimed novelty of the approaches described in Sects. 1.7.2 and 1.7.3, with respect to results available in the literature, not only dealing with classical BSS, but also with classical adaptive noise cancelation (ANC).

Several cumulant-based methods have been proposed for handling linear instantaneous mixtures of classical sources which are mutually independent and i.i.d, i.e.,

which have no temporal structure (or whose temporal structure is not exploited). This especially includes the COM2 [10], JADE [6], gradient-based [15] and fixed-point (FastICA) [32] kurtosis optimization methods. All these approaches resort to fourth-order cumulants, because it is well known that blind separation of i.i.d. sources cannot be achieved only using (first-order and) second-order statistics for such mixtures (see e.g. Sect. 7.4 of [33]). Similarly, the well-known Herault-Jutten network [34] uses fourth-order moments, or generalized moments (i.e., moments involving nonlinear functions of output signals) in this configuration.

On the contrary, the sources may be restored from their linear instantaneous mixtures by only using second-order statistics if they have temporal structure. This especially includes stationary sources which are autocorrelated over time: see e.g. the AMUSE method [51] also proposed by Fety [26], their SOBI extension [2] and the Molgedey-Schuster approach [37].

The relationship between all these classical BSS methods and our BQSS approaches of Sects. 1.7.2 and 1.7.3 may be interpreted as follows. The method proposed in Sect. 1.7.2 is able to solve the BQSS problem for i.i.d. sources by only using first-order and second-order moments of observations and separating system outputs, unlike in classical linear instantaneous BSS. The method described in Sect. 1.7.3, which is also applicable to i.i.d. sources, is only based on second-order statistics (and first-order ones, in the sense: *centered* moments), but one should keep in mind that these are statistics of a *nonlinearly transformed* version of the outputs of the separating system (also involving an observed signal). If comparing this approach to classical BSS, one should therefore also wonder whether classical BSS can be performed by only resorting to first-order and second-order statistics of a *transformed version* of the outputs of the separating system. The answer is positive: as mentioned above, some well-known classical BSS methods are based on a cost function C defined as the kurtosis of a (centered and scaled) output y_i of the separating system, i.e:

$$C = k(y_i) = E\{y_i^4\} - 3(E\{y_i^2\})^2. \tag{1.79}$$

Now allowing ourselves to consider a transformed version of y_i, we here introduce the signal $z_i = y_i^2$. The above cost function then reads

$$C = E\{z_i^2\} - 3(E\{z_i\})^2. \tag{1.80}$$

BSS is thus achieved by only using the first-order and second-order moments of this transformed output, as in our BQSS approach of Sect. 1.7.3. These classical BSS and BQSS methods however differ in the required combination of moments and in the nature of the transforms used for creating: (i) the signals involved in the separation criterion and (ii) the output signals of the separating system.

Our BQSS method of Sect. 1.7.3 may also be seen as a nonlinear extension of linear ANC, as will now be explained. In its simplest form, ANC combines mixing and separating stages which yield two signals, i.e: (i) an output signal which may be a linear instantaneous mixture of one unknown signal of interest and of one undesired

signal, and (ii) the undesired signal, measured alone. ANC then consists in blindly tuning a signal scale factor in the above possibly mixed output, so as to cancel the component of this output which corresponds to the undesired signal. Our method of Sect. 1.7.3 performs the following nonlinear extension of linear ANC: (i) we first introduced the signal z_3, which is derived as a nonlinear combination of available signals (see (1.69)), but which is also a linear instantaneous mixture of a signal of interest and of an undesired overall component (see (1.70)); (ii) we then created z_{12} as a nonlinear combination of the available signals (see (1.72)), that we designed so that it becomes equal to the undesired component of z_3, up to its tunable scale factor which appears in z_3 (see (1.70) and (1.71)). This then opens the way to the blind adaptation of the latter scale factor (so as to cancel it) that we eventually described in Sect. 1.7.3.

References

1. Almeida, L.B.: MISEP-linear and nonlinear ICA based on mutual information. J. Mach. Learn. Res. **4**, 1297–1318 (2003)
2. Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., Moulines, E.: A blind source separation technique using second-order statistics. IEEE Trans. Signal Process. **45**(2), 434–444 (1997)
3. Bennett, C.H., Shor, P.W.: Quantum information theory. IEEE Trans. Inf. Theory **44**(6), 2724–2742 (1998)
4. Bienvenu, G., Kopp, L.: Optimality of high resolution array processing using the eigensystem approach. IEEE Trans. Acoust. Speech Signal Process. **ASSP-31**(5), 1235–1247 (1983)
5. Brillouin, L.: Science and Information Theory. Academic Press, New York (1956)
6. Cardoso, J.-F., Souloumiac, A.: Blind beamforming for non Gaussian signals. IEE Proc. F **140**(6), 362–370 (1993)
7. Cardoso, J.-F.: Infomax and maximum likelihood for blind source separation. IEEE Signal Process. Lett. **4**(4), 112–114 (1997)
8. Cardoso, J.-F.: Blind signal separation: statistical principles. Proc. IEEE **86**(10), 2009–2025 (1998)
9. Chaouchi, C., Deville, Y., Hosseini, S.: Nonlinear source separation: a maximum likelihood approach for quadratic mixtures. In: Proceedings of the 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2010), Chamonix, France, 4–9 July 2010
10. Comon, P.: Independent component analysis, a new concept ? Signal Process. **36**(3), 287–314 (1994)
11. Comon, P., Jutten, C. (eds.): Handbook of Blind Source Separation. Independent Component Analysis and Applications. Academic Press, Oxford (2010)
12. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (1991)
13. Darbellay, G.A., Vajda, I.: Estimation of the information by an adaptive partitioning of the observation space. IEEE Trans. Inf. Theory **45**(4), 1315–1321 (1999)
14. De Lathauwer, L., De Moor, B., Vanderwalle, J.: Fetal electrocardiogram extraction by blind source subspace separation. IEEE Trans. Biomed. Eng. **47**(5), 567–572 (2000)
15. Delfosse, N., Loubaton, P.: Adaptive blind separation of independent sources: a deflation approach. Signal Process. **45**(1), 59–84 (1995)
16. Deville, Y., Damour, J., Charkani, N.: Multi-tag radio-frequency identification systems based on new blind source separation neural networks. Neurocomputing **49**, 369–388 (2002)
17. Deville, Y., Deville, A.: Blind separation of quantum states: estimating two qubits from an isotropic Heisenberg spin coupling model. In: Proceedings of the 7th International Conference

on Independent Component Analysis and Signal Separation (ICA 2007), vol. LNCS 4666, pp. 706–713. Springer, London, 9–12 Sept 2007. Erratum: replace two terms $E\{r_i\}E\{q_i\}$ in (33) of [17] by $E\{r_i q_i\}$, since q_i depends on r_i: see (5) in the current paper
18. Deville, Y.: Traitement du signal : signaux temporels et spatiotemporels—Analyse des signaux, théorie de l'information, traitement d'antenne, séparation aveugle de sources. Ellipses Editions Marketing, Paris (2011)
19. Deville, Y., Hosseini, S., Deville, A.: Effect of indirect dependencies on maximum likelihood and information theoretic blind source separation for nonlinear mixtures. Signal Process. **91**(4), 793–800 (2011)
20. Deville, Y.: ICA-based and second-order separability of nonlinear models involving reference signals: general properties and application to quantum bits. Signal Process. **92**(8), 1785–1795 (2012)
21. Deville, Y., Deville, A.: Classical-processing and quantum-processing signal separation methods for qubit uncoupling. Quantum Inf. Process. **11**(6), 1311–1347 (2012)
22. Deville, Y., Deville, A.: A quantum/classical-processing signal separation method for two qubits with cylindrical-symmetry Heisenberg coupling. In: Deloumeaux, P., Gorzalka, J.D. (eds.) Information Theory: New Research, pp. 145–170. Nova Science Publishers, Hauppauge, NY (2012). ISBN: 978-1-62100-325-0 (Chapter 5)
23. DiVincenzo, D.P.: Quantum computation. Science **270**, 255–261 (1995)
24. Duarte, L.T., Jutten, C.: A mutual information minimization approach for a class of nonlinear recurrent separating systems. In: IEEE International Workshop on Machine Learning for Signal Processing, pp. 122–127. Thessaloniki, Greece, 27–29 Aug 2007
25. Ehlers, F., Schuster, H.G.: Blind separation of convolutive mixtures and an application in automatic speech recognition in a noisy environment. IEEE Trans. Signal Process. **45**(10), 2608–2612 (1997)
26. Fety, L.: Méthodes de traitement d'antenne adaptées aux radiocommunications. Ph.D, ENST, Paris, France, June 3 (1988)
27. Gaeta, M., Lacoume, J.-L.: Sources separation without a priori knowledge: the maximum likelihood solution. In: Fifth European Signal Processing Conference (EUSIPCO-90), pp. 621–624. Barcelona, Spain, 18–21 Sept 1990
28. Guidara, R., Hosseini, S., Deville, Y.: Blind separation of nonstationary Markovian sources using an equivariant Newton-Raphson algorithm. IEEE Signal Process. Lett. **16**(5), 426–429 (2009)
29. Guidara, R., Hosseini, S., Deville, Y.: Maximum likelihood blind image separation using non-symmetrical half-plane Markov random fields. IEEE Trans. Image process. **18**(11), 2435–2450 (2009)
30. Hosseini, S., Jutten, C., Pham, D.T.: Markovian source separation. IEEE Trans. Signal Process. **51**(12), 3009–3019 (2003)
31. Hosseini, S., Deville, Y.: Blind maximum likelihood separation of a linear-quadratic mixture. In: Proceedings of the Fifth International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004), vol. LNCS 3195, pp. 694–701. Springer, Granada, Spain, 22–24 Sept 2004. Erratum: see also "Correction to "Blind maximum likelihood separation of a linear-quadratic mixture"", available on-line at http://arxiv.org/abs/1001.0863
32. Hyvarinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Comput. **9**, 1483–1492 (1997)
33. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
34. Jutten, C., Hérault, J.: Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture. Signal Process. **24**(1), 1–10 (1991)
35. Kendall, M., Stuart, A.: The Advanced Theory of Statistics, vol. 1. Charles Griffin, London, High Wycombe (1977)
36. Mokhtari, F., Babaie-Zadeh, M., Jutten, C.: Blind separation of bilinear mixtures using mutual information minimization. In: Proceedings of IEEE MLSP, France, Grenoble, 2–4 Sept 2009

37. Molgedey, L., Schuster, H.G.: Separation of a mixture of independent signals using time delayed correlation. Phys. Rev. Lett. **72**(23), 3634–3637 (1994)
38. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)
39. Pham, D.T.: Blind separation of instantaneous mixture of sources via an independent component analysis. IEEE Trans. Signal Process. **44**(11), 2768–2779 (1996)
40. Pham, D.T., Garat, P.: Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. IEEE Trans. Signal Process. **45**(7), 1712–1725 (1997)
41. Pham, D.-T., Cardoso, J.-F.: Blind separation of instantaneous mixtures of nonstationary sources. IEEE Trans. Signal Process. **49**(9), 1837–1848 (2001)
42. Pham, D.T.: Mutual information approach to blind separation of stationary sources. IEEE Trans. Inf. Theory **48**(7), 1935–1946 (2002)
43. Schmidt, R.: Multiple emitter location and signal parameter estimation. In: Proceedings of the RADC Spectrum Estimation Workshop, pp. 243–258. Rome,NY (1979)
44. Schmidt, R.: Multiple emitter location and signal parameter estimation. In: IEEE Transactions on Antennas and Propagation, vol. AP-34, no. 3, pp. 276–280, Mar 1986
45. Schrödinger, E.: What Is Life ?. Cambridge University Press, Cambridge (1944)
46. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana and Chicago (1949)
47. Shor, P.W.: Progress in quantum algorithms. Quantum Inf. Process. **3**(1–5), pp. 5–13 (2004)
48. Taleb, A., Jutten, C.: Source separation in post-nonlinear mixtures. IEEE Trans. Signal Process. **47**(10), 2807–2820 (1999)
49. Taleb, A.: A generic framework for blind source separation in structured nonlinear models. IEEE Trans. Signal Process. **50**(8), 1819–1830 (2002)
50. P. Tichavsky's home page: http://si.utia.cas.cz/Tichavsky.html
51. Tong, L., Liu, R., Soon, V.C., Huang, Y.-F.: Indeterminacy and identifiability of blind identification. IEEE Trans. Circuits Syst. **38**(5), 499–509 (1991)
52. Vandersypen, L.M.K., Chuang, I.L.: NMR techniques for quantum control and computation. Rev. Mod. Phys. **76**, 1037–1069 (2004)
53. Vandersypen, L.: Dot-To-Dot Design. In: IEEE Spectrum, pp. 34–39. Elsevier publishing co., New York (2007)
54. Van Trees, H.L.: Optimum Array Processing, Part IV of Detection, Estimation and Modulation Theory. Wiley, New York (2002)
55. White, A.G., Gilchrist, A.: Measuring two-qubit gates. J.Opt. Soc. Am. B **24**(2), 172–183 (2007)
56. Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Hearn, R.H., Zeidler, J.R., Dong, E., Goodlin, R.C.: Adaptive noise cancelling: principles and applications. Proc. IEEE **63**(12), 1692–1716 (1975)
57. Zhang, J., Khor, L.C., Woo, W.L., Dlay, S.S.: A maximum likelihood approach to nonlinear convolutive blind source separation.In: Proceedings of the Sixth International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006), vol. LNCS 3889, pp. 926–933. Springer, Charleston, SC, USA, 5–8 Mar 2006

Chapter 2
Blind Source Separation Based on Dictionary Learning: A Singularity-Aware Approach

Xiaochen Zhao, Guangyu Zhou, Wei Dai and Wenwu Wang

Abstract This chapter surveys recent works in applying sparse signal processing techniques, in particular, dictionary learning algorithms to solve the blind source separation problem. For the proof of concepts, the focus is on the scenario where the number of mixtures is not less than that of the sources. Based on the assumption that the sources are sparsely represented by some dictionaries, we present a joint source separation and dictionary learning algorithm (SparseBSS) to separate the noise corrupted mixed sources with very little extra information. We also discuss the singularity issue in the dictionary learning process, which is one major reason for algorithm failure. Finally, two approaches are presented to address the singularity issue.

2.1 Introduction

Blind source separation (BSS) has been investigated during the last two decades; many algorithms have been developed and applied in a wide range of applications including biomedical engineering, medical imaging, speech processing, astronomical

The first two authors made equal contribution to this chapter.

X. Zhao (✉) · G. Zhou · W. Dai
Imperial College London, London, UK
e-mail: xiaochen.zhao10@imperial.ac.uk

G. Zhou
e-mail: g.zhou11@imperial.ac.uk

W. Dai
e-mail: wei.dai1@imperial.ac.uk

W. Wang
University of Surrey, Surrey, UK
e-mail: w.wang@surrey.ac.uk

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_2,
© Springer-Verlag Berlin Heidelberg 2014

imaging, and communication systems. Typically, a linear mixture model is assumed where the mixtures $\boldsymbol{Z} \in \mathbb{R}^{r\times N}$ are described as $\boldsymbol{Z} = \boldsymbol{AS} + \boldsymbol{V}$. Each row of $\boldsymbol{S} \in \mathbb{R}^{s\times N}$ is a source and $\boldsymbol{A} \in \mathbb{R}^{r\times s}$ models the linear combinations of the sources. The matrix $\boldsymbol{V} \in \mathbb{R}^{r\times N}$ represents additive noise or interference introduced during mixture acquisition and transmission.

Usually in the BSS problem, the only known information is the mixtures $\boldsymbol{Z}$ and the number of sources. One needs to determine both the mixing matrix $\boldsymbol{A}$ and the sources $\boldsymbol{S}$, i.e., mathematically, one needs to solve

$$\min_{\boldsymbol{A},\boldsymbol{S}} \|\boldsymbol{Z} - \boldsymbol{AS}\|_F^2 .$$

It is clear that such a problem has an infinite number of solutions, i.e., the problem is ill-posed. In order to find the true sources and the mixing matrix (subject to scale and permutation ambiguities), it is often required to add extra constraints to the problem formulation. For example, a well-known method called independent component analysis (ICA) [1] assumes that the original sources are statistically independent. This has led to some widely used approaches such as Infomax [2], maximum likelihood estimation [3], the maximum a posterior (MAP) [4], and FastICA [1].

Sparsity prior is another property that can be used for BSS. Most natural signals are sparse under some dictionaries. The mixtures, viewed as a superposition of sources, are in general less sparse compared to the original sources. Based on this fact, the sparse prior has been used in solving the BSS problem from various perspectives since 2001, e.g., sparse ICA (SPICA) [5] and sparse component analysis (SCA) [6]. In this approach, there is typically no requirement that the original sources have to be independent. As a result, these algorithms are capable of dealing with highly correlated sources, for example, in separating two superposed identical speeches, with one being a few samples delayed version of the other. Jourjine et al. proposed an SCA-based algorithm in [7] aiming at solving the anechoic problem. SCA algorithms look for a sparse representation under predefined bases such as discrete cosine transform (DCT), wavelet, curvelet, etc. Morphological component analysis (MCA) [8] and its extended algorithms for multichannel cases, Multichannel MCA (MMCA) [9], and Generalized MCA (GMCA) [10], are also based on the assumption that the original sources are sparse in different bases instead of explicitly constructed dictionaries. However, these algorithms do not exhibit an outstanding performance since in most cases the predefined dictionaries are too general to offer sufficient details of sources when used in sparse representation.

A method to address this problem is to learn data-specific dictionaries. In [11], the authors advised to train a dictionary from the mixtures/corrupted-images and then decompose it into a few dictionaries according to the prior knowledge of the main components in different sources. This algorithm is used for separating images with different main frequency components (e.g., Cartoon and Texture images) and obtained satisfactory results in image denoising. Starck et al. proposed in [12] to learn dictionary from a set of exemplar images for each source. Xu et al. [13] proposed an algorithm, which allows the dictionaries to be learned from the sources or the

mixtures. In most BSS problems, however, dictionaries learned from the mixtures or from similar exemplar images rarely well represent the original sources.

To get more accurate separation results, the dictionaries should be adapted to the unknown sources. The motivation is clear from the assumption that the sources are sparsely represented by some dictionaries. The initial idea of learning dictionaries while separating the sources was suggested by Abolghasemi et al. [14]. They proposed a two-stage iterative process. In this process each source is equipped with a dictionary, which is learned in each iteration, right after the previous mixture learning stage. Considering the size of dictionaries being much larger than the mixing matrix, the main computational cost is on the dictionary learning stage. This two-stage procedure was further developed in Zhao et al. [15]. The method was termed as SparseBSS, which employs a joint optimization framework based on the idea of SimCO dictionary update algorithm [16]. By studying the optimization problem encountered in dictionary learning, the phenomenon of singularity in dictionary update was for the first time discovered. Furthermore, from the viewpoint of the dictionary redundancy, SparseBSS uses only one dictionary to represent all the sources, and is therefore computationally much more efficient than using multiple dictionaries as in [14]. This joint dictionary learning and source separation framework is the focus of this chapter. This framework can be extended potentially to a convolutive or underdetermined model, e.g., apply clustering method to solve the the ill-posed inverse problem in underdetermined model [13]; however, discussion on such an extension is beyond the scope of this chapter. In this chapter, we focus on overdetermined/even determined model.

The remainder of this chapter is organized as follows. Section 2.2 describes the framework of the BSS problem based on dictionary learning. The recently proposed algorithm SparseBSS is introduced and compared in detail with the related benchmark algorithm BMMCA. In Sect. 2.3, we briefly introduce the background of dictionary learning algorithms and then discuss the important observation of the singularity issue, which is a major reason for the failure of dictionary learning algorithms and hence dictionary learning-based BSS algorithms. Later, two available approaches are presented to address this problem. In Sect. 2.5, we conclude our work and discuss some possible extensions.

2.2 Framework of Dictionary Learning-Based BSS Problem

We consider the following linear and instantaneous mixing model. Suppose there are s source signals of the same length, denoted by $\boldsymbol{s}_1,\ \boldsymbol{s}_2, \ldots, \boldsymbol{s}_s$, respectively, where $\boldsymbol{s}_i \in \mathbb{R}^{1\times N}$ is a row vector to denote the ith source. Assume that these sources are linearly mixed into r observation signals denoted by $\boldsymbol{z}_1,\ \boldsymbol{z}_2, \ldots, \boldsymbol{z}_r$ respectively, where $\boldsymbol{z}_j \in \mathbb{R}^{1\times N}$. In the matrix format, denote $\boldsymbol{S} = \left[\boldsymbol{s}_1^T, \boldsymbol{s}_2^T, \ldots, \boldsymbol{s}_s^T\right]^T \in \mathbb{R}^{s\times N}$ and $\boldsymbol{Z} = \left[\boldsymbol{z}_1^T, \boldsymbol{z}_2^T, \ldots, \boldsymbol{z}_r^T\right]^T \in \mathbb{R}^{r\times N}$. Then the mixing model is given by

$$\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{V}, \tag{2.1}$$

where $\boldsymbol{A} \in \mathbb{R}^{r\times s}$ is the mixing matrix and $\boldsymbol{V} \in \mathbb{R}^{r\times N}$ is denoted as zero mean additive Gaussian noise. We also assume that $r \geq s$, i.e., the underdetermined case will not be discussed here.

2.2.1 Separation with Dictionaries Known in Advance

For some BSS algorithms, such as MMCA [9], orthogonal dictionaries $\boldsymbol{D}_i$'s are required to be known a priori. Each source $\boldsymbol{s}_i$ is assumed to be sparsely represented by a different $\boldsymbol{D}_i$. Hence, we have $\boldsymbol{s}_i = \boldsymbol{D}_i\boldsymbol{x}_i$ with $\boldsymbol{x}_i$'s being sparse. Given the observation $\boldsymbol{Z}$ and the dictionaries $\boldsymbol{D}_i$'s, MMCA [9] aims to estimate the mixing matrix and sources, based on the following form:

$$\min_{\boldsymbol{A},\boldsymbol{S}} \|\boldsymbol{Z} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + \sum_{i=1}^{n} \lambda_i \left\| \boldsymbol{s}_i \boldsymbol{D}_i^{\dagger} \right\|_1 . \tag{2.2}$$

Here $\lambda_i > 0$ is the weighting parameter determined by the noise deviation σ, $\|\cdot\|_F$ represents the Frobenius norm, $\|\cdot\|_1$ is the ℓ_1 norm and $\boldsymbol{D}_i^{\dagger}$ denotes the pseudo-inverse of $\boldsymbol{D}_i$. Predefined dictionaries generated from typical mathematical transforms, e.g., DCT, wavelets and curvelets, do not target particular sources, and thus do not always provide sufficiently accurate reconstruction and separation results. Elad et al. [11] designed a method to first train a redundant dictionary by K-SVD algorithm in advance, and then decompose it into a few dictionaries, one for each source. This method works well when the original sources have components that are largely different from each other under some unknown mathematical transformations (e.g. Cartoon and Texture images under the DCT transformation). Otherwise, the dictionaries found may not be appropriate in the sense that they may fit better the mixtures rather than the sources.

2.2.2 Separation with Unknown Dictionaries

2.2.2.1 SparseBSS Algorithm Framework

According to the authors' knowledge, BMMCA and SparseBSS are the two most recent BSS algorithms, which implement the idea of performing source separation and dictionary learning simultaneously. Due to space constraints, we focus on Sparse BSS in this chapter. In SparseBSS, one assumes that all the sources can be sparsely represented under the same dictionary. In order to obtain enough training samples for dictionary learning, multiple overlapped segments (patches) of the sources are taken. To extract small overlapped patches from the source image $\boldsymbol{s}_i$,

a binary matrix $\boldsymbol{P}_k \in \mathbb{R}^{n\times N}$ is defined as a patching operator[1] [15]. The product $\boldsymbol{P}_k \cdot \boldsymbol{s}_i^T \in \mathbb{R}^{n\times 1}$ is needed to obtain and vectorize the kth patch of size $\sqrt{n} \times \sqrt{n}$ taken from image $\mathcal{S}_i$. Denote $\boldsymbol{P} = [\boldsymbol{P}_1, \ldots, \boldsymbol{P}_K] \in \mathbb{R}^{n\times KN}$, where K is the number of patches taken from each image. Then the extraction of multiple sources $\boldsymbol{S}$ is defined as $\mathcal{P}\boldsymbol{S} = ([\boldsymbol{P}_1, \ldots, \boldsymbol{P}_K]) \cdot (\left[\boldsymbol{s}_1^T, \boldsymbol{s}_2^T, \ldots, \boldsymbol{s}_s^T\right] \otimes \boldsymbol{I}_K) = \boldsymbol{P} \cdot (\boldsymbol{S}^T \otimes \boldsymbol{I}_K) \in \mathbb{R}^{n\times Ks}$, where symbol $\otimes$ denotes the Kronecker product and $\boldsymbol{I}_K$ indicates the identity matrix. The computational cost associated with converting from images to patches is low. Each column of $\mathcal{P}\boldsymbol{S}$ represents one vectorized patch. We sparsely represent $\mathcal{P}\boldsymbol{S}$ by using only one dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times d}$ and a sparse coefficient matrix $\boldsymbol{X} \in \mathbb{R}^{d\times Ks}$, which suggests $\mathcal{P}\boldsymbol{S} \approx \boldsymbol{D}\boldsymbol{X}$. This is different from BMMCA, where multiple dictionaries are used for multiple sources.

With these notations, the BSS problem is formulated as the following joint optimization problem:

$$\min_{\boldsymbol{A},\boldsymbol{S},\boldsymbol{D},\boldsymbol{X}} \lambda \left\|\boldsymbol{Z} - \boldsymbol{A}\boldsymbol{S}\right\|_F^2 + \left\|\mathcal{P}^{\dagger}\left(\boldsymbol{D}\boldsymbol{X}\right) - \boldsymbol{S}\right\|_F^2. \tag{2.3}$$

The parameter λ is introduced to balance the measurement error and the sparse approximation error, and $\boldsymbol{X}$ is assumed to be sparse.

To find the solution of the above problem, we propose a joint optimization algorithm to iteratively update the following two pairs of variables $\{\boldsymbol{D}, \boldsymbol{X}\}$ and $\{\boldsymbol{A}, \boldsymbol{S}\}$ over two stages until a (local) minimizer is found. Note that in each stage there is only one pair of variables to be updated simultaneously by keeping the other pair fixed.

- Dictionary learning stage

$$\min_{\boldsymbol{D},\boldsymbol{X}} \left\|\boldsymbol{D}\boldsymbol{X} - \mathcal{P}\boldsymbol{S}\right\|_F^2, \tag{2.4}$$

- Mixture learning stage

$$\min_{\boldsymbol{A},\boldsymbol{S}} \lambda \left\|\boldsymbol{Z} - \boldsymbol{A}\boldsymbol{S}\right\|_F^2 + \left\|D\boldsymbol{X} - \mathcal{P}\boldsymbol{S}\right\|_F^2. \tag{2.5}$$

Without being explicit in (2.3), a sparse coding process is involved where greedy algorithms, such as orthogonal matching pursuit (OMP) [17] and subspace pursuit (SP), [18] are used to solve

$$\min_{\boldsymbol{X}} \left\|\boldsymbol{X}\right\|_0, \text{ s.t. } \left\|\boldsymbol{D}\boldsymbol{X} - \mathcal{P}\left(\boldsymbol{S}\right)\right\|_F^2 \leq \epsilon,$$

where $\|\boldsymbol{X}\|_0$ counts the number of nonzero elements in $\boldsymbol{X}$, the dictionary $\boldsymbol{D}$ is assumed fixed, and $\epsilon > 0$ is an upper bound on the sparse approximation error.

During the optimization, further constraints are made on the matrices $\boldsymbol{A}$ and $\boldsymbol{D}$. Consider the dictionary learning stage. Since the performance is invariant to scaling and permutations of the dictionary codewords (columns of $\boldsymbol{D}$), we follow the

[1] Note that in this chapter $\boldsymbol{P}_k$ is defined as a patching operator for image sources. The patching operator for audio sources can be similarly defined as well.

convention in the literature, e.g., [16], and enforce the dictionary to be updated on the set

$$\mathcal{D} = \left\{ \boldsymbol{D} \in \mathbb{R}^{n\times d} : \ \left\|\boldsymbol{D}_{:,i}\right\|_2 = 1, \ 1 \le i \le d \right\}, \tag{2.6}$$

where $\boldsymbol{D}_{:,i}$ stands for the ith column of $\boldsymbol{D}$. A detailed description of the advantage by adding this constraint can be found in [16]. Sparse coding, once performed, provides information about which elements of $\boldsymbol{X}$ are zeros and which are nonzeros. Define the sparsity pattern by $\Omega = \left\{(i,j) : \ \boldsymbol{X}_{i,j} \neq 0\right\}$, which is the index set of the nonzero elements of $\boldsymbol{X}$. Define $\mathcal{X}_\Omega$ as the set of all matrices conforming to the sparsity pattern Ω. This is the feasible set of the matrix $\boldsymbol{X}$. The optimization problem for the dictionary learning stage can be written as

$$\begin{aligned} \min_{\boldsymbol{D}\in\mathcal{D}} f_\mu(\boldsymbol{D}) &= \min_{\boldsymbol{D}\in\mathcal{D}} \ \min_{\boldsymbol{X}\in\mathcal{X}_\Omega} \ \|\boldsymbol{D}\boldsymbol{X} - \mathcal{P}(\boldsymbol{S})\|_F^2 + \mu \|\boldsymbol{X}\|_F^2, \\ &= \min_{\boldsymbol{D}\in\mathcal{D}} \ \min_{\boldsymbol{X}\in\mathcal{X}_\Omega} \ \left\| \begin{bmatrix} \mathcal{P}(\boldsymbol{S}) \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{D} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \boldsymbol{X} \right\|_F^2. \end{aligned} \tag{2.7}$$

The term $\mu \|\boldsymbol{X}\|_F^2$ introduces a penalty to alleviate the singularity issue. See more details in Sect. 2.3.3.

In the mixture learning stage, similar to the dictionary learning stage, we constrain the mixing matrix $\boldsymbol{A}$ in the set

$$\mathcal{A} = \left\{ \boldsymbol{A} \in \mathbb{R}^{r\times s} : \ \left\|\boldsymbol{A}_{:,i}\right\|_2 = 1, \ 1 \le i \le s \right\}. \tag{2.8}$$

This constraint is necessary. Otherwise, if the mixing matrix $\boldsymbol{A}$ is scaled by a constant c and the source $\boldsymbol{S}$ is inversely scaled by c^{-1}, then for any $\{\boldsymbol{A}, \boldsymbol{S}\}$ we can always find a solution $\{c\boldsymbol{A}, c^{-1}\boldsymbol{S} | c > 1\}$, which further decreases the objective function (2.3) from $\lambda \|\boldsymbol{Z} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + \|D\boldsymbol{X} - \mathcal{P}\boldsymbol{S}\|_F^2$ to $\lambda \|\boldsymbol{Z} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + c^{-2} \|D\boldsymbol{X} - \mathcal{P}\boldsymbol{S}\|_F^2$. Now if we view the sources $\boldsymbol{S} \in \mathbb{R}^{s\times n}$ as a "sparse" matrix with the sparsity pattern $\Omega' = \{(i,j) : \ 1 \le i \le s, \ 1 \le j \le N\}$. Then, the optimization problem for the mixture learning stage is exactly the same as that for the dictionary learning stage:

$$\begin{aligned} \min_{\boldsymbol{A}\in\mathcal{A}} f_\lambda(\boldsymbol{A}) &= \min_{\boldsymbol{A}\in\mathcal{A}} \ \min_{\boldsymbol{S}\in\mathbb{R}^{s\times n}} \ \lambda \|\boldsymbol{Z} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + \left\|\mathcal{P}^\dagger(\boldsymbol{D}\boldsymbol{X}) - \boldsymbol{S}\right\|_F^2 \\ &= \min_{\boldsymbol{A}\in\mathcal{A}} \ \min_{\boldsymbol{S}\in\mathcal{X}_{\Omega'}} \ \left\| \begin{bmatrix} \sqrt{\lambda}\boldsymbol{Z} \\ \mathcal{P}^\dagger(\boldsymbol{D}\boldsymbol{X}) \end{bmatrix} - \begin{bmatrix} \sqrt{\lambda}\boldsymbol{A} \\ \boldsymbol{I} \end{bmatrix} \boldsymbol{S} \right\|_F^2, \end{aligned} \tag{2.9}$$

where the fact that $\mathbb{R}^{s\times n} = \mathcal{X}_{\Omega'}$ has been used. As a result, the SimCO mechanism can be directly applied. Here, we do not require the prior knowledge of the scaling matrix in front of the true mixing matrix [10], as otherwise required in MMCA and GMCA algorithms.

To conclude this section, we emphasize the following treatment of the optimization problems (2.7) and (2.9). Both involve a joint optimization over two variables, i.e., $\boldsymbol{D}$ and $\boldsymbol{X}$ for (2.7) and $\boldsymbol{A}$ and $\boldsymbol{S}$ for (2.9). Note that if $\boldsymbol{D}$ and $\boldsymbol{A}$ are fixed, then

the optimal $\boldsymbol{X}$ and $\boldsymbol{S}$ can be easily computed by solving the corresponding least squares problems. Motivated by this fact, we write (2.7) and (2.9) as $\min_{\boldsymbol{D}\in\mathcal{D}} f_\mu(\boldsymbol{D})$ and $\min_{\boldsymbol{A}\in\mathcal{A}} f_\lambda(\boldsymbol{A})$, respectively, when $f_\mu(\boldsymbol{D})$ and $f_\lambda(\boldsymbol{A})$ are properly defined in (2.7) and (2.9). In this way, the optimization problems, at least from the surface, only involve one variable. This helps the discovery of the singularity issue and the developments of handling singularity. See Sect. 2.3 for details.

2.2.2.2 Implementation Details in SparseBSS

Most optimization methods are based on line search strategies. The dictionaries at the beginning and the end of the kth iteration, denoted by $\boldsymbol{D}^{(k)}$ and $\boldsymbol{D}^{(k+1)}$, respectively, can be related by $\boldsymbol{D}^{(k+1)} = \boldsymbol{D}^{(k)} + \alpha^{(k)}\boldsymbol{\eta}^{(k)}$ where $\alpha^{(k)}$ is an appropriately chosen step size and $\boldsymbol{\eta}^{(k)}$ is the search direction. The step size $\alpha^{(k)}$ can be determined by *Armijo condition* or *Golden selection* presented in [19]. The search direction $\boldsymbol{\eta}^{(k)}$ can be determined by a variety of gradient methods [19, 20]. The decision of $\boldsymbol{\eta}^{(k)}$ plays the key role, which directly affects the convergence rate of the whole algorithm. Generally speaking, a Newton direction is a preferred choice (compared with the gradient descent direction) [19]. In many cases, direct computation of the Newton direction is computationally prohibitive. Iterative methods can be used to search the Newton direction. Take the Newton Conjugate Gradient (Newton CG) method as an example. It starts with the gradient descent direction $\boldsymbol{\eta}_0$ and iteratively refines it toward the Newton direction. Denote the gradient of $f_\mu(\boldsymbol{D})$ as $\nabla f_\mu(\boldsymbol{D})$. Denote $\nabla_{\boldsymbol{\eta}}\left(\nabla f_\mu(\boldsymbol{D})\right)$ as the directional derivative of $\nabla f_\mu(\boldsymbol{D})$ along $\boldsymbol{\eta}$ [21]. In each line search step of the Newton CG method, instead of computing the Hessian $\nabla^2 f_\mu(\boldsymbol{D}) \in \mathbb{R}^{md\times md}$ explicitly, one only needs to compute $\nabla_{\boldsymbol{\eta}}\left(\nabla f_\mu(\boldsymbol{D})\right) \in \mathbb{R}^{m\times d}$. The required computational and storage resources are therefore much reduced.

When applying the Newton CG to minimize $f_\mu(\boldsymbol{D})$ in (2.7), the key computations are summarized below. Denote $\tilde{\boldsymbol{D}} = \begin{bmatrix}\boldsymbol{D}^T & \mu\boldsymbol{I}\end{bmatrix}^T$ and let $\Omega(:,j)$ be the index set of nonzero elements in $\boldsymbol{X}_{:,j}$. We consider $\tilde{\boldsymbol{D}}_i = \tilde{\boldsymbol{D}}_{:,\Omega(:,i)} \in \mathbb{R}^{(m+r)\times r}$ with $m > r$. Matrix $\tilde{\boldsymbol{D}}_i$ is a full column rank tall matrix. We denote

$$f_i(\tilde{\boldsymbol{D}}_i) = \min_{\boldsymbol{x}_i} \|\boldsymbol{y}_i - \tilde{\boldsymbol{D}}\boldsymbol{x}_i\|_2^2$$

and the optimal

$$\boldsymbol{x}_i^* = \arg\min_{\boldsymbol{x}_i} \|\boldsymbol{y}_i - \tilde{\boldsymbol{D}}\boldsymbol{x}_i\|_2^2.$$

Denote $\tilde{\boldsymbol{D}}_i^\dagger$ as the pseudo-inverse of $\tilde{\boldsymbol{D}}_i$. Then we have $\frac{\partial f}{\partial \boldsymbol{x}_i}\,|_{\boldsymbol{x}_i^*} = \boldsymbol{0}$, where $\boldsymbol{x}_i^* = \tilde{\boldsymbol{D}}_i^\dagger \boldsymbol{y}_i$, and $\nabla f_i(\tilde{\boldsymbol{D}}_i)$ can be written as

$$\nabla f_i(\tilde{\boldsymbol{D}}_i) = \frac{\partial f}{\partial \tilde{\boldsymbol{D}}_i} + \frac{\partial f}{\partial \boldsymbol{x}_i}\frac{\partial \boldsymbol{x}_i}{\partial \tilde{\boldsymbol{D}}_i} = -2(\boldsymbol{y}_i - \tilde{\boldsymbol{D}}_i\boldsymbol{x}_i^*)\boldsymbol{x}_i^{*T} + \boldsymbol{0} \tag{2.10}$$

To compute $\nabla_{\boldsymbol{\eta}}\left(\nabla f_i(\tilde{\boldsymbol{D}}_i)\right)$, we have

$$\begin{aligned}\nabla_{\boldsymbol{\eta}}\left(\nabla f_i(\tilde{\boldsymbol{D}}_i)\right) &= 2\nabla_{\boldsymbol{\eta}}\left(\tilde{\boldsymbol{D}}_i\boldsymbol{x}_i^* - \boldsymbol{y}_i\right)\boldsymbol{x}_i^{*T} + 2\left(\tilde{\boldsymbol{D}}_i\boldsymbol{x}_i^* - \boldsymbol{y}_i\right)\nabla_{\boldsymbol{\eta}}\boldsymbol{x}_i^{*T} \\ &= 2\nabla_{\boldsymbol{\eta}}\tilde{\boldsymbol{D}}_i\boldsymbol{x}_i^*\boldsymbol{x}_i^{*T} + 2\tilde{\boldsymbol{D}}_i\nabla_{\boldsymbol{\eta}}\boldsymbol{x}_i^*\boldsymbol{x}_i^{*T} + 2\left(\tilde{\boldsymbol{D}}_i\boldsymbol{x}_i^* - \boldsymbol{y}_i\right)\nabla_{\boldsymbol{\eta}}\boldsymbol{x}_i^{*T} \\ &= 2\boldsymbol{\eta}\boldsymbol{x}_i^*\boldsymbol{x}_i^{*T} + 2\tilde{\boldsymbol{D}}_i\nabla_{\boldsymbol{\eta}}\boldsymbol{x}_i^*\boldsymbol{x}_i^{*T} + 2\left(\tilde{\boldsymbol{D}}_i\boldsymbol{x}_i^* - \boldsymbol{y}_i\right)\nabla_{\boldsymbol{\eta}}\boldsymbol{x}_i^{*T}, \end{aligned} \tag{2.11}$$

where $\nabla_{\boldsymbol{\eta}}\boldsymbol{x}^*$ is relatively easy to obtain,

$$\nabla_{\boldsymbol{\eta}}x^* = -\left(\tilde{\boldsymbol{D}}^T\tilde{\boldsymbol{D}}\right)^{-1}\left(\left(\tilde{\boldsymbol{D}}^T\boldsymbol{\eta} + \boldsymbol{\eta}^T\tilde{\boldsymbol{D}}\right)\tilde{\boldsymbol{D}}^{\dagger} - \boldsymbol{\eta}^T\right)\boldsymbol{y}. \tag{2.12}$$

From the definition of $\tilde{\boldsymbol{D}}_i, \boldsymbol{D}_i$ is a submatrix of $\tilde{\boldsymbol{D}}_i$, therefore $\nabla f_i(\boldsymbol{D}_i)$ and $\nabla_{\boldsymbol{\eta}}\left(\nabla f_i(\boldsymbol{D}_i)\right)$ are also, respectively, submatrices of $\nabla f_i(\tilde{\boldsymbol{D}}_i)$ and $\nabla_{\boldsymbol{\eta}}\left(\nabla f_i(\tilde{\boldsymbol{D}}_i)\right)$, i.e., $\nabla f_i(\boldsymbol{D}_i) = \left(\nabla f_i\left(\tilde{\boldsymbol{D}}_i\right)\right)_{1:m,:}$ and $\nabla_{\boldsymbol{\eta}}\left(\nabla f_i(\boldsymbol{D}_i)\right) = \left(\nabla_{\boldsymbol{\eta}}\left(\nabla f_i(\tilde{\boldsymbol{D}}_i)\right)\right)_{1:m,:}$.

In addition, it is also worth noting that the SparseBSS model, using one dictionary to sparsely represent all the sources will get almost the same performance as using multiple but same-sized dictionaries when the dictionary redundancy d/n is large enough. As a result, it is reasonable to train only one dictionary for all the sources. An obvious advantage of using one dictionary is that the computational cost does not increase when the number of sources increases.

2.2.3 Blind MMCA and Its Comparison to SparseBSS

BMMCA [14] is another recently proposed BSS algorithm based on adaptive dictionary learning. Without knowing dictionaries in advance, the BMMCA algorithm also trains dictionaries from the observed mixture $\boldsymbol{Z}$. Inspired by the hierarchical scheme used in MMCA and the update method in K-SVD, the separation model in BMMCA is made up of a few rank-1 approximation problems, where each problem targets on the estimation of one particular source

$$\min_{\boldsymbol{A}_{:,i},\boldsymbol{s}_i,\boldsymbol{D}_i,\boldsymbol{X}_i} \lambda\left\|\boldsymbol{E}_i - \boldsymbol{A}_{:,i}\boldsymbol{s}_i\right\|_F^2 + \left\|\boldsymbol{D}_i\boldsymbol{X}_i - \mathcal{R}\boldsymbol{s}_i\right\|_2^2 + \mu\left\|\boldsymbol{X}_i\right\|_0. \tag{2.13}$$

Different from the operator $\mathcal{P}$ defined earlier in SparseBSS algorithm, the operator $\mathcal{R}$ in BMMCA is used to take patches from only one estimated image $\boldsymbol{s}_i$. $\boldsymbol{D}_i$ is the trained dictionaries for representing source $\boldsymbol{s}_i$. $\boldsymbol{E}_i$ is the residual which can be written as

$$\boldsymbol{E}_i = \boldsymbol{Z} - \sum_{j\neq i}\boldsymbol{A}_{:,j}\boldsymbol{s}_j. \tag{2.14}$$

Despite being similar in problem formulation, BMMCA and SparseBSS differ in terms of whether the sources share a single dictionary in dictionary learning. In the SparseBSS algorithm, only one dictionary is used to provide sparse representations for all sources. BMMCA requires multiple dictionaries, one for each source. In the mixing matrix update, BMMCA imitates the K-SVD algorithm by splitting the steps of update and normalization. Such two-step based approach does not bring the expected optimality of $\boldsymbol{A} \in \mathcal{A}$, thereby giving inaccurate estimation, while SparseBSS keeps $\boldsymbol{A} \in \mathcal{A}$ during the optimization process. In BMMCA, the authors claim that the ratio between the parameter λ and the noise standard deviation σ is fixed to 30, which will not guarantee good estimation results at various noise levels.

2.3 Dictionary Learning and the Singularity Issue

As is clear from previous discussions, dictionary learning plays an essential role in solving the BSS problem when the sparse prior is used, and hence is the focus of this section. We first briefly introduce the relevant background, then discuss an interesting phenomenon, the singularity issue in the dictionary update stage, and finally present two approaches to handle the singularity issue. For readers who are more interested in the SparseBSS algorithm themselves may consider this section as optional and skip to Sect. 2.4.

2.3.1 Brief Introduction to Dictionary Learning Algorithms

One of the earliest dictionary learning algorithms is the method of optimal directions (MOD) [22] proposed by Engan et al. The main idea is as follows: in each iteration, one first fixes the dictionary and uses OMP [17] or FOCUSS [23] to update the sparse coefficients, then fixes the obtained sparse coefficients and updates the dictionary in the next stage. MOD was later modified to iterative least squares algorithm (ILS-DLA) [24] and recursive least squares algorithm (RLS-DLA) [25]. Aharon et al. developed the K-SVD algorithm [26], which can be viewed as a generalization of the K-means algorithm. In each iteration, the first step is to update the sparse coefficients in the same way as in MOD. Then in the second step, one fixes the sparse pattern, and updates the dictionary and the nonzero coefficients simultaneously. In particular, the codewords in the dictionary are sequentially selected: the selected codeword and the corresponding row of the sparse coefficients are updated simultaneously by using singular value decomposition (SVD). More recently, Dai et al. [16] considered the dictionary learning problem from a new perspective. They formulated dictionary learning as an optimization problem on manifolds and developed simultaneous codeword optimization (SimCO) algorithm. In each iteration SimCO allows multiple codewords of the dictionary to be updated with corresponding rows of the

sparse coefficients jointly. This new algorithm can be viewed as a generalization of both MOD and K-SVD. Some other dictionary learning algorithms are also developed in the past decade targeting on various circumstances. For example, based on stochastic approximations, Mairal et al. [27] proposed an online algorithm to address the problem with large data sets.

Theoretical or in-depth analysis about the dictionary learning problem was mean time in progress as well. Gribonval et al. [28], Geng et al. [29], and Jenatton et al. [30] studied the stability and robustness of the objective function under different probabilistic modeling assumptions, respectively. In addition, Dai et al. observed in [16] that the dictionary update procedure may fail to converge to a minimizer. This is a common phenomenon happening in MOD, K-SVD, and SimCO. Dai et al. further observed that ill-conditioned dictionaries, rather than stationary dictionaries, are the major reason that has led to the failure of the convergence. To alleviate this problem, Regularized SimCO was proposed in [16]. Empirical performance improvement was observed. The same approach was also considered in [31], however, without detailed discussion on the singularity issue. More recently, the fundamental drawback of regularized SimCO was demonstrated using an artificial example [32]. To further handle the singularity issue, a Smoothed SimCO [33] was proposed by adding multiplicative terms rather than additive regularization terms to the objective function.

2.3.2 Singularity Issue and Its Impacts

In dictionary update stage of existing mainstream algorithms, singularity is observed as the major reason leading to failures [16, 33]. Simulations in [16] suggests that the mainstream algorithms fail mainly because of singular points in the objective function rather than non-optimal stationary points. As dictionary learning is an essential part of the aforementioned SparseBSS, the singularity issue also has negative impact on the overall performance of BSS. To explain the singularity issue in dictionary update, we first formally define the singular dictionaries.

Definition 1 *A dictionary $\boldsymbol{D} \in \mathbb{R}^{m\times d}$ is singular under a given sparsity pattern Ω if there exists an $i \in [n]$ such that the corresponding sub-dictionary $\boldsymbol{D}_i \triangleq \boldsymbol{D}_{:,\Omega(:,i)}$ is column rank deficient. Or equivalently, the minimum singular value of $\boldsymbol{D}_i$, denoted as $\lambda_{\min}(\boldsymbol{D}_i)$, is zero.*

A dictionary $\boldsymbol{D} \in \mathbb{R}^{m\times d}$ is said to be ill-conditioned under a given sparsity pattern Ω if there exists an $i \in [n]$ such that the condition number of the sub-dictionary $\boldsymbol{D}_i$ is large, or equivalently $\lambda_{\min}(\boldsymbol{D}_i)$ is close to zero.

Definition 2 [16] *Define the condition number of a dictionary $\boldsymbol{D}$* as:

$$\kappa(\boldsymbol{D}) = \max_{i\in[n]} \frac{\lambda_{\max}(\boldsymbol{D}_i)}{\lambda_{\min}(\boldsymbol{D}_i)},$$

where $\lambda_{\max}(\boldsymbol{D}_i)$ *and* $\lambda_{\min}(\boldsymbol{D}_i)$ *represent the maximum and the minimum singular value of the sub-dictionary* $\boldsymbol{D}_i$ *respectively.*

The word "singular" comes from the fact that $f(\boldsymbol{D}) = \min_{\boldsymbol{X}\in\mathcal{X}_\Omega} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2$ is not continuous at a singular dictionary[2] and the corresponding

$$\boldsymbol{X}(\boldsymbol{D}) \triangleq \arg\min_{\boldsymbol{X}\in\mathcal{X}_\Omega} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2$$

is not unique. The singularity of $f(\boldsymbol{D})$ leads to convergence problems. Benchmark dictionary update procedures may fail to find a globally optimal solution. Instead they converge to a singular point of $f(\boldsymbol{D})$, i.e., a singular dictionary.

Ill-conditioned dictionaries are in the neighborhood of singular ones. Algorithmically when one of the $\lambda_{\min}(\boldsymbol{D}_i)$s is ill-conditioned, the curvature of $f(\boldsymbol{D})$ is quite large and the value of the gradient fluctuates dramatically. This seriously affects the convergence rate of the dictionary update process.

Furthermore, ill-conditioned dictionaries also bring negative effect on the sparse coding stage. Denote $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$ as the ith column of $\boldsymbol{Y}$ and $\boldsymbol{X}$ respectively. Consider a summand of the formulation in sparse coding stage [16, 26], i.e.,

$$\min_{\boldsymbol{x}_i} \|\boldsymbol{y}_i - \boldsymbol{D}\boldsymbol{x}_i\|_F^2 + \|\boldsymbol{x}_i\|_0 .$$

An ill-conditioned $\boldsymbol{D}$ corresponds to a very large condition number, which breaks the restricted isometry condition (RIP) [34], and results in the unstable solutions: with small perturbations added on the training sample $\boldsymbol{Y}$, the solutions of $\boldsymbol{X}$ deviate significantly.

2.3.3 Regularized SimCO

The main idea of Regularized SimCO lies in the use of an additive penalty term to avoid singularity. Consider the objective function $f_\mu\left(\tilde{\boldsymbol{D}}\right)$ in (2.7),

$$\begin{aligned} f_\mu\left(\tilde{\boldsymbol{D}}\right) &= \min_{\boldsymbol{X}\in\mathcal{X}_\Omega} \|\boldsymbol{D}\boldsymbol{X} - \mathcal{P}(\boldsymbol{S})\|_F^2 + \mu\|\boldsymbol{X}\|_F^2 , \\ &= \min_{\boldsymbol{X}\in\mathcal{X}_\Omega} \left\| \begin{bmatrix} \mathcal{P}(\boldsymbol{S}) \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \boldsymbol{D} \\ \sqrt{\mu}\boldsymbol{I} \end{bmatrix} \boldsymbol{X} \right\|_F^2 . \end{aligned} \tag{2.15}$$

As long as $\mu \neq 0$ ($\mu > 0$ in our case), the block $\mu\boldsymbol{I}$ guarantees the full column rank of $\tilde{\boldsymbol{D}} = \left[\boldsymbol{D}^T \ \mu\boldsymbol{I}\right]^T$. Therefore, with the modified objective function $f_\mu\left(\tilde{\boldsymbol{D}}\right)$, there is

[2] An illustration: take $\boldsymbol{Y}$, $\boldsymbol{D}$, $\boldsymbol{X}$ as scalars. If $\boldsymbol{Y} \neq 0$, there exists a singular point at $\boldsymbol{D} = 0$ on $f(\boldsymbol{D}) = \min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2$, where $\boldsymbol{X}$ can be assigned as any real number.

no singular point so that gradient descent methods will only converge to stationary points.

This regularization technique is also applicable to MOD [16]. It is verified that this technique effectively mitigates the occurrence of ill-conditioned dictionary although at the same time some stationary points might be generated. To alleviate this problem, one can decrease gradually the regularization parameter μ during the optimization process [16]. In the end μ will decrease to zero. Nevertheless, it is still not guaranteed to converge to a global minimum. The explicit example constructed in [32] shows a failure of the Regularized SimCO. As a result, another method to address the singularity issue is introduced below.

2.3.4 Smoothed SimCO

Also aiming at handling the singularity issue, Smoothed SimCO [33] is to remove the singularity effect by adding multiplicative functions. The intuition is explained as follows. Write $f(\boldsymbol{D})$ into a summation of atomic functions

$$\begin{aligned} f(\boldsymbol{D}) &= \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2 \\ &= \sum_i \|\boldsymbol{Y}_{:,i} - \boldsymbol{D}_i \boldsymbol{X}_{\boldsymbol{\Omega}(:,i),i}\|_2^2 \\ &= \sum_i f_i(\boldsymbol{D}_i), \end{aligned} \tag{2.16}$$

where each $f_i(\boldsymbol{D}_i)$ is termed as an atomic function and $\boldsymbol{D}_i$ is defined in Definition 1. Let $\mathcal{I}$ be the index set corresponding to the $\boldsymbol{D}_i$'s of full column rank. Define an indicator function $\mathcal{X}_{\mathcal{I}}$ s.t. $\mathcal{X}_{\mathcal{I}}(i) = 1$ if $i \in \mathcal{I}$ and $\mathcal{X}_{\mathcal{I}}(i) = 0$ if $i \in \mathcal{I}^c$. Use $\mathcal{X}_{\mathcal{I}}(i)$ as a multiplicative modulation function and apply it to each $f_i(\boldsymbol{D}_i)$. Then one obtains

$$\bar{f}(\boldsymbol{D}) = \sum_i f_i(\boldsymbol{D}_i)\mathcal{X}_{\mathcal{I}}(i) = \sum_{i\in\mathcal{I}} f_i(\boldsymbol{D}_i). \tag{2.17}$$

This new function $\bar{f}$ is actually the best possible lower semi-continuous approximation of f and there is no new stationary point created.

Motivated from the above, we define

$$\tilde{f}(\boldsymbol{D}) = \sum_i f_i(\boldsymbol{D}_i) g\left(\lambda_{\min}(\boldsymbol{D}_i)\right), \tag{2.18}$$

where the shape of g is given in Fig. 2.1. The function g has the following properties: (1) $g(\lambda_{\min}) = 0$ for all $\lambda_{\min} \leq 0$; (2) $g(\lambda_{\min}) = 1$ for all $\lambda_{\min}(\boldsymbol{D}_i) > \delta > 0$, where δ is a threshold; (3) g is monotonically increasing; (4) g is second order differentiable. When using $\lambda_{\min}(\boldsymbol{D}_i)$ as the input variable for g and the positive threshold $\delta \to 0$,

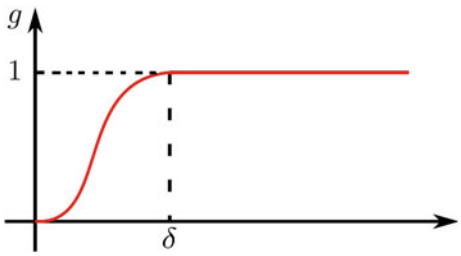


Fig. 2.1 A shape of function $g(\cdot)$

$\lambda_{\min}(\boldsymbol{D}_i)$ becomes an indicator function indicating whether $\boldsymbol{D}_i$ has a full column rank, i.e.,

$$\begin{cases} g(\lambda_{\min}(\boldsymbol{D}_i)) = 1 & \text{if } \boldsymbol{D}_i \text{ has full column rank;} \\ g(\lambda_{\min}(\boldsymbol{D}_i)) = 0 & \text{otherwise.} \end{cases}$$

The modulated objective function $\tilde{f}$ has several good properties, which do not exhibit in the regularized objective function (2.15). In particular, we have the following theorems.

Theorem 1 *Consider the smoothed objective function $\tilde{f}$ and the original objective function f defined in (2.18) and (2.16) ,respectively.*

1. *When $\delta > 0, \forall i, \tilde{f}(\boldsymbol{D})$ is continuous.*
2. *Consider the limit case where $\delta \to 0$ with $\delta > 0, \forall i$. The following statements hold:*
   a. *$\tilde{f}(\boldsymbol{D})$ and $f(\boldsymbol{D})$ differ only at the singular points.*
   b. *$\tilde{f}(\boldsymbol{D})$ is the best possible lower semi-continuous approximation of $f(\boldsymbol{D})$.*

Theorem 2 *Consider the smoothed objective function $\tilde{f}$ and the original objective function f defined in (2.18) and (2.16) ,respectively. For any $a \in \mathbb{R}$, define the lower level set $\mathcal{D}_f(a) = \{\boldsymbol{D} : f(\boldsymbol{D}) \leq a\}$. It is provable that when $\delta \to 0$, $\mathcal{D}_{\tilde{f}}(a)$ is the closure of $\mathcal{D}_f(a)$.*

In practice, we always choose a $\delta > 0$. The effect of a positive δ, roughly speaking, is to remove the barriers created by singular points, and replace them with "tunnels", whose widths are controlled by δ, to allow the optimization process to pass through. The smaller the δ is, the better $\tilde{f}$ approximates f, but the narrower the tunnels are, and the slower the convergence rate will be. As a result, the threshold δ should be properly chosen. A detailed discussion of choosing δ is presented in [32]. Compared with the choice of the parameter (μ) in the Regularized SimCO [16], the choice of the smoothing threshold δ is easier: one can simply choose a small $\delta > 0$ without decreasing it during the process.

As final remarks, Smoothed SimCO has several theoretical advantages over Regularized SimCO. However, the computations of $(\lambda_{\min}(\boldsymbol{D}_i))$'s introduce extra cost. The choice between these two methods will depend on the size of the problem under consideration.

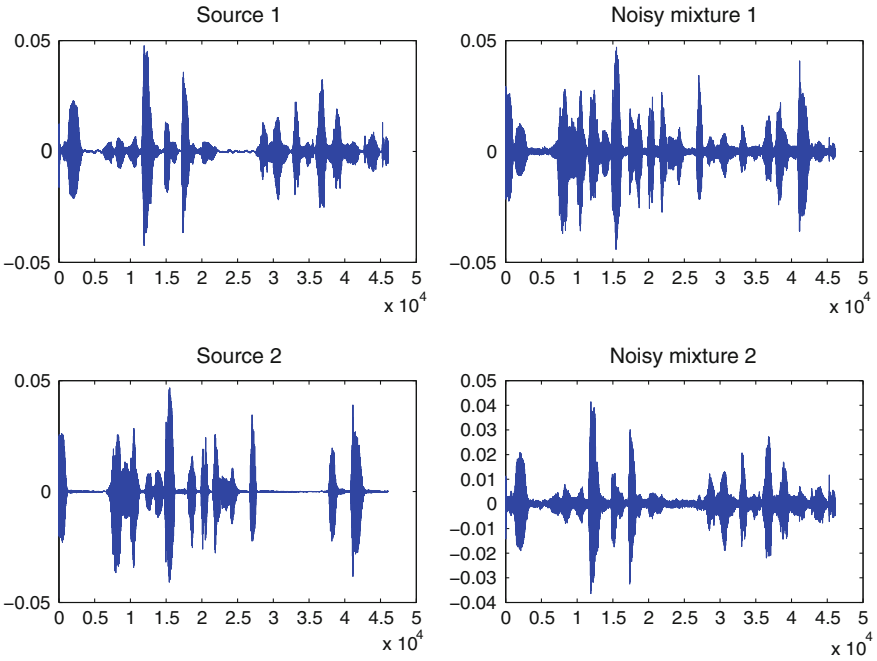


Fig. 2.2 Two speech sources and the corresponding noisy mixtures (20 dB Gaussian noise)

2.4 Algorithm Testing on Practical Applications

In this section, we present numerical results of the SparseBSS method compared with some other mainstream algorithms. We first focus on speech separation where an equal determined case will be considered. Then, we show an example for blind image separation, where we will consider an overdetermined case.

In the speech separation case two mixtures are used, which are the mixtures of two audio sources. Two male utterances in different languages are selected as the sources. The sources are mixed by a 2×2 random matrix $\boldsymbol{A}$ (with normalized columns). For the noisy case, a 20 dB Gaussian noise was added to the mixtures. See Fig. 2.2 for the sources and mixtures.

We compare SparseBSS with two benchmark algorithms including FastICA and QJADE [35]. The BSSEVAL toolbox [36] is used for the performance measurement. In particular, an estimated source $\hat{\boldsymbol{s}}$ is decomposed as $\hat{\boldsymbol{s}} = \boldsymbol{s}_{\text{target}} + \boldsymbol{e}_{\text{interf}} + \boldsymbol{e}_{\text{noise}} + \boldsymbol{e}_{\text{artif}}$, where $\boldsymbol{s}_{\text{target}}$ is the true source signal, $\boldsymbol{e}_{\text{interf}}$ denotes the interferences from other sources, $\boldsymbol{e}_{\text{noise}}$ represents the deformation caused by the noise, and $\boldsymbol{e}_{\text{artif}}$ includes all other artifacts introduced by the separation algorithm. Based on the decomposition, three performance criteria can be defined: the source-to-distortion ratio $\text{SDR} = 10 \log_{10} \frac{\|\boldsymbol{s}_{\text{target}}\|^2}{\|\boldsymbol{e}_{\text{interf}} + \boldsymbol{e}_{\text{noise}} + \boldsymbol{e}_{\text{artif}}\|^2}$, the source-to-artifact ratio

Table 2.1 Separation performance of the SparseBSS algorithm as compared to FastICA and QJADE

	△SDR	△SIR	△SAR
(a) The noiseless case			
QJADE	60.661	60.661	−1.560
FastICA	57.318	57.318	−0.272
SparseBSS	**69.835**	**69.835**	**1.379**
(b) The noisy case			
QJADE	7.453	58.324	−1.245
FastICA	7.138	40.789	−1.552
SparseBSS	**9.039**	**62.450**	**0.341**

The proposed SparseBSS algorithm performs better than the benchmark algorithms. Table 2.1a. For the same algorithm, the $\triangle SDR$ and $\triangle SIR$ are the same in noiseless case. The $\triangle SDRs$ and $\triangle SIRs$ for all the tested algorithms are large and similar, suggesting that all the compared algorithms perform very well. The artifact introduced by SparseBSS is small as its $\triangle SAR$ is positive. Table 2.1b. In the presence of noise with SNR = 20 dB, SparseBSS excels the other algorithms in $\triangle SDR$, $\triangle SIR$, and $\triangle SAR$. One interesting phenomenon is that the $\triangle SDRs$ are much smaller than those in the noiseless case, implying that the distortion introduced by the noise is trivial. However, SparseBSS still has better performance

$\text{SAR} = 10\log_{10}\frac{\|\boldsymbol{s}_{\text{target}}+\boldsymbol{e}_{\text{interf}}+\boldsymbol{e}_{\text{noise}}\|^2}{\|\boldsymbol{e}_{\text{artif}}\|^2}$, and the source-to-interference ratio $\text{SIR} = 10\log_{10}\frac{\|\boldsymbol{s}_{\text{target}}\|^2}{\|\boldsymbol{e}_{\text{interf}}\|^2}$. Among them, the SDR measures the overall performance (quality) of the algorithm, and the SIR focuses on the interference rejection. We investigate the gains of SDRs, SARs, and SIRs from the mixtures to the estimated sources. For example, $\triangle SDR = SDR_{out} - SDR_{in}$, where SDR_{out} is calculated from its definition and SDR_{in} is obtained by letting $\hat{\boldsymbol{s}} = \boldsymbol{Z}$ with the same equation. The results (in dB) are summarized in Table 2.1.

The selection of λ is an important practical issue since it is related to the noise level and largely affects the algorithm performance. From the optimization formulation (2.3), it is clear that with a fixed SNR, different choices of λ may give different separation performance. To show this, we use the estimation error $\left\|\boldsymbol{A}_{\text{true}} - \hat{\boldsymbol{A}}\right\|_F^2$ of the mixing matrix to measure the separation performance, where $\boldsymbol{A}_{\text{true}}$ and $\hat{\boldsymbol{A}}$ are the true and estimated mixing matrices, respectively. The simulation results are presented in Fig. 2.3. Consistent with the intuition, simulations suggest that the smaller the noise level, the larger the optimal value of λ. The results in Fig. 2.3 help in setting λ when the noise level is known a priori.

Next, we show an example for blind image separation, where we consider an overdetermined case. The mixed images are generated from two source images using a 4×2 full rank column normalized mixing matrix $\boldsymbol{A}$ with its elements generated randomly according to a Gaussian process. The mean squared errors (MSEs) are used to compare the reconstruction performance of the candidate algorithms when no noise is added. MSE is defined as $MSE = (1/N)\,\|\chi - \tilde{\chi}\|_F^2$, where χ is the source image and $\tilde{\chi}$ is the reconstructed image. The lower the MSE, the better the reconstruction

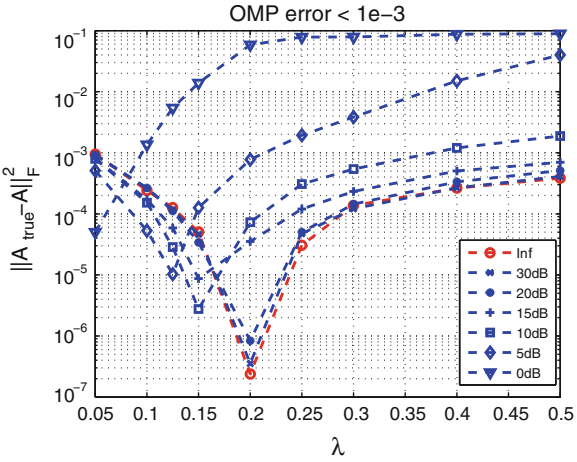


Fig. 2.3 Relation of the parameter λ to the estimation error of the mixing matrix under different noise levels. The signal-to-noise ratio (SNR) is defined as $\rho = 10\log_{10}\|\boldsymbol{AS}\|_F^2 / \|\boldsymbol{V}\|_F^2$ dB

Table 2.2 Achieved MSEs of the algorithms in a noiseless case

	FastICA	GMCA	BMMCA	SparseBSS
Lena	8.7489	4.3780	3.2631	**3.1346**
Boat	18.9269	**6.3662**	12.5973	6.6555

performance. Table 2.2 illustrates the results of four tested algorithms. For the noisy case, a Gaussian white noise is added to the four mixtures with $\sigma = 10$. We use the Peak Signal-to-Noise Ratio (PSNR) to measure the reconstruction quality, which is defined as, $PSNR = 20\log_{10}(MAX/\sqrt{MSE})$, where MAX indicates the maximum possible pixel value of the image, (e.g., $MAX = 255$ for a uint-8 image). Higher PSNR indicates better quality. The noisy observations are illustrated in Fig. 2.4b.[3]

Finally, we show another example of blind image separation to demonstrate the importance of the singularity-aware process. In this example, we use two classic images *Lena* and *Texture* as the source images (Fig. 2.6a). Four noiseless mixtures were generated from the sources. The separation results are shown in Fig. 2.6b and c. Note that images like *Texture* contain a lot of frequency components corresponding to a particular frequency. Hence, an initial dictionary with more codewords corresponding to the particular frequency may perform better for the estimation of these images. Motivated by this, in Fig. 2.6b the initial dictionary is generated from an overcomplete DCT dictionary, but contains more high frequency codewords. Such choice

[3] For the BMMCA test, a better performance was demonstrated in [14]. We point out that here a different true mixing matrix is used. And furthermore, in our tests the patches are taken with a 50 % overlap (by shifting 4 pixels from the current patch to the next) while in [14] the patches are taken by shifting only one pixel from the current patch to the next.

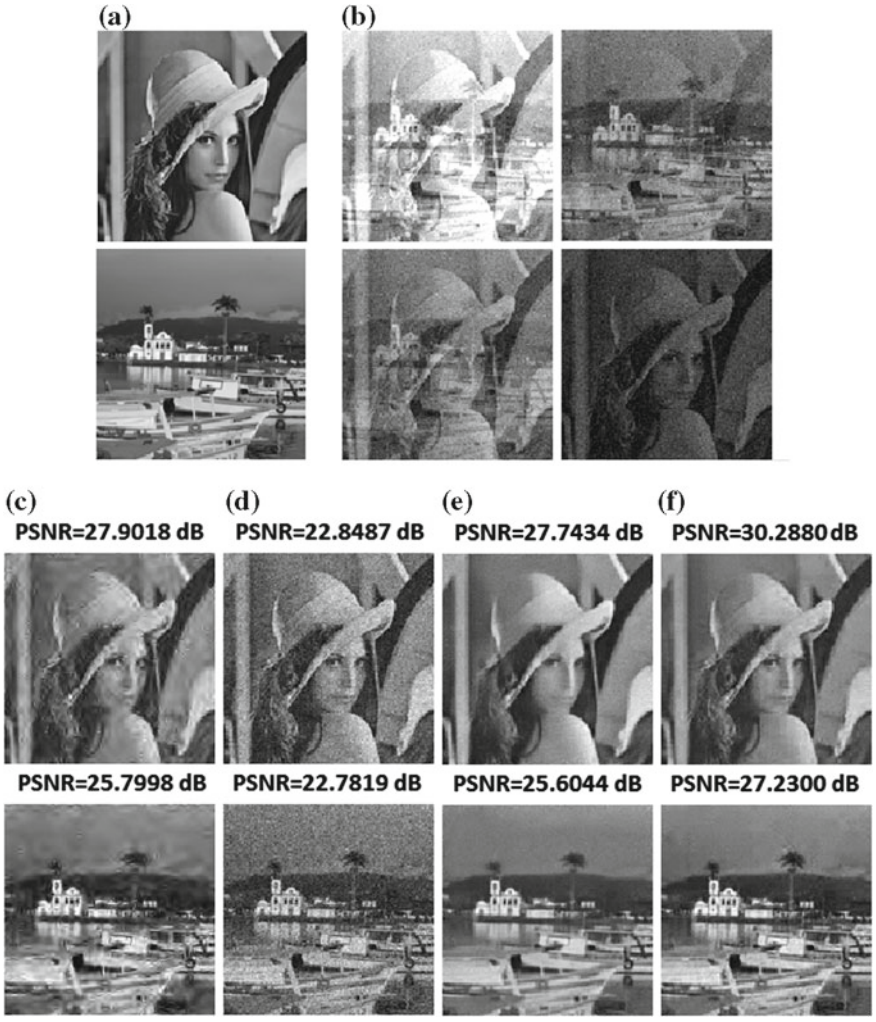


Fig. 2.4 Two classic images, *Lena* and *Boat* were selected as the source images, which are shown in (**a**). The mixtures are shown in (**b**). The separation results are shown in (**c**–**f**). We compared SparseBSS with other benchmark algorithms: FastICA [37], GMCA [10], and BMMCA [14]. We set the overlap percentage equal to 50 % for both BMMCA and SparseBSS. The recovered source images by the SparseBSS tend to be less blurred compared to the other three algorithms

can lead to better separation results. At the same time, the very similar dictionary codewords may introduce the risk of singularity issue (Fig. 2.5).

The major difference between Fig. 2.6b and c is that: in Fig. 2.6b the Regularized SimCO process ($\mu = 0.05$) is introduced, while in Fig. 2.6c there is no regularization term in the dictionary learning stage. As one can see from the numerical results, Fig. 2.6b performs much better than Fig. 2.6c. By checking the condition number

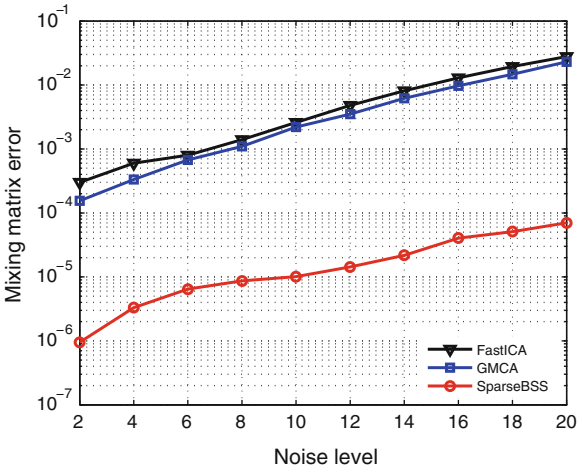


Fig. 2.5 Compare the performance of estimating the mixing matrix for all the methods in different noise standard deviation σs. In this experiment, σ varies from 2 to 20. The performance of GMCA is better than that of FastICA. The curve for BMMCA is not available as the setting for the parameters is too sophisticated and inconsistent for different σ to obtain a good result. SparseBSS outperforms the compared algorithms

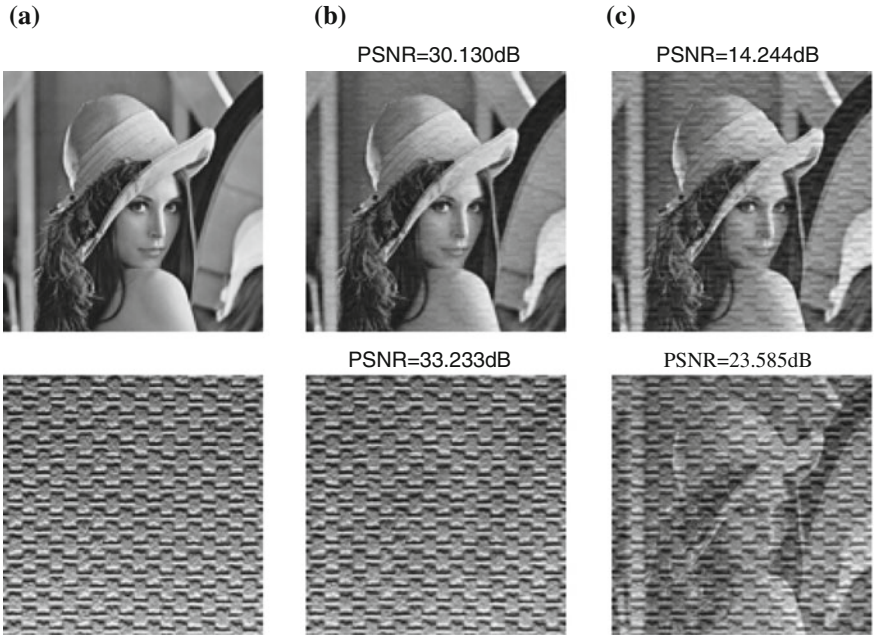


Fig. 2.6 The two source images *Lena* and *Texture* are shown in (**a**). The separation results are shown in (**b**) and (**c**). The comparison results demonstrate the importance of the singularity -aware process

when the regularized term is not introduced ($\mu = 0$), the value stays at a high level as expected (larger than 40 in this example). This confirms the necessity of considering the singularity issue in BSS and the effectiveness of the proposed singularity-aware approach.

2.5 Conclusions and Prospective Extensions

In conclusion, we briefly introduced a development of the blind source separation algorithms based on dictionary learning. In particular, we focus on the SparseBSS algorithm and the optimization procedures. The singularity issue might lead to the failure of these algorithms. At the same time there are still some open questions to be addressed.

In dictionary learning, it remains open how to find an optimum choice of the redundancy factor $\tau = d/n$ of the over-complete dictionary. A higher redundancy factor leads to either more sparse representation or more precise reconstruction. Moreover, one has to consider the computational capabilities when implementing the algorithms. From this point of view, it is better to keep the redundancy factor low. In the simulation, we have used a 64 by 256 dictionary, which gives the redundancy factor $\tau = 256/64 = 4$. This choice is empirical: the sparse representation results are good and the computational cost is limited. A rigorous analysis on the selection of τ is still missing.

The relation between the parameters λ, ϵ, and noise standard deviation σ is also worth investigating. As presented in the first experiment on blind audio separation, the relation between λ and σ is discussed when the error bound ϵ is fixed in the sparse coding stage. One can roughly estimate the value of the parameter λ assuming the noise level is known a priori. Similar investigation is undertaken in [14], where the authors claim that when $\lambda \approx \sigma/30$, the algorithm achieved similar reconstruction performance under various σ's. From another perspective, the error bound ϵ is proportional to the noise standard deviation. It turns out that once a well-approximated relation between ϵ and σ is obtained, one may get more precise estimation of parameter λ, rather than keeping ϵ fixed. This analysis, therefore, is counted as another open question.

References

1. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
2. Bell, J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Comput. **7**, 1129–1159 (1995)
3. Gaeta, M., Lacoume, J.L.: Source separaion without prior knowledge: the maximum likelihood solution. In: Proceedings of European Signal Processing Conference, pp. 621–624 (1990)

4. Belouchrani, A., Cardoso, J.F.: Maximum likelihood source separation for discrete sources. In: Proceedings of European Signal Processing Conference, pp. 768–771 (1994)
5. Bronstein, M., Zibulevsky, M., Zeevi, Y.: Sparse ica for blind separation of transmitted and reflected images. Int. J. Imaging Sci. Technol. **15**, 84–91 (2005)
6. Gribonval, R., Lesage, S.: A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges. In: Proceedings of European Symposium on Artificial, Neural Networks, pp. 323–330 (2006)
7. Jourjine, A., Rickard, S., Yilmaz, O.: Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2985–2988 (2000)
8. Starck, J., Elad, M., Donoho, D.: Redundant multiscale transforms and their application for morphological component analysis. Adv. Imaging Electron. Phys. **132**, 287–348 (2004)
9. Bobin, J., Moudden, Y., Starck, J., Elad, M.: Morphological diversity and source separation. IEEE Sign. Process. Lett. **13**(7), 409–412 (2006)
10. Bobin, J., Starck, J., Fadili, J., Moudden, Y.: Sparsity and morphological diversity in blind source separation. IEEE Trans. Image Process. **16**(11), 2662–2674 (2007)
11. Eladl, M.: Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, 1st edn. Springer, New York (2010). Incorporated
12. Peyré, G., Fadili, J., Starck, J.-L.: Learning adapted dictionaries for geometry and texture separation. In: Proceedings of SPIE Wavelet XII, vol. 6701, p. 67011T (2007)
13. Xu, T., Wang, W., Dai, W.: Sparse coding with adaptive dictionary learning for underdetermined blind speech separation. Speech Commun. **55**(3), 432–450 (2013)
14. Abolghasemi, V., Ferdowsi, S., Sanei, S.: Blind separation of image sources via adaptive dictionary learning. IEEE Trans. Image Process. **21**(6), 2921–2930 (2012)
15. Zhao, X., Xu, T., Zhou, G., Wang, W., Dai, W.: Joint image separation and dictionary learning. In: Accepted by 18th International Conference on Digital Signal Processing, Santorini, Greece (2013)
16. Dai, W., Xu, T., Wang, W.: Simultaneous codeword optimization (simco) for dictionary update and learning. IEEE Trans. Sign. Process. **60**(12), 6340–6353 (2012)
17. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, pp. 40–44 (1993)
18. Dai, W., Milenkovic, O.: Subspace pursuit for compressive sensing signal reconstruction. IEEE Trans. Inf. Theory **55**(5), 2230–2249 (2009)
19. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (1999)
20. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**, 303–353 (1999)
21. Hildebrand, F.B.: Advanced Calculus for Applications. Prentice-Hall, Upper Saddle River (1976)
22. Engan, K., Aase, S.O., Husoy, J.H.: Method of optimal directions for frame design. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 2443–2446 (1999)
23. Gorodnitsky, I.F., George, J.S., Rao, B.D.: Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm. Electroencephalogr. Clin. Neurophysiol. **95**, 231–251 (1995)
24. Engan, K., Skretting, K., Husoy, J.: Family of iterative is-based dictionary learning algorithms, ils-dla, for sparse signal representation. Digit. Sign. Process. **17**(1), 32–49 (2007)
25. Skretting, K., Engan, K.: Recursive least squares dictionary learning algorithm. IEEE Trans. Sign. Process. **58**(4), 2121–2130 (2010)
26. Aharon, M., Elad, M., Brucketein, A.: K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Sign. Process. **54**(11), 4311–4322 (2006)
27. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. **11**, 19–60 (2010)
28. Gribonval, R., Schnass, K.: Dictionary identification: sparse matrix-factorisation via l1-minimisation. CoRR, vol. abs/0904.4774 (2009)

29. Geng, Q., Wang, H., Wright, J.: On the local correctness of l1 minimization for dictionary learning. CoRR, vol. abs/1101.5672 (2011)
30. Jenatton, R., Gribonval, R., Bach, F.: Local stability and robustness of sparse dictionary learning in the presence of noise. CoRR (2012)
31. Yaghoobi, M., Blumensath, T., Davies, M.E.: Dictionary learning for sparse approximations with the majorization method. IEEE Trans. Sign. Process. **57**(6), 2178–2191 (2009)
32. Zhao, X., Zhou, G., Dai, W.: Dictionary learning: a singularity problem and how to handle it (in preparation)
33. Zhao, X., Zhou, G., Dai, W.: Smoothed SimCO for dictionary learning: handling the singularity issue. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (2013)
34. Candes, E., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory **51**(12), 4203–4215 (2005)
35. Cardoso, J.F., Souloumiac, A.: Blind beamforming for non-Gaussian signals. IEE Proc. **140**(6), 362–370 (1993)
36. Vincent, E., Gribonval, R., Fevotte, C.: Performance measurement in blind audio source separation. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1462–1469 (2006)
37. Hyvarinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Networks **10**(3), 626–634 (1999)

Chapter 3
Performance Study for Complex Independent Component Analysis

Benedikt Loesch and Bin Yang

Abstract The goal of independent component analysis (ICA) is to decompose observed signals into components as independent as possible. In linear instantaneous blind source separation, ICA is used to separate linear instantaneous mixtures of source signals into signals that are as close as possible to the original signals. In the estimation of the so-called demixing matrix one has to distinguish two different factors:

1. Variance of the estimated inverse mixing matrix in the noiseless case due to randomness of the sources.
2. Bias of the demixing matrix from the inverse mixing matrix:

This chapter studies both factors for circular and noncircular complex mixtures. It is important to note that the complex case is not directly equivalent to the real case of twice larger dimension. In the derivations, we aim to clearly show the connections and differences between the complex and real cases. In the first part of the chapter, we derive a closed-form expression for the CRB of the demixing matrix for instantaneous noncircular complex mixtures. We also study the CRB numerically for the family of noncircular complex generalized Gaussian distributions (GGD) and compare it to simulation results of several ICA estimators. In the second part, we consider a linear noisy noncircular complex mixing model and derive an analytic expression for the

Sections 3.1.1 and 3.2 of this chapter are based on our previous journal publication [35]. © 2013 IEEE. Reprinted, with permission, from Loesch and Yang [35].
Sections 3.3.1–3.3.3 of this chapter are based on our previous conference publication [34]. First published in the Proceedings of the 20th European Signal Processing Conference (EUSIPCO-2012) in 2012, published by EURASIP.

B. Loesch (✉) · B. Yang
Institute of Signal Processing and System Theory, University of Stuttgart, Stuttgart, Germany
e-mail: benedikt.loesch@iss.uni-stuttgart.de

B. Yang
e-mail: bin.yang@iss.uni-stuttgart.de

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_3,
© Springer-Verlag Berlin Heidelberg 2014

demixing matrix of ICA based on the Kullback-Leibler divergence (KLD). We show that for a wide range of both the shape parameter and the noncircularity index of the GGD, the signal-to-interference-plus-noise ratio (SINR) of KLD-based ICA is close to that of linear MMSE estimation. Furthermore, we show how to extend our derivations to the overdetermined case ($M > N$) with circular complex noise.

3.1 Introduction

The goal of independent component analysis (ICA) is to decompose observed signals into components as independent as possible. In linear instantaneous blind source separation, ICA is used to separate linear instaneous mixtures of source signals into signals which are as close as possible to the original signals. In the estimation of the so-called demixing matrix, one has to distinguish two different factors:

1. Variance of the estimated inverse mixing matrix in the noiseless case due to randomness of the sources. This variance can be lower bounded by the Cramér-Rao bound for ICA derived for the real case in [41, 45] and for the circular and noncircular complex case in [33, 35].
2. Bias of the demixing matrix from the inverse mixing matrix: As already noted in [16], the presence of noise leads to a bias of the demixing matrix from the inverse mixing matrix. Often a bias of an estimator is considered to be unwanted, but in the case of noisy ICA the bias of the demixing matrix from the inverse mixing matrix actually leads to a reduced noise level in the separated signals and hence it can be considered to be desired.

This chapter studies both factors for circular and noncircular complex mixtures. It is important to note that the complex case is not directly equivalent to the real case of twice larger dimension [19]. In the derivations, we aim to clearly show the connections and differences between the complex and real cases.

In many practical applications such as audio processing in frequency-domain or telecommunication, the signals are complex. While many publications focus on circular[1] complex signals (as traditionally assumed in signal processing), [4, 36, 44] provide a good overview of applications with noncircular complex signals and discuss how to properly deal with noncircularity. Many signals of practical interest are noncircular. Digital modulation schemes[2] usually produce noncircular complex baseband signals, since the symbol constellations in the complex plane are only rotationally symmetric for a discrete set of rotation angles but not any arbitrary real rotation angle as necessary for circularity [2]. Another source of noncircularity is an imbalance between the in-phase and quadrature (I/Q) components of communication signals. Noncircularity can also be exploited in feature extraction in electrocardio-

[1] See Sect. 3.1.1 for a definition.

[2] Examples of digital modulation schemes are phase shift keying (PSK), pulse amplitude modulation (PAM) or quadrature amplitude modulation (QAM).

grams (ECGs) and in the analysis of functional magnetic resonance imaging (fMRI) [4]. Moreover, the theory of noncircularity has found applications in acoustics and optics [44].

Although a large number of different algorithms for complex ICA have been proposed [7, 10, 14, 17, 18, 20, 29, 30, 38, 39], the CRB for the complex demixing matrix has only been derived recently in [33, 35]. General conditions regarding identifiability, uniqueness, and separability for complex ICA can be found in Eriksson and Koivunen [19]. Yeredor [48] provides a performance analysis for the strong uncorrelating transform (SUT) in terms of the interference-to-signal ratio matrix. However, since the SUT uses only second-order statistics, the results from [48] do not apply for ICA algorithms exploiting also the non-Gaussianity of the sources. As discussed in [3, 4], many ICA approaches exploiting non-Gaussianity of the sources are intimately related and can be studied under the umbrella of a maximum likelihood framework whose asymptotic performance reaches the CRB if the assumed distribution of the sources matches the true distribution.

The structure of the separation problem changes substantially if we account for additive noise. As discussed in [12], the mixing model is no longer equivariant and the likelihood contrast can no longer be assimilated to mutual information. Furthermore, the ML estimate of the source signals is no longer a linear function of the observations [23]. Source estimation from noisy mixtures can be classified into linear and nonlinear separation. In linear ICA, the presence of noise leads to a bias in the estimation of the mixing matrix. Douglas et al. [16] introduced measures to reduce this bias. Cardoso [8] showed that the performance of noisy source separation depends on the distribution of the sources, the signal-to-noise ratio (SNR) and the mixing matrix. Davies [13] showed for the real case that it is not meaningful to estimate both the mixing matrix and the full covariance matrix of the noise from the data. Koldovsky and Tichavsky [27, 28] drew parallels between linear minimum mean squared error (MMSE) estimation and ICA for the real data case. Up to now, closed-form expressions for the bias of the ICA solution in the complex case have not been derived except for the recent work of Loesch and Yang [34].

After a review of notation for complex-valued signals, complex ICA, and the CRB for a complex parameter vector in Sect. 3.1.1, we derive a closed-form expression for the CRB of the demixing matrix for instantaneous noncircular complex mixtures in Sect. 3.2. We first introduce the signal model and the assumptions in Sect. 3.2.1 and then derive the CRB for the complex demixing matrix in Sect. 3.2.2. Section 3.2.3 discusses the circular complex case and noncircular complex Gaussian case as two special cases of the CRB. In Sect. 3.2.4, we study the CRB numerically for family of noncircular complex generalized Gaussian distributions[3] (GGD) and compare it to simulation results of several ICA estimators.

In Sect. 3.3, we consider a linear noisy noncircular complex mixing model and derive an analytic expression for the demixing matrix of ICA based on the Kullback-Leibler divergence (KLD) [34]. This expression contains the circular complex and

[3] See Sect. 3.2.4 for a definition.

real case as special cases. The derivation is done using a perturbation analysis valid for small noise variance.[4] In Sect. 3.3.3, we show that for a wide range of both the shape parameter and the noncircularity index of the GGD, the signal-to-interference-plus-noise ratio (SINR) of KLD-based ICA is close to that of linear MMSE estimation. We also discuss the situations where the two solutions differ. Furthermore, we extend our derivations to the overdetermined case ($M > N$) with circular complex noise in Sect. 3.3.4.

Compared to our previous journal and conference publications [33–35], we extend the performance study to a larger number of ICA algorithms and extend the results for noisy mixtures to the overdetermined case.

3.1.1 Notations for Complex-Valued Signals

3.1.1.1 Complex Random Vector

Let $\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I \in \mathbb{C}^N$ be a complex random *column* vector with a corresponding probability density function (pdf) defined as the pdf $\tilde{p}(\mathbf{x}_R, \mathbf{x}_I)$ of the real part $\mathbf{x}_R$ and imaginary part $\mathbf{x}_I$ of $\mathbf{x}$. Since $\mathbf{x}_R = \frac{\mathbf{x}+\mathbf{x}^*}{2}$ and $\mathbf{x}_I = \frac{\mathbf{x}-\mathbf{x}^*}{2j}$, we can rewrite the pdf $\tilde{p}(\mathbf{x}_R, \mathbf{x}_I)$ as a function of $\mathbf{x}$ and $\mathbf{x}^*$, i.e., $\tilde{p}(\mathbf{x}_R, \mathbf{x}_I) = p(\mathbf{x}, \mathbf{x}^*)$. In the following, we will use $p(\mathbf{x})$ as a short notation for $p(\mathbf{x}, \mathbf{x}^*)$. The covariance matrix of $\mathbf{x}$ is

$$\mathrm{cov}(\mathbf{x}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^H\right]. \tag{3.1}$$

The pseudo-covariance matrix of $\mathbf{x}$ is

$$\mathrm{pcov}(\mathbf{x}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^T\right]. \tag{3.2}$$

$(\cdot)^T$ and $(\cdot)^H$ stand for transpose and complex conjugate transpose of a vector or matrix. The augmented covariance matrix of $\mathbf{x}$ is the covariance matrix of the augmented vector $\underline{\mathbf{x}} = \left[\mathbf{x}^T\ \mathbf{x}^H\right]^T$:

$$\mathrm{cov}(\underline{\mathbf{x}}) = \begin{bmatrix} \mathrm{cov}(\mathbf{x}) & \mathrm{pcov}(\mathbf{x}) \\ \mathrm{pcov}(\mathbf{x})^* & \mathrm{cov}(\mathbf{x})^* \end{bmatrix}. \tag{3.3}$$

$\mathbf{x}$ is called circular if $p(\mathbf{x}e^{j\alpha}) = p(\mathbf{x})\ \forall\alpha \in \mathbb{R}$. Otherwise it is called noncircular. Actually, for a random variable s, the circularity definition $p(se^{j\alpha}) = p(s)\ \forall\alpha \in \mathbb{R}$ is much stronger than the second-order circularity given by $\gamma = \mathrm{E}\left[s^2\right] = 0$. There exist noncircular complex random variables with $\gamma = 0$. For simplicity, however,

[4] For a large noise variance σ^2 the theoretical analysis cannot fully describe the behavior of KLD-based ICA since we only take into account terms of order σ^2. However, simulation results show that KLD-based ICA still performs similarly to linear MMSE estimation.

we use the second-order noncircularity index $\gamma = \mathrm{E}\left[s^2\right]$ to quantify noncircularity in the remainder of this chapter.

3.1.1.2 Complex Gradient

Let a complex *column* parameter vector $\boldsymbol{\theta} = \boldsymbol{\theta}_R + j\boldsymbol{\theta}_I \in \mathbb{C}^M$, its real and imaginary part $\boldsymbol{\theta}_R, \boldsymbol{\theta}_I \in \mathbb{R}^M$, and a real scalar cost function $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \tilde{f}(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I) \in \mathbb{R}$ be given. For ease of notation, we will also use the simplified notation $f(\boldsymbol{\theta})$ instead of $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$. Instead of calculating the derivatives of $\tilde{f}(\cdot)$ with respect to $\boldsymbol{\theta}_R$ and $\boldsymbol{\theta}_I$, the Wirtinger calculus computes the partial derivatives of $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, treating $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ as two independent variables [21, 44]. The complex gradient vectors $\nabla_{\boldsymbol{\theta}} f$ and $\nabla_{\boldsymbol{\theta}^*} f$ are given by

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} f &= \frac{\partial f}{\partial \boldsymbol{\theta}} = \frac{1}{2}\left(\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} - j\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I}\right) \in \mathbb{C}^M, \\ \nabla_{\boldsymbol{\theta}^*} f &= \frac{\partial f}{\partial \boldsymbol{\theta}^*} = \frac{1}{2}\left(\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} + j\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I}\right) \in \mathbb{C}^M. \end{aligned} \tag{3.4}$$

The stationary points of $f(\cdot)$ and $\tilde{f}(\cdot)$ are given by $\left(\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} = \mathbf{0} \text{ and } \frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I} = \mathbf{0}\right)$ or $\frac{\partial f}{\partial \boldsymbol{\theta}} = \mathbf{0}$ or $\frac{\partial f}{\partial \boldsymbol{\theta}^*} = \mathbf{0}$. The direction of steepest descent of a real function $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ is given by $-\frac{\partial f}{\partial \boldsymbol{\theta}^*}$ and not $-\frac{\partial f}{\partial \boldsymbol{\theta}}$ [6]. Note that $-\frac{\partial f}{\partial \boldsymbol{\theta}^*}$ is the direction of steepest descent for $\boldsymbol{\theta}$ and not for $\boldsymbol{\theta}^*$.

As long as the real and imaginary part of a complex function $g(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = g_R(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I) + j g_I(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I)$ are differentiable, the Wirtinger derivatives $\frac{\partial g}{\partial \boldsymbol{\theta}} = \frac{\partial g_R}{\partial \boldsymbol{\theta}} + j\frac{\partial g_I}{\partial \boldsymbol{\theta}}$ and $\frac{\partial g}{\partial \boldsymbol{\theta}^*} = \frac{\partial g_R}{\partial \boldsymbol{\theta}^*} + j\frac{\partial g_I}{\partial \boldsymbol{\theta}^*}$ also exist [43]. Furthermore, we note that the Wirtinger derivatives defined in (3.4) are also valid for partial derivatives of f with respect to a parameter *matrix* $\boldsymbol{\Theta}$. In this chapter, we will also use real derivatives which we denote as $(\cdot)'$ wherever possible.

3.1.1.3 Cramér-Rao Bound for a Complex Parameter Vector

Assume that L complex observations of $\mathbf{x}$ are iid with the pdf $p(\mathbf{x}; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is an N-dimensional complex parameter vector. In principle, it would be possible to derive the CRB for complex parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_R + j\boldsymbol{\theta}_I$ by considering the real CRB of the $2N$-dimensional real composite vector $\bar{\boldsymbol{\theta}} = \left[\boldsymbol{\theta}_R^T\ \boldsymbol{\theta}_I^T\right]^T$:

$$\mathrm{cov}(\bar{\boldsymbol{\theta}}) = \begin{bmatrix} \mathrm{cov}(\boldsymbol{\theta}_R) & \mathrm{cov}(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I) \\ \mathrm{cov}(\boldsymbol{\theta}_I, \boldsymbol{\theta}_R) & \mathrm{cov}(\boldsymbol{\theta}_I) \end{bmatrix} \geq L^{-1}\mathbf{J}_{\bar{\boldsymbol{\theta}}}^{-1}, \tag{3.5}$$

where $\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{y} - \mathrm{E}[\mathbf{y}])^T\right]$ denotes the cross-covariance matrix of $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{J}_{\bar{\boldsymbol{\theta}}} = \mathrm{E}\left[\left\{\nabla_{\bar{\boldsymbol{\theta}}} \ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})\right\}\left\{\nabla_{\bar{\boldsymbol{\theta}}} \ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})\right\}^T\right]$ is the real Fisher information matrix (FIM) and $\nabla_{\bar{\boldsymbol{\theta}}} \ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})$ is the real gradient vector of $\ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})$.

However, it is often more convenient to directly work with the complex CRB introduced in this section: The complex FIM of $\boldsymbol{\theta}$ is defined as

$$\mathscr{J}_{\boldsymbol{\theta}} = \begin{bmatrix} \mathscr{I}_{\boldsymbol{\theta}} & \mathscr{P}_{\boldsymbol{\theta}} \\ \mathscr{P}_{\boldsymbol{\theta}}^* & \mathscr{I}_{\boldsymbol{\theta}}^* \end{bmatrix}, \tag{3.6}$$

where $\mathscr{I}_{\boldsymbol{\theta}} = \mathrm{E}\left[\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\}\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\}^H\right]$ is called the information matrix and $\mathscr{P}_{\boldsymbol{\theta}} = \mathrm{E}\left[\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\}\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\}^T\right]$ the pseudo-information matrix.

The inverse of the FIM of $\boldsymbol{\theta}$ gives, under some regularity conditions, a lower bound for the augmented covariance matrix of an unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ [42, 44]

$$\begin{bmatrix} \mathrm{cov}(\hat{\boldsymbol{\theta}}) & \mathrm{pcov}(\hat{\boldsymbol{\theta}}) \\ \mathrm{pcov}(\hat{\boldsymbol{\theta}})^* & \mathrm{cov}(\hat{\boldsymbol{\theta}})^* \end{bmatrix} \geq \left(L\,\mathscr{J}_{\boldsymbol{\theta}}\right)^{-1} = L^{-1}\begin{bmatrix} \mathscr{I}_{\boldsymbol{\theta}} & \mathscr{P}_{\boldsymbol{\theta}} \\ \mathscr{P}_{\boldsymbol{\theta}}^* & \mathscr{I}_{\boldsymbol{\theta}}^* \end{bmatrix}^{-1}. \tag{3.7}$$

Note that the complex CRB (3.7) can be transformed to the corresponding real CRB (3.5) by using the transform $\mathbf{J}_{\bar{\boldsymbol{\theta}}}^{-1} = \frac{1}{2}\mathbf{T}\mathscr{J}_{\boldsymbol{\theta}}^{-1}\mathbf{T}^{-1}$ [42], where $\mathbf{T} = \frac{1}{2}\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ -j\mathbf{I} & j\mathbf{I} \end{bmatrix}$ is a $2N \times 2N$ matrix and $\mathbf{I}$ is the $N \times N$ identity matrix.

By using the block matrix inversion lemma [22], we get from (3.7)

$$\begin{bmatrix} \mathrm{cov}(\hat{\boldsymbol{\theta}}) & \mathrm{pcov}(\hat{\boldsymbol{\theta}}) \\ \mathrm{pcov}(\hat{\boldsymbol{\theta}})^* & \mathrm{cov}(\hat{\boldsymbol{\theta}})^* \end{bmatrix} \geq L^{-1}\begin{bmatrix} \mathbf{R}_{\boldsymbol{\theta}}^{-1} & -\mathbf{R}_{\boldsymbol{\theta}}^{-1}\mathbf{Q}_{\boldsymbol{\theta}} \\ -\mathbf{Q}_{\boldsymbol{\theta}}^H\mathbf{R}_{\boldsymbol{\theta}}^{-1} & \mathbf{R}_{\boldsymbol{\theta}}^{-*} \end{bmatrix} \tag{3.8}$$

with $\mathbf{R}_{\boldsymbol{\theta}} = \mathscr{I}_{\boldsymbol{\theta}} - \mathscr{P}_{\boldsymbol{\theta}}\mathscr{I}_{\boldsymbol{\theta}}^{-*}\mathscr{P}_{\boldsymbol{\theta}}^*$ and $\mathbf{Q}_{\boldsymbol{\theta}} = \mathscr{P}_{\boldsymbol{\theta}}\mathscr{I}_{\boldsymbol{\theta}}^{-*}$. $\mathbf{A}^{-*}$ is a short notation for $\left(\mathbf{A}^{-1}\right)^* = (\mathbf{A}^*)^{-1}$. Often we are interested in the bound for $\mathrm{cov}(\hat{\boldsymbol{\theta}})$ only, which can be obtained from (3.8) as

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1} = L^{-1}(\mathscr{I}_{\boldsymbol{\theta}} - \mathscr{P}_{\boldsymbol{\theta}}\mathscr{I}_{\boldsymbol{\theta}}^{-*}\mathscr{P}_{\boldsymbol{\theta}}^*)^{-1}. \tag{3.9}$$

Note that (3.9) gives a bound solely on the covariance matrix of an unbiased estimator. If an estimator reaches that bound, i.e., $\mathrm{cov}(\hat{\boldsymbol{\theta}}) = L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1}$, it does not imply that it also reaches the general CRB defined in (3.7). Only if the pseudo-information matrix $\mathscr{P}_{\boldsymbol{\theta}}$ vanishes, $\mathrm{cov}(\hat{\boldsymbol{\theta}}) = L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1}$ implies that $\hat{\boldsymbol{\theta}}$ reaches the CRB (3.7).

Sometimes, we are interested in introducing constraints on some or all of the complex parameters. The constrained CRB can be derived by following the steps in either [42] or [24]. If the unconstrained Fisher information matrix is singular, only the constrained CRB from [24] can be applied.

3.2 Cramér-Rao Bound for Complex ICA

For the performance analysis of ICA algorithms, it is useful to have a lower bound for the covariance matrix of estimators of the demixing matrix $\mathbf{W}$. The Cramér-Rao bound (CRB) is a lower bound on the covariance matrix of any unbiased estimator of a parameter vector. A closed-form expression for the CRB of the demixing matrix for real instantaneous ICA has been derived recently in [41, 45] which we summarized in Appendix 1. However, in many practical applications such as audio processing in frequency-domain or telecommunication, the signals are complex and hence we need the CRB for complex ICA.

3.2.1 Signal Model and Assumptions

Throughout this section, we assume an instantaneous complex linear square noiseless mixing model

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{3.10}$$

where $\mathbf{x} \in \mathbb{C}^N$ are N linear combinations of the N source signals $\mathbf{s} \in \mathbb{C}^N$. We make the following assumptions:

A1. The mixing matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ is deterministic and invertible.

A2. $\mathbf{s} = [s_1, \ldots, s_N]^T \in \mathbb{C}^N$ are N independent random variables with zero mean, unit variance $\mathrm{E}\left[|s_i|^2\right] = 1$ and second-order noncircularity index $\gamma_i = E[s_i^2] \in [0, 1]$.[5] Since $\gamma_i \in \mathbb{R}$, the real and imaginary part of s_i are uncorrelated. $\gamma_i \neq 0$ if and only if the variances of the real and imaginary part of s_i differ.
The probability density functions (pdfs) $p_{s_i}(s_i)$ of different source signals s_i can be identical or different. $p_{s_i}(s_i)$ is continuously differentiable with respect to s_i and s_i^* in the sense of Wirtinger derivatives [46] which have been shortly reviewed in Sect. 3.1.1. All required expectations exist.

The task of ICA is to demix the signals $\mathbf{x}$ by a linear demixing matrix $\mathbf{W} \in \mathbb{C}^{N \times N}$

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} \tag{3.11}$$

such that $\mathbf{y}$ is "as close to $\mathbf{s}$" as possible according to some metric.

The ideal solution for $\mathbf{W}$ is $\mathbf{A}^{-1}$, neglecting scaling, phase, and permutation ambiguity [19]. If we know the pdfs $p_{s_i}(s_i)$ perfectly, there is no scaling ambiguity. Due to the "working" assumption $\gamma_i \in [0, 1]$ (see Appendix 2), there is no phase ambiguity for noncircular sources ($\gamma_i > 0$) [1, 37]. A phase ambiguity occurs only for circular sources ($\gamma_i = 0$). Noncircular sources s_i which do not comply with the assumption $\gamma_i \in [0, 1]$ can be transformed according to $s_i e^{j\alpha_i}$ such that $\gamma_i \in [0, 1]$.

[5] Due to the inherent scaling ambiguity between the mixing matrix $\mathbf{A}$ and the source signals $\mathbf{s}$, without loss of generality, we can scale $\mathbf{s}$ and accordingly $\mathbf{A}$ such that $\mathrm{E}\left[|s_i|^2\right] = 1$ and $\gamma_i \in [0, 1]$.

In general, a complex source signal s can be described by the following statistical properties:

- non-Gaussianity,
- noncircularity,
- nonwhiteness, i.e., $s(t_1)$ and $s(t_2)$ are dependent for different time instants $t_1 \neq t_2$,
- nonstationarity, i.e., the statistical properties of $s(t)$ change over time.

In this section, we focus on noncircular complex source signals with independent and identically distributed (iid) time samples. An extension to temporally non-iid sources, i.e., to incorporate nonstationarity and nonwhiteness of the sources, has been given in [35].

Two temporally iid sources can be separated by ICA

- if at least one of the two sources is non-Gaussian or
- if both sources are Gaussian but differ in noncircularity [19].

3.2.2 Derivation of the Cramér-Rao Bound

We form the parameter vector

$$\boldsymbol{\theta} = \mathrm{vec}(\mathbf{W}^T) = [\mathbf{w}_1^T, \ldots, \mathbf{w}_N^T]^T \in \mathbb{C}^{N^2} \tag{3.12}$$

where $\mathbf{w}_i^T$ denotes the i-th row vector of $\mathbf{W}$. The $\mathrm{vec}(\cdot)$ operator stacks the columns of its argument into one long column vector. Given the pdfs $p_{s_i}(s_i)$ of the complex source signals s_i and the complex linear transform $\mathbf{x} = \mathbf{A}\mathbf{s}$, it is easy to derive the pdf of $\mathbf{x}$ as $p(\mathbf{x};\boldsymbol{\theta}) = |\det(\mathbf{W})|^2 \prod_{i=1}^N p_{s_i}(\mathbf{w}_i^T\mathbf{x})$. Here, in the derivation of the CRB, $\mathbf{W}$ is a short notation for $\mathbf{A}^{-1}$ and not the demixing matrix which would contain permutation, scaling, and phase ambiguity. By using matrix derivatives [2, 3, 21], we obtain

$$\frac{\partial}{\partial \mathbf{W}^H} \ln p(\mathbf{x};\boldsymbol{\theta}) = \mathbf{A}^* - \mathbf{x}^*\boldsymbol{\varphi}^T(\mathbf{W}\mathbf{x}) = \mathbf{A}^*(\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s}))^* \tag{3.13}$$

where $\boldsymbol{\varphi}(\mathbf{s}) = [\varphi_1(s_1), \ldots, \varphi_N(s_N)]^T$ and $\varphi_i(s_i)$ is defined as

$$\varphi_i(s_i) = -\frac{\partial}{\partial s_i^*} \ln p_{s_i}(s_i) = -\frac{1}{2}\frac{1}{p_{s_i}(s_i)}\left[\frac{\partial p_{s_i}(s_i)}{\partial s_{i,R}} + j\frac{\partial p_{s_i}(s_i)}{\partial s_{i,I}}\right]. \tag{3.14}$$

Since $\boldsymbol{\theta} = \mathrm{vec}(\mathbf{W}^T)$, we get

$$\nabla_{\boldsymbol{\theta}^*} \ln p_{\mathbf{x}}(\mathbf{x};\boldsymbol{\theta}) = \mathrm{vec}\left(\frac{\partial}{\partial \mathbf{W}^H} \ln p_{\mathbf{x}}(\mathbf{x};\boldsymbol{\theta})\right) = \left[(\mathbf{I}\otimes\mathbf{A})\,\mathrm{vec}\left(\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}(\mathbf{s})^H\right)\right]^*, \tag{3.15}$$

where $\mathbf{A} \otimes \mathbf{B} = \left[a_{ij}\mathbf{B}\right]$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. Hence, the information and pseudo-information matrix in (3.6) become

$$\begin{aligned}\mathscr{I}_{\boldsymbol{\theta}} &= \left((\mathbf{I}\otimes\mathbf{A})\mathrm{E}\left[\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H\right](\mathbf{I}\otimes\mathbf{A}^H)\right)^* \\ &= \left((\mathbf{I}\otimes\mathbf{A})\mathbf{M}_1(\mathbf{I}\otimes\mathbf{A}^H)\right)^*, \qquad (3.16)\\ \mathscr{P}_{\boldsymbol{\theta}} &= \left((\mathbf{I}\otimes\mathbf{A})\mathrm{E}\left[\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T\right](\mathbf{I}\otimes\mathbf{A}^T)\right)^* \\ &= \left((\mathbf{I}\otimes\mathbf{A})\mathbf{M}_2(\mathbf{I}\otimes\mathbf{A}^T)\right)^*, \qquad (3.17)\end{aligned}$$

where

$$\begin{aligned}\mathbf{M}_1 &= \mathrm{E}\left[\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H\right] \\ \text{and}\quad \mathbf{M}_2 &= \mathrm{E}\left[\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T\right]. \qquad (3.18)\end{aligned}$$

3.2.2.1 Induced CRB for the Gain Matrix G = WA

Since the so-called gain matrix $\mathbf{G} = \mathbf{WA}$ is a linear function of $\mathbf{W}$, the CRB for $\mathbf{W}$ "induces" a bound for $\mathbf{G}$. For simplicity, we first derive this induced CRB (iCRB) for $\mathbf{G} = \mathbf{WA} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ which is independent of the mixing matrix $\mathbf{A}$. Later we will obtain the CRB for $\mathbf{W}$ from the iCRB for $\mathbf{G}$.[6] When $\hat{\mathbf{G}} = \hat{\mathbf{W}}\mathbf{A}$ denotes the estimated gain matrix, the diagonal elements $\hat{G}_{ii}$ should be close to 1. They reflect how well we can estimate the power of each source signal. The off-diagonal elements $\hat{G}_{ij}$ should be close to 0 and reflect how well we can suppress interfering components. We define the corresponding stacked parameter vector

$$\boldsymbol{\vartheta} = \mathrm{vec}(\mathbf{G}^T) = \mathrm{vec}(\mathbf{A}^T\mathbf{W}^T) = (\mathbf{I}\otimes\mathbf{A}^T)\,\mathrm{vec}(\mathbf{W}^T) = (\mathbf{I}\otimes\mathbf{A}^T)\boldsymbol{\theta}. \qquad (3.19)$$

The covariance matrix of $\hat{\boldsymbol{\vartheta}} = \mathrm{vec}((\hat{\mathbf{W}}\mathbf{A})^T)$ is given by $\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) = (\mathbf{I}\otimes\mathbf{A}^T)\,\mathrm{cov}(\hat{\boldsymbol{\theta}})\,(\mathbf{I}\otimes\mathbf{A}^*)$ where $\hat{\boldsymbol{\theta}} = \mathrm{vec}(\hat{\mathbf{W}}^T)$. By combining (3.9) with (3.16) and (3.17), we get

$$\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) \geq L^{-1}(\mathbf{I}\otimes\mathbf{A}^T)(\mathscr{I}_{\boldsymbol{\theta}} - \mathscr{P}_{\boldsymbol{\theta}}\mathscr{I}_{\boldsymbol{\theta}}^{-*}\mathscr{P}_{\boldsymbol{\theta}}^{*})^{-1}(\mathbf{I}\otimes\mathbf{A}^*) = L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1} \qquad (3.20)$$

[6] Some authors [5, 15, 47] prefer the so-called expected interference-to-source ratio (ISR) matrix whose elements $\overline{\mathrm{ISR}}_{ij}$ are defined (for $i \neq j$ and unit variance sources) as $\overline{\mathrm{ISR}}_{ij} = \mathrm{E}\left[\frac{|G_{ij}|^2}{|G_{ii}|^2}\right]$, where G_{ii} denotes the diagonal elements and G_{ij} the off-diagonal elements of $\mathbf{G}$. To compute $\overline{\mathrm{ISR}}_{ij}$, usually $G_{ii} \approx 1$ (i.e., $\mathrm{var}(G_{ii}) \ll 1$) is assumed such that $\overline{\mathrm{ISR}}_{ij} \approx \mathrm{var}(G_{ij})$. In this section, we do not use the ISR matrix but instead directly derive the iCRB for $\mathbf{G}$.

with

$$\mathbf{R}_{\vartheta} = (\mathbf{M}_1 - \mathbf{M}_2 \mathbf{M}_1^{-*} \mathbf{M}_2^*)^*. \tag{3.21}$$

As shown in [35], $\mathbf{R}_{\vartheta}$ can be calculated as

$$\mathbf{R}_{\vartheta} = \sum_{i=1}^{N} d_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} a_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} b_{ij} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \tag{3.22}$$

where $d_i = \frac{(\eta_i - 1)^2 - |\beta_i - 1|^2}{\eta_i - 1} \in \mathbb{R}$, $a_{ij} = \kappa_i - \frac{|\gamma_j \xi_i|^2}{\kappa_i} - \frac{1}{\kappa_j} \in \mathbb{R}$ and $b_{ij} = -\left(\frac{\gamma_j^* \xi_i^*}{\kappa_i} + \frac{\gamma_i \xi_j}{\kappa_j}\right) = b_{ji}^* \in \mathbb{C}$. $\mathbf{L}_{ij}$ in (3.22) denotes an $N \times N$ matrix with a 1 at the (i, j) position and 0's elsewhere.

The parameters η_i, κ_i, β_i, ξ_i and γ_j are defined as

$$\eta_i = \mathrm{E}\left[|s_i|^2 |\varphi_i(s_i)|^2\right] > 1, \tag{3.23}$$

$$\kappa_i = \mathrm{E}\left[|\varphi_i(s_i)|^2\right] \geq 1, \tag{3.24}$$

$$\beta_i = \mathrm{E}\left[s_i^2 (\varphi_i^*(s_i))^2\right] \in \mathbb{C}, \tag{3.25}$$

$$\xi_i = \mathrm{E}\left[(\varphi_i^*(s_i))^2\right] \in \mathbb{C}, \tag{3.26}$$

$$\gamma_j = \mathrm{E}\left[s_j^2\right] \in \mathbb{R}. \tag{3.27}$$

Properties and other equivalent forms of these parameters can be found in the appendix of [35].

$\mathbf{R}_{\vartheta}$ has a special sparse structure which is illustrated below for $N = 3$:

$$\mathbf{R}_{\vartheta} = \left[\begin{array}{ccc:ccc:ccc} d_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{12} & 0 & b_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{13} & 0 & 0 & 0 & b_{13} & 0 & 0 \\ \hdashline 0 & b_{21} & 0 & a_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & d_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{23} & 0 & b_{23} & 0 \\ \hdashline 0 & 0 & b_{31} & 0 & 0 & 0 & a_{31} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & b_{32} & 0 & a_{32} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d_3 \end{array}\right] ss$$

The i-th diagonal element of the i-th diagonal block is $\mathbf{R}_{\vartheta}[i, i]_{(i,i)} = d_i$. The j-th diagonal element of the i-th diagonal block is $\mathbf{R}_{\vartheta}[i, i]_{(j,j)} = a_{ij}$. The (j, i) element of the $[i, j]$ block is $\mathbf{R}_{\vartheta}[i, j]_{(j,i)} = b_{ij}$. All remaining elements are 0. By permuting rows and columns of $\mathbf{R}_{\vartheta}$, it can be brought into a block-diagonal form. Then it

consists only of 1×1 blocks with elements d_i and 2×2 blocks $\begin{bmatrix} a_{ij} & b_{ij} \\ b_{ji} & a_{ji} \end{bmatrix}$. Hence, $\mathbf{R}_{\boldsymbol{\vartheta}}$ can be easily inverted resulting in a block-diagonal matrix where all 1×1 and 2×2 blocks are individually inverted as long as $d_i \neq 0$ and $a_{ij}a_{ji} - b_{ij}b_{ji} \neq 0$. The result is

$$\begin{aligned}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1} &= \sum_{i=1}^{N} \frac{1}{d_i} \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{a_{ji}}{a_{ij}a_{ji} - b_{ij}b_{ji}} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} \\ &\quad + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{-b_{ij}}{a_{ij}a_{ji} - b_{ij}b_{ji}} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \\ &= \sum_{i=1}^{N} f_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} g_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} h_{ij} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \end{aligned} \tag{3.28}$$

with

$$f_i = \frac{1}{d_i} = \frac{\eta_i - 1}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \tag{3.29}$$

$$g_{ij} = \frac{a_{ji}}{a_{ij}a_{ji} - b_{ij}b_{ji}} = \frac{\kappa_j(\kappa_i\kappa_j - 1) - |\gamma_i\xi_j|^2\kappa_i}{(\kappa_i\kappa_j - 1)^2 + |\gamma_i\gamma_j\xi_i\xi_j - 1|^2 - 1 - \kappa_i^2|\gamma_i\xi_j|^2 - \kappa_j^2|\gamma_j\xi_i|^2}, \tag{3.30}$$

$$h_{ij} = \frac{-b_{ij}}{a_{ij}a_{ji} - b_{ij}b_{ji}} = \frac{\gamma_j^*\xi_i^*\kappa_j + \gamma_i\xi_j\kappa_i}{(\kappa_i\kappa_j - 1)^2 + |\gamma_i\gamma_j\xi_i\xi_j - 1|^2 - 1 - \kappa_i^2|\gamma_i\xi_j|^2 - \kappa_j^2|\gamma_j\xi_i|^2}. \tag{3.31}$$

This means that $\mathrm{var}(\hat{G}_{ii})$ and $\mathrm{var}(\hat{G}_{ij})$ of $\hat{\mathbf{G}} = \hat{\mathbf{W}}\mathbf{A}$ are lower bounded by the (i, i)-th and (j, j)-th element of the (i, i)-th block of $L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$:

$$\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{L} f_i = \frac{1}{L} \frac{\eta_i - 1}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \tag{3.32}$$

$$\mathrm{var}(\hat{G}_{ij}) \geq \frac{1}{L} g_{ij} = \frac{1}{L} \frac{\kappa_j(\kappa_i\kappa_j - 1) - |\gamma_i\xi_j|^2\kappa_i}{(\kappa_i\kappa_j - 1)^2 + |\gamma_i\gamma_j\xi_i\xi_j - 1|^2 - 1 - \kappa_i^2|\gamma_i\xi_j|^2 - \kappa_j^2|\gamma_j\xi_i|^2}. \tag{3.33}$$

Note that $L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$ is the iCRB for $\boldsymbol{\vartheta}$ as in (3.9). In order to get the complete iCRB for $\begin{bmatrix} \boldsymbol{\vartheta} \\ \boldsymbol{\vartheta}^* \end{bmatrix}$ as in (3.8), we would also need $\mathbf{P}_{\boldsymbol{\vartheta}} = -\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}\mathbf{Q}_{\boldsymbol{\vartheta}} = -\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}\mathbf{M}_2^*\mathbf{M}_1^{-1}$.

It can be shown in a similar way

$$\mathbf{P}_{\boldsymbol{\vartheta}} = \sum_{i=1}^{N} \tilde{f}_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\tilde{g}_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \tilde{h}_{ij} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \right) \tag{3.34}$$

has the same form as $\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$ in (3.28) with

$$\tilde{f}_i = -\frac{f_i(\beta_i - 1)^*}{\eta_i - 1} = \frac{-(\beta_i - 1)^*}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \tag{3.35}$$

$$\tilde{g}_{ij} = -\frac{g_{ij}\gamma_j^* \xi_i^* + h_{ij}}{\kappa_i} = \frac{-(\kappa_j^2 - |\gamma_i \xi_j|^2)\gamma_j^* \xi_i^* \gamma_i \xi_j}{(\kappa_i \kappa_j - 1)^2 + |\gamma_i \gamma_j \xi_i \xi_j - 1|^2 - 1 - \kappa_i^2 |\gamma_i \xi_j|^2 - \kappa_j^2 |\gamma_j \xi_i|^2}, \tag{3.36}$$

$$\tilde{h}_{ij} = -\frac{g_{ij} + \gamma_i^* \xi_j^* h_{ij}}{\kappa_j} = \frac{1 - \kappa_i \kappa_j - (\gamma_j \xi_i \gamma_i \xi_j)^*}{(\kappa_i \kappa_j - 1)^2 + |\gamma_i \gamma_j \xi_i \xi_j - 1|^2 - 1 - \kappa_i^2 |\gamma_i \xi_j|^2 - \kappa_j^2 |\gamma_j \xi_i|^2}. \tag{3.37}$$

Note that according to (3.28) and (3.34) the iCRB for $\mathbf{G} = \mathbf{WA}$ has a nice decoupling property: the iCRB for G_{ii} only depends on the distribution of source i and the iCRB for G_{ij} only depends on the distribution of sources i and j and not on any other sources. Note that (3.32) and (3.33) cannot be used as a bound for real ICA since the FIM would be singular.

3.2.2.2 CRB for the Demixing Matrix W

Starting with the iCRB $L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$ for the stacked gain matrix $\boldsymbol{\vartheta} = \mathrm{vec}((\mathbf{WA})^T) = (\mathbf{I} \otimes \mathbf{A}^T) \cdot \mathrm{vec}(\mathbf{W}^T)$, it is now straightforward to derive the CRB for the stacked demixing matrix $\boldsymbol{\theta} = \mathrm{vec}(\mathbf{W}^T) = (\mathbf{I} \otimes \mathbf{A}^T)^{-1}\boldsymbol{\vartheta} = (\mathbf{I} \otimes \mathbf{W}^T)\boldsymbol{\vartheta}$. Since $\boldsymbol{\theta}$ is a linear function of $\boldsymbol{\vartheta}$,

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1} = L^{-1}(\mathbf{I} \otimes \mathbf{W}^T)\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}(\mathbf{I} \otimes \mathbf{W}^*) \tag{3.38}$$

holds for any unbiased estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$. See [35] for a more detailed expression of the CRB for $\mathbf{W}$.

3.2.3 Special Cases of the iCRB

In the previous section, we derived the iCRB for the gain matrix $\mathbf{G} = \mathbf{WA}$ for the general complex case. Below, we study some special cases of the iCRB.

3.2.3.1 Case A: All Sources Are Circular Complex

If all sources are circular complex, $\gamma_i = 0$ and $\beta_i = \eta_i$ [35]. Due to the phase ambiguity in circular complex ICA, the Fisher information for the diagonal elements G_{ii} is 0 and hence their iCRB does not exist. However, we can constrain G_{ii} to be real and derive the constrained CRB [24] for G_{ii}: As noted at the end of Sect. 3.2.2.1, G_{ii} is decoupled from G_{ij} and G_{jj} and hence it is sufficient to consider the constrained CRB for G_{ii} alone.

The constrained CRB for G_{ii} is given by [35]

$$\operatorname{var}(\hat{G}_{ii}) \geq \frac{1}{4L(\eta_i - 1)}. \tag{3.39}$$

The bound in (3.39) is valid for a phase-constrained G_{ii} such that $G_{ii} \in \mathbb{R}$. Equation (3.39) looks similar to the real case (3.90) except for a factor of 4 since η_i is defined using Wirtinger derivatives instead of real derivatives.

For $\operatorname{var}(\hat{G}_{ij})$ we get from (3.33)

$$\operatorname{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\kappa_j}{\kappa_i \kappa_j - 1}, \tag{3.40}$$

which again looks the same as in the real case (3.91). However, in the complex case, κ_i is defined using the Wirtinger derivative instead of real derivative. Furthermore, in the complex case κ measures the non-Gaussianity and noncircularity whereas in the real case κ measures only the non-Gaussianity.

If source i and j are both circular Gaussian, $\kappa_i = \kappa_j = 1$ and $\operatorname{var}(\hat{G}_{ij}) \to \infty$. This corresponds to the known fact that circular complex Gaussian sources cannot be separated by ICA.

3.2.3.2 Case B: All Sources are Noncircular Complex Gaussian

If all sources are noncircular Gaussian with different $\gamma_i \in \mathbb{R}$, it can be shown using the expressions for κ, ξ, η and β in (3.86)–(3.89) with $c = 1$ that

$$\operatorname{var}(\hat{G}_{ii}) \geq \frac{1}{L} \frac{1}{4\gamma_i^2}, \tag{3.41}$$

$$\begin{aligned} \operatorname{var}(\hat{G}_{ij}) &\geq \frac{1}{L} \frac{\gamma_i^2 + \gamma_j^2 - 2\gamma_i^2\gamma_j^2}{(\gamma_j^2 - \gamma_i^2)^2}(1 - \gamma_i^2) \\ &= \frac{1 - \gamma_i^2}{2L} \left[\frac{1 - \gamma_i\gamma_j}{(\gamma_i - \gamma_j)^2} + \frac{1 + \gamma_i\gamma_j}{(\gamma_i + \gamma_j)^2} \right]. \end{aligned} \tag{3.42}$$

Note that (3.42) is exactly the same result as obtained in [48] for the performance analysis of the SUT, i.e., our result shows that for noncircular Gaussian sources the SUT is indeed asymptotically optimal.

If all sources are noncircular Gaussian with identical γ_i, it can be shown that the iCRB for G_{ij} does not exist because $\gamma_j^2 - \gamma_i^2 \to 0$. This confirms the result obtained in [19, 29] which showed that ICA fails for two or more noncircular Gaussian signals with same γ_i.

3.2.4 Results for Generalized Gaussian Distribution

In order to verify the CRB derived in the previous sections, we now study complex ICA with noncircular complex generalized Gaussian distributed (GGD) sources. We choose this family of parametric pdf since it enables an analytical calculation of the CRB. The pdf of such a noncircular complex source s with zero mean, variance $\mathrm{E}[|s|^2] = 1$ and noncircularity index $\gamma \in [0, 1]$ can be written as [40]

$$p(s, s^*) = \frac{c\alpha \cdot \exp\left(-\left[\frac{\alpha/2}{\gamma^2-1}\left(\gamma s^2 + \gamma s^{*2} - 2ss^*\right)\right]^c\right)}{\pi\Gamma(1/c)(1-\gamma^2)^{1/2}},$$

where $\alpha = \Gamma(2/c)/\Gamma(1/c)$ and $\Gamma(\cdot)$ is the Gamma function. The shape parameter $c > 0$ varies the form of the pdf from super-Gaussian ($c < 1$) to sub-Gaussian ($c > 1$). For $c = 1$, the pdf is Gaussian. $0 \leq \gamma \leq 1$ controls the noncircularity of the pdf. The four parameters κ, β, η, ξ required to calculate the CRB are derived in Appendix 1. For the simulation study, we consider $N = 3$ sources with random mixing matrices $\mathbf{A}$. The real and imaginary part of all elements of $\mathbf{A}$ are independent and uniformly distributed in $[-1, 1]$. We conducted 100 experiments with different random matrices $\mathbf{A}$ and consider the following different ICA estimators[7]: Complex ML-ICA [29], complex ICA by entropy bound minimization (ICA-EBM) [30], noncircular complex ncFastICA (ncFastICA) [39], adaptable complex maximization of non-Gaussianity (ACMN) [38] and strong uncorrelating transform (SUT) [18, 44]. The properties and assumptions of the five different ICA algorithms are summarized in Table 3.1.

We want to compare the separation performance of ICA with respect to the iCRB and hence we define the performance metric as in [45]: After running an ICA algorithm, we correct the permutation ambiguity of the estimated demixing matrix and calculate the signal-to-interference ratio (SIR) averaged over all N sources:

$$\mathrm{SIR} = \frac{1}{N}\sum_{i=1}^{N} \frac{\mathrm{E}\left[|G_{ii}|^2\right]}{\sum_{j\neq i}\mathrm{E}\left[|G_{ij}|^2\right]} = \frac{1}{N}\sum_{i=1}^{N} \frac{1+\mathrm{var}(G_{ii})}{\sum_{j\neq i}\mathrm{var}(G_{ij})}. \tag{3.43}$$

[7] Note that many alternative ICA estimators such as [7, 10, 14, 17, 20] exist.

Table 3.1 Considered separation algorithms and their properties

Algorithm	Pdf model/φ	Separation principle
Complex ML-ICA (ML-ICA)	True pdf of the sources	Non-Gaussianity and noncircularity
Complex ICA by Entropy Bound Minimization (ICA-EBM)	Adaptive	Non-Gaussianity and noncircularity
Noncircular Complex FastICA (ncFastICA)	Fixed φ	Non-Gaussianity
Adaptable Complex Maximization of Non-Gaussianity (ACMN)	Adaptive φ	Non-Gaussianity
Strong Uncorrelating Transform (SUT)	–	noncircularity

In (3.43), the averaging over simulation trials takes place before taking the ratio.

In practice, the accuracy of the estimated demixing matrix depends not only on the optimization *cost function* but also on the optimization *algorithm* used to implement the estimator: In some rare cases, complex ML-ICA based on natural-gradient ascent converges to a local maximum of the likelihood and yields a lower SIR value than ICA-EBM. To overcome this problem, we initialized ML-ICA from the solution obtained by ICA-EBM which is close to the optimal solution.

3.2.4.1 Case A: All Sources Are Identically Distributed

First, we study the performance when all sources are identically distributed with the same shape parameter c and the same noncircularity index γ. Figure 3.1 shows the results: The SIR given by the iCRB increases with increasing non-Gaussianity ($c \to \infty$ or $c \to 0$). For $c \approx 1$, SIR is low since (nearly) Gaussian sources with the same noncircularity index γ cannot be separated by ICA. For $c \neq 1$, the SIR also increases with increasing noncircularity γ, but much slower since all sources have the same noncircularity γ. Clearly, all ICA algorithms work quite well except for $c \approx 1$ (Gaussian). ML-ICA (Fig. 3.1b) achieves the best performance followed by ICA-EBM (Fig. 3.1c) and ACMN (Fig. 3.1f). ncFastICA with kurtosis cost function achieves better performance for sub-Gaussian sources ($c > 1$) than for super-Gaussian sources ($c < 1$), whereas ncFastICA with square root (sqrt) nonlinearity works better for super-Gaussian sources than for sub-Gaussian sources. However, as also mentioned in [30], the square root nonlinearity leads overall to the best performance and hence we only consider ncFastICA with this nonlinearity in the following. As expected, SUT fails since it only uses noncircularity for separation and hence we do not show the results. The reason why ML-ICA outperforms ICA-EBM is that ML-ICA uses nonlinearities matched to the source distributions while ICA-EBM uses a linear combination of prespecified nonlinear functions. Note that ICA-EBM allows one to select nonlinearities for approximating the source entropy. Hence if prior knowledge about the source distributions is available, it can be incorporated into ICA-EBM thus improving its performance.

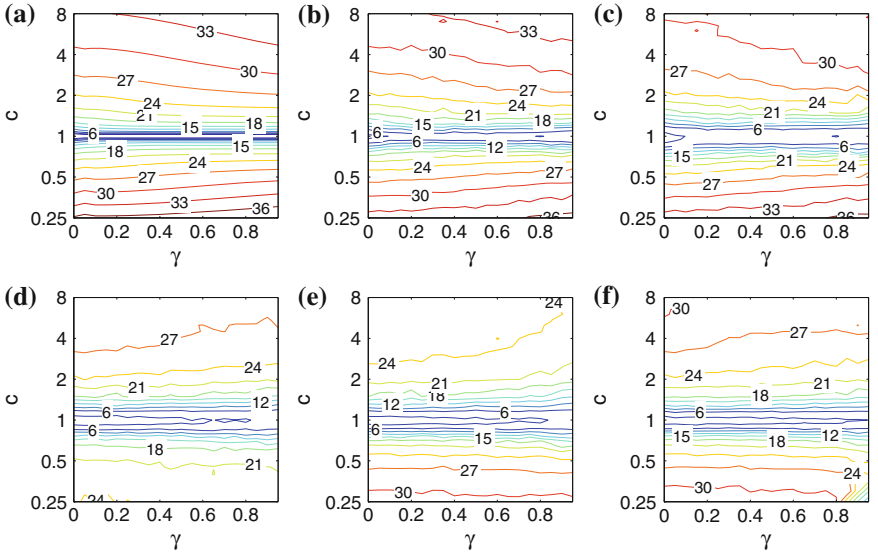


Fig. 3.1 Comparison of signal-to-interference ratio [dB] of different ICA estimators with CRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_i = c$ and $\gamma_i = \gamma$. iCRB (**a**), ML-ICA (**b**), ICA-EBM (**c**), ncFastICA kurtosis (**d**), ncFastICA sqrt (**e**), ACMN (**f**)

3.2.4.2 Case B: All Sources Have Different Shape Parameters and Different Noncircularities

Now we study the performance when the sources follow a GGD with different shape parameters $c_1 = 1, c_2 = c, c_3 = 1/c$ and different noncircularity indices $\gamma_i = (i-1)\Delta\gamma$. Figure 3.2 shows that the SIR given by the iCRB increases both with increasing non-Gaussianity of source 2 and 3 (i.e., $c < 1$) as well as increasing difference in noncircularity indices $\Delta\gamma$. ML-ICA achieves again the best performance, followed by ICA-EBM. The reason is again that ML-ICA uses for each source s_i a nonlinearity $\varphi_i(s_i)$ matched to its pdf $p_{s_i}(s_i)$ whereas the nonlinearities used in ICA-EBM are fixed a priori. Although ncFastICA and ACMN exploit the noncircularity of the sources to improve the convergence, their cost function only uses non-Gaussianity and not noncircularity. This is reflected clearly in Fig. 3.2 since performance for ncFastICA and ACMN is almost constant for different $\Delta\gamma$. SUT uses only noncircularity for separation, and hence performance is almost constant for different c. SUT can work quite well, as long as $\Delta\gamma$ is large enough. Only ML-ICA and ICA-EBM use both non-Gaussianity and noncircularity and hence the contour lines in Fig. 3.2b, c resemble those of the CRB Fig. 3.2a.

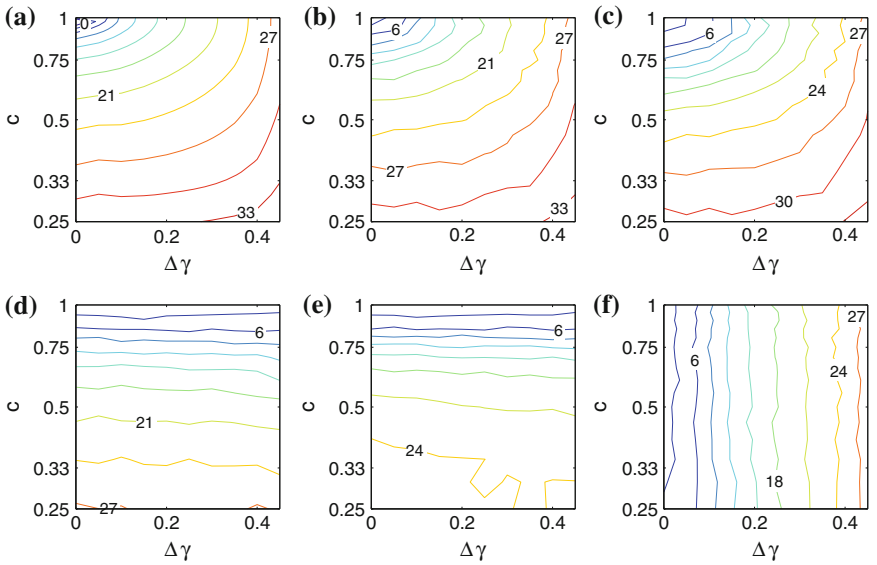


Fig. 3.2 Comparison of signal-to-interference ratio [dB] of different ICA estimators with iCRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_1 = 1, c_2 = c, c_3 = 1/c, \gamma_i = (i-1)\Delta\gamma$. iCRB (**a**), ML-ICA (**b**), ICA-EBM (**c**), ncFastICA sqrt (**d**), ACMN (**e**), SUT (**f**)

3.2.4.3 Performance as a Function of the Sample Size

Here, we study the performance as a function of sample size L. Clearly, Fig. 3.3 shows that for circular *non-Gaussian* sources and limited sample size L, ML-ICA achieves the best performance followed by ACMN and then ICA-EBM. The reason why ACMN outperforms ICA-EBM for circular sources could be the fact that ACMN needs to adapt less parameters since it uses only non-Gaussianity. As expected, SUT fails since it only uses noncircularity for separation. For circular super-Gaussian sources (Fig. 3.3a), ACMN and ncFastICA perform almost the same. For sub-Gaussian sources (Fig. 3.3b), the sqrt nonlinearity is sub-optimal as shown in the larger error of ncFastICA. Figure 3.4a shows results for noncircular *Gaussian sources* with distinct noncircularity indices: SUT and ML-ICA perform equally well since for noncircular Gaussian sources they are equivalent and asymptotically optimal. ICA-EBM approaches the performance of SUT and ML-ICA for large enough sample size. ncFastICA and ACMN which use only non-Gaussianity for separation fail. Figure 3.4b, c shows results for noncircular super-Gaussian ($c = 0.5$) and sub-Gaussian ($c = 6$) sources with distint noncircularity indices: With limited sample size, ML-ICA achieves again the best performance followed by ICA-EBM. For a large sample size ($L \geq 1000$) and a wide range of distributions including strongly super-Gaussian but excluding strongly sub-Gaussian sources, ICA-EBM comes close to the performance of ML-ICA, see Figs. 3.1, 3.3, 3.4. The reason for

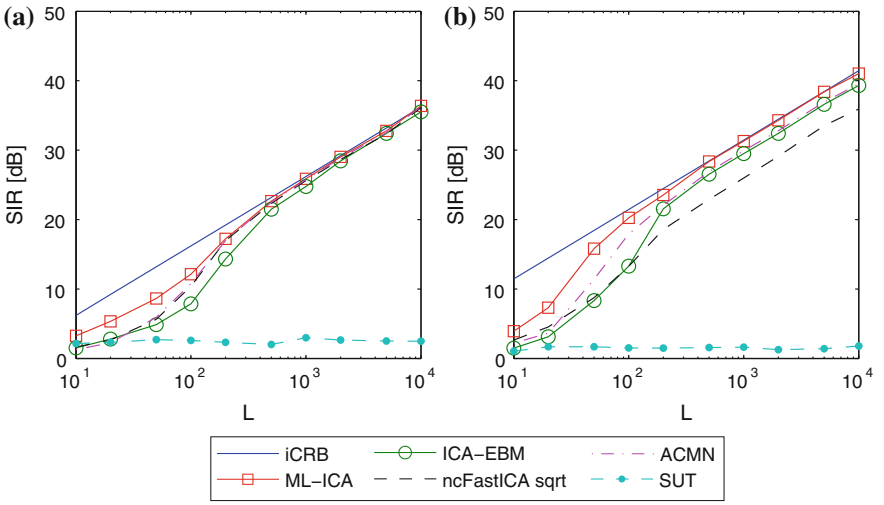


Fig. 3.3 Performance as a function of sample size L, circular GGD sources. $c = 0.5$ (**a**), $c = 6$ (**b**)

this behavior is that ML-ICA uses nonlinearities matched to the source distributions while ICA-EBM uses a linear combination of prespecified nonlinear functions. These could be extended to improve performance for strongly sub-Gaussian sources. The performance of ncFastICA and ACMN is quite far from that given by the iCRB since these two algorithms do not use noncircularity for separation. For signals with distinct noncircularity indices, SUT can achieve decent separation, but for strongly non-Gaussian signals the performance is quite far from that given by the iCRB (see also Fig. 3.2).

3.2.5 Conclusion

In this section, we have derived the CRB for the noncircular complex ICA problem with temporally iid sources. The induced CRB (iCRB) for the gain matrix, i.e., the demixing-mixing-matrix product, depends on the distribution of the sources through five parameters, which can be easily calculated. The derived bound is valid for the general noncircular complex case and contains the circular complex and the noncircular complex Gaussian case as two special cases. The iCRB reflects the phase ambiguity in circular complex ICA. In that case, we derived a constrained CRB for a phase-constrained demixing matrix. Simulation results using five ICA algorithms have shown that for sources following a noncircular complex generalized Gaussian distribution, some algorithms can achieve a signal-to-interference ratio (SIR) close to that of the CRB. Among the studied algorithms, complex ML-ICA and ICA-EBM perform best. The complex ML-ICA algorithm, which uses for each source a nonlinearity matched to its pdf, outperforms ICA-EBM especially for small

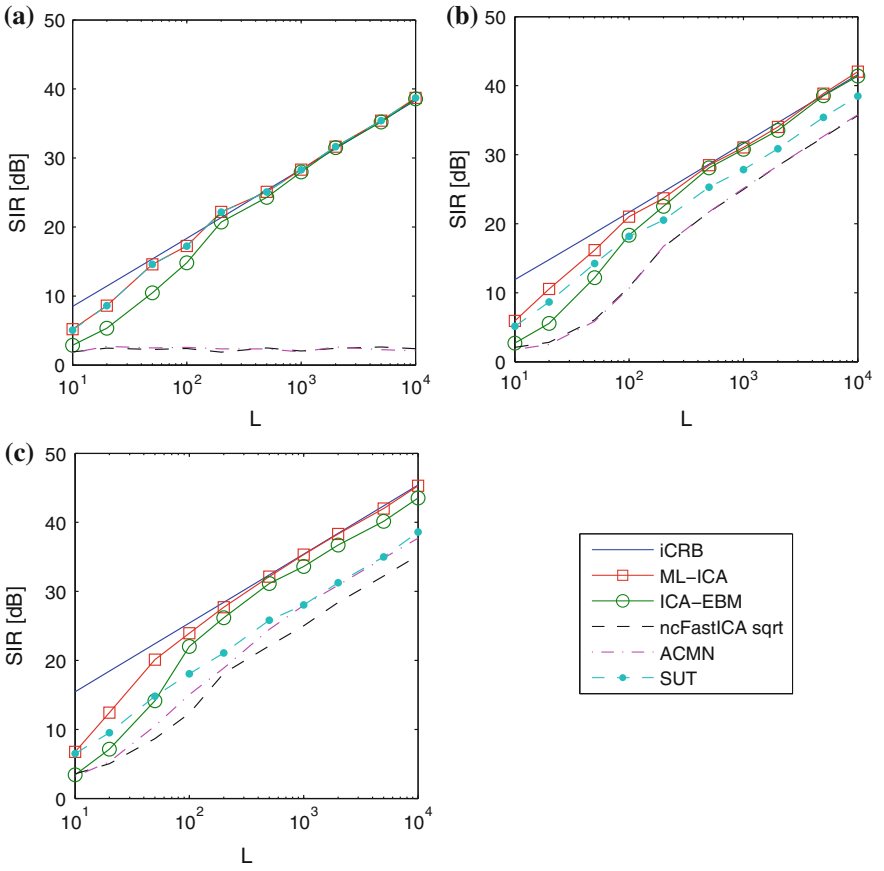


Fig. 3.4 Performance as a function of sample size L, noncircular GGD sources with $c_i = c$ and $\gamma_i = (i-1)\Delta\gamma$. $c = 1$, $\Delta\gamma = 0.45$ (**a**), $c = 0.5$, $\Delta\gamma = 0.45$ (**b**), $c = 6$, $\Delta\gamma = 0.45$ (**c**)

sample sizes. However, for ML-ICA the pdfs of the sources must be known whereas no such knowledge is required for ICA-EBM. Hence, for practical applications where the pdfs of the sources might be unknown ICA-EBM is an adequate algorithm whose performance comes quite close to the iCRB for large enough sample size L.

3.3 Solution of Linear Complex ICA in the Presence of Noise

In this section, we study the bias of the demixing matrix in linear noisy ICA from the inverse mixing matrix. We first derive the ICA solution for the general complex determined case. We then show how the circular complex case and the real case can be derived as special cases. Next, we verify the results using simulations. Finally, we extend our derivations to the overdetermined case with circular complex noise.

3.3.1 Signal Model and Assumptions

We assume the linear noisy mixing model

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{v} \tag{3.44}$$

where $\mathbf{x} \in \mathbb{C}^N$ are N linear combinations of N original signals $\mathbf{s} \in \mathbb{C}^N$ with additive noise $\mathbf{v} \in \mathbb{C}^N$. Here, all signals are modeled as temporally iid. In addition to the assumptions A1 and A2 (invertibility of mixing matrix $\mathbf{A}$ and assumptions about the source signals $\mathbf{s}$) defined in Sect. 3.2.1, we make the following two assumptions regarding the noise $\mathbf{v}$:

1. $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathbb{C}^N$ are N random variables with zero mean and the covariance matrix $E[\mathbf{v}\mathbf{v}^H] = \sigma^2 \mathbf{R}_\mathbf{v}$. $\sigma^2 = \frac{1}{N}\mathrm{tr}\left[\mathrm{E}(\mathbf{v}\mathbf{v}^H)\right]$ is the average variance of $\mathbf{v}$ and $\mathrm{tr}(\mathbf{R}_\mathbf{v}) = N$. $\bar{\mathbf{R}}_\mathbf{v} = \frac{1}{\sigma^2}\mathrm{E}[\mathbf{v}\mathbf{v}^T]$ is the normalized pseudo-covariance matrix. $\bar{\mathbf{R}}_\mathbf{v} = \mathbf{0}$ if $\mathbf{v}$ is circular complex. The pdf of $\mathbf{v}$ is arbitrary but assumed to be symmetric, i.e., $p_\mathbf{v}(\mathbf{v}) = p_\mathbf{v}(-\mathbf{v})$. This implies $\mathrm{E}(\prod_{i=1}^N v_i^{k_i}(v_i^*)^{\tilde{k}_i}) = 0$ for $\sum_{i=1}^N \left(k_i + \tilde{k}_i\right)$ odd.
2. $\mathbf{s}$ and $\mathbf{v}$ are independent.

The task of noisy linear ICA is to demix the signals $\mathbf{x}$ by a linear transform $\mathbf{W} \in \mathbb{C}^{N\times N}$

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} + \mathbf{W}\mathbf{v} \tag{3.45}$$

so that $\mathbf{y}$ is "as close to $\mathbf{s}$" as possible according to some metric.

3.3.2 KLD-Based ICA for Determined Case

We focus on the ICA solution based on the KLD

$$D_{\mathrm{KL}}(\mathbf{W}) = \int p_\mathbf{y}(\mathbf{y};\mathbf{W}) \ln \frac{p_\mathbf{y}(\mathbf{y};\mathbf{W})}{p_\mathbf{s}(\mathbf{y})} d\mathbf{y} \tag{3.46}$$

where $p_\mathbf{y}(\mathbf{y};\mathbf{W})$ is the pdf of $\mathbf{y}$. It depends on the pdf of observation $\mathbf{x}$, i.e., on the pdf of the original source signals $\mathbf{s}$ and noise $\mathbf{v}$, as well as on the demixing matrix $\mathbf{W}$. $p_\mathbf{s}(\mathbf{s}) = \prod_{i=1}^N p_{s_i}(s_i)$ is the assumed pdf of the original signals. We assume that we have perfect knowledge about the distribution of the original signals and $p_\mathbf{s}(\mathbf{s})$ is identical to the true pdf $p_\mathbf{s}^0(\mathbf{s})$ of $\mathbf{s}$. The KLD is known to have the following properties:

- $D_{\mathrm{KL}}(\mathbf{W}) \geq 0$ for any $p_\mathbf{y}(\mathbf{y};\mathbf{W})$ and $p_\mathbf{s}(\mathbf{y})$.
- $D_{\mathrm{KL}}(\mathbf{W}) = 0$ iff $p_\mathbf{y}(\mathbf{y};\mathbf{W}) = p_\mathbf{s}(\mathbf{y})$.

This means, minimizing the KLD with respect to $\mathbf{W}$ is equivalent to making the pdf of the demixed signals $\mathbf{y}$ as similar as possible to the pdf of the source signals $p_\mathbf{s}(\mathbf{s})$.

Since we assume $p_{\mathbf{s}}(\mathbf{s}) = \prod_{i=1}^{N} p_{s_i}(s_i)$, minimizing KLD corresponds to making (a) y_i as independent as possible and (b) y_i to have a pdf as close as possible to $p_{s_i}(s_i)$. This has been stated as "total mismatch = deviation from independence + marginal mismatch" by Cardoso in [9]. The ICA solution $\mathbf{W}_{\text{ICA}}$ for the demixing matrix based on KLD is given by

$$\mathbf{W}_{\text{ICA}} = \arg\min_{\mathbf{W}} D_{\text{KL}}(\mathbf{W}). \tag{3.47}$$

In the following, we will first derive the ICA solution for the general noncircular complex case. The circular complex case and the real case are discussed as two special cases.

3.3.2.1 General Noncircular Complex Case

The KLD cost function of a complex demixing matrix $\mathbf{W}$ is a function of the real and imaginary part of $\mathbf{W}$. Using the Wirtinger calculus (see [21, 44] and the summary in Sect. 3.1.1.2), we can also write it as a function of $\mathbf{W}$ and $\mathbf{W}^*$:

$$D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*) = \int p_{\mathbf{y}}(\mathbf{y}, \mathbf{y}^*; \mathbf{W}, \mathbf{W}^*) \ln \frac{p_{\mathbf{y}}(\mathbf{y}, \mathbf{y}^*; \mathbf{W}, \mathbf{W}^*)}{p_{\mathbf{s}}(\mathbf{y}, \mathbf{y}^*)} d\mathbf{y}. \tag{3.48}$$

The derivative $\frac{\partial D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)}{\partial \mathbf{W}^*}$ of the KLD cost function in (3.48) is given by [21]

$$\frac{\partial D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)}{\partial \mathbf{W}^*} = -\mathbf{W}^{-H} + \text{E}\left[\boldsymbol{\varphi}(\mathbf{y}, \mathbf{y}^*)\mathbf{x}^H\right], \tag{3.49}$$

where $\boldsymbol{\varphi}(\mathbf{y}, \mathbf{y}^*) = [\varphi_1(y_1, y_1^*), \ldots, \varphi_N(y_N, y_N^*)]^T$ and $\varphi_i(s_i, s_i^*) = -\frac{\partial \ln p_{s_i}(s_i, s_i^*)}{\partial s_i^*}$. The derivative $\frac{\partial}{\partial s^*}$ is also defined using the Wirtinger calculus.

A necessary condition for minimizing $D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)$ at $\mathbf{W} = \mathbf{W}_{\text{ICA}}$ is

$$\left.\frac{\partial D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)}{\partial \mathbf{W}^*}\right|_{\mathbf{W}=\mathbf{W}_{\text{ICA}}} \stackrel{!}{=} \mathbf{0} \quad \text{or} \quad \text{E}(\boldsymbol{\varphi}(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^H) \stackrel{!}{=} \mathbf{I} \tag{3.50}$$

with $\mathbf{y}_{\text{ICA}} = \mathbf{W}_{\text{ICA}}\mathbf{x} = \mathbf{W}_{\text{ICA}}\mathbf{A}\mathbf{s} + \mathbf{W}_{\text{ICA}}\mathbf{v} = \hat{\mathbf{y}} + \mathbf{W}_{\text{ICA}}\mathbf{v}$. An equivalent condition to $\text{E}(\boldsymbol{\varphi}(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^H) \stackrel{!}{=} \mathbf{I}$ in (3.50) is

$$\text{E}(\boldsymbol{\varphi}(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^H)^* = \text{E}(\boldsymbol{\varphi}^*(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^T) \stackrel{!}{=} \mathbf{I} \tag{3.51}$$

which we will use in the following to facilitate comparison with Sect. 3.2.

The properties of the ICA solution based on KLD are:

- $\mathbf{W}_{\text{ICA}} = \mathbf{A}^{-1}$ if $\sigma^2 = 0$ (no noise) and $p_{\mathbf{s}}(\mathbf{s}) = p_{\mathbf{s}}^0(\mathbf{s})$.
- To compute $\mathbf{W}_{\text{ICA}}$, we do not need to know $\mathbf{A}$ or $\mathbf{s}$, but the pdf $p_{\mathbf{s}}(\mathbf{s}) = \prod_{i=1}^{N} p_{s_i}(s_i)$ is required. All $p_{s_i}(s_i)$ must either be non-Gaussian or Gaussian with distinct noncircularity indices.

- No permutation ambiguity if $p_{s_i}(\cdot) \neq p_{s_j}(\cdot)\ \forall i \neq j$.
- There is no scaling ambiguity if $p_{s_i}(s_i) = p_i^0(s_i)$ is known $\forall i$. Only a phase ambiguity remains if $p_{s_i}(s_i)$ is circular.

As shown in Appendix 2, the ICA solution for the general noncircular complex case can be derived approximately using a two-step perturbation analysis for low noise and is given by

$$\mathbf{W}_{\text{ICA}} = (\mathbf{I} + \sigma^2\mathbf{C})\mathbf{A}^{-1} + \mathcal{O}(\sigma^4). \tag{3.52}$$

The elements of $\mathbf{C}$ can be obtained from (3.97) and (3.98). If $p_{\mathbf{s}}(\mathbf{s})$ is symmetric in the real or imaginary part of $\mathbf{s}$, they are given by (3.99) and (3.100).

For comparison, we consider the linear MMSE estimator

$$\mathbf{W}_{\text{MMSE}} = \mathbf{A}^H \left(\mathbf{A}\mathbf{A}^H + \sigma^2\mathbf{R}_{\mathbf{v}}\right)^{-1} \tag{3.53}$$

$$= \left[\mathbf{I} - \sigma^2\mathbf{R}_{-1}\right]\mathbf{A}^{-1} + \mathcal{O}(\sigma^4). \tag{3.54}$$

where the last line is a first-order Taylor series expansion in σ^2 and $\mathbf{R}_{-1} = \mathbf{A}^{-1}\mathbf{R}_{\mathbf{v}}\mathbf{A}^{-H}$. Comparing (3.54) with (3.52) we see that $\mathbf{W}_{\text{ICA}}$ and $\mathbf{W}_{\text{MMSE}}$ are similar if $\mathbf{C} \approx -\mathbf{R}_{-1}$.

3.3.2.2 Circular Complex Case

We assume now that the source signals $\mathbf{s}$ and the noise $\mathbf{v}$ are circular. Hence, both the noncircularity index of the sources γ and the pseudo-covariance matrix $\bar{\mathbf{R}}_{\mathbf{v}}$ are zero. As a consequence, (3.99) and (3.100) simplify to

$$\begin{aligned} C_{ii} &= -\frac{\kappa_i + \lambda_i}{1 + \rho_i + \delta_i}\left[\mathbf{R}_{-1}\right]_{ii} \in \mathbb{R}, \\ C_{ij} &= -\frac{\kappa_j(\kappa_i - 1)}{\kappa_i\kappa_j - 1}\left[\mathbf{R}_{-1}\right]_{ij} \in \mathbb{C} \quad (i \neq j). \end{aligned} \tag{3.55}$$

3.3.2.3 Real Case

For real signals and noise, we have

$$\gamma_i = 1, \quad \mathbf{R}_{\mathbf{v}} = \bar{\mathbf{R}}_{\mathbf{v}}. \tag{3.56}$$

In the derivation of $\mathbf{W}_{\text{ICA}}$ we have considered Taylor series expansions of $\boldsymbol{\varphi}(\mathbf{y})$ using Wirtinger derivatives. The Wirtinger derivatives $\partial/\partial s$ and $\partial/\partial s^*$ of $\varphi(s) \in \mathbb{R}$ are now identical (see (3.4)) and hence

$$\xi_i = \kappa_i, \quad \rho_i = \delta_i, \quad \lambda_i = \omega_i = \tau_i. \tag{3.57}$$

Furthermore, the Wirtinger derivatives of $\varphi(s) \in \mathbb{R}$ are identical to the real derivatives except for a factor of $\frac{1}{2}$ (see (3.4)). Hence it holds

$$\kappa_i = \frac{\mathring{\kappa}_i}{2}, \quad \rho_i = \frac{\mathring{\rho}_i}{2}, \quad \lambda_i = \frac{\mathring{\lambda}_i}{4}, \tag{3.58}$$

where $\mathring{\kappa}_i$, $\mathring{\rho}_i$ and $\mathring{\lambda}_i$ are defined using real derivatives of $\varphi(s)$, denoted by $(\cdot)'$:

$$\begin{aligned}
\mathring{\kappa}_i &= \mathrm{E}(\varphi_i'(s_i)) = \int \frac{d}{ds_i}\left(\frac{-p_{s_i}'(s_i)}{p_{s_i}(s_i)}\right) p_i^0(s_i) ds_i, \\
\mathring{\rho}_i &= \mathrm{E}(\varphi_i'(s_i) s_i^2) = \int \frac{d}{ds_i}\left(\frac{-p_{s_i}'(s_i)}{p_{s_i}(s_i)}\right) s_i^2 p_i^0(s_i) ds_i, \\
\mathring{\lambda}_i &= \mathrm{E}(\varphi_i''(s_i) s_i) = \int \frac{d^2}{ds_i^2}\left(\frac{-p_{s_i}'(s_i)}{p_{s_i}(s_i)}\right) s_i p_i^0(s_i) ds_i.
\end{aligned} \tag{3.59}$$

Using (3.56)–(3.58), we get from (3.99) and (3.100)

$$\begin{aligned}
C_{ii} &= -\frac{\mathring{\kappa}_i + \frac{1}{2}\mathring{\lambda}_i}{1 + \mathring{\rho}_i} \left[\mathbf{R}_{-1}\right]_{ii} = -M_{ii} \left[\mathbf{R}_{-1}\right]_{ii}, \\
C_{ij} &= -\frac{\mathring{\kappa}_j(\mathring{\kappa}_i - 1)}{\mathring{\kappa}_i \mathring{\kappa}_j - 1} \left[\mathbf{R}_{-1}\right]_{ij} = -M_{ij} \left[\mathbf{R}_{-1}\right]_{ij} \quad (i \neq j).
\end{aligned} \tag{3.60}$$

where $M_{ii} = \frac{\mathring{\kappa}_i + \mathring{\lambda}_i/2}{1+\mathring{\rho}_i}$ and $M_{ij} = \frac{\mathring{\kappa}_j(\mathring{\kappa}_i - 1)}{\mathring{\kappa}_i \mathring{\kappa}_j - 1}$. Note that (3.60) corresponds to the results in [32].

3.3.3 Results for Complex Generalized Gaussian Distribution

We study KLD-ICA for $N = 3$ sources with spatially white Gaussian noise with $E[\mathbf{v}\mathbf{v}^H] = \sigma^2 \mathbf{I}$ and the square mixing matrix $\mathbf{A} = [a_{mn}]$, where $a_{mn} = e^{-j\pi m \sin\theta_n}$ and $\theta_n = -60°, 0°, 60°$. As proposed in [26], we use the signal-to-interference-plus-noise ratio (SINR) to evaluate separation performance. For spatially uncorrelated noise, we compute the SINR for a given demixing matrix $\mathbf{W}$ by averaging the SINR for each source i

$$\mathrm{SINR} = \frac{1}{N} \sum_{i=1}^{N} \frac{|[\mathbf{WA}]_{ii}|^2}{\sum_{j \neq i} |[\mathbf{WA}]_{ij}|^2 + \sigma^2 \sum_j |\mathbf{W}_{ij}|^2}. \tag{3.61}$$

The term $|[\mathbf{WA}]_{ii}|^2$ reflects the power of the desired source i in the demixed signal y_i. The term $\sum_{j \neq i} |[\mathbf{WA}]_{ij}|^2$ corresponds to the power of the interfering signals

$j \neq i$ in the demixed signal y_i and $\sigma^2 \sum_j |\mathbf{W}_{ij}|^2$ is the noise power in the demixed signal y_i. For the remainder of this section, the signal-to-noise ratio (SNR) is defined as that before the mixing process and not at the sensors, i.e., SNR $= \frac{\mathrm{E}[s^2]}{\sigma^2} = \frac{1}{\sigma^2}$. It can be shown that among all linear demixing matrices $\mathbf{W}$, $\mathbf{W}_{\mathrm{MMSE}}$ from (3.53) is the one which maximizes the SINR [28]. We compare the SINR of the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$ from (3.52), the average SINR of $\hat{\mathbf{W}}_{\mathrm{ICA}}$ obtained from 100 runs of KLD-based ICA using L samples and the SINR of $\mathbf{W}_{\mathrm{MMSE}}$ from (3.53). The ICA algorithm is initialized with $\mathbf{W} = \mathbf{I}$ and performs gradient descent using the relative gradient [12], i.e., postmultiplies the gradient of KLD (3.49) by $\mathbf{W}^H\mathbf{W}$. We normalize each row of the relative gradient, resulting in an adaptive step size for each source. In the derivation of the theoretical solution $\mathbf{W}_{\mathrm{ICA}}$, we evaluated all expectations exactly. Hence $\mathbf{W}_{\mathrm{ICA}}$ only accounts for the bias from $\mathbf{A}^{-1}$ but not for estimation variance whereas $\hat{\mathbf{W}}_{\mathrm{ICA}}$ contains both factors.

In the following, all sources are GGD with the same shape parameter $c_i = c$. The noncircular complex GGD with zero mean and $E[|s|^2] = 1$ has already been introduced in Sect. 3.2.4. By integration in polar coordinates, it can be shown that

$$\kappa = \int \frac{\partial \varphi^*}{\partial s^*} p_{\mathbf{s}}^0(s) ds = \frac{c^2 \Gamma(2/c)}{(1-\gamma^2)\Gamma^2(1/c)}, \tag{3.62}$$

$$\delta = \int \frac{\partial \varphi^*}{\partial s^*} s s^* p_{\mathbf{s}}^0(s) = \frac{2c + (1-c)\gamma^2}{2(1-\gamma^2)}, \tag{3.63}$$

$$\rho = \int \frac{\partial \varphi^*}{\partial s} s^2 p_{\mathbf{s}}^0(s) ds = -\frac{2c - 2 + (1-3c)\gamma^2}{2(1-\gamma^2)}, \tag{3.64}$$

$$\xi = \int \frac{\partial \varphi^*}{\partial s} p_{\mathbf{s}}^0(s) ds = -\gamma\kappa, \tag{3.65}$$

$$\lambda = \int \frac{\partial^2 \varphi^*}{\partial s \partial s^*} s p_{\mathbf{s}}^0(s) ds = (c-1)\kappa, \tag{3.66}$$

$$\omega = \int \frac{\partial^2 \varphi^*}{(\partial s)^2} s p_{\mathbf{s}}^0(s) ds = -\frac{3}{2}(c-1)\gamma\kappa, \tag{3.67}$$

$$\tau = \int \frac{\partial^2 \varphi^*}{(\partial s^*)^2} s p_{\mathbf{s}}^0(s) ds = -\frac{1}{2}(c-1)\gamma\kappa. \tag{3.68}$$

Note that there exists a relationship between these parameters and the ones in the derivation of the CRB in Sect. 3.2: κ and ξ are identical. Using Corollary 2 from [35], we furthermore get

$$\delta = \eta - 1 \quad \text{and} \quad \rho = \beta - 2 \tag{3.69}$$

where $\eta = \mathrm{E}\left[|s|^2 |\varphi(s)|^2\right]$ and $\beta = \mathrm{E}\left[s^2 (\varphi^*(s))^2\right]$ have been defined in (3.23) and (3.25) in the previous section. These relationships hold not only for GGD but for all source distributions.

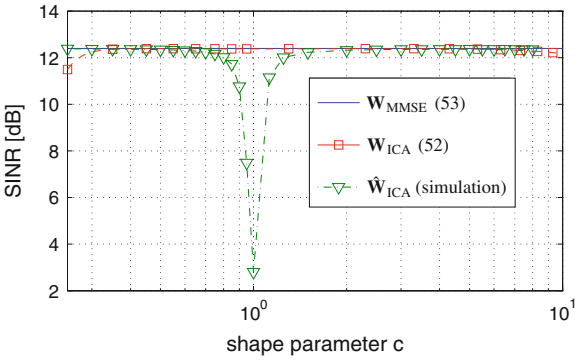


Fig. 3.5 SINR for circular complex GGD signals and circular complex noise, SNR = 10 dB, $L = 10^4$ samples

3.3.3.1 Circular Complex Case

For a circular complex GGD, $\gamma = 0$ and hence we get $\kappa = \frac{c^2\Gamma(2/c)}{\Gamma^2(1/c)}$, $\delta = c$, $\rho = c-1$, $\lambda = (c-1)\kappa$ and $\xi = \omega = \tau = 0$. Figure 3.5 shows that for a wide range of the shape parameter c, both the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$ and its estimate $\hat{\mathbf{W}}_{\mathrm{ICA}}$ obtained by running KLD-ICA using $L = 10^4$ samples achieve an SINR close to that of the MMSE solution $\mathbf{W}_{\mathrm{MMSE}}$. Note that for c close to 1, the SINR of the theoretical solution $\mathbf{W}_{\mathrm{ICA}}$ is not achievable in practice, since all sources become Gaussian and the CRB approaches infinity for $c \to 1$ (see Sect. 3.2 and [35]). Hence estimation of $\mathbf{W}$ becomes impossible. This is reflected in Fig. 3.5: The SINR for $\hat{\mathbf{W}}_{\mathrm{ICA}}$ estimated by KLD-ICA decreases for $c \to 1$.

Note that for strongly non-Gaussian sources ($c \ll 1$ or $c \gg 1$) the SINR of the theoretical solution $\mathbf{W}_{\mathrm{ICA}}$ might be smaller than that for $\hat{\mathbf{W}}_{\mathrm{ICA}}$ because $\mathbf{W}_{\mathrm{ICA}}$ is based on a Taylor series expansion up to order σ^2. For strongly non-Gaussian sources, higher-order terms become important. These are implicitly taken into account by $\hat{\mathbf{W}}_{\mathrm{ICA}}$ but not by $\mathbf{W}_{\mathrm{ICA}}$.

3.3.3.2 Noncircular Complex Case

First, we study the performance with circular noise, i.e., $\mathbf{R}_{\mathbf{v}} = \mathbf{I}$ and $\bar{\mathbf{R}}_{\mathbf{v}} = \mathbf{0}$, and SNR of 10 dB. The SINR of the MMSE solution $\mathbf{W}_{\mathrm{MMSE}}$ is 12.4 dB. Figure 3.6 shows that for a wide range of the shape parameter c and the noncircularity index γ, the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$ achieves an SINR close to that of MMSE. Comparing Fig. 3.6a, b, we note that the contour plot for the simulation using $L = 10^3$ samples differs from the contour plot for the theoretical ICA solution. One reason is that for noncircular sources with the same noncircularity index $\gamma_i = \gamma$, the estimation variance increases for $c \to 1$ (see Sect. 3.2 and [35]). Hence, in the simulation the SINR decreases in the vicinity of $c = 1$. Furthermore, the smaller sample size of

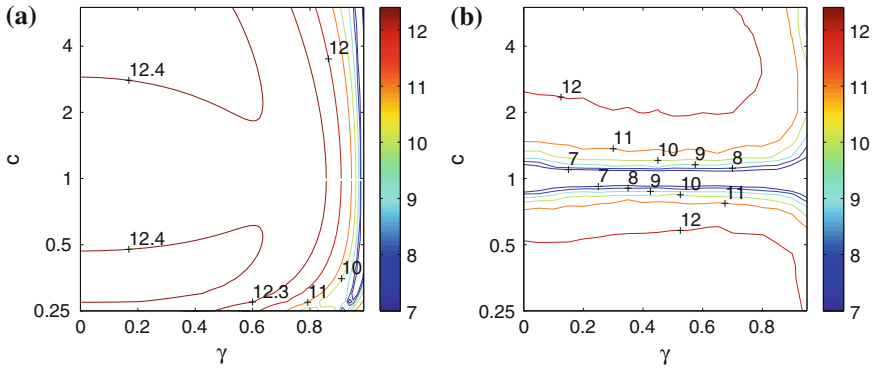


Fig. 3.6 SINR [dB] of ICA solution for noncircular complex GGD signals with $\gamma_i = \gamma$, circular complex noise and SNR = 10 dB. $\mathbf{W}_{\mathrm{ICA}}$ (52) (**a**), $\hat{\mathbf{W}}_{\mathrm{ICA}}$ (simulation, $L = 10^3$ samples) (**b**)

$L = 10^3$ leads to a larger variance of $\hat{\mathbf{W}}_{\mathrm{ICA}}$ which is not reflected in the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$. With a larger sample size the SINR of $\mathbf{W}_{\mathrm{ICA}}$ would be much closer to that of $\mathbf{W}_{\mathrm{MMSE}}$. However, Fig. 3.6b shows that even with a limited sample size KLD-ICA can still achieve SINR performance quite close to that of MMSE except for $c \approx 1$.

Now, we consider the case where sources are noncircular complex with $\gamma_1 = 0.5$, $\gamma_{2,3} = 0.5 \pm \Delta\gamma$ and the noise $\mathbf{v}$ is noncircular with $\mathbf{R}_{\mathbf{v}} = \mathbf{I}$ and $\bar{\mathbf{R}}_{\mathbf{v}} = 0.5 \cdot \mathbf{I}$, i.e., $\gamma_{\mathrm{noise}} = 0.5$. Figure 3.7 shows decreasing SINR values for $c \to 1$ and $\Delta\gamma \to 0$ since in that region $|\Re C_{ij}|$ in (3.100) becomes large if sources or noise are noncircular. However, except for this region, the SINR of the theoretical ICA solution (Fig. 3.7a) is still close to that of MMSE (12.4 dB). The form of the contour plot for the simulation (Fig. 3.7b) is similar to that of the theoretical solution but shows slightly lower SINR performance especially for $c \approx 1$ and small $\Delta\gamma$. This is again due to increasing estimation variance for $c \to 1$ and small $\Delta\gamma$ (see Sect. 3.2 and [35]). Nevertheless, the performance obtainable in simulations can still be considered good as long as c is not close to 1 or $\Delta\gamma$ is sufficiently large. Finally, we want to note that in Fig. 3.7 the decrease in SINR for strongly noncircular (large $\Delta\gamma$), non-Gaussian ($c \neq 1$) sources is caused by the noncircularity of the noise. The reason is that the MMSE (or maximum SINR) and the minimum KLD criterion yield different demixing matrices $\mathbf{W}$ for noncircular noise: As can be seen from (3.52), (3.97) and (3.98), $\mathbf{W}_{\mathrm{ICA}}$ depends both on the noncircularity of the sources ($\gamma_i \neq 0$) as well as on the noncircularity of the noise ($\bar{\mathbf{R}}_{\mathbf{v}} \neq \mathbf{0}$) whereas $\mathbf{W}_{\mathrm{MMSE}}$ from (3.54) only depends on the normal covariance matrix of the noise $\mathbf{R}_{\mathbf{v}}$. This is due to the different cost functions: Minimization of KLD makes the pdf of the demixed signals as similar to the assumed pdf of the sources as possible whereas MMSE minimizes the expected quadratic error between the demixed signals and the original sources. For circular noise, the difference between $\mathbf{W}_{\mathrm{ICA}}$ and $\mathbf{W}_{\mathrm{MMSE}}$ in terms of SINR is much smaller.

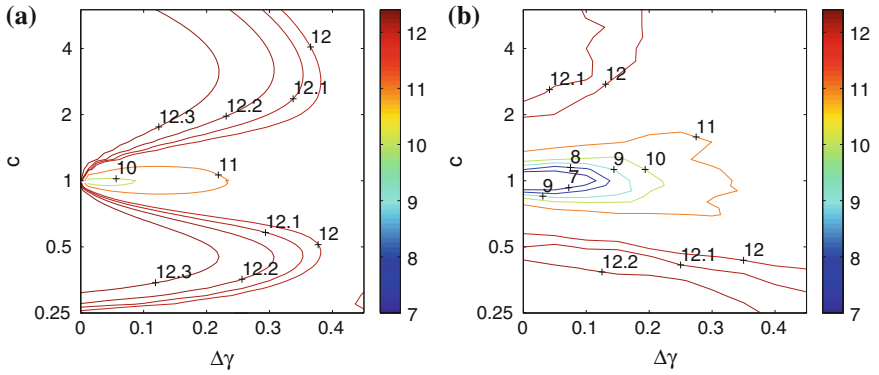


Fig. 3.7 SINR [dB] of ICA solution for noncircular complex GGD signals with $\gamma_1 = 0.5$, $\gamma_{2,3} = \gamma_1 \pm \Delta\gamma$, noncircular complex noise, and SNR $= 10$ dB. $\mathbf{W}_{\mathrm{ICA}}$ (52) (**a**), $\hat{\mathbf{W}}_{\mathrm{ICA}}$ (simulation, $L = 10^3$ samples) (**b**)

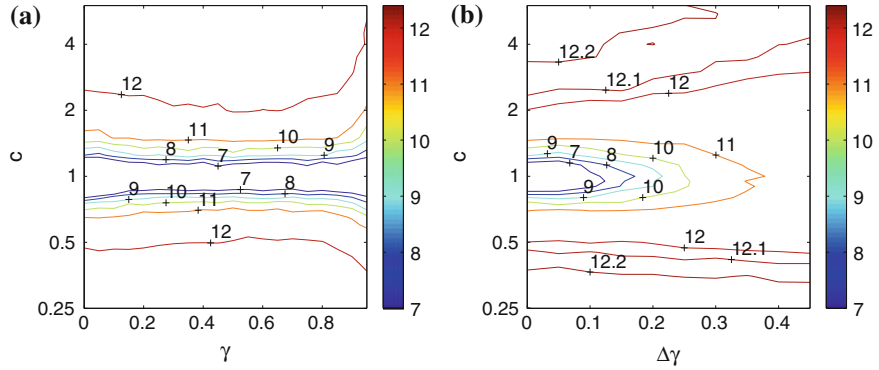


Fig. 3.8 SINR [dB] of ICA-EBM solution for noncircular complex GGD signals, $L = 10^3$ samples and SNR $= 10$ dB. $\gamma_i = \gamma$, circular complex noise(**a**), $\gamma_1 = 0.5$, $\gamma_{2,3} = \gamma_1 \pm \delta\gamma$ (**b**)

In summary, the results in this subsection have shown that

- in many cases the theoretical solution $\mathbf{W}_{\mathrm{ICA}}$ of KLD-ICA can achieve an SINR close to the optimum attainable by the MMSE demixing matrix $\mathbf{W}_{\mathrm{MMSE}}$.
- for sources following a GGD, $\hat{\mathbf{W}}_{\mathrm{ICA}}$ obtained by running KLD-ICA with a finite amount of samples L can achieve an SINR quite close to that of $\mathbf{W}_{\mathrm{MMSE}}$ except for (nearly) Gaussian sources with similar noncircularity indices.
- for strongly noncircular, non-Gaussian sources and noncircular noise, the minimization of the KLD and of the MSE yield different solutions.

Although we assumed that we perfectly know the distributions of the sources, other approaches such as ICA-EBM [30] exist which do not require such knowledge. As shown in Fig. 3.8, simulation results using ICA-EBM show similar SINR performance as KLD-ICA (see Figs. 3.6b, 3.7b).

3.3.4 Extension to Overdetermined Case

ICA algorithms for the overdetermined case have already been studied in a number of publications (see e.g., [11, 25, 49, 50]). In the overdetermined case, $\mathbf{x} \in \mathbb{C}^M$ with $M > N$. In the noiseless case we can select any N rows of $\mathbf{x}$ to perform ICA as long as the corresponding square mixing matrix $\widetilde{\mathbf{A}}$ is invertible. When we consider noisy mixtures, this does not hold since the information contained in the $M - N$ additional observations is useful to improve demixing. Hence, we need to consider the KLD for $M > N$. In this case, the demixing matrix $\mathbf{W}$ can be decomposed as $\mathbf{W} = \begin{bmatrix}\mathbf{W}_1 & \mathbf{W}_2\end{bmatrix}$, where $\mathbf{W}_1 \in \mathbb{C}^{N\times N}$ and $\mathbf{W}_2 \in \mathbb{C}^{N\times(M-N)}$. We define an auxiliary vector $\bar{\mathbf{y}} \in \mathbb{C}^M$:

$$\bar{\mathbf{y}} = \begin{bmatrix}\mathbf{W}_1 & \mathbf{W}_2 \\ \mathbf{0} & \mathbf{I}_{M-N}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1 \\ \mathbf{x}_2\end{bmatrix} = \breve{\mathbf{W}}\mathbf{x} = \begin{bmatrix}\mathbf{y} \\ \mathbf{x}_2\end{bmatrix} \tag{3.70}$$

Then we calculate $p_{\mathbf{y}}(\mathbf{y}; \mathbf{W})$ by

$$p_{\bar{\mathbf{y}}}(\bar{\mathbf{y}}; \breve{\mathbf{W}}) = \frac{1}{|\det(\breve{\mathbf{W}})|^2} p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{|\det(\mathbf{W}_1)|^2} p_{\mathbf{x}}(\mathbf{x}), \tag{3.71}$$

$$p_{\mathbf{y}}(\mathbf{y}; \mathbf{W}) = \frac{1}{|\det(\mathbf{W}_1)|^2} \int p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}_2. \tag{3.72}$$

since the linear transformation of a complex random vector yields $|\det(\mathbf{W})|^2$ instead of $|\det(\mathbf{W})|$ in the real case (see [4, 42]).

Using the above steps we obtain the modified KLD for $M > N$

$$D_{\mathrm{KL}}(\mathbf{W}) = -\ln|\det(\mathbf{W}_1)|^2 - \sum_{i=1}^{N} \mathrm{E}\left[\ln p_{s_i}(y_i)\right] + \text{const.} \tag{3.73}$$

instead of $D_{\mathrm{KL}}(\mathbf{W}) = -\ln|\det(\mathbf{W})|^2 - \sum_{i=1}^{N} \mathrm{E}\left[\ln p_{s_i}(y_i)\right] + \text{const.}$ for $M = N$.

To derive $\mathbf{W}_{\mathrm{ICA}}$ for $M > N$, we could now perform a similar Taylor series expansion as for $M = N$. However, it is more convenient to reduce the overdetermined case $M > N$ to the determined case by applying a linear transform to the data to condense all information about the source signals in the first N observations and by applying another transform to decorrelate the noise terms in the first N observations from those in the remaining $M - N$ observations. The result of these two transforms has a similar effect as a dimension reduction using principal component analysis (PCA) except that the correlation matrix of the observations is only block-diagonal instead of diagonal. To derive $\mathbf{W}_{\mathrm{ICA}}$, we can then combine the solution for the determined case with the linear transforms. Note that this approach is only used for the analysis of KLD-based ICA for the overdetermined case because it simplifies the theoretical derivation. In ICA applications, the transforms are done implicitly by the algorithm itself.

The first step of this procedure is to use the orthogonal transform $\mathbf{Q}$ defined by the decomposition $\mathbf{A} = \mathbf{Q}^H \begin{bmatrix} \bar{\mathbf{A}}_1 \\ \mathbf{0} \end{bmatrix}$ to condense all information about the source signals in the first N observations:

$$\begin{aligned} \bar{\mathbf{x}} &= \mathbf{Q}\mathbf{x} = \mathbf{Q}(\mathbf{A}\mathbf{s} + \mathbf{v}) \\ &= \mathbf{Q}\mathbf{Q}^H \begin{bmatrix} \bar{\mathbf{A}}_1 \\ \mathbf{0} \end{bmatrix} \mathbf{s} + \mathbf{Q}\mathbf{v} = \begin{bmatrix} \bar{\mathbf{A}}_1 \\ \mathbf{0} \end{bmatrix} \mathbf{s} + \bar{\mathbf{v}} = \begin{bmatrix} \bar{\mathbf{A}}_1 \mathbf{s} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{v}}_1 \\ \bar{\mathbf{v}}_2 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \end{bmatrix}, \end{aligned} \tag{3.74}$$

$$\mathbf{R}_{\bar{\mathbf{v}}} = \frac{1}{\sigma^2} \mathrm{E}\left(\bar{\mathbf{v}}\bar{\mathbf{v}}^H\right) = \mathbf{Q}\mathbf{R}_{\mathbf{v}}\mathbf{Q}^H = \begin{bmatrix} \mathbf{R}_{\bar{\mathbf{v}}_{11}} & \mathbf{R}_{\bar{\mathbf{v}}_{12}} \\ \mathbf{R}_{\bar{\mathbf{v}}_{21}} & \mathbf{R}_{\bar{\mathbf{v}}_{22}} \end{bmatrix}. \tag{3.75}$$

Note that $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$ may be correlated, i.e., $\bar{\mathbf{x}}_2 = \bar{\mathbf{v}}_2$ is useful for the processing of $\bar{\mathbf{x}}_1 = \bar{\mathbf{A}}_1\mathbf{s} + \bar{\mathbf{v}}_1$ to reduce the impact of $\bar{\mathbf{v}}_1$. Hence, we decorrelate the noise terms $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$ by a second transform $\mathbf{T} = \begin{bmatrix} \mathbf{I}_N & -\mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1} \\ \mathbf{0} & \mathbf{I}_{M-N} \end{bmatrix}$:

$$\widetilde{\mathbf{x}} = \mathbf{T}\bar{\mathbf{x}} = \mathbf{T}\begin{bmatrix} \bar{\mathbf{A}}_1 \mathbf{s} \\ \mathbf{0} \end{bmatrix} + \mathbf{T}\bar{\mathbf{v}} = \begin{bmatrix} \bar{\mathbf{A}}_1 \mathbf{s} \\ \mathbf{0} \end{bmatrix} + \widetilde{\mathbf{v}} = \begin{bmatrix} \bar{\mathbf{A}}_1 \mathbf{s} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{v}}_1 \\ \widetilde{\mathbf{v}}_2 \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{x}}_1 \\ \widetilde{\mathbf{x}}_2 \end{bmatrix}, \tag{3.76}$$

$$\mathbf{R}_{\widetilde{\mathbf{v}}} = \frac{1}{\sigma^2} \mathrm{E}\left(\widetilde{\mathbf{v}}\widetilde{\mathbf{v}}^H\right) = \begin{bmatrix} \mathbf{R}_{\bar{\mathbf{v}}_{11}} - \mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1}\mathbf{R}_{\bar{\mathbf{v}}_{21}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\bar{\mathbf{v}}_{22}} \end{bmatrix}. \tag{3.77}$$

$\widetilde{\mathbf{v}}_1$ and $\widetilde{\mathbf{v}}_2$ are now uncorrelated and $\widetilde{\mathbf{x}}_2 = \widetilde{\mathbf{v}}_2$ does not contain any second-order information useful for the processing of $\widetilde{\mathbf{x}}_1 = \bar{\mathbf{A}}_1\mathbf{s} + \widetilde{\mathbf{v}}_1$.

The separated signals $\mathbf{y}$ are now obtained by

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{Q}^H\mathbf{T}^{-1}\widetilde{\mathbf{x}} = \widetilde{\mathbf{W}}\widetilde{\mathbf{x}} = \begin{bmatrix} \widetilde{\mathbf{W}}_1 & \widetilde{\mathbf{W}}_2 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{x}}_1 \\ \widetilde{\mathbf{x}}_2 \end{bmatrix} = \widetilde{\mathbf{W}}_1\widetilde{\mathbf{x}}_1 + \widetilde{\mathbf{W}}_2\widetilde{\mathbf{x}}_2 = \mathbf{y}_1 + \mathbf{y}_2 \tag{3.78}$$

with $\widetilde{\mathbf{W}} = \mathbf{W}\mathbf{Q}^H\mathbf{T}^{-1}$. The noise-only contribution $\mathbf{y}_2 = \widetilde{\mathbf{W}}_2\widetilde{\mathbf{x}}_2 = \widetilde{\mathbf{W}}_2\widetilde{\mathbf{v}}_2$ to $\mathbf{y}$ is uncorrelated to $\mathbf{y}_1 = \widetilde{\mathbf{W}}_1\widetilde{\mathbf{x}}_1$. Hence, it is sufficient to consider the first N observations $\widetilde{\mathbf{x}}_1$ to derive the ICA solution for $\widetilde{\mathbf{W}}_1$.

Considering the KLD (3.73) for the transformed demixing model $\mathbf{y} = \begin{bmatrix} \widetilde{\mathbf{W}}_1 & \widetilde{\mathbf{W}}_2 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{x}}_1 \\ \widetilde{\mathbf{x}}_2 \end{bmatrix}$, we get

$$D_{\mathrm{KL}}(\widetilde{\mathbf{W}}) = -\ln|\det(\widetilde{\mathbf{W}}_1)| - \sum_{i=1}^{N} \mathrm{E}\left[\ln p_{s_i}(y_i)\right] + \text{const.} \tag{3.79}$$

with $\widetilde{\mathbf{W}} = \begin{bmatrix} \widetilde{\mathbf{W}}_1 & \widetilde{\mathbf{W}}_2 \end{bmatrix}$. The real derivatives of $D_{\mathrm{KL}}(\widetilde{\mathbf{W}})$ with respect to $\widetilde{\mathbf{W}}_1$ and $\widetilde{\mathbf{W}}_2$ are given by

$$\frac{\partial D_{\mathrm{KL}}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}_1} = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{x}}_1^H\right] - \widetilde{\mathbf{W}}_1^{-H} = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{y}}_1^H - \mathbf{I}\right]\widetilde{\mathbf{W}}_1^{-H} \stackrel{!}{=} \mathbf{0} \tag{3.80}$$

$$\frac{\partial D_{\mathrm{KL}}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}_2} = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{x}}_2^H\right] = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{y}}_2^H\right]\widetilde{\mathbf{W}}_2^{-H} \stackrel{!}{=} \mathbf{0} \tag{3.81}$$

A perturbation analysis of (3.81) at $\mathbf{y} = \widetilde{\mathbf{W}}_1\bar{\mathbf{A}}_1\mathbf{s}$ yields $\widetilde{\mathbf{W}}_2 = \mathscr{O}(\sigma^4)$ due to $\widetilde{\mathbf{x}}_2 = \widetilde{\mathbf{v}}_2$. Hence, $\mathbf{y}$ is given by $\mathbf{y} = \widetilde{\mathbf{W}}_1\widetilde{\mathbf{x}}_1 + \mathscr{O}(\sigma^4)$. The solution for $\widetilde{\mathbf{W}}_1$ is similar to the case of $M = N$

$$\widetilde{\mathbf{W}}_{1,\mathrm{ICA}} = (\mathbf{I} + \sigma^2\mathbf{C})\bar{\mathbf{A}}_1^{-1} + \mathscr{O}(\sigma^4) \tag{3.82}$$

where the elements of $\mathbf{C}$ can be computed from (3.97) and (3.98) with $\bar{\mathbf{R}}_{-1} = \mathbf{0}$ and $\mathbf{R}_{-1} = (\bar{\mathbf{A}}_1^{-1}(\mathbf{R}_{\bar{\mathbf{v}}_{11}} - \mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1})\bar{\mathbf{A}}_1^{-H})$.

Finally, we need to combine $\widetilde{\mathbf{W}}_{1,\mathrm{ICA}}$, $\mathbf{T}$ and $\mathbf{Q}$ to form the final solution:

$$\begin{aligned}\widetilde{\mathbf{W}}_{\mathrm{ICA}} &= \left[\widetilde{\mathbf{W}}_{1,\mathrm{ICA}}\ \mathbf{0}\right] + \mathscr{O}(\sigma^4),\\ \mathbf{W}_{\mathrm{ICA}} &= \widetilde{\mathbf{W}}_{\mathrm{ICA}}\mathbf{T}\mathbf{Q} = \widetilde{\mathbf{W}}_{1,\mathrm{ICA}}\left[\mathbf{I} - \mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1}\right]\mathbf{Q} + \mathscr{O}(\sigma^4).\end{aligned} \tag{3.83}$$

Note that for the case of noncircular complex noise, the presented transformation does not work since we would need to take into account the pseudo-covariance matrix of the noise.

3.3.4.1 Results for Circular Complex GGD

Here, we study the performance for the overdetermined case with $M = 6$ sensors, $N = 3$ sources, and circular complex noise. The sources follow a circular complex GGD distribution with identical shape parameters c. Similar to Sect. 3.3.3, we use the mixing matrix $\mathbf{A} = [a_{mn}]$ with $a_{mn} = e^{-j\pi m \sin\theta_n}$ with $\theta_n = -60°, 0°, 60°$. We first consider spatially uncorrelated noise with $\mathbf{R}_{\mathbf{v}} = \mathbf{I}$. Figure 3.9a shows that for a wide range of the shape parameter c, both the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$ and its estimate $\hat{\mathbf{W}}_{\mathrm{ICA}}$ obtained by running KLD-ICA using $L = 10^4$ samples achieve an SINR close to that of the MMSE solution $\mathbf{W}_{\mathrm{MMSE}}$. Furthermore, note that additional sensors can improve the SINR of the demixed signals: Using only the first $M = 3$ sensors, $\mathbf{W}_{\mathrm{MMSE}}$ achieves an SINR of 12.4 dB (see Fig. 3.5), whereas with $M = 6$ sensors it achieves an SINR of 17.4 dB.

When the noise $\mathbf{v}$ is correlated with the normalized correlation matrix

$$\mathbf{R}_{\mathbf{v}} = \begin{bmatrix} 1.00+0.00j & 0.62+0.23j & 0.44-0.16j & 0.46+0.11j & -0.09+0.26j & -0.03+0.09j \\ 0.62-0.23j & 1.00+0.00j & 0.56+0.06j & 0.47-0.13j & 0.44+0.18j & -0.09+0.26j \\ 0.44+0.16j & 0.56-0.06j & 1.00+0.00j & 0.52+0.09j & 0.47-0.13j & 0.46+0.11j \\ 0.46-0.11j & 0.47+0.13j & 0.52-0.09j & 1.00+0.00j & 0.56+0.06j & 0.44-0.16j \\ -0.09-0.26j & 0.43-0.18j & 0.47+0.13j & 0.56-0.06j & 1.00+0.00j & 0.62+0.23j \\ -0.03-0.09j & -0.09-0.26j & 0.46-0.11j & 0.44+0.16j & 0.62-0.23j & 1.00+0.00j \end{bmatrix},$$

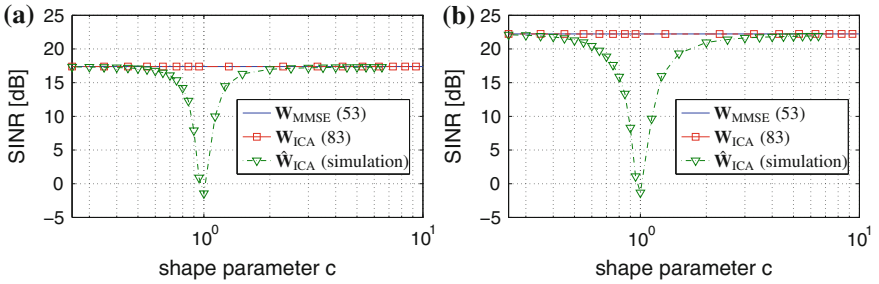


Fig. 3.9 SINR for overdetermined case with circular complex GGD signals and circular complex noise, SNR = 10 dB, $L = 10^4$ samples. $\mathbf{R}_{\mathbf{v}} = \mathbf{I}$ (**a**), $\mathbf{R}_{\mathbf{v}} \neq \mathbf{I}$ (**b**)

$\mathbf{W}_{\mathrm{MMSE}}$ achieves an SINR of 22.2 dB for an SNR of 10 dB and all $M = 6$ sensors. With the first $M = 3$ sensors, it achieves only an SINR of 13.3 dB. Compared to the case of uncorrelated noise, the form of the SINR curve for $\hat{\mathbf{W}}_{\mathrm{ICA}}$ changes slightly but it is still quite close to that of $\mathbf{W}_{\mathrm{MMSE}}$ except for $c \approx 1$ (see Fig. 3.9b).

3.4 Conclusion

We have derived an analytic expression for the demixing matrix of KLD-based ICA for the low noise regime. We have considered the general noncircular complex determined case. The solution for the circular complex and real case can be derived as special cases. Furthermore, we have shown how to reduce the overdetermined case $M > N$ to the determined case. Although the KLD and MMSE solutions differ, linear demixing based on these two criteria yields demixed signals with similar SINR in many cases. In practice, however, not only the bias studied in this chapter but also the variance of the estimate are important for SINR. For the noiseless case, the variance of the estimated demixing matrix is lower bounded by the CRB derived in Sect. 3.2 and [35].

Appendix 1

Values of κ, ξ, β, η for Complex GGD

The pdf of a noncircular complex GGD with zero mean, variance $\mathrm{E}[|s|^2] = 1$ and noncircularity index $\gamma \in [0, 1]$ is given by

$$p(s, s^*) = \frac{c\alpha \cdot \exp\left(-\left[\frac{\alpha/2}{\gamma^2-1}\left(\gamma s^2 + \gamma s^{*2} - 2ss^*\right)\right]^c\right)}{\pi\,\Gamma(1/c)(1-\gamma^2)^{1/2}}, \tag{3.84}$$

where $\alpha = \Gamma(2/c)/\Gamma(1/c)$ and $\Gamma(\cdot)$ is the Gamma function. The function $\varphi(s, s^*) = -\frac{\partial}{\partial s^*} \ln p(s, s^*)$ is then given by

$$\varphi(s, s^*) = \frac{2c(\alpha/2)^c}{(\gamma^2 - 1)^c} \left(\gamma s^2 + \gamma (s^*)^2 - 2ss^*\right)^{c-1} (\gamma s^* - s). \tag{3.85}$$

By integration in polar coordinates, it can be shown that κ, ξ, β and η are given by:

$$\kappa = \mathrm{E}\left[|\varphi(s)|^2\right] = \frac{c^2 \Gamma(2/c)}{(1-\gamma^2)\Gamma^2(1/c)}, \tag{3.86}$$

$$\xi = \mathrm{E}\left[(\varphi^*(s))^2\right] = -\frac{c^2 \gamma \Gamma(2/c)}{(1-\gamma^2)\Gamma^2(1/c)} = -\gamma\kappa, \tag{3.87}$$

$$\eta = \mathrm{E}\left[|s|^2 |\varphi(s)|^2\right] = \frac{(c+1)\cdot(2-\gamma^2)}{2(1-\gamma^2)}, \tag{3.88}$$

$$\beta = \mathrm{E}\left[s^2(\varphi^*(s))^2\right] = \frac{(c+1)\cdot(2-3\gamma^2)}{2(1-\gamma^2)}. \tag{3.89}$$

Induced CRB for Real ICA

Here, we briefly review the iCRB for real ICA [41, 45]. In the following, all real quantities q are denoted as $\mathring{q}$. In the derivation of the iCRB for the real case $\mathring{\varphi}(\mathring{s}) = -\partial \ln p(\mathring{s})/\partial \mathring{s}$ and the parameters $\mathring{\kappa} = E[\mathring{\varphi}^2(\mathring{s})]$, $\mathring{\eta} = \mathrm{E}[\mathring{s}^2\mathring{\varphi}^2(\mathring{s})] = 2 + \mathrm{E}\left[\mathring{s}^2 \frac{\partial \mathring{\varphi}(\mathring{s})}{\partial \mathring{s}}\right]$ are defined using real derivatives. In [41, 45] it was shown that

$$\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{L(\mathring{\eta}_i - 1)}, \tag{3.90}$$

$$\mathrm{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\mathring{\kappa}_j}{\mathring{\kappa}_i \mathring{\kappa}_j - 1}. \tag{3.91}$$

Appendix 2

Here we derive an analytic expression for $\mathbf{W}_{\mathrm{ICA}}$ in the presence of noise by using a perturbation analysis. Motivated by $\mathbf{W}_{\mathrm{ICA}} \overset{\sigma^2=0}{=} \mathbf{A}^{-1}$, we assume that $\mathbf{W}_{\mathrm{ICA}}$ can be written as $\mathbf{W}_{\mathrm{ICA}} = \mathbf{A}^{-1} + \sigma^2\mathbf{B} + \mathscr{O}(\sigma^4)$ and derive $\mathbf{B}$ by a two-step perturbation analysis:

1. Taylor series approximation of $\mathrm{E}(\boldsymbol{\varphi}^*(\mathbf{y})\mathbf{y}^T)$ in (3.51) at $\mathbf{y} = \hat{\mathbf{y}} = \mathbf{W}_{\mathrm{ICA}}\mathbf{A}\mathbf{s}$,
2. Taylor series approximation of the result of the above step by exploiting $\mathbf{W}_{\mathrm{ICA}} = \mathbf{A}^{-1} + \sigma^2\mathbf{B} + \mathscr{O}(\sigma^4)$ and $\hat{\mathbf{y}} = \mathbf{s} + \sigma^2\mathbf{B}\mathbf{A}\mathbf{s} + \mathscr{O}(\sigma^4) = \mathbf{s} + \sigma^2\mathbf{C}\mathbf{s} + \mathscr{O}(\sigma^4) = \mathbf{s} + \sigma^2\mathbf{b} + \mathscr{O}(\sigma^4)$ with $\mathbf{C} = \mathbf{B}\mathbf{A}$ and $\mathbf{b} = \mathbf{C}\mathbf{s} = [b_1, \ldots, b_N]^T$.

In this way, we determine explicitely the deviation $\sigma^2\mathbf{B}$ of $\mathbf{W}_{\mathrm{ICA}}$ from the inverse solution $\mathbf{A}^{-1}$.

The general Taylor series expansion of $\varphi^*(y) \widehat{=} \varphi^*(y, y^*)$ is given as

$$\begin{aligned}\varphi^*(y, y^*) &= \varphi^*(\hat{y}, \hat{y}^*) + \frac{\partial\varphi^*}{\partial y}\Delta y + \frac{\partial\varphi^*}{\partial y^*}\Delta y^* + \frac{1}{2}\left(\frac{\partial^2\varphi^*}{(\partial y)^2}(\Delta y)^2 + \frac{\partial^2\varphi^*}{(\partial y^*)^2}(\Delta y^*)^2\right) \\ &\quad + \frac{\partial^2\varphi^*}{\partial y\partial y^*}\Delta y\Delta y^* + \ldots \\ &= \varphi^*(\hat{y}, \hat{y}^*) + \varpi(y, y^*)\Delta y + \vartheta(y, y^*)\Delta y^* \\ &\quad + \frac{1}{2}\left(\nu(y, y^*)(\Delta y)^2 + \zeta(y, y^*)(\Delta y^*)^2\right) + \epsilon(y, y^*)\Delta y\Delta y^* + \ldots \end{aligned} \tag{3.92}$$

with $\varpi(y, y^*) = \frac{\partial\varphi^*}{\partial y}$, $\vartheta(y, y^*) = \frac{\partial\varphi^*}{\partial y^*}$, $\nu(y, y^*) = \frac{\partial^2\varphi^*}{(\partial y)^2}$, $\zeta(y, y^*) = \frac{\partial^2\varphi^*}{(\partial y^*)^2}$ and $\epsilon(y, y^*) = \frac{\partial^2\varphi^*}{\partial y\partial y^*}$. To simplify notation, we will drop the dependence of $\varphi^*(\cdot)$, $\varpi(\cdot)$, $\vartheta(\cdot)$, $\nu(\cdot)$, $\zeta(\cdot)$, $\epsilon(\cdot)$ on y^* and keep only the dependence on y in the following.

Let

$$\rho_i = \mathrm{E}\left[\varpi_i(s_i)s_i^2\right], \qquad \delta_i = \mathrm{E}\left[\vartheta_i(s_i)s_i^*s_i\right], \tag{3.93}$$

$$\kappa_i = \mathrm{E}\left[\vartheta_i(s_i)\right], \qquad \xi_i = \mathrm{E}\left[\varpi_i(s_i)\right], \tag{3.94}$$

$$\omega_i = \mathrm{E}\left[\nu_i(s_i)s_i\right], \qquad \tau_i = \mathrm{E}\left[\zeta_i(s_i)s_i\right], \tag{3.95}$$

$$\lambda_i = \mathrm{E}\left[\epsilon_i(s_i)s_i\right], \qquad \gamma_i = \mathrm{E}[s_i^2]. \tag{3.96}$$

As shown in [31, 34], $\mathbf{W}_{\mathrm{ICA}} = \mathbf{A}^{-1} + \sigma^2\mathbf{C}$, where the elements of $\mathbf{C}$ can be computed from

$$\rho_i C_{ii} + \delta_i C_{ii}^* + C_{ii} = -(\kappa_i + \lambda_i)\left[\mathbf{R}_{-1}\right]_{ii} - (\xi_i + \frac{1}{2}\omega_i)\left[\bar{\mathbf{R}}_{-1}\right]_{ii} - \frac{1}{2}\tau_i\left[\bar{\mathbf{R}}_{-1}\right]_{ii}^*. \tag{3.97}$$

and

$$\begin{aligned}\gamma_j\xi_i C_{ij} + \kappa_i C_{ij}^* + C_{ji} &= -\kappa_i\left[\mathbf{R}_{-1}\right]_{ij}^* - \xi_i\left[\bar{\mathbf{R}}_{-1}\right]_{ij}, \\ \gamma_i\xi_j C_{ji} + \kappa_j C_{ji}^* + C_{ij} &= -\kappa_j\left[\mathbf{R}_{-1}\right]_{ji}^* - \xi_j\left[\bar{\mathbf{R}}_{-1}\right]_{ji}. \end{aligned} \tag{3.98}$$

with the transformed noise covariance matrix $\mathbf{R}_{-1} = \mathbf{W}\mathbf{R}_{\mathbf{v}}\mathbf{W}^H = \mathbf{A}^{-1}\mathbf{R}_{\mathbf{v}}\mathbf{A}^{-H} + \mathcal{O}(\sigma^2)$ and the transformed noise pseudo-covariance matrix $\bar{\mathbf{R}}_{-1} = \mathbf{W}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{W}^T = \mathbf{A}^{-1}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{A}^{-T} + \mathcal{O}(\sigma^2)$. Note that $\mathbf{R}_{-1}^H = \mathbf{R}_{-1}$ and $\bar{\mathbf{R}}_{-1}^T = \bar{\mathbf{R}}_{-1}$.

If $p(s, s^*)$ is symmetric in the real part $\Re s$ or imaginary part $\Im s$ of s, i.e., $p(-\Re s, \Im s) = p(\Re s, \Im s)$ or $p(\Re s, -\Im s) = p(\Re s, \Im s)$, the parameters κ_i, ρ_i, δ_i, λ_i, ξ_i, ω_i, τ_i are real. For $\rho_i + 1 \pm \delta_i \neq 0$, we then get from (3.97)

$$\Re C_{ii} = -\frac{(\kappa_i + \lambda_i)\left[\mathbf{R}_{-1}\right]_{ii} + (\xi_i + \frac{1}{2}(\omega_i + \tau_i))\left[\Re\bar{\mathbf{R}}_{-1}\right]_{ii}}{\rho_i + 1 + \delta_i},$$
$$\Im C_{ii} = -\frac{(\xi_i + \frac{1}{2}(\omega_i - \tau_i))\left[\Im\bar{\mathbf{R}}_{-1}\right]_{ii}}{\rho_i + 1 - \delta_i}. \tag{3.99}$$

For $(\gamma_j\xi_i + \kappa_i)(\gamma_i\xi_j + \kappa_j) \neq 1$ and $(\gamma_j\xi_i - \kappa_i)(\gamma_i\xi_j - \kappa_j) \neq 1$, we obtain from (3.98)

$$\Re C_{ij} = \frac{(\kappa_j - \kappa_i(\gamma_i\xi_j + \kappa_j))\left[\Re\mathbf{R}_{-1}\right]_{ij} + (\xi_j - \xi_i(\gamma_i\xi_j + \kappa_j))\left[\Re\bar{\mathbf{R}}_{-1}\right]_{ij}}{(\gamma_j\xi_i + \kappa_i)(\gamma_i\xi_j + \kappa_j) - 1},$$
$$\Im C_{ij} = \frac{(\kappa_j + \kappa_i(\gamma_i\xi_j - \kappa_j))\left[\Im\mathbf{R}_{-1}\right]_{ij} + (\xi_j - \xi_i(\gamma_i\xi_j - \kappa_j))\left[\Im\bar{\mathbf{R}}_{-1}\right]_{ij}}{(\gamma_j\xi_i - \kappa_i)(\gamma_i\xi_j - \kappa_j) - 1}. \tag{3.100}$$

References

1. Adali, T., Li, H.: A practical formulation for computation of complex gradients and its application to maximum likelihood ICA. In: Proc. IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 2, pp. II-633-II-636 (2007)
2. Adali, T., Li, H.: Complex-valued adaptive signal processing, ch. 1. In: T. Adali, S. Haykin (eds.) Adaptive Signal Processing: Next Generation Solutions, pp. 1–85. Wiley, New York (2010)
3. Adali, T., Li, H., Novey, M., Cardoso, J.F.: Complex ica using nonlinear functions. IEEE Trans. Signal Process. **56**(9), 4536–4544 (2008)
4. Adali, T., Schreier, P., Scharf, L.: Complex-valued signal processing: the proper way to deal with impropriety. IEEE Trans. Signal Process. **59**(11), 5101–5125 (2011)
5. Anderson, M., Li, X.L., Rodriquez, P.A., Adali, T.: An effective decoupling method for matrix optimization and its application to the ICA problem. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 1885–1888 (2012)
6. Brandwood, D.H.: A complex gradient operator and its application in adaptive array theory. IEE Proc. **130**, 11–16 (1983)
7. Cardoso, J., Souloumiac, A.: Blind beamforming for non-gaussian signals. Radar Signal Process. IEE Proc. F **140**(6), 362–370 (1993)
8. Cardoso, J.F.: On the performance of orthogonal source separation algorithms. In: Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 776–779 (1994)
9. Cardoso, J.F.: Blind signal separation: statistical principles. Proc. IEEE **86**(10), 2009–2025 (1998)
10. Cardoso, J.F., Adali, T.: The maximum likelihood approach to complex ICA. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 673–676 (2006)
11. Cichocki, A., Sabala, I., Choi, S., Orsier, B., Szupiluk, R.: Self adaptive independent component analysis for sub-gaussian and super-gaussian mixtures with unknown number of sources and additive noise. In: Proceedings of 1997 International Symposium on Nonlinear Theory and its Applications (NOLTA-97), vol. 2, pp. 731–734 (1997)
12. Comon, P., Jutten, C. (eds.): Handbook of Blind Source Separation: Independent Component Analysis and Applications, 1st edn. Elsevier, Amsterdam (2010)
13. Davies, M.: Identifiability issues in noisy ica. IEEE Signal Process. Lett. **11**(5), 470–473 (2004)

14. De Lathauwer, L., De Moor, B.: On the blind separation of non-circular sources. In: Proceedings of the European Signal Processing Conference (EUSIPCO), vol. 2, pp. 99–102. Toulouse, France (2002)
15. Doron, E., Yeredor, A., Tichavsky, P.: Cramér-Rao-induced bound for blind separation of stationary parametric gaussian sources. IEEE Signal Process. Lett. **14**(6), 417–420 (2007)
16. Douglas, S., Cichocki, A., Amari, S.: A bias removal technique for blind source separation with noisy measurements. Eletron. Lett. **34**(14), 1379–1380 (1998)
17. Douglas, S.C.: Fixed-point algorithms for the blind separation of arbitrary complex-valued non-gaussian signal mixtures. EURASIP J. Appl. Signal Process. **2007**(1), Article ID 36,525 (2007)
18. Eriksson, J., Koivunen, V.: Complex-valued ICA using second order statistics. pp. 183–192 (2004)
19. Eriksson, J., Koivunen, V.: Complex random vectors and ICA models: identifiability, uniqueness, and separability. IEEE Trans. Inf. Theory **52**(3), 1017–1029 (2006)
20. Fiori, S.: Neural independent component analysis by maximum-mismatch learning principle. Neural Netw. **16**(8), 1201–1221 (2003)
21. Hjørungnes, A.: Complex-Valued Matrix Derivatives. Cambridge University Press, Cambridge (2011)
22. Horn, R.A., Johnson, C.R.: Matrix analysis, 1st publis. (1985), 10th print edn. Cambridge University Press, Cambridge (1999)
23. Hyvärinen, A.: Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. Neurocomputing **22**, 49–67 (1998)
24. Jagannatham, A., Rao, B.: Cramér-Rao lower bound for constrained complex parameters. IEEE Signal Process. Lett. **11**(11), 875–878 (2004)
25. Joho, M., Mathis, H., Lambert, R.H.: Overdetermined blind source separation: using more sensors than source signals in a noisy mixture. In: Proceedings of the International Conference on Independent Component Analysis and Blind Source Separation (ICA), pp. 81–86 (2000)
26. Koldovsky, Z., Tichavsky, P.: Methods of fair comparison of performance of linear ICA techniques in presence of additive noise. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 873–876 (2006)
27. Koldovsky, Z., Tichavsky, P.: Asymptotic analysis of bias of Fast ICA-based algorithms in the presence of additive noise. Technical Report 2181, UTIA, AV CR (2007a)
28. Koldovsky, Z., Tichavsky, P.: Blind instantaneous noisy mixture separation with best interference-plus-noise rejection. In: Proceedings of the International Conference on Independent Component Analysis and Blind Source Separation (ICA), pp. 730–737 (2007b)
29. Li, H., Adali, T.: Algorithms for complex ML ICA and their stability analysis using Wirtinger calculus. IEEE Trans. Signal Process. 58(12) (2010)
30. Li, X.L., Adali, T.: Complex independent component analysis by entropy bound minimization. IEEE Trans. Circ. Syst. I Regul. Pap. **57**(7), 1417–1430 (2010)
31. Loesch, B.: Complex blind source separation with audio applications. Ph.D. thesis, University of Stuttgart (2013). http://www.hut-verlag.de/9783843911214.html
32. Loesch, B., Yang, B.: On the relation between ICA and MMSE based source separation. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 3720–3723 (2011)
33. Loesch, B., Yang, B.: Cramér-Rao bound for circular complex independent component analysis. In: Proceedings of the International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA), pp. 42–49 (2012a)
34. Loesch, B., Yang, B.: On the solution of circular and noncircular complex KLD-ICA in the presence of noise. In: Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 1479–1483 (2012b)
35. Loesch, B., Yang, B.: Cramér-Rao bound for circular and noncircular complex independent component analysis. IEEE Trans. Signal Process. **61**(2), 365–379 (2013)
36. Mandic, D.P., Goh, V.S.L.: Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear, and Neural Models, 1st edn. Wiley, Chichester (2009)

37. Novey, M., Adali, T.: ICA by maximization of nongaussianity using complex functions. In: Proceedings of the IEEE Workshop on Machine Learning for Signal Processing (MLSP), pp. 21–26 (2005)
38. Novey, M., Adali, T.: Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm. In: Proceedings of the IEEE Workshop on Machine Learning for Signal Processing (MLSP), pp. 79–84 (2006)
39. Novey, M., Adali, T.: On extending the complex FastICA algorithm to noncircular sources. IEEE Trans. Signal Process. **56**(5), 2148–2154 (2008)
40. Novey, M., Adali, T., Roy, A.: A complex generalized gaussian distribution—characterization, generation, and estimation. IEEE Trans. Signal Process. **58**(3), 1427–1433 (2010)
41. Ollila, E., Kim, H.J., Koivunen, V.: Compact Cramér-Rao bound expression for independent component analysis. IEEE Trans. Signal Process. **56**(4), 1421–1428 (2008)
42. Ollila, E., Koivunen, V., Eriksson, J.: On the Cramér-Rao bound for the constrained and unconstrained complex parameters, pp. 414–418 (2008)
43. Remmert, R.: Theory of Complex Functions. Graduate Texts in Mathematics. Springer, New York (1991)
44. Schreier, P.J., Scharf, L.L.: Statistical signal processing of complex-valued data: The theory of improper and noncircular signals. Cambridge University Press, Cambridge (2010)
45. Tichavsky, P., Koldovsky, Z., Oja, E.: Performance analysis of the Fast ICA algorithm and Cramér-Rao bounds for linear independent component analysis. IEEE Trans. Signal Process. **54**(4) (2006)
46. Wirtinger, W.: Zur formalen theorie der funktionen von mehr komplexen veränderlichen. Math. Ann. **97**(1), 357–375 (1927)
47. Yeredor, A.: Blind separation of gaussian sources with general covariance structures: bounds and optimal estimation. IEEE Trans. Signal Process. **58**(10), 5057–5068 (2010)
48. Yeredor, A.: Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources. IEEE Trans. Signal Process. **60**(1), 478–483 (2012)
49. Zhang, L.Q., Cichocki, A., Amari, S.: Natural gradient algorithm for blind separation of overdetermined mixture with additive noise. IEEE Signal Process. Lett. **6**(11), 293–295 (1999)
50. Zhu, X.L., Zhang, X.D., Ye, J.M.: A generalized contrast function and stability analysis for overdetermined blind separation of instantaneous mixtures. Neural Comput. **18**(3), 709–728 (2006)

Chapter 4
Subband-Based Blind Source Separation and Permutation Alignment

Bo Peng and Wei Liu

Abstract The aim of this chapter is to present the fundamental ideas of subband-based convolutive blind source separation (BSS) employing filter banks, in particular with a focus on the inherent permutation alignment problem associated with this approach, and bring attention to the most recent developments in this area, including the joint BSS approach in solving the convolutive mixing problem.

4.1 Introduction to the Convolutive Mixing Problem

Blind source separation (BSS) has been studied extensively in the past 2 decades, with the "cocktail party problem" as the most representative example [7, 14]. The BSS problem was initially formulated by a linear instantaneous mixing model and based on this model the independent component analysis was introduced [8], which exploits the statistical independence of the source signals, and compensates for the lack of prior knowledge in the mixing model. A plethora of algorithms were proposed and developed afterward, including the Informax approach [4], natural gradient algorithm using Kullback–Leibler divergence [2], fastICA [15], and linear predictor-based algorithms [33–35, 37], etc.

Later, the instantaneous mixing model was extended and the convolutive mixing model accounting for delay and reflection was considered, which is often encountered in acoustic mixing problems [30–32]. However, it also appears in wireless communications when there are multipath effects in the channel model and some biomedical problems [59]. Blind deconvolution techniques were proposed to solve these problems by extending the existing instantaneous algorithms in the time domain [3].

B. Peng · W. Liu (✉)
Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, UK
e-mail: w.liu@sheffield.ac.uk

B. Peng
e-mail: peng.bo@outlook.com

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_4,
© Springer-Verlag Berlin Heidelberg 2014

The convolutive mixing model can also be transformed into the frequency-domain after frequency decomposition of the mixed signals, in order to improve the convergence rate and reduce the computation time. This is possible since convolutive mixing in the time domain corresponds to an instantaneous one in the frequency domain and instantaneous BSS can be performed at all the frequencies in parallel [57]. Depending on the methods of frequency decomposition, it is termed as frequency-domain BSS if the standard discrete Fourier transform (DFT) technique is used [43, 54, 61], or it can be termed as subband-based BSS if a more general filter banks system is used [11, 41, 44].

Because the instantaneous algorithm is applied individually at each frequency band, there arise subband permutation and scaling ambiguities for both the DFT and filter banks-based approaches. The scaling ambiguities will cause spectral distortions, but can be mitigated to some degree by normalizing the separation matrices [43]. However, the permutation ambiguity leads to permutation misalignment between different frequencies/subbands, which is usually termed as the permutation problem. Without permutation alignment, the following synthesis stage will remix the separated signals and cause serious performance loss [16]. For frequency/subband-domain BSS, there are mainly two approaches to solve the permutation problem. One relies on the direction of arrival (DOA) angles of the source signals, which are estimated from the obtained separation matrices [23]. However, it is difficult to have accurate estimations when the signals arrive from many different directions due to multipath propagation, and the geometric information of the sensors is also required, which could be unavailable in some BSS problems. The other one exploits the correlation properties between separated signals at adjacent frequency bands [43]. This is a popular method which can be employed in many frequency/subband-based BSS problems without requiring additional assumptions, and performed as a separate permutation alignment stage after applying instantaneous BSS algorithms. However, when the correlation between adjacent frequency components becomes insufficient, the alignment results will be less reliable and eventually affect the overall separation performance.

The correlation-based alignment is a synchronizing process applied immediately after the separation for each subband. Alternatively, we can avoid the permutation problem at the beginning of the separation process, by employing joint BSS algorithms. A joint BSS algorithm exploits the mutual information between multiple data sets, and is designed to separate the source signals for different data sets, while still maintaining their correct order. Several methods were proposed for achieving this goal, including independent vector analysis [21, 22, 24], multiset canonical correlation analysis (M-CCA) [19, 28], and the method based on generalized joint diagonalization of cumulant matrices [27]. In this chapter, we will introduce this concept as another method for solving the permutation problem.

This chapter is organized as follows. In Sect. 4.2, a general formulation of the convolutive mixing model is provided, explaining the purpose of transforming the problem into the frequency/subband domain, followed by a review of the filter banks system for subband decomposition, including the DFT as a special case. Then in Sect. 4.3, the subband permutation problem is discussed and various approaches

for subband alignment are studied. In Sect. 4.4, details about the recently proposed approach based on intersubband correlation maximization are provided. Simulation results are given in Sect. 4.5, followed by a summary in Sect. 4.6.

4.2 Overview of Subband-Based BSS

4.2.1 Real-Valued and Complex-Valued Filter Banks

4.2.1.1 Basics of Filter Banks

In digital signal processing, if the sampling frequency at different parts of the system changes, it is often referred to as a multirate system. A typical example is the subband system that employs filter banks, which consists of decimators, expanders and two sets of filters, as shown in Fig. 4.1 [1, 40, 58, 62]. The first set of filters consisting of $h_m[n], m = 0, \ldots, M-1$, is called the analysis bank, while the second set consisting of $f_m[n], m = 0, \ldots, M-1$, called the synthesis bank. Both the analysis and the synthesis filters are a series of bandpass filters, with each set covering the fullband from 0 to 2π in normalized frequency.

The input fullband signal $x[n]$ is first split into multiple subband channels by the analysis filters, and each subband signal only occupies a small bandwidth of the original fullband one, which can be sampled at a lower rate due to the reduced bandwidth. This lower sampling rate is achieved by a decimator (also called downsampler) with a factor of N, which retains only every Nth sample of its input. After the required processing, these decimated signals are interpolated by an expander (also called an upsampler) with the same factor N before passing through the synthesis filters, and reconstructed back to its original sampling rate.

For a general M-channel filter banks system with a decimation factor of N, the signal in each subband channel can be decimated by up to $1/M$ of the original sampling rate, which is referred to as a critically sampled system with $M = N$ and an oversampled one when $M > N$. Performing signal processing in subbands instead of the fullband usually has the advantage of lower computational complexity due to a reduced data rate, and a faster convergence rate for adaptive algorithms. Moreover, since the problem can be solved at subbands in parallel, each subband task can then be performed on different processors if necessary, providing the additional advantage of fast real-time implementation.

In practice, the filters in the analysis and synthesis banks have nonideal frequency responses. So filter bank designs often focus on minimization of the stopband energy, in order to reduce the aliasing level after the decimation operation. Also the overall filter banks system itself is usually required not to introduce any distortion to its input signal other than some delay and scaling effect, which is referred to as the perfect reconstruction (PR) condition in filter banks design. The PR condition as well as the minimization of the stopband energy will be discussed in Sect. 4.4.1.

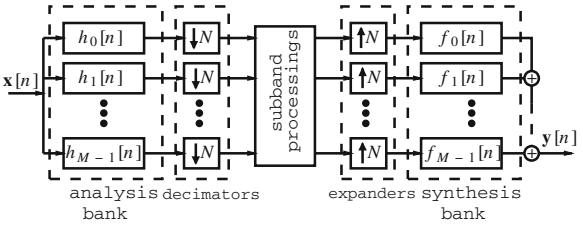


Fig. 4.1 General structure of an M-channel filter banks system with a decimation factor N

4.2.1.2 Modulated Filter Banks

In a modulated filter banks system, the filters in the analysis and the synthesis banks are obtained by modulating a single lowpass filter, so that the system can be designed and implemented more efficiently. Various modulation schemes have been proposed, including DFT [5], modified DFT [18], generalized DFT (GDFT) [64] and discrete cosine transform (DCT) [6, 50, 58].

For an M-channel cosine-modulated filter banks (CMFBs) system, it can be obtained by combining $2M$ complex filters using exponential modulations, which cancel the imaginary parts. The aliasing components of the system caused by decimation can by approximately canceled by the synthesis filters. The expressions for analysis and synthesis filters are given in (4.1), (4.2), and details of derivation can be found in [58].

$$h_m[n] = 2p_0[n]\cos\left[(m+0.5)\frac{\pi}{M}\left(n-\frac{L_p-1}{2}\right)+\theta_m\right] \tag{4.1}$$

$$f_m[n] = 2p_0[n]\cos\left[(m+0.5)\frac{\pi}{M}\left(n-\frac{L_p-1}{2}\right)-\theta_m\right] \tag{4.2}$$

$$\text{for } m = 0, 1, \ldots, M-1.$$

When subband processing is not limited to real-valued operations, we can consider GDFT filter banks, where the analysis filters and the synthesis filters are derived by modulating a prototype filter $p_0[n]$, given by

$$h_m[n] = p_0[n]\cdot \mathrm{e}^{j\frac{2\pi}{M}(m+m_0)(n+n_0)}, \tag{4.3}$$

$$f_m[n] = h_m^*[L_p-n], \tag{4.4}$$

$$\text{for } n = 0, 1, \ldots, L_p-1 \quad \text{and } m = 0, 1, \ldots, M-1,$$

where m_0 and n_0 are offsets for the frequency and time indices, respectively. The parameter m_0 allows the GDFT filter banks to have an analogous frequency response to the DFT filter banks. For example, when $m_0 = 0.5$ and M is even, we will have a special case where the first $M/2$ subbands are all located within the frequency range $[0, \pi]$, as shown in Fig. 4.2. The center of each analysis filter is located at $\left(\frac{2m\pi}{M}+\frac{\pi}{M}\right)$

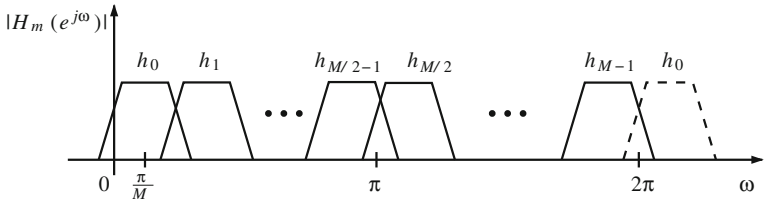


Fig. 4.2 The arrangement of the analysis filters for M-channel generalized DFT filter banks

and filter banks with this arrangement is often referred to as odd-stacked filter banks [9].

Because of symmetry of the frequency responses imposed by the odd-stacked arrangement, the first and the last $M/2$ analysis filters are conjugately related [12]. So if the input is real-valued, only subband operations in the first $\frac{(2m+1)\pi}{M}$ channels need to be processed. Similarly, the first $M/2$ synthesis filters are also complex conjugate of the remaining half. So at the synthesis stage, only $M/2$ subbands need to be processed, and the fullband signal can be recovered by taking the real part of the sum of the outputs from the first $M/2$ channels.

Moreover, by changing the value of n_0, the group delay of the analysis filters can be altered. If the group delay is constant, the filters will have the linear phase property, and avoid the distortion in phase. So when the prototype filter is a real-valued FIR filter with linear phase, a good choice for the time offset is $n_0 = \frac{L_p-1}{2}$. With this choice, the modulation sequence $t[n] = \mathrm{e}^{j\frac{2\pi}{M}(m+m_0)(n+n_0)}$, $n = 0, 1, \ldots, N-1$ will become symmetric, and the analysis filters will have linear phase after modulation.

4.2.2 Blind Source Separation in Frequency Domain

The convolutive mixing model arises when considering an acoustic scenario: there are noticeable delays during the propagation of sound between the speakers and the microphones, and for an indoor environment there will be reflections and additional delays due to the multipath effect. It assumes that each propagation channel is a linear time invariant system, and can be modeled by an FIR filter. For the multichannel model with N_s sources and N_r mixtures/microphones, the convolutive mixing model is given by

$$x_i[n] = \sum_{j=1}^{N_s} \sum_{l=0}^{L_{\mathbf{A}}-1} a_{ij}[l] s_j[n-l], \tag{4.5}$$

where $s_j[n]$ denotes the jth source signal, $x_i[n]$ is the ith mixed signal, and $L_{\mathbf{A}}$ is the number of taps for each of the FIR filters.

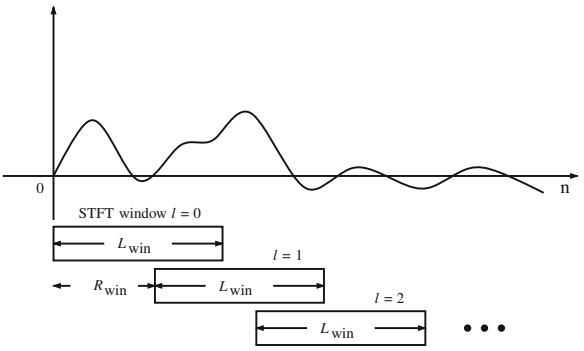


Fig. 4.3 Operation of the STFT using a sliding window

As already mentioned, transforming the separation problem into the frequency domain will greatly simplify the solution. The transform can be achieved by applying the short-time discrete Fourier transform (STFT) to the mixed signals and then the convolutive mixing model in (4.5) is changed to

$$X_j(\omega, k) = \sum_{i=1}^{N_s} A_{ji}(\mathrm{e}^{j\omega}) S_i(\omega, k), (i = 1, \ldots, N_s), \tag{4.6}$$

where $A_{ji}(\mathrm{e}^{j\omega})$ is the frequency-domain representation of $a_{ji}(n)$, and $S_i(\omega, k)$ and $X_j(\omega, k)$ are the time–frequency representations of s_i and x_i at frame index k, respectively.

The STFT process is illustrated in Fig. 4.3. Suppose the window's length is L_{win}. Then a L_{win}-point DFT is applied to the data samples within the sliding window. After the DFT, the window is then advanced by R_{win} samples and another L_{win}-point DFT is applied to the next frame. The output of the signal transformed by the sliding-window DFT is shown in Fig. 4.4, where T_{win} groups of data are obtained, given by

$$T_{\text{win}} = \lceil \frac{L_s - L_{\text{win}}}{R_{\text{win}}} \rceil, \tag{4.7}$$

where L_s denotes the total number of samples for each of the mixed signals and $\lceil \cdot \rceil$ is the ceiling function.

Thus, after transforming the mixed signals into the frequency domain, we can have the demixing model at frequency ω as follows,

$$\mathbf{Y}(\omega, k) = \mathbf{W}(\omega)\mathbf{X}(\omega, k), \tag{4.8}$$

where $\mathbf{Y}(\omega, k)$ and $\mathbf{X}(\omega, k)$ are the frequency transformations of $\mathbf{y}[n]$ and $\mathbf{x}[n]$ at frequency ω, respectively, with $\mathbf{y}[n]$ being the vector representing the recovered

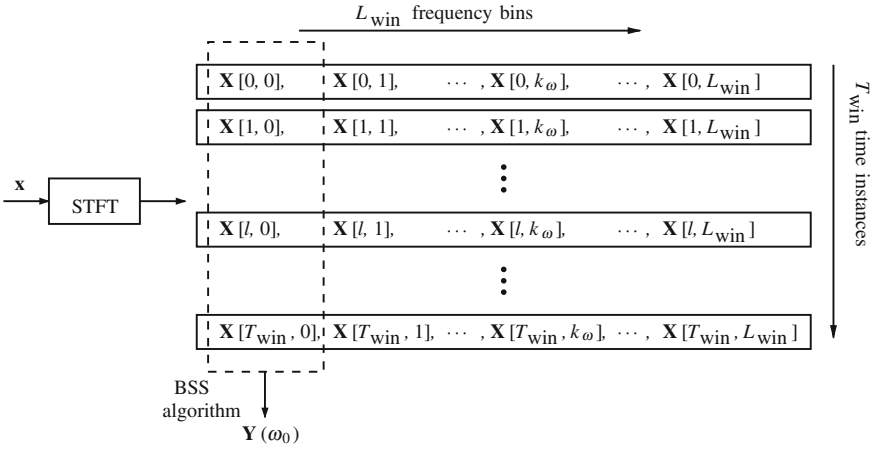


Fig. 4.4 Implementation of sliding-window discrete Fourier transform for frequency-domain BSS

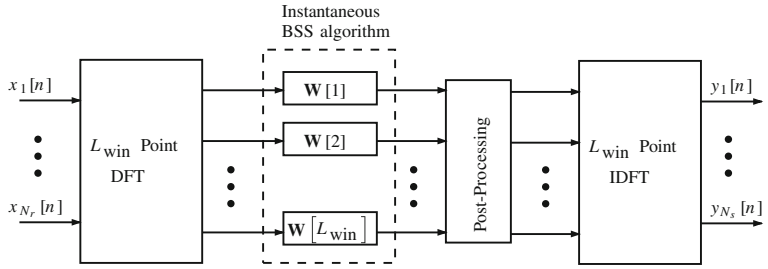


Fig. 4.5 Flow chart of the sliding-window DFT-based frequency-domain BSS

time-domain signals. The separation matrix $\mathbf{W}(\omega)$ can be obtained by a standard instantaneous BSS algorithm at each frequency.

After applying the instantaneous BSS algorithms, post-processing will be required to mitigate the scaling and permutation ambiguities for all frequencies. Finally, the inverse of the sliding-window DFT is applied across all the frequencies to retrieve the fullband output. Alternatively, inverse DFT can be applied to the separation matrix $\mathbf{W}(\omega)$, leading to an time-domain separation matrix $\mathbf{W}$, whose entry will be an FIR filter with L_{win} taps.

The overall structure of the frequency-domain BSS is illustrated in Fig. 4.5. The frequency-domain BSS converts a complicated multichannel deconvolution problem into a number of instantaneous mixing problems, and the BSS algorithm at each frequency becomes easy to converge, which is due to the reduced demixing filter length at each subband and also reduced condition number of the covariance matrix of the corresponding decimated signals, as in the general subband adaptive filtering case. In addition, since BSS algorithms can be applied simultaneously for different frequency components, the computational time would be reduced. However, for the

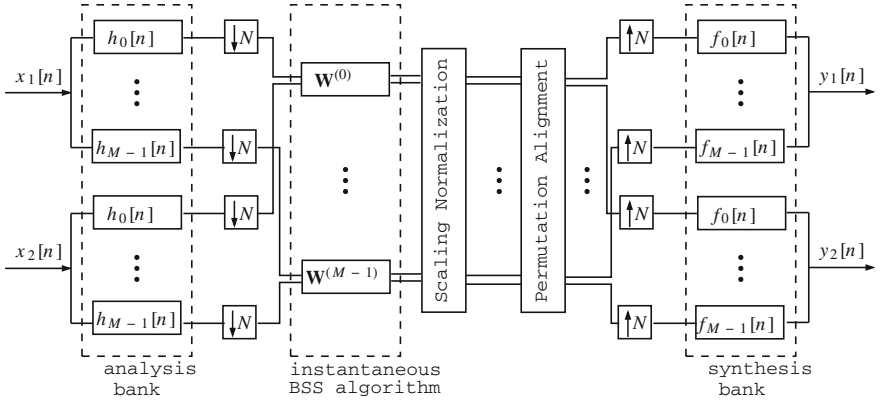


Fig. 4.6 Structure of subband-based BSS with two sources and two mixtures ($N_s = 2$, $N_r = 2$)

sliding window-based frequency-domain BSS, the design of the window function is limited to the same length as the DFT, while for filter banks, the length of the prototype filter can be any value larger or equal to the subband channel number. Therefore, a general subband implementation is needed to provide a robust solution to deal with different problems.

4.2.3 Subband Decomposition and Mixing Model

Instead of using the sliding-window DFT, we can use filter banks to achieve a more flexible frequency decomposition. Applying subband decomposition to the fullband mixture and performing the BSS operation at each subband leads to the subband-based BSS structure shown in Fig. 4.6. The analysis bank has M filters, which split each of the mixed signals into M subbands

$$x_{\text{full}}^{(m)}[n] = \sum_{l=0}^{Lp-1} h_m[l]x[n-l], \quad \text{for } m = 0, \ldots, M-1, \tag{4.9}$$

where h_m denotes the mth analysis filter and its length is L_p. Each subband signal is then downsampled by the decimator with the rate of N

$$x^{(m)}[n] = x_{\text{full}}^{(m)}[Nn]. \tag{4.10}$$

The decomposed subband data is illustrated in Fig. 4.7, which is similar to the frequency decomposition in Fig. 4.4, but with a different time–frequency relationship.

If the decimation factor is sufficiently large compared with the length of the channel impulse response a_{ji}, (4.5) can be further simplified as [41]

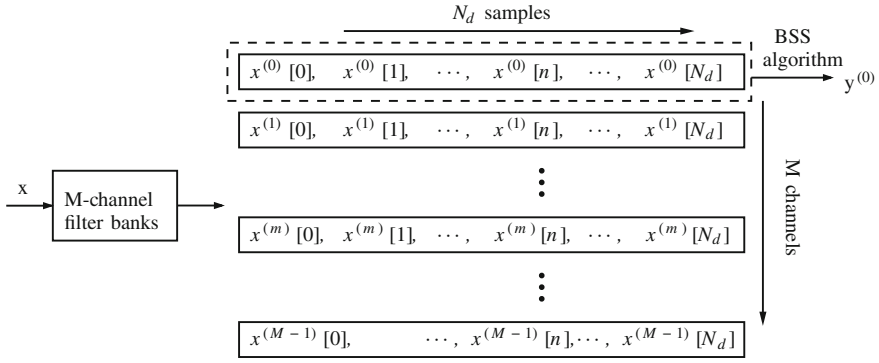


Fig. 4.7 Implementation of filter banks with subband-based BSS

$$\mathbf{x}^{(m)}[n] = \mathbf{A}^{(m)}\mathbf{s}^{(m)}[n], \tag{4.11}$$

where $\mathbf{x}^{(m)}[n] = \left[x_1^{(m)}[n], \ldots, x_{N_r}^{(m)}[n]\right]^T$ is the mth subband components of the fullband mixed signals, $\mathbf{s}^{(m)}[n] = \left[s_1^{(m)}[n], \ldots, s_{N_s}^{(m)}[n]\right]^T$ is the mth subband components of the fullband source signals, and $\mathbf{A}^{(m)}$ is the corresponding $N_r \times N_s$ instantaneous mixing matrix. Therefore at each subband, an individual instantaneous BSS problem is formed and instantaneous BSS algorithms can be employed to obtain $\mathbf{W}^{(m)}$. The separated signal at the mth subband is given by

$$\mathbf{u}^{(m)}[n] = \mathbf{W}^{(m)}\mathbf{x}^{(m)}[n]. \tag{4.12}$$

The subband-based BSS system experiences the same scaling and permutation ambiguities as the frequency-domain BSS, and post-processing is also required. After post-processing, we obtain scaled and permutation aligned subband signal $\mathbf{y}^{(m)}$, which is then upsampled back to its original sampling frequency

$$\mathbf{y}_{\text{full}}^{(m)}[n] = \begin{cases} \mathbf{y}^{(m)}[n/N], & n = 0, \pm N, \pm 2N, \ldots. \\ 0 & \text{otherwise.} \end{cases} \tag{4.13}$$

And the fullband separation results are obtained after the synthesis bank, given by

$$\mathbf{y}[n] = \sum_{m=0}^{M-1} \sum_{l=0}^{L_p-1} f_m[l]\mathbf{y}_{\text{full}}^{(m)}[n-l], \tag{4.14}$$

where f_m denotes the mth synthesis filter of length L_p.

The frequency-domain BSS and the subband-based BSS use the same principle to solve the convolutive mixing problem, which offers better convergence and

requires less computation time than the time-domain blind deconvolution algorithms. However, this is achieved at the expense of scaling and permutation ambiguities at subbands, which could seriously reduce the overall performance of the system.

4.3 Permutation Alignment

4.3.1 Ambiguities in Subband-Based BSS

In this section, solutions to the scaling and permutation problems in subband-based BSS will be introduced and explained. As the subband-based BSS and frequency-domain BSS have similar structures, the methods can be directly applied to the frequency-domain BSS.

Following (4.12), the mixing-demixing relationship for each subband can be formulated

$$\mathbf{W}^{(m)}\mathbf{x}^{(m)}[n] = \mathbf{P}^{(m)}\,\mathbf{D}^{(m)}\mathbf{s}^{(m)}[n], \tag{4.15}$$

where $\mathbf{P}^{(m)}$ is the permutation matrix for the mth subband and $\mathbf{D}^{(m)}$ is a diagonal matrix, whose entries along the diagonal are real scalar coefficients. The permutation and scaling matrices $\mathbf{P}^{(m)}$ and $\mathbf{D}^{(m)}$ are random, and the uncertainties in them need to be addressed when the fullband signals are reconstructed at the synthesis stage. However, the scaling problem and permutation problem have different effect toward the overall performance, and can be considered and solved separately.

The scaling ambiguity would cause each of the subband components with unequal scaling. Assume there is no permutation ambiguity, i.e.,$\mathbf{P}^{(m)} = \mathbf{I}$, for $m = 0, \ldots, M-1$, we can rewrite (4.14) as

$$\left(y_i^{(m)}[n]\right)_{(\text{scaled})} = \sum_{m=0}^{M-1} d_i^{(m)} \sum_{l=0}^{L_p-1} f_m[l]s_i^{(m)}[n-l], \tag{4.16}$$

where $d_i^{(m)}$ is the (i,i)th entry of the scaling matrix $\mathbf{D}^{(m)}$. So the result of (4.16) is a filtered version of the original source s_i, and the frequency response of this distortion filter is given by $d_i^{(m)}$ for $m = 0, \ldots, M-1$. One efficient solution to reduce this distortion is to normalize the separation filter $\mathbf{W}^{(m)}$ for $m = 0, \ldots, M-1$, which will suppress the variance in $d_i^{(m)}$. However, please note that the scaling ambiguity for the overall separating system will never be solved due to the blind nature of the problem.

Although in acoustic applications, a filtered output will cause a downgrade in audio perception, the scaling ambiguity will not affect the separation performance because each $(\mathbf{y}^{(m)})_{(\text{scaled})}$ represents a fully recovered signal without any other mixtures. However, the permutation ambiguity can result in "remixing" of the subband components, i.e.,

$$\left(y_i^{(m)}[n]\right)_{(\text{perm})} = \sum_{m=0}^{M-1} \sum_{l=0}^{L_p-1} f_m[l] s_{\delta(i)^{(m)}}^{(m)}[n-l], \tag{4.17}$$

where $\delta(i)^{(m)} = (1, \ldots, N_s)$ denotes the index number of the nonzero entry at the ith row of the permutation matrix $\mathbf{P}^{(m)}$. Because the permutation matrices are random and very likely different, subband signals corresponding to different sources would be mixed again. The permutation ambiguity will adversely affect the overall separation result, and the effect will become more significant when the number of the sources increases.

4.3.1.1 Method to Mitigate the Scaling Problem

The mixed signals can be viewed as the combination of N_s independent signals,

$$\mathbf{x}^{(m)}[n] = \mathbf{v}_1^{(m)}[n] + \mathbf{v}_2^{(m)}[n] + \cdots + \mathbf{v}_{N_s}^{(m)}[n], \tag{4.18}$$

where $\mathbf{v}_i^{(m)}[n] = [v_{1i}^{(m)}[n], v_{2i}^{(m)}[n], \ldots, v_{N_r i}^{(m)}[n]]^T$, N_r is the total number of mixtures and we assume $N_r = N_s$ here. Then we can have [43]

$$\begin{aligned} \mathbf{x}^{(m)}[n] &= \left(\mathbf{W}^{(m)}\right)^{-1} \mathbf{W}^{(m)} \mathbf{x}^{(m)}[n] = \left(\mathbf{W}^{(m)}\right)^{-1} \mathbf{I}\mathbf{W}^{(m)} \mathbf{x}^{(m)}[n] \\ &= \left(\mathbf{W}^{(m)}\right)^{-1} \left(\mathbf{E}_1 + \cdots + \mathbf{E}_{N_s}\right) \mathbf{W}^{(m)} \mathbf{x}^{(m)}[n] \end{aligned} \tag{4.19}$$

and therefore

$$\mathbf{v}_i^{(m)}[n] = \left(\mathbf{W}^{(m)}\right)^{-1} \mathbf{E}_i \mathbf{W}^{(m)} \mathbf{x}^{(m)}[n] \tag{4.20}$$

for $i = 1, \ldots, N_s$, where $\mathbf{I}$ is an identity matrix, $\mathbf{E}_i$ is a matrix with all elements being zeros except at the ith row and the ith column, which has a value of 1. For example, when $N_s = 3$, we have

$$\mathbf{I} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

From (4.15) we have

$$\left(\mathbf{W}^{(m)}\right)^{-1} = \mathbf{A}^{(m)} \cdot \left(\mathbf{P}^{(m)}\mathbf{D}^{(m)}\right)^{-1} = \mathbf{A}^{(m)} \left(\mathbf{D}^{(m)}\right)^{-1} \left(\mathbf{P}^{(m)}\right)^{T}. \tag{4.21}$$

Substituting (4.21) into (4.20), since $\mathbf{D}^{(m)}$ is a diagonal matrix, we have

$$\begin{aligned}\mathbf{v}_i^{(m)}[n] &= \mathbf{A}^{(m)}\left(\mathbf{D}^{(m)}\right)^{-1}\left(\mathbf{P}^{(m)}\right)^T\mathbf{E}_i\mathbf{P}^{(m)}\mathbf{D}^{(m)}\mathbf{s}^{(m)}[n] \\ &= \mathbf{A}^{(m)}\mathbf{P}_i^{(m)}\mathbf{s}^{(m)}[n],\end{aligned} \tag{4.22}$$

where $\mathbf{P}_i^{(m)} = \left(\mathbf{P}^{(m)}\right)^T\mathbf{E}_i\mathbf{P}^{(m)}$ is a matrix whose elements are all zeros but with one element along the diagonal being 1. Its position is unknown due to the permutation ambiguity. However, (4.22) shows this method in (4.20) can normalize the separation matrix and prevent the BSS algorithm from introducing additional attenuation. Assume there is a scaling factor $\alpha^{(m)}$. We can then have

$$\begin{aligned}\mathbf{v}_i^{(m)}[n] &= \left(\alpha^{(m)}\mathbf{W}^{(m)}\right)^{-1}\mathbf{E}_i\left(\alpha^{(m)}W^{(m)}\right)\mathbf{x}^{(m)}[n] \\ &= \left(\mathbf{W}^{(m)}\right)^{-1}\mathbf{E}_i\mathbf{W}^{(m)}\mathbf{x}^{(m)}[n],\end{aligned} \tag{4.23}$$

which cancels $\alpha^{(m)}$ from the subband output. This technique is useful for frequency/-subband-based BSS, where the signals from different frequencies/subbands are normalized to the level before applying the BSS. However, the scaling caused by the mixing filter is arbitrary and can not be solved due to the blind nature of the problem.

4.3.2 DOA-Based Permutation Alignment

This approach exploits the relationship between the coefficients of the separation filters and the beamforming theory, which was first proposed in [23], and further improved in [17, 51]. Since in most BSS problems, multiple sensors are used for receiving the signals, the receiving end becomes an array system as shown in Fig. 4.8. And if the distance between any two sensors is less than half of the wavelength of the signal, there will be no spatial aliasing. In beamforming theory [40], if we assume that there is no reverberations, the coefficients of the mixing matrix can be approximated for each frequency as follows [40]

$$a_{ij}(\omega) = \mathrm{e}^{j\frac{\omega}{c}d_i\sin(\theta_j)}, \tag{4.24}$$

where θ_j is the arriving angle of the jth source, c is the propagation velocity and d_i is the location of the ith sensor. Note that the angle θ_j is an unknown variable. Given the separation matrix obtained, we have the following transfer function for the jth source

$$U(\theta_j,\omega) = \sum_j^{N_s} w_{ji}(\omega)a_{ij}(\omega) = \sum_j^{N_s} w_{ji}(\omega)\mathrm{e}^{j\frac{\omega}{c}d_i\sin(\theta_j)}, \tag{4.25}$$

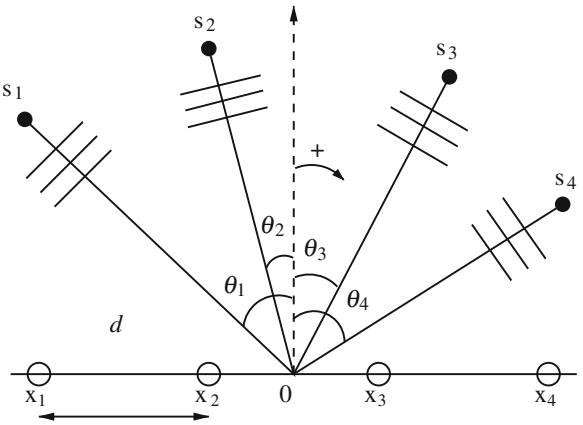


Fig. 4.8 The receiving array and signals in a BSS problem

which forms a directivity pattern. Because each row of the separation matrix extracts one source from the mixture and suppresses the other sources, for each row of the separation matrix, it corresponds to a different directivity pattern. The angles of nulls in the directivity pattern indicate the possible directions of the interfering source signals. For the same source at a fixed location, its angle of null should be similar in the directivity patterns at different frequencies. So those frequency components with similar directivity patterns will be from the same source, which clears the permutation ambiguity.

However, this method have limitations as the following assumptions are required for it to work:

1. There is no reverberation or multipath effect during the propagation.
2. The distance between any two sensors needs to be known, and must be less than half wavelength to avoid the spatial aliasing problem.
3. The sources arrive from fixed angles, and these angles must be different from each other.
4. The sources should be far away from the receiver/sensor, to satisfy the far field condition in beamforming.

In [56], the authors also pointed out that this method is not accurate enough, which becomes a big problem when there are more than two sources and the DOAs of theses sources are close to each other. In practical problems such as separation of speech signals, the multipath effect will be inevitable for an indoor environment, which further reduces the reliability of this method. So in [56], the permutation alignment based on interfrequency correlation is also used to improve the result.

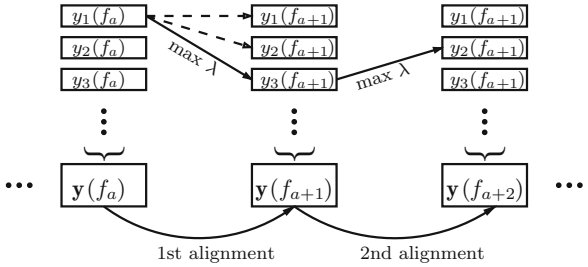


Fig. 4.9 Example of an alignment procedure during frequency-domain BSS. The *solid line* connects the matched pair with maximum interfrequency correlation

4.3.3 Correlation-Based Permutation Alignment

This approach exploits the dependence between the separated frequency components, which is firstly proposed in [43], assuming that there will be correlation between two adjacent frequency components if they are from the same source signal. This assumption is viable even if the source signals are nonstationary, such as in the case of a speech signal, and therefore it is widely used in frequency-domain or subband-based blind speech separation.

The dependence between two frequency components at frequency f_a and f_b can be measured by calculating their normalized correlation, given by

$$\lambda(f_a, f_b) = \frac{E\{(y(f_a) - \mu(f_a))(y(f_b) - \mu(f_b))\}}{\sigma_{f_a}\sigma_{f_b}}, \tag{4.26}$$

where $y_{(f_a)}$ and $y_{(f_b)}$ denote the separated components at frequency f_a and f_b, respectively, and μ and σ denote the mean and standard deviation of the component. Because the mixing model often assumes that different sources are not correlated with each other, the separated frequency components of different sources will have a very small value or even zero for λ, while two separated components from the same source will have a larger value of λ.

An example of a basic alignment procedure is shown in Fig. 4.9. We can start from one component $y_1(f_a)$ at frequency f_a, and calculate the normalized correlations between $y_1(f_a)$ and all the outputs at f_{a+1}. Based on the interfrequency dependence assumption, the pair with largest normalized correlation is from the same source, e.g., $y_1(f_a)$ and $y_1(f_{a+1})$ in the graph. The mapping is then continued between frequency f_{a+1} and f_{a+2}, and extended to the rest of the frequencies. After completing the same procedure to all other components, a complete relationship among different frequency components is obtained and the permutation ambiguity is cleared.

Compared to the method based on DOA estimation, this one is much more flexible, as it does not constrain the BSS algorithm in any way, and the only assumption about the interfrequency dependence can be well met in most BSS problems. On the other hand, this method heavily depends the second-order statistics of the separated

outputs. So when the separation is only partially achieved and the output component still contains the mixture from other source signals, the dependence/independence assumption may not hold well, which will cause incorrect alignment decisions. Furthermore, since the mapping is performed in a sequential manner, one incorrect alignment will propagate to the remaining frequencies.

To reduce errors in permutation alignment, alternative measures of the interfrequency dependence are proposed. In [43], the calculation of the normalized correlation is based on the envelops of the frequency components, which can be obtained by applying a sliding window function. In [53], the correlation coefficients of signal power ratios are used to measure the interfrequency dependence. In [42], each frequency component is modeled by generalized Gaussian distribution, and the variance among the distributions is calculated and used for clustering the components.

There are also approaches to minimize the error propagation and improve the robustness. In [60], a two-step permutation alignment scheme is proposed to minimize the error propagation. During the first step, the conventional bin-wise permutation alignment is applied across all the frequencies, and a group of components with strong interfrequency dependence is selected, which is supposed to have a more accurate alignment. At the second step, this group is merged with neighboring frequency components based on the same correlation criterion. This approach achieved good results in simulations, but a good choice of this starting region becomes very important to the final results. A similar method is used in [53], where the frequency components are first divided into subgroups by clustering process. In [56], the correlation approach is combined with the DOA approach, as the latter one is supposed to be robust against local errors but lack of accuracy.

4.3.4 Joint BSS by Subband Based M-CCA

As mentioned in the Sect. 4.1, we can extend the subband-based BSS structure and consider joint separation of signals at multiple subbands. The advantage of joint BSS in subbands is that no permutation alignment is needed, as the algorithm will automatically align the separated subband signals belonging to the same source.

When a number of data sets are collected and there are dependencies among them, joint analysis is usually preferred to exploit the mutual information among them. This concept has been used in many applications. For example, during the estimation of brain activation, the information can be extracted from a number of functional magnetic resonance (fMRI) data of different subjects [26, 29]. Similarly during frequency decomposition of acoustic signals, dependence exists among signals from different frequencies [20, 21].

This feature can be exploited by multiset canonical correlation analysis (M-CCA) [19, 28], which estimates the linear relationship of data sets by maximizing their correlation [13]. It only relies on the second-order statistics of the signals and has been proved to be an efficient algorithm for separation [36, 38, 39].

After passing through the analysis bank, the subband signals will be preprocessed by a whitening operation; then the M-CCA based on maximizing the sum of squared correlation (SSQCOR) is employed. At the kth stage, the criterion to recover the kth source is given by [28]

$$[\mathbf{w}_k^{(0)}, \ldots, \mathbf{w}_k^{(M-1)}] = \underset{\mathbf{W}_k}{\operatorname{argmax}} \left\{ \sum_{m,n=1}^{M} |\hat{r}_k^{(m,n)}|^2 \right\}, \tag{4.27}$$

$$\text{subject to} \quad \mathbf{w}_k^{(m)} \perp \left\{ \mathbf{w}_1^{(m)}, \ldots, \mathbf{w}_{k-1}^{(m)} \right\}, \tag{4.28}$$

$$\left\| \mathbf{w}_k^{(m)} \right\| = 1, \quad \text{for } m = 0, \ldots, M-1 \tag{4.29}$$

where

$$\hat{r}_k^{(m,n)} = \operatorname{corr}\left(\mathbf{w}_k^{(m)} \mathbf{x}^{(m)}, \mathbf{w}_k^{(n)} \mathbf{x}^{(n)} \right). \tag{4.30}$$

In the context of BSS, $\mathbf{w}_k^{(m)}$ denotes the kth row vector of the separation matrix applied to the mth subband. The objective function (4.27) with two constraints (4.28) and (4.29) can be solved by forming a Lagrangian function with respect to the separation matrix for each of the subbands. The optimum values of $\mathbf{w}_k$ is then obtained by setting its partial derivative function to zero, which leads to the solution to a generalized eigenvalue problem that is updated for each stage [19]. The procedure is repeated until the last signal is recovered.

Equation (4.30) can be further derived as

$$\begin{aligned} \hat{r}_k^{(m,n)} &= \operatorname{corr}\left(\mathbf{w}_k^{(m)} \mathbf{A}^{(m)} \mathbf{s}^{(m)}, \mathbf{w}_k^{(n)} \mathbf{A}^{(n)} \mathbf{s}^{(n)} \right) \\ &= \operatorname{corr}\left(\mathbf{t}_k^{(m)} \mathbf{s}^{(m)}, \mathbf{t}_k^{(n)} \mathbf{s}^{(n)} \right) = \mathbf{t}_k^{(m)} \Lambda^{(m,n)} \mathbf{t}_k^{(n)}, \end{aligned} \tag{4.31}$$

where $\Lambda^{(m,n)}$ is the correlation matrix of the source signals $\mathbf{s}^{(m)}$ and $\mathbf{s}^{(n)}$, $\mathbf{A}^{(m)}$ is the equivalent instantaneous mixing matrix of $\mathbf{A}$ for the mth subband, and $\mathbf{t}_k^{(m)}$ is the kth row vector of the global mixing-demixing matrix $\mathbf{T}^{(m)}$ at the mth subband

$$\mathbf{T}^{(m)} = \mathbf{W}^{(m)} \cdot \mathbf{A}^{(m)} = [(\mathbf{t}_1^{(m)})^{\mathrm{T}}, \ldots, (\mathbf{t}_k^{(m)})^T, \ldots, (\mathbf{t}_{N_s}^{(m)})^T]^T. \tag{4.32}$$

For a satisfactory separation result, the M-CCA would require $\Lambda^{(m,n)}$ having a form close to a diagonal matrix, whose diagonal entries are the correlation values between the matched sources from $s_i^{(m)}$ and $s_i^{(n)}$, $i = 1, \ldots, N_s$ [28]. For speech signals decomposed by filter banks, this assumption can be enhanced by using the prototype filter optimized for the intersubband correlation [45–49], which will be discussed in the following section.

4.4 Design of GDFT Filter Banks for Subband BSS

4.4.1 Review of Filter Banks Design

In Sect. 4.2, we have introduced the formulation for CMFBs and GDFT filter banks. The complex-valued filter banks have been shown to have lower aliasing errors than the real-valued ones. In this section, we will briefly review the basic ideas for designing GDFT filter banks, and then focus on a specific design method for improving the performance of subband-based BSS.

4.4.1.1 Aliasing in Filter Banks System

The main purpose of the analysis filters in the filter banks is to control the subband bandwidth and minimize the components exceeding the required frequency band, which will alias into the baseband and cause severe distortion to the decimated subband signals.

At the mth subband of a filter banks system, the signal after decimation can be formulated by the following equation

$$X^{(m)}(z) = \frac{1}{N} H_m(z^{1/N}) X(z^{1/N}) + \frac{1}{N} \sum_{n=1}^{N-1} H_m(z^{1/N} \mathrm{e}^{-j2\pi n/N}) X(z^{1/N} \mathrm{e}^{-j2\pi n/N}), \tag{4.33}$$

where N is the decimation factor, $X^{(m)}(z)$ is the z-transform of the decomposed signal at the mth subband, and $H_m(z)$ is the z-transform of the mth analysis filter. The first term at the right hand of (4.33) denotes the desired subband signal, and the second term denotes the sum of $(N-1)$ aliasing components, which are the frequency shifted versions of the original subband signal after decimation.

The analysis filters are designed to minimize subband distortion when shifted signals overlap with the baseband signal [52]. Thus these aliasing components will degrade the performance of the required subband processing [26]. In the context of BSS, it will not only affect the subband separation result, but also destroy the correlation between adjacent subband signals, rendering the correlation-based permutation alignment approach even less effective.

By a very large oversampling ratio M/N, or equivalently a large subband sampling frequency, the aliasing component will eventually be reduced to a great degree, but it will also lead to an increase of the computational load. In addition, it causes the band-edge effect if the bandwidth of the guard-band is too wide [26]. The large spectral dynamic range at the band edge will lead to an ill-conditioned autocorrelation matrix of the subband signal, and even affect the convergence rate if the autocorrelation matrix is used in the following subband processing. In our simulations, an oversampling ratio of $\frac{4}{3}$ is used.

Based on (4.33), we further derive the output of the whole subband-based BSS system,

$$Y_i(z) = \frac{1}{N}\sum_{m=0}^{M-1} G_m(z)\hat{\mathbf{w}}_i^{(m)} F_m(z)\mathbf{X}(z) + \frac{1}{N}\sum_{m=0}^{M-1} G_m(z)\hat{\mathbf{w}}_i^{(m)} \times \sum_{l=1}^{N-1} F_m(z\mathrm{e}^{\frac{-j2\pi l}{N}})\mathbf{X}(z\mathrm{e}^{\frac{-j2\pi l}{N}}), \tag{4.34}$$

for $i = 1, \ldots, N_s$, where $\mathbf{X} = [X_1(z), \ldots, X_{N_s}(z)]$ is the z-transform of received signals, $Y_i(z)$ denotes the z-transform of the ith separated signal, and $F_m(z)$ and $G_m(z)$ are the z-transform of the mth analysis and synthesis filters. The vector $\hat{\mathbf{w}}_i^{(m)}$ is the ith row of the matrix $\hat{\mathbf{W}}^{(m)}$, which is the equivalent separation matrix after scaling normalization and permutation alignment at the mth subband. The first part on the right hand side of (4.34) is the transfer function between the source and the output and the second part represents the aliasing components from all frequencies.

4.4.1.2 Reducing the Aliasing Error

For an M-channel filter banks system, the cut-off frequency has to be at least $\omega_p = \pi/M$ to cover the fullband, and the transition band is between $\frac{\pi}{M}$ and $\omega_s = \frac{\pi}{N}$. To reduce the magnitude of the second term of (4.34) and also minimize the aliasing components in (4.33), oversampling is often considered, i.e., $N < M$. At the same time, the stopband energy of the prototype filter is minimized, written as

$$E_s = \int_{\omega_s}^{\pi} \left|P_0(\mathrm{e}^{j\omega})\right|^2 \mathrm{d}\omega = \int_{\omega_s}^{\pi} \left|\sum_{n=0}^{L_p-1} p_0[n]\mathrm{e}^{j\omega n}\right|^2 \mathrm{d}\omega. \tag{4.35}$$

The signal to aliasing ratio (SAR) can be used to measure the aliasing components caused by the energy at the stopband [63], given by

$$\mathrm{SAR} = \frac{\int_0^{\pi/N} \left|P_0(\mathrm{e}^{j\omega})\right|^2 \mathrm{d}\omega}{\int_{\pi/N}^{\pi} \left|P_0(\mathrm{e}^{j\omega})\right|^2 \mathrm{d}\omega}. \tag{4.36}$$

4.4.1.3 Perfect Reconstruction Condition

When the stopband energy is minimized in (4.35) and we adopt the oversampling structure to reduce the aliasing component, the distortion of the subband-based BSS will be governed by the first part of (4.34), given by

$$E_d = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{1}{N}\sum_{m=0}^{M-1} G_m(\mathrm{e}^{j\omega})\hat{\mathbf{w}}_i^{(m)} F_m(\mathrm{e}^{j\omega})\mathbf{X}(\mathrm{e}^{j\omega}) - S_i(\mathrm{e}^{j\omega})\right|^2 \mathrm{d}\omega, \tag{4.37}$$

where $S_i(\mathrm{e}^{j\omega})$ is the ith source signal.

In BSS, knowledge of the mixing filter $A^{(m)}$ is not available, and each of the separated signals is always subject to an arbitrary filtering effect. In addition, the

separation vector $\mathbf{w}_i^{(m)}$ will not always converge to the ideal one. Thus, the separated subband signals will retain residues from other sources. Using $X_{\text{int}}(e^{j\omega})$ to denote the interference components and the scaler $\beta_i(e^{j\omega})$ for the attenuation caused by the overall filtering effect between the ith source and the ith receiver at frequency ω, we have

$$\hat{\mathbf{w}}_i^{(m)}\mathbf{X}(e^{j\omega}) = \beta_i(e^{j\omega})S_i(e^{j\omega}) - X_{\text{int}}(e^{j\omega}). \tag{4.38}$$

Since the analysis and the synthesis filters are derived by the same low-pass filter, we can substitute $\left|P_0(e^{j(\omega-\omega_k)}\right|^2 = G_k(e^{j\omega})F_k(e^{j\omega})$ and (4.38) into (4.37). Therefore,

$$\begin{aligned}
E_d &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_k)}\right|^2\hat{\mathbf{w}}_i^{(m)}\mathbf{X}(e^{j\omega}) - S_i(e^{j\omega})\right|^2 \mathrm{d}\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_k)}\right|^2 - 1\right)\left(\beta_i(e^{j\omega})S_i(e^{j\omega}) - X_{\text{int}}(e^{j\omega})\right)\right. \\
&\qquad \left. - X_{\text{int}}(e^{j\omega}) - (1-\beta_i(e^{j\omega}))S_i(e^{j\omega})\right|^2 \mathrm{d}\omega \\
&\le \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_k)}\right|^2 - 1\right)\left(S_i(e^{j\omega}) - X_{\text{int}}(e^{j\omega})\right)\right|^2 \\
&\quad + \left|X_{\text{int}}(e^{j\omega})\right|^2 + \left|(1-\beta_i(e^{j\omega}))S_i(e^{j\omega})\right|^2 \mathrm{d}\omega,
\end{aligned} \tag{4.39}$$

where $H(e^{j\omega})$ is the frequency response of the prototype filter, $\omega_m = 2\pi(m + 1/2)/M$, and β_i is an unknown scaling coefficient, determined by the mixing filters. Thus, the value of $\left|(1-\beta_i(e^{j\omega}))S_i(e^{j\omega})\right|^2$ is also unknown.

Now assume for a perfect separation, i.e., $\beta_i = 1$ and the interference component $\left|X_{\text{int}}(e^{j\omega})\right|^2$ is eliminated. Then only the first part of the final expression of (4.39) remains, which can be further transformed into

$$\begin{aligned}
E_{d_1} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(e^{j(\omega-\omega_m)}\right|^2 - 1\right)\left(S_i(e^{j\omega}) - X_{\text{int}}(e^{j\omega})\right)\right|^2 \\
&\le \max_{\omega}\left|S_i(e^{j\omega}) - X_{\text{int}}(e^{j\omega})\right|^2 \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{1}{N}\sum_{k=0}^{M-1}\left|P_0(e^{j(\omega-\omega_m)}\right|^2 - 1\right)\mathrm{d}\omega,
\end{aligned} \tag{4.40}$$

which forms the classical power complimentary condition for the prototype filter and E_{d_1} can be minimized by adopting the PR condition

$$\frac{1}{N}\sum_{m=0}^{M-1}\left|P_0(\mathrm{e}^{j(\omega-\omega_m)})\right|^2 = 1. \tag{4.41}$$

However, as the separating matrix $W^{(m)}$ can only be approximated by the inverse of the mixing filter at each subband subject to an arbitrary scaling function by the BSS algorithm, the assumption of $\beta_i = 1$ and $\left|X_{\text{int}}(\mathrm{e}^{j\omega})\right|^2 = 0$ is not practical and the PR condition is not really necessary in the context of subband-based BSS. Therefore, instead of removing the PR condition completely, we can adopt a relaxed condition on the passband energy of the prototype filter, given by

$$E_p = \frac{1}{N_p}\sum_{k=1}^{N_p}\left|\left|P_0(\mathrm{e}^{jw_k})\right|^2 - 1\right|^2 = \frac{1}{N_p}\sum_{k=1}^{N_p}\left|\left|\sum_{n=0}^{n=L_p-1} h_0[n]\mathrm{e}^{-j\omega_k n}\right|^2 - 1\right|^2, \tag{4.42}$$

where N_p is the number of frequency points selected, and the frequency points $[w_1, \ldots, w_{N_p}] \in (0, \pi/M)$. During optimization, only a small value for N_p is needed.

4.4.2 Subband Correlation Maximization

The relaxed PR condition requires much fewer number of constraints at the passband of the prototype filter. The extra design freedom provides chance to further reduce energy at the stopband, and also provide more space to introduce a new optimization criterion specific for the BSS application.

As mentioned, the mutual information between subbands is important to both permutation alignment in a standard BSS and successful operation of the recent proposed joint BSS. For a general fullband input signal $x[n]$, the intersubband correlation $\lambda^{(m,m+1)}(l)$ at time lag l between subband channels m and $(m+1)$ can be determined as [25]

$$\lambda^{(m,m+1)}(l) \propto E\left[h_m(n+l)\cdot h_{m+1}(n)\right] * E\left[x[n+l]\cdot x[n]\right], \tag{4.43}$$

where $*$ denotes the convolution operation and $h_m(n)$ and $h_{m+1}(n)$ are the coefficients of the mth and the $(m+1)$th analysis filters. Both the statistical property of the input signal and the filter coefficients affect the value of $\lambda^{(m,m+1)}(l)$. By carefully choosing the filter's coefficients, the value of $\lambda^{(m,m+1)}(l)$ can be increased.

In [47], the cost function for intersubband correlation is proposed, in which the correlation $\bar{r}$ over all M channels is calculated by (4.44), (4.45) and (4.46).

$$r^{(m,m+1)} = \arg \max_{l\in[-p,\ldots,p]} \left\{ |\lambda^{(m,m+1)}(l)| \right\}, \tag{4.44}$$

$$\lambda^{(m,m+1)}(l) = \frac{\sum_{n=0}^{\infty} \left[q^{(m)}[n+l]\right]\left[q^{(m+1)}[n]\right]}{\sigma_q^{(m)} \cdot \sigma_q^{(m+1)}}, \tag{4.45}$$

$$\bar{r} = \frac{1}{M-1} \sum_{m=1}^{M-1} r^{(m,m+1)}, \tag{4.46}$$

where p is a small positive integer defining the range of the time lag over which the correlation is considered, $\lambda^{(m,m+1)}(l)$ is the normalized correlation between the mth and the $(m+1)$th subbands with an offset l, $q^{(m)}[n+l]$ is the mth channel decimated signal for a general input $q[n]$ at time index $n+l$, $q[n]$ is modeled as zero-mean wide sense stationary white Gaussian, and $\sigma_q^{(m)}$ is the standard deviation of $q^{(m)}[n]$.

Because the magnitude of the normalized correlation is always smaller than 1, the objective function for minimization can be formulated as

$$\Phi_{\text{corr}} = 1 - \bar{r}. \tag{4.47}$$

For the optimized design of the GDFT prototype filter, the optimization of $p_0[n]$ is formulated in (4.48), which minimizes both the stopband energy E_s given in (4.35) and Φ_{corr}, constrained by the frequency response at the passband defined in (4.42)

$$\min_{h[n],0\le n\le L_p} \quad (1-\alpha)E_s + \alpha \cdot \Phi_{\text{corr}} \quad \text{subject to} \quad E_p < \varepsilon_p, \tag{4.48}$$

where ε_p is a small value set to be the upper limit of the passband distortion error E_p and α is the weighting factor between E_s and Φ_{corr}.

Equation (4.48) is similar to the design of CMFBs in [47]. However, for the reason stated in the previous section, the original PR condition is replaced by a soft constraint on the passband response of the prototype filter. As a result, the aliasing error is expected to be reduced significantly by replacing cosine modulation by GDFT modulation, which translates into further increased intersubband correlation, so that an improved performance in subband permutation alignment can be obtained.

Moreover, for the two components in the cost function, if we want to increase the level of cross-correlation, the stopband attenuation for the designed prototype filter has to be smaller, which may increase the aliasing level after decimation and as a result reduce the cross-correlation between the adjacent subbands after decimation. On the other hand, smaller attenuation at the stopband also undermines the assumption that after subband decomposition, the convolutive mixing problem has been transformed into an instantaneous one. One important note is that, even if we have the same PR condition, the same stopband attenuation and the same overlapped area between adjacent subbands as the existing designs without correlation maximization, the design in (4.48) will at least have an effect of redistributing the correlation value among different time lags and focusing the overall correlation at a specific time lag,

Table 4.1 Parameters of the design example for the optimized filter banks

$M = 64$	$N = 48$	$w_s = 1.96\pi/N$
$w_p = 1.9\pi/M$	$l = 2$	$L_p = 384$
$\varepsilon_p = 10^{-3}$	$\alpha = 10^{-2}$	$N_p = 4$

so that we can use the correlation at that time lag for more effective permutation alignment.

One issue with the choice of the oversampled GDFT filter banks is the values of M and N. In theory, there are mainly two factors to consider in determining the values of M and N. First, they should be large enough to make sure that after subband decomposition, the convolutive mixing problem has been transformed into a series of instantaneous mixing problems. In this case, their values are actually determined by the complexity of the unknown fullband mixing filters in the original convolutive mixing problem. However, a large value for M and N increases the computational complexity of the system and reduces the data length of the decomposed subband signals, with the latter one leading to less accurate estimation of their statistics and cross-correlation, and as a result a degraded overall performance. It is extremely difficult, if not impossible, to determine their optimum values and for now they can only be chosen empirically. The same problem exists in the frequency-domain BSS method, i.e., how to choose the right length of the DFT operation.

For oversampled GDFT filter banks, another problem is the ratio between M and N. A larger ratio M/N gives more overlapped area between adjacent subbands, and leaves more degrees of freedom for cross-correlation maximization. However, this also results in higher computational complexity for the same value of M.

4.4.3 Filter Banks Design Examples

An example prototype filter is designed based on the optimized design with the parameters listed in Table 4.1 and the resultant frequency response shown in Fig. 4.10b. For comparison, the prototype filter for conventional GDFT filter banks of $M = 64$ and $N = 48$ is also designed, and the frequency response is shown in Fig. 4.10a. The response in Fig. 4.10b has a small ripple around the passband edge, as the PR condition is relaxed. In return, it has a wider bandwidth for signal to pass and a steeper transition band before reaching the aliasing margin at π/N.

The improvement due to the new design can be evaluated using (4.36), in which the SAR is calculated. The optimized prototype filter has a ratio of 29.80 dB, while the conventional one has a ratio of 26.99 dB.

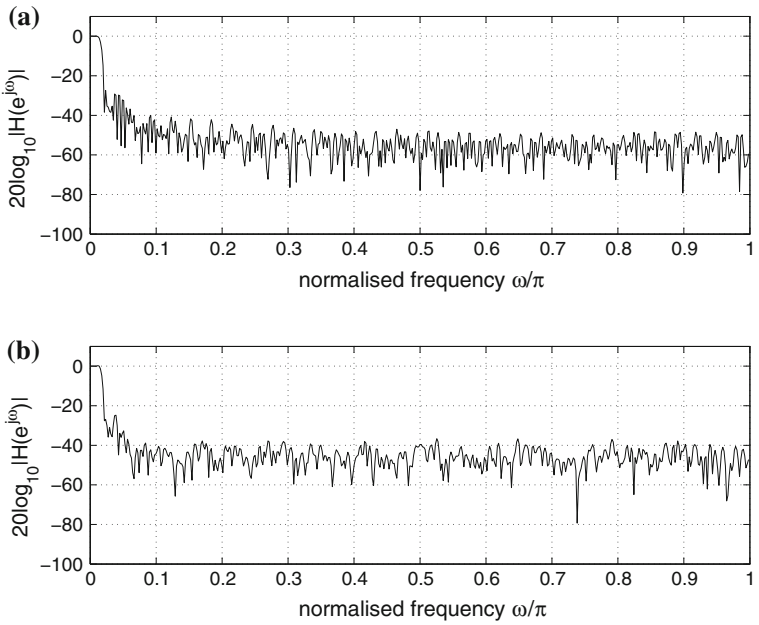


Fig. 4.10 Frequency response of two prototype filters of a 64-channel GDFT filter banks system

4.4.3.1 Filter Bank Reconstruction

Despite of the relaxation of the PR condition, the BSS-optimized design can still retain the original waveform of the input to a large degree. This is illustrated in Fig. 4.11, where a unit impulse was the input of the oversampled GDFT filter banks, and the prototype filter in Fig. 4.10b was used. Because the length of the prototype filter is $L_p = 384$, the filter banks have introduced a delay of L_p, after which the magnitude attenuated impulse can be observed.

4.4.3.2 Intersubband Correlation

For the permutation alignment approach based on intersubband correlation, the correlation value is calculated between adjacent subband signals, where the combination with largest correlation is chosen. The result of correlation optimization can be demonstrated by considering the simplest problem with two source signals ($N_s = 2$). In order to show the improvement in alignment, two source signals s_1 and s_2 without mixing are used. The difference between subband correlation for two different combinations is used for evaluation, which is defined as

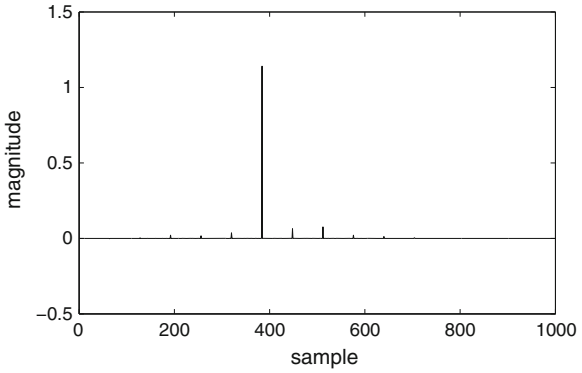


Fig. 4.11 The output of the optimized GDFT filter banks for an unit impulse as the input

$$\begin{aligned}\Theta_{\text{match}}^{(m)} &= \text{corr}\,(s_1^{(m)}, s_1^{(m+1)}) + \text{corr}\,(s_2^{(m)}, s_2^{(m+1)}),\\ \Theta_{\text{unmat}}^{(m)} &= \text{corr}\,(s_1^{(m)}, s_2^{(m+1)}) + \text{corr}\,(s_2^{(k)}, s_1^{(m+1)}),\\ \Theta_s^{(m)} &= \Theta_{\text{match}}^{(m)} - \Theta_{\text{unmat}}^{(m)},\end{aligned} \tag{4.49}$$

for $m = 0, \ldots, M-2$, where Θ_{match} represents the intersubband correlation of the matched subband signals, Θ_{unmat} represents the intersubband correlation of mismatched subband signals, and corr $(\cdot)$ calculates the normalized cross-correlation, as defined by (4.43). $\Theta_s^{(m)}$ for $m = 0, \ldots, M-2$ can be viewed as a special case in the process of permutation alignment when the BSS algorithm has fully recovered the source signals at each subband, and for correct alignments, $\Theta_s^{(m)}$ should be larger than zero.

Figure 4.12 shows the results for the optimized GDFT filter banks with two speech signals as the sources. For comparison, results based on conventional GDFT filter banks are also shown in Fig. 4.12, where the same sets of signals are applied. The result shows that correct permutation alignment can be achieved for the subband signals if complete separation can be achieved. From Fig. 4.12, we can clearly observe an improvement for the optimized design. When the optimized filter banks are used, $\Theta_s^{(m)}$ is always larger than 0.13; in contrast, the intersubband correlations using the conventional design drop below 0.1 and become very close to zero for some subbands. While a positive value of $\Theta_s^{(m)}$ guarantees correct permutation alignment, a large positive value of $\Theta_s^{(m)}$ will provide a margin of safety, which becomes important when subband separation is not achieved completely.

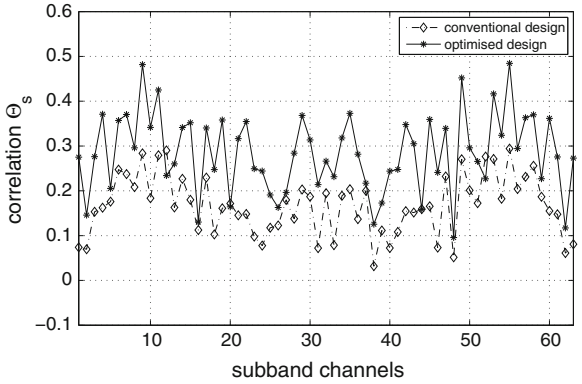


Fig. 4.12 Intersubband correlation of speech signals in two GDFT filter banks systems

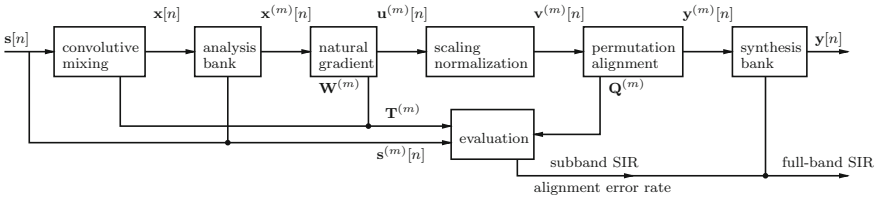


Fig. 4.13 Structure of the simulation process of the subband-based BSS system

4.5 Simulation Results

In this section, MATLAB-based experiments are performed to simulate the subband convolutive BSS system, as shown in Fig. 4.13. The source signals $\mathbf{s}[n]$ are recorded speech signals sampled at 8 KHz, which are mixed by convolution operations with mixing FIR filters. They are frequency decomposed at the analysis bank, where GDFT-modulated filter banks are considered.

The prototype filter for the GDFT filter banks is designed by the BSS-optimized method, with parameters listed at Table 4.1 and the frequency response is shown in Fig. 4.10b. It is compared with the conventional design using the method in [12].

The modified iterative natural gradient algorithm is applied to each subband channel to obtain the separating matrices [10]. Since the subband signals are complex-valued, the following nonlinear function is used [55]

$$\varphi(\mathbf{u}^{(k)}[p]) = \left[\frac{u_1^{(k)}[p]}{\left|u_1^{(k)}[p]\right|}, \ldots, \frac{u_{N_s}^{(k)}[p]}{\left|u_{N_s}^{(k)}[p]\right|} \right]^T . \tag{4.50}$$

For each subband, to mitigate the scaling ambiguity problem, the minimal distortion principle explained in Sect. 4.3.1.1 is used. For the permutation ambiguity problem, we will align the subbands by the correlation-based approach.

4.5.1 Evaluation Criteria

Because the original signals and the mixing filters are all known in our simulation process, we can quantitatively evaluate the separation performance. The signal to interference ratio (SIR) at the mth subband for each separated signal is calculated using the following equation

$$\mathrm{SIR}^{(m)} = \frac{\sum_{i=1}^{N_s} \left\| t_{ii}^{(m)} \cdot s_i^{(m)} \right\|^2}{\sum_{i=1}^{N_s} \sum_{j=1, j\neq i}^{N_s} \left\| t_{ij}^{(m)} \cdot s_j^{(m)} \right\|^2}, \tag{4.51}$$

where $s_i^{(m)}$ is obtained by passing the ith source signal through the mth analysis filter, and $t_{ji}^{(m)}$ is the (j, i)-th entry of the combined mixing-demixing matrix $\mathbf{A}^{(m)}$ for the mth subband. We also consider the SIR at each of the outputs, and for the ith output,

$$\mathrm{SIR}_i^{(m)} = \frac{\left\| t_{ii}^{(m)} \cdot s_i^{(m)} \right\|^2}{\sum_{i=1}^{N_s} \sum_{j=1, j\neq i}^{N_s} \left\| t_{ij}^{(m)} \cdot s_j^{(m)} \right\|^2}, \tag{4.52}$$

To evaluate the permutation alignment result, we use the source signals as reference, which are decomposed into M subbands by the same filter banks. Misalignment between the reference and separated signals at adjacent subband means a permutation alignment error has occurred.

4.5.2 Three-Microphone Three-Source Scenario

In our simulation, we consider a BSS problem with three speakers and three receivers. Two sets of mixing filters are randomly generated, which are 10 and 20 tap long, respectively, as shown in Figs. 4.14 and 4.15. The speech signals shown in Fig. 4.16 are used as source signals, which are 7 s long and sampled at 8 KHz.

Figures 4.17 and 4.18 show the subband SIR result when the mixing filters of length 10 were used, and both the optimized and the conventional GDFT filter banks produced good results at lower frequencies, where an SIR level at around 5 dB can be achieved. However, as we can see at subband $m = 21$ of the conventional system, misalignment occurs and the subband SIRs for the three

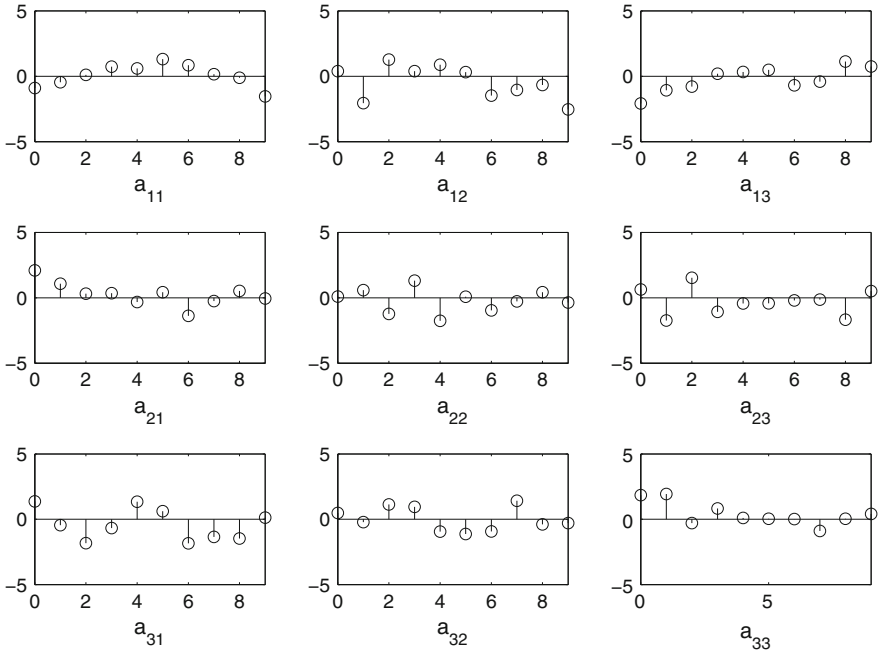


Fig. 4.14 Impulse responses of the mixing filters for a three-speaker–three-receiver system, and each filter is 10 tap long

outputs were $(-20.1, 5.24, -4.72)$ dB. By comparing the separated signals with the source signals, we can obtain the SIRs for correct permutation, which are $(18.05, 5.24, 2.07)$ dB. As demonstrated by the examples in Sect. 4.4, when the subband SIRs are at a low level, i.e., 2.07 dB in this case, a misalignment may occur because of the presence of interference signals. Similarly for $m = 22$, the subband SIRs are $(-23.09, -7.07, -20.20)$ dB and the fixed SIRs are $(21.19, 9.40, 1.42)$ dB, which caused a second misalignment because of the same reason.

In contrast, when the optimized design was used, the subband SIR were $(12.75, 7.63, 4.2)$ dB and $(17.07, 6.69, 2.84)$ dB; correct alignment is obtained as the optimized design can enhance the inter-subband correlation between matched subband signals.

In Figs. 4.19 and 4.20, mixing filters of length $= 20$ were used for the convolutive mixtures. As the mixing filters become longer with a more complicated frequency response, the separation is expected to become more difficult as the number of estimated coefficients has been increased. This change can be observed from the two figures, as the subband separation performance become worse and more misalignments are present for both GDFT filter banks. For the conventional design, the misalignment appeared at $m = (5, 9, 14, 15, 23, 40, 48, 49, 54, 58)$, and permutation errors propagate between these subbands, which severely distorted the separation results. However, as the optimized design is more robust to the reduction

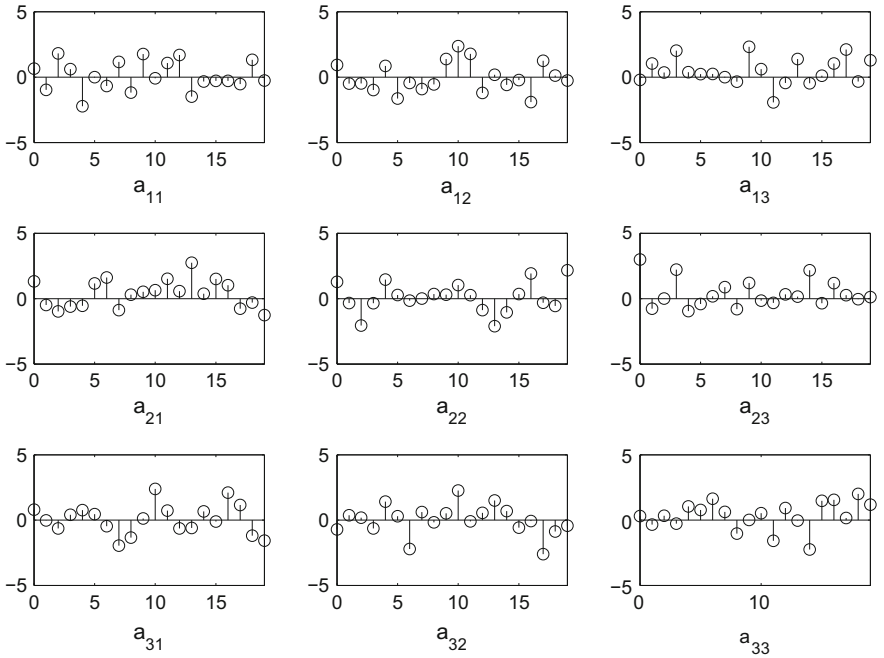


Fig. 4.15 Impulse responses of the mixing filter for a three-speaker–three-receiver system, and each filter is 20 tap long

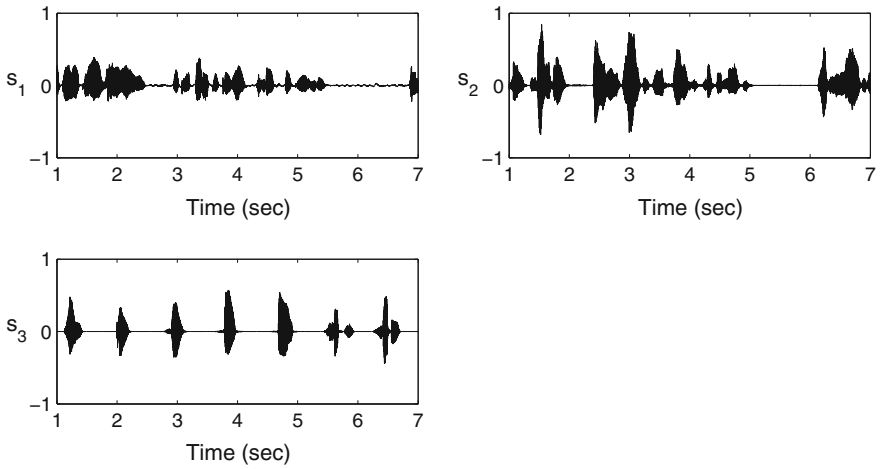


Fig. 4.16 The three speech signals used

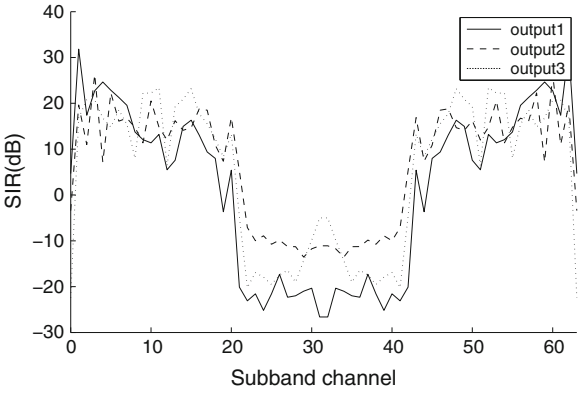


Fig. 4.17 The subband SIR for the three outputs using the conventional oversampled GDFT modulated filter banks for mixing filters of length = 10

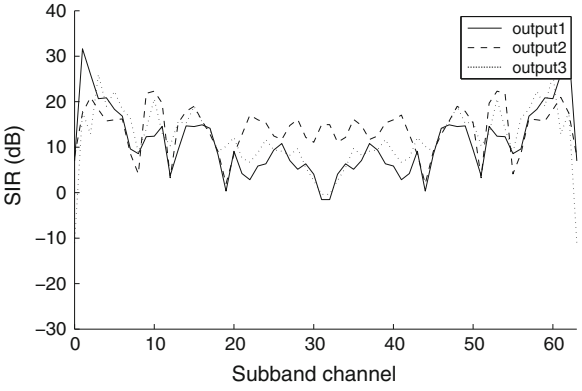


Fig. 4.18 The subband SIR for the three outputs using the optimized oversampled GDFT modulated filter banks for mixing filters of length = 10

of subband SIR, misalignment only occurred at subbands with lowest SIR (at $m = 15, 23, 40, 48$). Since the first misalignment occurred at a higher subband, and the energy of speech signals is normally focused on lower frequencies, the impact of permutation errors is less significant.

The fullband overall SIR values can be obtained by passing the subband components $t_{ii}^{(m)} \cdot s_i^{(m)}$ and $t_{ij}^{(m)} \cdot s_j^{(m)}$ through the synthesis filters and calculating the ratio after the summation. Table 4.2 summarizes the results.

The optimized and the conventional GDFT filter banks have similar results when the subband SIRs are high, where permutation alignment can be correctly achieved. However, when the separation difficulty increases, which is usually because of the changes in the source signals or the mixing filters, permutation misalignment may

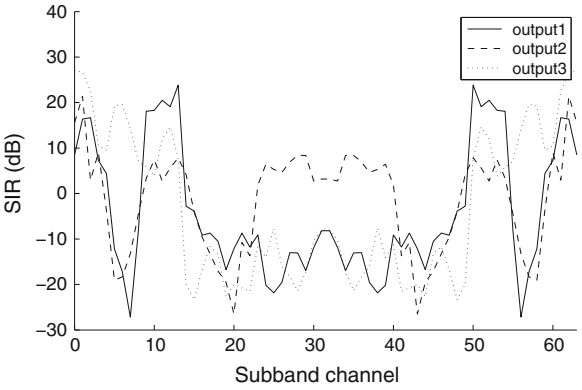


Fig. 4.19 The subband SIR for the three outputs using the conventional oversampled GDFT filter banks for mixing filters of length = 20

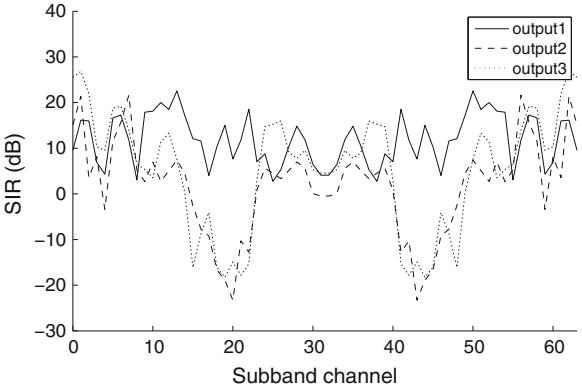


Fig. 4.20 The subband SIR for the three outputs using the optimized oversampled GDFT filter banks for mixing filters of length = 20

occur. For the optimized design, it is quite robust and correct alignment can still be obtained even when the subband SIR is relatively low.

Permutation error is a good performance indicator of the subband-based BSS, and when the subband system has zero permutation errors, the BSS algorithm can reach its full potential; while the optimized prototype filter is designed to improve the alignment result, misalignment may still occur at a few subbands. Although occurrence of the misalignment is hardly predicted and a few misalignment may propagate to a larger number of subbands, the optimized design can generally improve the overall separation result. If the misalignment occurs at higher frequencies, the main components of the source speech signals can still be recovered correctly.

Table 4.2 Simulation results: averaged fullband SIR values for each output, the number of permutation errors and permutation misalignments

	SIR1 (dB)	SIR2 (dB)	SIR3 (dB)	Permutation error	Misalignment
Mixing tap = 10					
Conventional GDFT	12.55	15.1	17.9	24/64	4/64
Optimized GDFT	20.4	16.8	17.24	0	0
Mixing tap = 20					
Conventional GDFT	2.98	1.99	13.22	20/64	10/64
Optimized GDFT	12.56	6.33	13.53	16/64	2/64

4.6 Chapter Summary

In this chapter, we have studied the subband-based BSS problem and methods for removing permutation ambiguity after frequency/subband decomposition. We first reviewed the fundamental ideas for filter banks design, including minimization of the stopband energy and the PR condition. The scaling indeterminacy of the BSS problem were analyzed, and it has been shown that the benefit brought by the PR condition is quite limited in the context of BSS. Thus, in Sect. 4.4, a relaxed PR condition was employed in the filter banks design, which only considers a small number of frequencies in the passband. The resultant additional design freedom can be exploited by adopting a recently proposed design criterion based on intersubband correlation maximization. The optimized GDFT filter banks have increased the cross-correlation between matched subband signals. This improvement is especially useful when there are still interferences present in separated subband signals. Simulation results have shown clear improvement in the robustness of the alignment process when the optimized GDFT filter banks were used, resulting in a reduced number of alignment errors and a much improved overall separation performance.

References

1. Akansu, A., Haddad, R.: Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets. Academic Press, Boston (1992)
2. Amari, S.: Natural gradient works efficiently in learning. Neural Comput. **10**, 251–276 (1998)
3. Amari, S., Douglas, S., Cichocki, A., Yang, H.: Multichannel blind deconvolution and equalization using the natural gradient. In: First IEEE Signal Processing Workshop on Signal Processing Advances in, Wireless Communications, pp. 101–104 (1997)
4. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Comput. **7**(6), 1129–1159 (1995)
5. Bellanger, M., Daguet, J.: TDM-FDM transmultiplexer: digital polyphase and FFT. IEEE Trans. Commun. **22**(9), 1199–1205 (1974)
6. Chan, S.C., Liu, W., Ho, K.L.: Multiplier-less perfect reconstruction modulated filter banks with sum-of-powers-of-two coefficients. IEEE Signal Process. Lett. **8**, 163–166 (2001)
7. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing. Wiley, New York (2003)

8. Comon, P.: Independent component analysis, a new concept? Signal Process. **36**(3), 287–314 (1994)
9. Crochiere, R.E., Rabiner, L.R.: Multirate Digital Signal Processing. Prentice Hall, Englewood Cliffs (1983)
10. Douglas, S., Gupta, M.: Scaled natural gradient algorithms for instantaneous and convolutive blind source separation. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. II-637–640 (2007)
11. Grbic, N., Tao, X.J., Nordholm, S., Claesson, I.: Blind signal separation using overcomplete subband representation. IEEE Trans. Speech Audio Process. **9**(5), 524–533 (2001)
12. Harteneck, M., Weiss, S., Stewart, R.W.: Design of near perfect reconstruction oversampled filter banks for subband adaptive filters. IEEE Trans. Circuits Syst. II: Analog Digital Signal Process. **46**, 1081–1085 (1999)
13. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3–4), 321–377 (1936)
14. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
15. Hyvärinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Comput. **9**, 1483–1492 (1997)
16. Ikram, M., Morgan, D.: Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 1041–1044. Istanbul, Turkey (2000)
17. Ikram, M., Morgan, D.: Permutation inconsistency in blind speech separation: investigation and solutions. IEEE Trans. Speech Audio Process. **13**(1), 1–13 (2005)
18. Karp, T., Fliege, N.: Modified DFT filter banks with perfect reconstruction. IEEE Trans. Circuits Syst. II: Analog Digital Signal Process. **46**(11), 1404–1414 (1999)
19. Kettenring, J.R.: Canonical analysis of several sets of variables. Biometrika **58**(3), 433–451 (1971)
20. Kim, T.: Real-time independent vector analysis for convolutive blind source separation. IEEE Trans. Circuits Syst. I Regul. Pap. **57**(7), 1431–1438 (2010)
21. Kim, T., Attias, H.T., Lee, S.Y., Lee, T.W.: Blind source separation exploiting higher-order frequency dependencies. IEEE Trans. Audio Speech Lang. Process. **15**(1), 70–79 (2007)
22. Kim, T., Lee, I., Lee, T.W.: Independent vector analysis: definition and algorithms. In: Fortieth Asilomar Conference on Signals, Systems and Computers (ACSSC'06), pp. 1393–1396 (2006)
23. Kurita, S., Saruwatari, H., Kajita, S., Takeda, K., Itakura, F.: Evaluation of blind signal separation method using directivity pattern under reverberant conditions. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'00), vol. 5, pp. 3140–3143 (2000)
24. Lee, J.H., Lee, T.W., Jolesz, F.A., Yoo, S.S.: Independent vector analysis (IVA): multivariate approach for fMRI group study. NeuroImage **40**(1), 86–109 (2008)
25. Lee, K.A., Gan, W.S.: On the subband orthogonality of cosine-modulated filter banks. IEEE Trans. Circuits Syst. II Express Briefs **53**(8), 677–681 (2006)
26. Lee, K.A., Gan, W.S., Kuo, S.M.: Subband Adaptive Filtering: Theory and Implementation. Wiley, New York (2009)
27. Li, X.L., Adalı, T., Anderson, M.: Joint blind source separation by generalized joint diagonalization of cumulant matrices. Signal Process. **91**(10), 2314–2322 (2011)
28. Li, Y.O., Adalı, T., Wang, W., Calhoun, V.D.: Joint blind source separation by multi-set canonical correlation analysis. IEEE Trans. Signal Process. **57**(10), 3918–3929 (2009)
29. Li, Y.O., Eichele, T., Calhoun, V., Adalı, T.: Group study of simulated driving fMRI data by multiset canonical correlation analysis. J. Signal Process. Syst. **68**(1), 31–48 (2012)
30. Liu, W.: Blind beamforming for multi-path wideband signals based on frequency invariant transformation. In: Proceedings of the International Symposium on Communications, Control and Signal Processing (ISCCSP). Limassol, Cyprus (2010)
31. Liu, W.: Wideband beamforming for multi-path signals based on frequency invariant transformation. Int. J. Autom. Comput. **9**, 420–428 (2012)

32. Liu, W., Mandic, D.P.: Semi-blind source separation for convolutive mixtures based on frequency invariant transformation. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 285–288. Philadelphia, USA (2005)
33. Liu, W., Mandic, D.P., Cichocki, A.: A class of novel blind source extraction algorithms based on a linear predictor. In: Proceedings of IEEE International Symposium on Circuits and Systems, pp. 3599–3602. Kobe, Japan (2005)
34. Liu, W., Mandic, D.P., Cichocki, A.: Blind second-order source extraction of instantaneous noisy mixtures. IEEE Trans. Circuits Syst. II Express Briefs **53**(9), 931–935 (2006)
35. Liu, W., Mandic, D.P., Cichocki, A.: Blind source extraction of instantaneous noisy mixtures using a linear predictor. In: Proceedings of IEEE International Symposium on Circuits and Systems, pp. 4199–4202. Kos, Greece (2006)
36. Liu, W., Mandic, D.P., Cichocki, A.: Analysis and online realization of the cca approach for blind source separation. IEEE Trans. Neural Networks **18**(5), 1505–1510 (2007)
37. Liu, W., Mandic, D.P., Cichocki, A.: Blind source extraction based on a linear predictor. IET Signal Process. **1**(1), 29–34 (2007)
38. Liu, W., Mandic, D.P., Cichocki, A.: Blind source separation based on generalised canonical correlation analysis and its adaptive realization. In: Proceedings of International Congress on Image and Signal Processing, vol. 5, pp. 417–421. Hainan, China (2008)
39. Liu, W., Mandic, D.P., Cichocki, A.: A dual-linear predictor approach to blind source extraction for noisy mixtures. In: Proceedings of IEEE Workshop on Sensor Array and Multichannel Signal Processing, pp. 515–519. Darmstadt, Germany (2008)
40. Liu, W., Weiss, S.: Wideband Beamforming: Concepts and Techniqeus. Wiley, Chichester, UK (2010)
41. Low, S.Y., Nordholm, S., Togneri, R.: Convolutive blind signal separation with post-processing. IEEE Trans. Speech Audio Process. **12**(5), 539–548 (2004)
42. Mazur, R., Mertins, A.: An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models. IEEE Trans. Acoustics Speech Lang. Process. **17**(1), 117–126 (2009)
43. Murata, N., Ikeda, S., Ziehe, A.: An approach to blind source separation based on temporal structure of speech signals. Neurocomputing **41**(1–4), 1–24 (2001)
44. Park, H.M., Dhir, C.S., Oh, S.H., Lee, S.Y.: A filter bank approach to independent component analysis for convolved mixtures. Neurocomputing **69**(16–18), 2065–2077 (2006)
45. Peng, B., Liu, W., Mandic, D.P.: An improved solution to the subband blind source separation permutation problem based on optimized filter banks. In: Proceedings of the International Symposium on Communications, Control and Signal Processing, pp. 1–4. Limassol, Cyprus (2010)
46. Peng, B., Liu, W., Mandic, D.P.: Novel design of oversampled GDFT filter banks for application to subband based blind source separation. In: Proceedings of the IEEE Statistical Signal Processing Workshop (SSP), pp. 637–640. Nice, France (2011)
47. Peng, B., Liu, W., Mandic, D.P.: Reducing permutation error in subband-based convolutive blind separation. IET Signal Process. **6**(1), 34–44 (2012)
48. Peng, B., Liu, W., Mandic, D.P.: Design of oversampled generalised discrete fourier transform filter banks for application to subband-based blind source separation. IET Signal Process. **7**(9), 843–853 (2013)
49. Peng, B., Liu, W., Mandic, D.P.: Subband-based joint blind source separation for convolutive mixtures employing m-cca. In: Proceedings of the Constantinides International Workshop on Signal Processing. London, UK (2013)
50. Ramstad, T.A., Tanem, J.P.: Cosine-modulated analysis-synthesis filterbank with critical sampling and perfect reconstruction. In: Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, pp. 1789–1792 (1991)
51. Saruwatari, H., Kawamura, T., Shikano, K.: Blind source separation based on fast-convergence algorithm using ica and array signal processing. IEEE Trans. Audio Speech Lang. Process. **14**, 666–678 (2001)

52. Sathe, V., Vaidyanathan, P.: Effects of multirate systems on the statistical properties of random signals. IEEE Trans. Signal Process. **41**(1), 131 (1993).
53. Sawada, H., Araki, S., Makino, S.: Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS. In: IEEE International Symposium on Circuits and Systems (ISCAS 2007), pp. 3247–3250 (2007)
54. Sawada, H., Araki, S., Makino, S.: Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. IEEE Trans. Audio Speech Lang. Process. **19**(3), 516–527 (2011)
55. Sawada, H., Mukai, R., Araki, S., Makino, S.: Polar coordinate based nonlinear function for frequency-domain blind source separation. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, pp. I-1001–I-1004 (2002)
56. Sawada, H., Mukai, R., Araki, S., Makino, S.: A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Trans. Speech Audio Process. **12**(5), 530–538 (2004)
57. Smaragdis, P.: Blind separation of convolved mixtures in the frequency domain. Neurocomputing **22**(1–3), 21–34 (1998)
58. Vaidyanathan, P.P.: Multirate Systems and Filter Banks. Prentice Hall, Englewood Cliffs (1993)
59. Vinjamuri, R., Crammond, D., Kondziolka, D., Lee, H.N., Mao, Z.H.: Extraction of sources of tremor in hand movements of patients with movement disorders. IEEE Trans. Inf. Technol. Biomed. **13**(1), 49–56 (2009)
60. Wang, L., Ding, H., Yin, F.: A region-growing permutation alignment approach in frequency-domain blind source separation of speech mixtures. IEEE Trans. Audio Speech Lang. Process. **19**(3), 549–557 (2011)
61. Wang, W., Sanei, S., Chambers, J.: Penalty function based joint diagonalisation approach for convolutive blind separation of nonstationary sources. IEEE Trans. Signal Process. **53**(5), 1654–1669 (2005)
62. Weiss, S., Stewart, R.W.: On Adaptive Filtering in Oversampled Subbands. Shaker Verlag, Aachen (1998)
63. Weiss, S., Stewart, R.W., Stenger, A., Rabenstein, R.: Performance limitations of subband adaptive filters. In: Proceedings of EUSIPCO, vol. III, pp. 1245–1248 (1998)
64. Wilbur, M.R., Davidson, T.N., Reilly, J.P.: Efficient design of oversampled NPR GDFT filterbanks. IEEE Trans. Signal Process. **52**(7), 1947–1963 (2004)

Chapter 5
Frequency Domain Blind Source Separation Based on Independent Vector Analysis with a Multivariate Generalized Gaussian Source Prior

Yanfeng Liang, Syed Mohsen Naqvi, Wenwu Wang and Jonathon A. Chambers

Abstract Independent vector analysis (IVA) is designed for retaining the dependency contained in each source vector, while removing the dependency between different source vectors during the source separation process. It can theoretically avoid the permutation problem inherent to independent component analysis (ICA). The dependency in each source vector is maintained by adopting a multivariate source prior instead of a univariate source prior. In this chapter, a multivariate generalized Gaussian distribution is proposed to be the source prior, which can exploit the energy correlation within each source vector. It can preserve the dependency between different frequency bins better to achieve an improved separation performance, and is suitable for the whole family of IVA algorithms. Experimental results on real speech signals confirm the advantage of adopting the new source prior on three types of IVA algorithms.

5.1 Introduction

Blind source separation (BSS) aims to separate specific signals from observed mixtures with very limited prior knowledge, and has been researched over recent decades and has wide potential applications, such as in biomedical signal processing, image

Y. Liang (✉) · S. M. Naqvi · J. A. Chambers
Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK
e-mail: y.liang2@lboro.ac.uk

S. M. Naqvi
e-mail: S.M.R.Naqvi@lboro.ac.uk

J. A. Chambers
e-mail: J.A.Chambers@lboro.ac.uk

Wenwu Wang
University of Surrey, Guildford, Surrey, GU2 7XH, UK
e-mail: W.Wang@surrey.ac.uk

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_5,
© Springer-Verlag Berlin Heidelberg 2014

processing, speech processing, and communication systems [1, 2]. A classical BSS problem is the machine cocktail party problem, which was proposed by Colin Cherry in 1953 [3, 4]. His drive was a machine to mimic the ability of a human to extract a target speech signal from microphone measurements acquired in a room environment.

In order to solve the BSS problem, a statistical signal processing method, i.e., independent component analysis (ICA), is proposed to exploit the non-Gaussianity of the signals [5]. It works efficiently to solve the instantaneous BSS problem. However, the problem becomes convolutive BSS problem in a room environment due to the reflections from the ceiling, floor, and walls. The length of the room impulse response is typically on the order of thousands of samples, which leads to huge computational cost when using time domain methods. Therefore, frequency domain methods have been proposed to reduce the computational cost due to the convolution operation in the time domain becomes multiplication in the frequency domain provided the block length of the transform is substantially larger than the length of the time domain filter [6, 7]. When the mixtures are transformed into the frequency domain by using the discrete Fourier transform (DFT), the instantaneous ICA can be applied in each frequency bin to separate the signals. However, the permutation ambiguity inherent to ICA becomes more severe because of the potential misalignment of the separated sources at different frequency bins. In this case, when the separated sources are transformed back to the time domain, the separation performance will be poor. Therefore, various methods have been proposed to mitigate the permutation problem [7]. However, most of them use extra information such as source geometry or prior knowledge of the source structure, and pre or post processing is needed for all of these methods which introduces additional complexity and delay.

Recently, independent vector analysis (IVA) has been proposed to solve the permutation problem naturally during the learning process without any pre or post processing [8]. It can theoretically avoid the permutation problem by retaining the dependency in each individual source vector while removing the dependency between the source vectors of different signals [9, 10]. The main difference between ICA algorithms and IVA algorithms is the nonlinear score function. For conventional ICA algorithms, the nonlinear score function is a univariate function which only uses the data in each frequency bin to update the unmixing matrix. However, the nonlinear score function for IVA is a multivariate function, which can use the data from all the frequency bins. Therefore, it can exploit the inter-frequency dependencies to mitigate the permutation problem.

There are three state-of-the-art types of IVA algorithms, which are the natural gradient IVA (NG-IVA), the fast fixed-point IVA (FastIVA) and the auxiliary function based IVA (AuxIVA). NG-IVA adopts the Kullback-Leibler divergence between the joint probability density function and the product of marginal probability density functions of the individual source vectors as the cost function, and the natural gradient method is used to minimize the cost function [9]. FastIVA is a fast form of IVA which uses Newton's method to update the unmixing matrix [11]. AuxIVA uses the auxiliary function technique to converge quickly without introducing tuning parameters and can guarantee the objective function decreases monotonically [12]. There are also

several other IVA algorithms, which are based on these three IVA algorithms. The adaptive step size IVA algorithm, which is based on the NG-IVA algorithm, can automatically select the step size to achieve a faster convergence [13]. The audio-video based IVA method combines video information with FastIVA to obtain a faster and better separation performance in noisy and reverberant room environments [14]. And IVA methods which exploit the activity and dynamic structure of the sources to achieve improved separation performance have also been proposed [15, 16].

The nonlinear score function of IVA is used to preserve inter-frequency dependencies for individual sources [9]. Because the nonlinear score function is derived from the source prior, an appropriate source model is needed. For the original IVA algorithms, a multivariate Laplace distribution is adopted as the source prior. It is a spherically symmetric distribution, which implies the dependencies between different frequency bins are all the same. In order to describe a variable dependency structure, a chain-type overlapped source model has been proposed [17]. Similarly, a harmonic structure dependency model has been proposed [18]. A Gaussian mixture model can also be adopted as the source prior, whose advantage is that it enables the IVA algorithms to separate a wider class of signals [19, 20]. Most of these source models assume the covariance matrix of each source vector is an identity matrix due to the orthogonal Fourier basis. This implies that there is no second order correlation between different frequency bins. Although recently a multivariate Gaussian source prior has been proposed to introduce the second order correlation [21], it is only applicable when there are large second order correlations such as in functional magnetic resonance imaging (fMRI) studies. For the frequency domain IVA algorithms, higher order correlation information between different frequency bins is still missing and should be exploited.

In this chapter, a multivariate generalized Gaussian distribution is adopted as the source prior. It has heavier tails compared with the original multivariate Laplacian distribution, which makes the IVA algorithms derived from it more robust to outliers. It can also preserve the dependency across different frequency bins in a similar way as when the original multivariate Laplacian distribution is used to derive an IVA algorithm. Moreover, the nonlinear source function derived from this new source prior can additionally introduce the energy correlation within each source vector. Therefore, it contains more informative dependency structure and can thereby better preserve the dependencies between different frequency bins to achieve an improved separation performance.

The structure of this chapter is as follows. In Sect. 5.2, the original IVA is introduced. In Sect. 5.3, the energy correlation within a frequency domain speech signal is introduced. Then a multivariate generalized Gaussian distribution is proposed to be the source prior and analyzed in Sect. 5.4. Three types of IVA algorithms with the proposed source prior are discussed in Sect. 5.5. The experimental results are shown in Sect. 5.6, and finally the conclusions are drawn in Sect. 5.7.

5.2 Independent Vector Analysis

In this chapter, we mainly focus on the IVA algorithms used in the frequency domain. The noise-free model in the frequency domain is described as:

$$\mathbf{x}(k) = \mathbf{H}(k)\mathbf{s}(k) \tag{5.1}$$

$$\hat{\mathbf{s}}(k) = \mathbf{W}(k)\mathbf{x}(k) \tag{5.2}$$

where $\mathbf{x}(k) = [x_1(k), \ldots, x_m(k)]^T$ is the observed signal; $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ is the source signal; $\hat{\mathbf{s}}(k) = [\hat{s}_1(k), \ldots, \hat{s}_n(k)]^T$ is the estimated signal. They are all in the frequency domain and $(\cdot)^T$ denotes vector transpose. The index $k = 1, 2, \ldots, K$ denotes the k-th frequency bin, and K is the number of frequency bins; m is the number of microphones and n is the number of sources. $\mathbf{H}(k)$ is the mixing matrix whose dimension is $m \times n$, and $\mathbf{W}(k)$ is the unmixing matrix whose dimension is $n \times m$. In this chapter, we assume that the number of sources is the same as the number of microphones, i.e., $m = n$.

Independent vector analysis is proposed to avoid the permutation problem by retaining the inter-frequency dependencies for each source while removing the dependencies between different sources. It theoretically ensures that the alignment of the separated signals are consistent across the frequency bins. The IVA method adopts the Kullback-Leibler divergence [9] between the joint probability density function $p(\hat{\mathbf{s}}_1 \ldots \hat{\mathbf{s}}_n)$ and the product of marginal probability density functions of the individual source vectors $\prod q(\hat{\mathbf{s}}_i)$ as the cost function

$$\begin{aligned} J &= KL\Big(p\left(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n\right) || \prod q\left(\hat{\mathbf{s}}_i\right)\Big) \\ &= \int p\left(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n\right) \log \frac{p\left(\hat{\mathbf{s}}_1 \cdots \hat{\mathbf{s}}_n\right)}{\prod q\left(\hat{\mathbf{s}}_i\right)} d\hat{\mathbf{s}}_1 \cdots d\hat{\mathbf{s}}_n \\ &= \text{const} - \sum_{k=1}^{K} \log |\det\left(\mathbf{W}(k)\right)| - \sum_{i=1}^{n} E\left[\log q\left(\hat{\mathbf{s}}_i\right)\right] \end{aligned} \tag{5.3}$$

where $E[\cdot]$ denotes the statistical expectation operator, and $\det(\cdot)$ is the matrix determinant operator. The dependency between different source vectors should be removed but the inter-relationships between the components within each vector can be retained, when the cost function is minimized. These inter-frequency dependencies are modelled by the probability density function of the source.

Traditionally, the scalar Laplacian distribution is widely used as the source prior for the frequency domain ICA-based approaches. The resultant nonlinear score function is a univariate function, which can not preserve the inter-frequency dependencies because it is only associated with each individual frequency bin. In order to keep the inter-frequency dependencies of each source vector, a multivariate Laplacian distribution is adopted as the source prior for IVA, which can be written as

$$q(\mathbf{s}_i) \propto \exp\left(-\sqrt{(\mathbf{s}_i - \boldsymbol{\mu}_i)^{\dagger}\Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)}\right) \tag{5.4}$$

where $(\cdot)^{\dagger}$ denotes the Hermitian transpose, $\boldsymbol{\mu}_i$ and Σ_i are respectively the mean vector and covariance matrix of the i-th source. Then the nonlinear score function can be derived according to this source prior. We assume that the mean vector is a zero vector and the covariance matrix is a diagonal matrix due to the orthogonality of the Fourier basis, which implies that each frequency bin sample is uncorrelated with the others. As such, the resultant nonlinear score function to extract the i-th source at the k-th frequency bin can be obtained as:

$$\begin{aligned}\varphi^{(k)}\left(\hat{s}_i(1)\cdots\hat{s}_i(k)\right) &= -\frac{\partial \log\left(q\left(\hat{s}_i(1)\cdots\hat{s}_i(k)\right)\right)}{\partial \hat{s}_i(k)} \\ &= \frac{\partial\sqrt{\sum_{k=1}^{K}\left|\frac{\hat{s}_i(k)}{\sigma_i(k)}\right|^2}}{\partial \hat{s}_i(k)} = \frac{\hat{s}_i(k)}{(\sigma_i(k))^2\sqrt{\sum_{k=1}^{K}\left|\frac{\hat{s}_i(k)}{\sigma_i(k)}\right|^2}}\end{aligned} \tag{5.5}$$

where $\sigma_i(k)$ denotes the standard deviation of the i-th source at the k-th frequency bin. This is a multivariate function, and the dependency between the frequency bins is thereby accounted for in learning. When the natural gradient method is used to minimize the cost function, the unmixing matrix update equation is:

$$\Delta\mathbf{W}(k) = \left(\mathbf{I} - E\left[\left(\varphi^{(k)}(\hat{\boldsymbol{s}})\right)\hat{\mathbf{s}}^{*}(k)\right]\right)\mathbf{W}(k) \tag{5.6}$$

where $\mathbf{I}$ is the identity matrix, and $(\cdot)^{*}$ denotes the conjugate operators. $\varphi^{(k)}(\hat{\boldsymbol{s}})$ is the nonlinear score function

$$\varphi^{(k)}(\hat{\boldsymbol{s}}) = \left[\varphi^{(k)}(\hat{\boldsymbol{s}}_{\mathbf{1}}), \ldots, \varphi^{(k)}(\hat{\boldsymbol{s}}_{\boldsymbol{n}})\right]^{T} \tag{5.7}$$

5.3 Energy Correlation Within a Frequency Domain Speech Signals

In the derivation of the original IVA algorithms little attention was focused upon the correlation information between different frequency bins due to the orthogonal Fourier basis. However, the higher order information exists and could be introduced to exploit the dependency between different frequency bins and better preserve the inter-frequency dependency. The correlation of squares of components is discussed in [22], which can be used to exploit the dependency between different components. For the frequency domain speech signals, the energy correlation between different

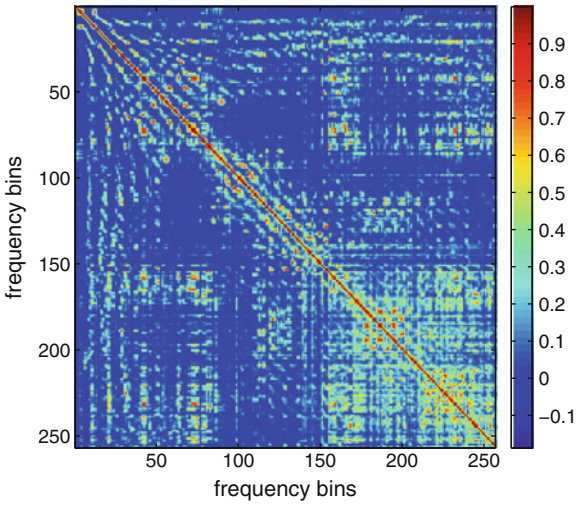


Fig. 5.1 The frequency domain energy correlation of the speech signal "si1039.wav"

frequency bins is such square correlation, which can be defined as

$$\text{cov}\left(|s_i(a)|^2, |s_i(b)|^2\right) = E\left[|s_i(a)|^2|s_i(b)|^2\right] - E\left[|s_i(a)|^2\right]E\left[|s_i(b)|^2\right] \quad (5.8)$$

The use of such energy correlation has seldom been highlighted in the original IVA algorithms. In order to show the energy correlation within the frequency domain speech signals, we choose a particular speech signal "si10390.wav" from the TIMIT database [23], with 8 kHz sampling frequency and 1,024 DFT length. Then the matrix of energy correlation coefficients between different frequency bins is plotted as shown in Fig. 5.1. Figure 5.1 is just part of the whole matrix of energy coefficient, which corresponds to the frequency bins from 1 to 256. The high frequency part is omitted due to the limited energy which leads to large correlation coefficients.

It is shown in Fig. 5.1, besides the information on the diagonal, there are many information on the off-diagonal elements, which is correspond to the energy correlation between different frequency bins. It indicates that there are energy correlation as defines in Eq. (5.8), which also leads to that $E\left[|s_i(a)|^2|s_i(b)|^2\right]$ is not equal to zero for many points and this information should be used to help to further exploit the dependency between different frequency bins.

5.4 Multivariate Generalized Gaussian Source Prior

In this section, we propose a particular multivariate generalized Gaussian source prior as the source prior, from which a new nonlinear score function can be derived to introduce the energy correlation information to improve the separation performance.

The source prior we proposed belongs to the family of multivariate generalized Gaussian distributions which takes the form

$$q(\mathbf{s}_i) \propto \exp\left(-\left(\frac{(\mathbf{s}_i - \boldsymbol{\mu}_i)^{\dagger}\Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)}{\alpha}\right)^{\beta}\right) \tag{5.9}$$

when $\alpha = 1$ and $\beta = \frac{1}{2}$, it is the multivariate Laplace distribution adopted by the original IVA algorithm [9].

Now we assume that $\alpha = 1$, the mean vector is a zero vector and the covariance matrix is an identity matrix due to the orthogonality of the Fourier basis and scaling adjustment. Then, the source prior takes the general form

$$p(\mathbf{s}_i) \propto \exp\left(-\left(\sum_{k=1}^{K}|s_i(k)|^2\right)^{\beta}\right) \tag{5.10}$$

where we constrain $0 < \beta < 1$ to obtain a super-Gaussian distribution to describe the speech signals. The nonlinear score function based on this new source prior is

$$\varphi^{(k)}\left(\hat{s}_i(1)\ldots\hat{s}_i(k)\right) = \frac{2\beta\hat{s}_i(k)}{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^{1-\beta}} = \frac{2\beta\hat{s}_i(k)}{\left(\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2\right)^{\frac{1-\beta}{2}}} \tag{5.11}$$

In order to introduce the energy correlation, the root needs to be odd, otherwise the square will be cancelled. Then the denominator can be expanded as

$$\left(\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2\right)^{\frac{1-\beta}{2}} = \left(\sum_{k=1}^{K}|\hat{s}_i(k)|^4 + \sum_{a\neq b}c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2\right)^{\frac{1-\beta}{2}} \tag{5.12}$$

which contains cross items $\sum_{a\neq b}c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2$ corresponding to energy correlation between different frequency bins, and c_{ab} is a scalar constant between the a-th and b-th frequency bins.

Thus the following condition must be satisfied

$$\frac{1-\beta}{2} = \frac{1}{2I+1} \tag{5.13}$$

where I is positive integer. Then we can obtain the condition for β

$$\beta = \frac{2I-1}{2I+1} \tag{5.14}$$

On the other hand, β is the shape parameter of the generalized multivariate Gaussian distribution. In order to make the proposed source prior more robust to outliers compared with the original source prior, β should be less than the 1/2, which is correspondent to the original source prior. Thus

$$\frac{2I-1}{2I+1} < \frac{1}{2} \tag{5.15}$$

Finally, $I = 1$ is the only solution, and the associated β is 1/3. Thus the appropriate generalized Gaussian distribution takes the form

$$q(\mathbf{s}_i) \propto \exp\left(-\sqrt[3]{(\mathbf{s}_i - \boldsymbol{\mu}_i)^\dagger \Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)}\right) \tag{5.16}$$

We next show that this source prior can also preserve the inter-frequency dependencies within each source vector in a similar manner to the original source prior for IVA [9].

We begin with the definition of a K-dimensional random variable

$$\mathbf{s}_i = v^{\frac{3}{4}}\boldsymbol{\xi}_i + \boldsymbol{\mu}_i \tag{5.17}$$

where v is a scalar random variable, and $\boldsymbol{\xi}_i$ obeys a generalized Gaussian distribution which has the form:

$$p(\boldsymbol{\xi}_i) \propto \exp\left(-\left(\frac{\boldsymbol{\xi}_i^\dagger \Sigma_i^{-1}\boldsymbol{\xi}_i}{2\sqrt{2}}\right)^{\frac{2}{3}}\right). \tag{5.18}$$

If v has a Gamma distribution of the form:

$$p(v) \propto v^{\frac{1}{2}} \exp\left(-\frac{v}{2}\right) \tag{5.19}$$

then the proposed source prior can be achieved by integrating the joint distribution of $\mathbf{s}_i$ and v over v as follows:

$$\begin{aligned} q(\mathbf{s}_i) &= \int_0^\infty q(\mathbf{s}_i|v)p(v)dv \\ &= \alpha_1 \int_0^\infty v^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\frac{\left((\mathbf{s}_i - \boldsymbol{\mu}_i)^\dagger \Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)\right)^{\frac{2}{3}}}{v} + v\right)\right) dv \\ &= \alpha_2 \exp\left(-\sqrt[3]{(\mathbf{s}_i - \boldsymbol{\mu}_i)^\dagger \Sigma_i^{-1}(\mathbf{s}_i - \boldsymbol{\mu}_i)}\right) \end{aligned} \tag{5.20}$$

where α_1 and α_2 are both normalization terms. Therefore, equation (5.20) confirms that the new source prior has the dependency generated by v.

In Lee's paper [24], the source priors suitable for IVA are discussed. A general form of source prior is described as:

$$q(\mathbf{s}_i) \propto \exp\left(-(\|\mathbf{s}_i\|_p)^{\frac{1}{L}}\right) = \exp\left(-\left(\sum_k |s_i^{(k)}|^p\right)^{\frac{1}{pL}}\right) \tag{5.21}$$

where $\|\cdot\|_p$ denotes the l_p norm, and L is termed as the sparseness parameter. He suggested that the spherical symmetry assumption is suitable for modeling the frequency components of speech, i.e. $p = 2$, and through certain experimental studies found that the best separation performance can be achieved when $L = 7$.

Our new proposed source prior also belongs to this family. If we choose $p = 2$ to make it spherically symmetric, and choose $L = \frac{3}{2}$, the proposed source prior can be obtained. However, our detailed experimental results show that the improvement of performance is not robust when $L = 7$ as mentioned in [24], while the NG-IVA which adopts our new source prior can consistently achieve improved separation performance.

5.5 IVA Algorithms with the New Source Prior

5.5.1 NG-IVA with the New Source Prior

Applying this new source prior to derive the nonlinear score function of the NG-IVA algorithm with the assumption that the mean vector is zero and the covariance matrix is an identity matrix, we can obtain

$$\varphi^{(k)}\left(\hat{s}_i(1)\ldots\hat{s}_i(k)\right) = \frac{2\hat{s}_i(k)}{3\sqrt[3]{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2}}. \tag{5.22}$$

If we expand the equation under the cubic root, it can be written as:

$$\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2 = \sum_{k=1}^{K}|\hat{s}_i(k)|^4 + \sum_{a\neq b} c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2 \tag{5.23}$$

which contains cross items $\sum_{a\neq b} c_{ab}|\hat{s}_i(a)|^2|\hat{s}_i(b)|^2$. These terms are related to the energy correlation between different components within each source vector, and capture the level of interdependency between different frequency bins. Thus, this

new multivariate nonlinear function can provide a more informative model of the dependency structure. Moreover, it can better describe the speech model.

5.5.2 FastIVA with the New Source Prior

Fast fixed-point independent vector analysis is a fast form of IVA algorithm. Newton's method is adopted in the update, which converges quadratically and is free from selecting an efficient learning rate [11]. The contrast function used by FastIVA is as follows:

$$J = \sum_{i=1}^{n} \left(E\left[F\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) \right] - \sum_{k=1}^{K} \lambda_i^{(k)} \left(\mathbf{w}_i(k)^{\dagger} \mathbf{w}_i(k) - 1 \right) \right) \tag{5.24}$$

where $\mathbf{w}_i^{\dagger}$ is the i-th row of the unmixing matrix $\mathbf{W}$, and λ_i is the i-th Lagrange multiplier. $F(\cdot)$ is the nonlinear function, which can take on several different forms as discussed in [11]. It is a multivariate function of the summation of the desired signals in all frequency bins. With normalization, the learning rule is:

$$\begin{aligned} \mathbf{w}_i(k) \leftarrow & E\left[F'\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) + |\hat{s}_i(k)|^2 F''\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) \right] \mathbf{w}_i(k) \\ & - E\left[(\hat{s}_i(k))^* F'\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) \mathbf{x}^k \right] \end{aligned} \tag{5.25}$$

where $F'(\cdot)$ and $F''(\cdot)$ denote the derivative and second derivative of $F(\cdot)$, respectively. If this is used for all sources, an unmixing matrix $\mathbf{W}(k)$ can be constructed, which must be decorrelated with

$$\mathbf{W}(k) \leftarrow \left(\mathbf{W}(k)(\mathbf{W}(k))^{\dagger} \right)^{-1/2} \mathbf{W}(k). \tag{5.26}$$

When the multivariate Laplacian distribution is used as the source prior for the FastIVA algorithm, with the zero mean and unity variance assumptions, the nonlinear function takes the form

$$F\left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right) = \left(\sum_{k=1}^{K} |\hat{s}_i(k)|^2 \right)^{\frac{1}{2}}. \tag{5.27}$$

When the new multivariate generalized Gaussian distribution is used as the source prior, with the same assumptions, the nonlinear function becomes:

$$F\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)=\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^{\frac{1}{3}}. \tag{5.28}$$

Therefore, the first derivative becomes:

$$F'\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)=\frac{2}{3\sqrt[3]{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2}}. \tag{5.29}$$

It is very similar to Eq. (5.22), and it also contains cross terms which can exploit the energy correlation between different frequency bins. Thus, the FastIVA algorithm with the new source prior is likely to help improve the separation performance.

5.5.3 AuxIVA with the New Source Prior

AuxIVA adopts the auxiliary function technique to avoid the step size tuning [25]. In the auxiliary function technique, an auxiliary function is designed for optimization. During the learning process, the auxiliary function is minimized in terms of auxiliary variables. The auxiliary function technique can guarantee monotonic decrease of the cost function, and therefore provides effective iterative update rules [12].

The contrast function for AuxIVA is derived from the source prior [25]. For the original IVA algorithm,

$$G(\hat{\mathbf{s}}_i)=G_R(r_i)=r_i \tag{5.30}$$

where $r_i=\|\hat{\mathbf{s}}_i\|_2$.

By using the proposed multivariate generalized Gaussian source prior, we can obtain the following contrast function

$$G(\hat{\mathbf{s}}_i)=G_R(r_i)=r_i^{\frac{2}{3}}. \tag{5.31}$$

The update rules contain two parts, i.e., the auxiliary variable updates and unmixing matrix updates. In summary, the update rules are as follows:

$$r_i=\sqrt{\sum_{k=1}^{K}|\mathbf{w}_i^{\dagger}(k)\mathbf{x}(k)|^2} \tag{5.32}$$

$$V_i(k)=E\left[\frac{G'_R(r_i)}{r_i}\mathbf{x}(k)\mathbf{x}(k)^{\dagger}\right] \tag{5.33}$$

$$\mathbf{w}_i(k) = \left(W(k)V_i(k)\right)^{-1}\mathbf{e}_i \tag{5.34}$$

$$\mathbf{w}_i(k) = \frac{\mathbf{w}_i(k)}{\sqrt{\mathbf{w}_i^{\dagger}(k)V_i(k)\mathbf{w}_i(k)}}. \tag{5.35}$$

In Eq. (5.34), $\mathbf{e}_i$ is a unity vector, the i-th element of which is unity.

During the update process of the auxiliary variable $V_i(k)$, we notice that $\frac{G'_R(r_i)}{r_i}$ is used to keep the dependency between different frequency bins for source i. In this chapter, as we defined previously, $G_R(r_i) = r_i^{\frac{2}{3}}$. Therefore

$$\frac{G'_R(r_i)}{r_i} = \frac{2}{3r_i^{\frac{4}{3}}} = \frac{2}{3\sqrt[3]{\left(\sum_{k=1}^{K}|\hat{s}_i(k)|^2\right)^2}} \tag{5.36}$$

which has the same form as Eq. (5.29). The update rules also contain the fourth order terms to exploit the energy correlation within the frequency domain speech signal vectors and should thereby help to achieve a better separation performance.

5.6 Experiments

In this section, we used all three state-of-the-art IVA algorithms with the proposed multivariate generalized Gaussian source prior to separate the mixtures obtained in a reverberant room environment. The speech signals were chosen from the TIMIT dataset [23], and each of them was approximately 7 s long. The image method was used to generate the room impulse responses, and the dimensions of the room were $7 \times 5 \times 3$ m^3. The DFT length was set to be 1,024, and the reverberation time RT60 = 200 ms. We used a 2×2 mixing case, and the microphone positions are [3.48, 2.50, 1.50] and [3.52, 2.50, 1.50] m respectively. The sampling frequency was 8 kHz. The separation performance was evaluated objectively by the signal-to-distortion ratio (SDR) and signal-to-interference ratio (SIR) [26]. Figure 5.2 is the plan view of the experimental setting.

5.6.1 NG-IVA Algorithms Comparison

In the first experiment, two different speech signals were chosen randomly from the TIMIT dataset and were convolved into two mixtures. Then the NG-IVA method with original source prior, the NG-IVA method with our proposed source prior and NG-

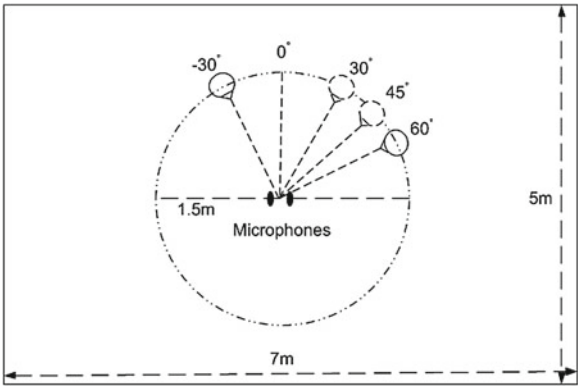


Fig. 5.2 Plan view of the experiment setting in the room environment with two microphones and two sources

Table 5.1 Separation performance comparison in SDR (dB)

	Original	Proposed	Lee's
Mixture 1	12.27	12.90	4.74
Mixture 2	18.13	18.47	18.34
Mixture 3	8.88	11.83	11.41
Mixture 4	15.57	16.92	5.95
Mixture 5	18.10	18.69	15.44
Mixture 6	18.81	19.58	3.71
Mixture 7	15.94	16.59	8.63
Mixture 8	15.29	15.75	16.03
Mixture 9	18.58	19.05	17.35
Mixture 10	18.80	19.31	0.78

IVA with Lee's source prior where the sparseness parameter $L = 7$, were all used to separate the mixtures, respectively. Then the source positions were changed to repeat the simulation. For every pair of speech signals, three different azimuth angles for the sources relative to the normal to the microphone array were set for testing, these angles were selected from $30°$, $45°$, $60°$, and $-30°$. After that, we chose another pair of speech signals to repeat the above simulations. We used ten different pairs of speech signals totally, and repeated the simulation 30 times at different positions. Tables 5.1 and 5.2 show the average separation performance for each pair of speech signals in terms of SDR and SIR in dB.

The experimental results indicate clearly that NG-IVA with the proposed source prior can consistently improve the separation performance. Although the NG-IVA with Lee's source prior can get improvement results sometimes, the separation improvement is not consistent, in some cases there is essentially no separation such as mixtures 1, 6, and 10. Even though it can achieve better separation than original NG-IVA, it is still no better than the proposed method. Only for mixture 8, does

Table 5.2 Separation performance comparison in SIR (dB)

	Original	Proposed	Lee's
Mixture 1	14.08	14.84	5.62
Mixture 2	19.57	19.86	19.81
Mixture 3	10.72	13.74	13.19
Mixture 4	16.98	18.46	7.16
Mixture 5	20.14	20.47	16.94
Mixture 6	20.30	20.98	4.35
Mixture 7	17.88	18.40	10.73
Mixture 8	19.88	20.41	20.61
Mixture 9	20.75	20.89	18.80
Mixture 10	20.28	20.60	1.48

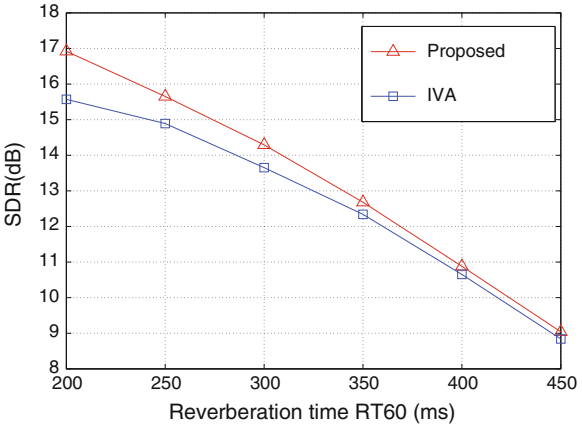


Fig. 5.3 Separation comparison in terms of SDR between original and proposed NG-IVA algorithms as a function of reverberation time

it achieve the best separation performance. Therefore, among all these three algorithms, the NG-IVA with the proposed source prior is the best method, because it can consistently achieve better separation performance. The average SDR improvement and SIR improvement are approximately 0.9 and 0.8 dB, respectively compared with the original NG-IVA algorithm.

Then we used the NG-IVA algorithms with the proposed source prior to separate the mixtures obtained from different reverberant room environments. Two speech signals were selected from the TIMIT dataset randomly and convolved into two mixtures. The azimuth angles for the sources relative to the normal to the microphone array were set as 60° and −30°. Both the original NG-IVA and the proposed method were used to separate the mixtures. The results are shown in Figs. 5.3 and 5.4, which show the separation performance comparisons in different reverberant environments. Figures 5.3 and 5.4 show the SDR and SIR comparison, respectively.

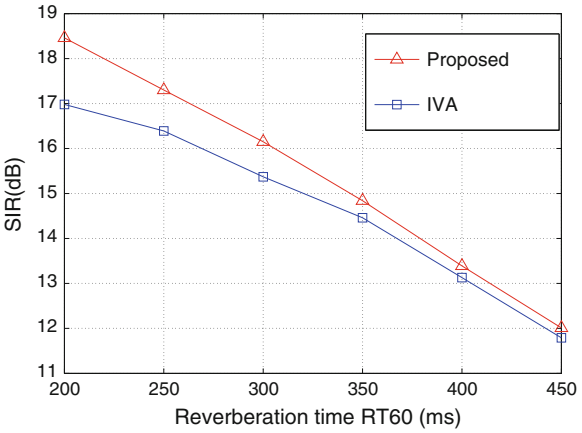


Fig. 5.4 Separation comparison in terms of SIR between original and proposed NG-IVA algorithms as a function of reverberation time

They indicate that the proposed algorithm can consistently improve the separation performance in different reverberant environments, up to a reverberation time of 450 ms. The advantage reduces with increasing reverberation time RT60 due to the greater challenge in extracting the individual source vectors.

5.6.2 FastIVA Algorithms Comparison

In the second experiment, all the experimental settings and the processes are all the same as the first experiment. Here we randomly selected five pairs of speech signals from the TIMIT dataset and convolved them into mixtures. The original FastIVA algorithm and the FastIVA algorithm with the proposed source prior were used to separate the speech mixtures. Then the source positions were changed to repeat the experiment, the average separation performance comparison is shown in Table 5.3. It shows that the separation performance can be improved by adopting the proposed source prior. The average SDR improvement and SIR improvement both are approximately 0.6 dB.

We also compared the separation performance of these two algorithms in different reverberant room environments as in the first experiment. The SDR and SIR comparisons are shown in Figs. 5.5 and 5.6. in terms of SDR and SIR comparison, respectively. The results show that the FastIVA algorithm with the proposed source prior can improve the separation performance, but again the advantage is reduced with increasing reverberation time RT60.

Table 5.3 Separation performance comparison in terms of SDR and SIR measures in dB

Mixtures	Mixture 1	Mixture 2	Mixture 3	Mixture 4	Mixture 5
Original FastIVA (SDR)	17.77	19.48	14.75	18.12	16.79
Proposed FastIVA (SDR)	18.04	20.63	15.08	18.88	17.31
Original FastIVA (SIR)	19.32	21.01	17.04	19.80	19.18
Proposed FastIVA (SIR)	19.59	22.04	17.31	20.51	19.74

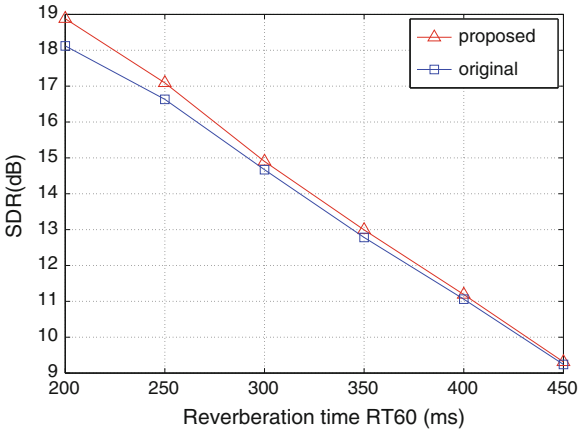


Fig. 5.5 Separation comparison in terms of SDR between original and proposed FastIVA algorithms as a function of reverberation time

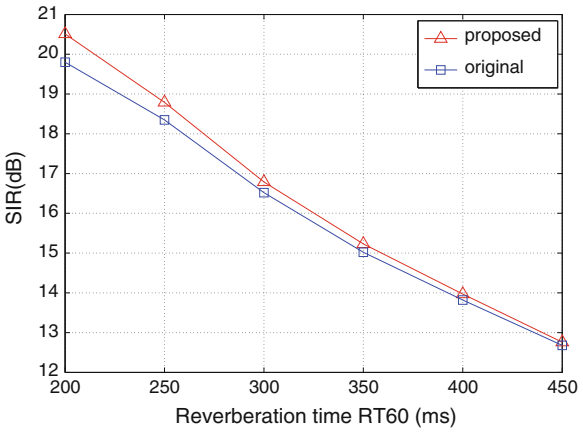


Fig. 5.6 Separation comparison in terms of SIR between original and proposed FastIVA algorithms as a function of reverberation time

Table 5.4 Separation performance comparison in terms of SDR and SIR measures in dB

Mixtures	Mixture 1	Mixture 2	Mixture 3	Mixture 4	Mixture 5
Original AuxIVA (SDR)	12.13	14.62	9.86	19.23	18.64
Proposed AuxIVA (SDR)	14.82	16.30	12.45	19.92	19.50
Original AuxIVA (SIR)	14.06	16.72	11.59	20.54	20.12
Proposed AuxIVA (SIR)	17.26	18.42	14.58	21.20	20.90

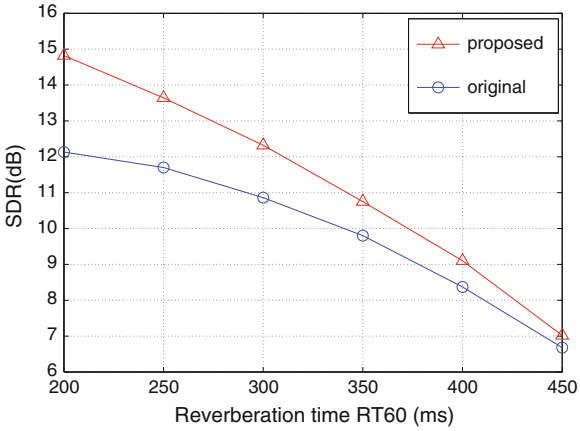


Fig. 5.7 Separation comparison in terms of SDR between original and proposed AuxIVA algorithms as a function of reverberation time

5.6.3 AuxIVA Algorithms Comparison

In the third experiment, the separation performance of AuxIVA with original source prior and AuxIVA with proposed source prior were compared. Again five different pairs of speech signals were used, and the simulation was repeated 15 times at different positions. Table 5.4 shows the average separation performance for each pair of speech signals in terms of SDR and SIR. The average SDR and SIR improvements are approximately 1.7 and 1.9 dB, respectively. The results confirm the advantage of the proposed AuxIVA method which can better preserve the dependency between different frequency bins of each source and thereby achieve a better separation performance.

Then we also tested the robustness of the proposed AuxIVA method in different reverberant room environments. The experimental settings are all the same as previous two experiments. The results are shown in Figs. 5.7 and 5.8, which show the separation performance comparison in different reverberant environments. It indicates that the AuxIVA algorithm with the proposed source prior can consistently improve the separation performance in different reverberant environments as the other two kinds of IVA algorithms.

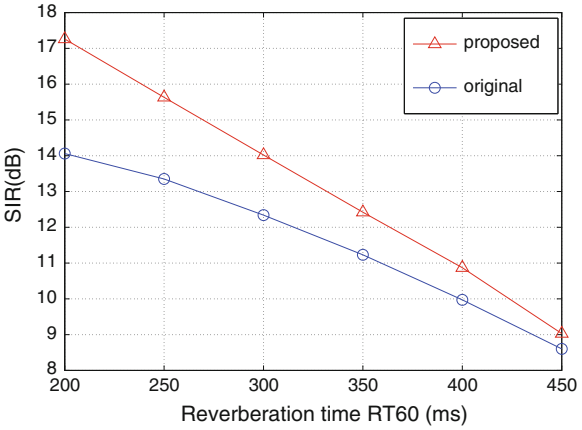


Fig. 5.8 Separation comparison in terms of SIR between original and proposed AuxIVA algorithms as a function of reverberation time

Examining the results for all the three algorithms, our proposed source prior offers the maximum improvement in the AuxIVA algorithm. However, it is difficult to make a general recommendation, which is the best algorithm due to the variability of performance with different speech signals and mixing environments.

5.7 Conclusions

In this chapter, a specific multivariate generalized Gaussian distribution was adopted as the source prior for IVA. This new source prior can better preserve the inter-frequency dependencies as compared to the original multivariate Laplace source prior, and is more robust to outliers. When the proposed source prior was used in IVA algorithms, it introduces energy correlation commonly found in frequency domain speech signals to improve the learning process and enhance separation. Three state-of-the-art types of IVA algorithms with the new source prior, i.e., NG-IVA, FastIVA, and AuxIVA, were all analyzed, and the experimental results confirm the advantage of adopting the new source prior particularly for low reverberation environment.

Acknowledgments Some of the material of this chapter is under review for publication in *Signal Processing* as "Independent Vector Analysis with a Generalized Multivariate Gaussian Source Prior for Frequency Domain Blind Source Separation," in October 2013.

References

1. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. Wiley, New York (2003)
2. Comon, P., Jutten, C.: Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press, Oxford (2009)
3. Cherry, C.: Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. **25**, 975–979 (1953)
4. Cherry, C., Taylor, W.: Some further experiments upon the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. **26**, 554–559 (1954)
5. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
6. Parra, L., Spence, C.: Convolutive blind separation of non-stationary sources. IEEE Trans. Speech Audio Process. **8**, 320–327 (2000)
7. Pedersen, M.S., Larsen, J., Kjems, U., Parra, L.C.: A survey of convolutive blind source separation methods. In: Handbook on Speech Processing and Speech Communication. Springer, New York (2007)
8. Kim, T., Lee, I., Lee, T.-W.: Independent vector analysis: definition and algorithms. In: Fortieth Asilomar Conference on Signals, Systems and Computers 2006. Asilomar, USA (2006)
9. Kim, T., Attias, H., Lee, S., Lee, T.: Blind source separation exploiting higher-order frequency dependencies. IEEE Trans. Audio Speech Lang. Process. **15**, 70–79 (2007)
10. Kim, T.: Real-time independent vector analysis for convolutive blind source separation. IEEE Trans. Circuits Syst. **57**, 1431–1438 (2010)
11. Lee, I., Kim, T., Lee, T.-W.: Fast fixed-point independent vector analysis algorithms for convolutive blind source separation. Signal Process. **87**, 1859–1871 (2007)
12. Ono, N.: Stable and fast update rules for independent vector analysis based on auxiliary function technique. In: 2011 IEEE WASPAA. New Paltz, USA (2011)
13. Liang, Y., Naqvi, S.M., Chambers, J.: Adaptive step size indepndent vector analysis for blind source separation. In: 17th International Conference on Digital Signal Processing. Corfu, Greece (2011)
14. Liang, Y., Naqvi, S.M., Chambers, J.: Audio video based fast fixed-point independent vector analysis for multisource separation in a room environment. EURASIP J. Adv. Signal Process. **2012**, 183 (2012)
15. Masnadi-Shirazi, A., Zhang, W., Rao, B.D.: Glimpsing IVA: A framework for overcomplete/complete/undercomplete convolutive source separation. IEEE Trans. Audio Speech Lang. Process. **18**, 1841–1855 (2010)
16. Ono, T., Ono, N., Sagayama, S.: User-guided indpendent vector analysis with source activity tuning. In: ICASSP 2012. Kyoto, Japan (2012)
17. Lee, I., Jang, G.-J., Lee, T.-W.: Independent vector analysis using densities represented by chain-like overlapped cliques in graphical models for separation of convolutedly mixed signals. Electron. Lett. **45**, 710–711 (2009)
18. Choi, C.H., Chang, W., Lee, S.Y.: Blind source separation of speech and music signals using harmonic frequency dependent independent vector analysis. Electron. Lett. **48**, 124–125 (2012)
19. Lee, I., Hao, J., Lee, T.W.: Adaptive independent vector analysis for the separation of convoluted mixtures using EM algorithm. In: ICASSP 2008. USA, Las Vegas (2008)
20. Hao, J., Lee, I., Lee, T.W.: Independent vector analysis for source separation using a mixture of Gaussian prior. Neural Comput. **22**, 1646–1673 (2010)
21. Anderson, M., Adali, T., Li, X.-L.: Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis. IEEE Trans. Signal Process. **60**, 1672–1682 (2012)
22. Hyvärinen, A.: Independent component analysis: recent advances. Philos. Transact. A Math. Phys. Eng. Sci. **371**(1984), 1–19 (2012)
23. Garofolo, J.S., et al.: TIMIT acoustic-phonetic continuous speech corpus. Linguistic Data Consortium, Philadelphia (1993)

24. Lee, I., Lee, T.W.: On the assumption of spherical symmetry and sparseness for the frequency-domain speech model. IEEE Trans. Audio Speech Lang. Process. **15**, 1521–1528 (2007)
25. Ono, N., Miyabe, S.: Auxiliary-function-based independent component analysis for super-Gaussian source. In: LVA/IVA 2010. St. Malo, France (2010)
26. Vincent, E., Fevotte, C., Gribonval, R.: Performance measurement in blind audio source separation. IEEE Trans. Audio Speech Lang. Process. **14**, 1462–1469 (2006)

Chapter 6
Sparse Component Analysis: A General Framework for Linear and Nonlinear Blind Source Separation and Mixture Identification

Yannick Deville

Abstract In this chapter, we consider two closely related data processing tasks. The first one is Blind Source Separation (BSS), which consists in estimating a set of unknown source data (one-dimensional signals, images, ...) from observed mixtures of these data, while the mixing operator has unknown parameter values. The second task is Blind Mixture Identification (BMI), which aims at estimating these unknown parameter values of the mixing operator. We provide a unified view and describe the latest extensions of the general framework that we have been developing for BSS and BMI since the beginning of the 2000s. This framework yields a wide range of BSS/BMI methods applicable to various types of sources (one-dimensional signals, images, ...) mixed according to various models (linear instantaneous, anechoic, full convolutive, nonlinear and especially linear-quadratic), possibly with non-negativity or sum-to-one constraints. This framework is based on the concept of joint sparsity of the source data, considered in various domains (original temporal or spatial domain, transformed representation in time-frequency or time-scale/wavelet domain, ...). More precisely, the proposed methods essentially require a few tiny zones, in mixed signals or in their transformed versions, where only one of the source "signals" is active, i.e., nonzero. They therefore set very limited constraints on source sparsity and could then be considered as "quasi-non-sparse component analysis" methods. Besides, unlike Independent Component Analysis methods, they are suited to correlated sources. We also discuss their application to various data processing functions, ranging from audio signal separation to unmixing of hyperspectral remote sensing images.

Y. Deville (✉)
Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, UPS-CNRS-OMP, 14 Avenue Edouard Belin, 31400 Toulouse, France
e-mail: yannick.deville@irap.omp.eu

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_6,
© Springer-Verlag Berlin Heidelberg 2014

6.1 Introduction

In this chapter, we consider two closely related data processing tasks. The first one is Blind Source Separation (BSS), which is a generic signal processing problem, where the term "signal" is to be understood in a broad sense: it may especially refer to one-dimensional (1D) series (e.g., depending on a time or wavelength variable) or to two-dimensional (2D) data (e.g., images), but also to more general types of data. BSS consists in estimating a set of unknown "source signals" from observed mixtures of these data, while the mixing operator is most often only partly known: it is known to belong to a given linear or nonlinear class, but it has unknown parameter values. The second task that we consider is Blind Mixture Identification (BMI), which consists in estimating the above-mentioned unknown parameter values of the mixing operator. Although the emphasis is often put on BSS in the literature, BMI is also of interest in various applications (see e.g., Karoui et al. [28]), and many so-called "BSS methods" in fact achieve both BSS and BMI.

The BSS/BMI field emerged in the 1980s and, for about two decades, the main available class of methods for achieving BSS/BMI was Independent Component Analysis (ICA), which was especially introduced in [11] and is described in detail in such books as [12, 19, 23]. A second class of methods, namely Sparse Component Analysis (SCA), then started to emerge at the end of the 1990s (see e.g., [33]), and became prominent during the 2000s (see e.g., the description of this field in Chap. 10 of [12] or the introduction to this domain in Chap. 11 of [19]). The early SCA methods which then became popular e.g., include (i) DUET, which was introduced by Jourjine et al. in 2000 in [24] and then detailed in [55] and (ii) TiFROM, that we proposed in 2001 in [1] and then extended in [3, 42]. The TiFROM approach is based on the exploitation of "zones", in mixed signals or in their transformed versions, where only one of the source signals is "active", i.e., has nonzero components (this corresponds to the concept of joint sparsity of source signals). We then widely extended this approach based on "single-source zones". We thus introduced a general framework for developing sparsity-based BSS/BMI methods, which are applicable to different types of source data and mixing operators.

Detailed descriptions of some of these methods were provided in a few journal papers: Abrard and Deville [3] and [18] respectively, present the versions of TiFROM and TiFCorr suited to linear instantaneous mixtures of one-dimensional signals, Puigt and Deville [42] describes an extension of TiFROM applicable to mixing models which also include time delays, and [25] concerns methods intended for a specific class of linear instantaneous mixtures of 2D signals (this configuration is especially faced in remote sensing applications, as explained further in this chapter). Throughout the last decade, we performed a large number of additional investigations, to further analyze the properties of these methods, to develop alternatives for similar configurations and also to create other methods for different types of data (1D signals, images) or mixing operators. At this stage, the corresponding results are only available in short conference papers.

In this chapter, we aim at proceeding much further than the above scattered contributions, both by providing the reader with a unified view of this general framework for developing sparsity-based BSS/BMI methods, and by explicitly considering various specific applications of this approach to different types of source data and mixing models. Tests results also illustrate the high performance achieved by these methods. It should therefore be clear that the scope of this chapter is at an intermediate level:

- It is much wider than the presentation of a single SCA method, i.e., this chapter mainly describes a whole class of such methods and suggests how various extensions may be further developed.
- However, we here do not aim at providing a review of the complete SCA field: instead, this chapter is focused on (i) the class of methods that we developed by taking advantage of single-source analysis zones, (ii) variants and extensions of these methods then proposed by other research groups, (iii) some other types of methods which are somewhat related to those that we describe hereafter. For already available reviews of the complete SCA field, the reader may e.g., refer to Chap. 10 of [12].

Throughout this chapter, we progress from standard and/or simple types of source signals and mixing models, toward more advanced ones. The remainder of this chapter is therefore organized as follows. The first sections deal with 1D sources, mixed according to increasingly complex operators. In Sect. 6.2, we consider linear instantaneous mixtures of these 1D sources, which is the mostly used configuration in the literature. The extension of the above methods to convolutive mixtures is relatively natural and is therefore more briefly presented in Sect. 6.3, where we first focus on attenuation-delay (or anechoic) mixtures and then extend the discussion to general convolutive mixtures. Section 6.4 is devoted to nonlinear mixtures, with main emphasis on linear-quadratic instantaneous mixtures. We then more briefly discuss the case of 2D sources, mixed according to a linear instantaneous model, also considering the situation when the sources values and/or mixing coefficients are constrained to meet some properties, as in the field of remote sensing. Finally, conclusions are drawn from this presentation in Sect. 6.6 and a specific topic is considered in the appendix.

6.2 Linear Instantaneous Mixtures of 1D Sources

6.2.1 Problem Statement, Definitions and First Assumptions

6.2.1.1 Original Signal Representation

The BSS/BMI problem and considered source properties were only defined in general terms in Sect. 6.1. We now detail their version corresponding to the types of sources and mixtures addressed in this section. In their original representation domain, all

considered signals are one-dimensional, i.e., they depend on a single scalar variable. This variable is here denoted as t, since it is a time variable in most applications. However, it should be clear that this section also applies to cases when this variable t has another nature.[1] Depending whether this variable belongs to a continuous of discrete subset of $\mathbb{R}$, the considered signals will be referred to as continuous-time or discrete-time signals.

The mixing model between such 1D signals considered at this stage of our presentation is the so-called determined linear instantaneous (or memoryless) mixture,[2] which may be defined as follows. The values of the P observed mixed signals $x_1(t), \ldots, x_P(t)$ at time t only depend on the values of N unknown source signals $s_1(t), \ldots, s_N(t)$ at the same time t (instantaneous mixture, as opposed e.g. to the time shifts considered in Sect. 6.3.1). Moreover, these observed signals $x_i(t)$ are linear combinations of the source signals $s_j(t)$ (linear mixture), i.e.,

$$x_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t) \qquad \forall i \in \{1, \ldots, P\} \tag{6.1}$$

where the values of the mixing coefficients a_{ij} are unknown. The signals and mixing coefficients may here be real-valued or complex-valued. Eq. (6.1) may also be expressed in matrix form as

$$x(t) = As(t) \tag{6.2}$$

where

$$s(t) = [s_1(t), \ldots, s_N(t)]^T \tag{6.3}$$

$$x(t) = [x_1(t), \ldots, x_P(t)]^T, \tag{6.4}$$

^{T} stands for transpose and A is the unknown $P \times N$ matrix consisting of the mixing coefficients a_{ij} . Finally, determined mixtures correspond to the case when the number P of observed signals is equal to the number N of source signals (underdetermined and overdetermined mixtures are considered in Sect. 6.2.4.1). The mixing matrix A is then square. Moreover, in this chapter, we constrain it to meet the following condition (which here means that it is invertible):

---

[1] For instance, one of the possible approaches to signals encountered in the field of remote sensing (Earth observation) consists in considering each signal associated with a given spatial position as a 1D function of wavelength (instead of time), representing the fraction of incident light power reflected by the considered spatial location of the scene at each wavelength. This signal representation is e.g., defined in [25]. However, the class of BSS methods studied in this chapter is not very well suited to this signal representation in remote sensing, since the signals then most often do not meet the sparsity assumptions required by these SCA methods. For remote sensing data, another representation based on 2D source signals is therefore also used, as, e.g., explained in [25] and briefly discussed in Sect. 6.5.

[2] The considered mixtures are noiseless here. The influence of observation noise is discussed further in this chapter.

Assumption 6.1 A is a full-column-rank matrix.

Starting from a set of vectors $x(t)$ of observed signals at some times t, BSS consists in estimating the corresponding vectors $s(t)$ of source signals, up to the well-known indeterminacies of linear instantaneous BSS, namely a permutation and scale factors. The corresponding BMI problem consists in estimating the above matrix A.

6.2.1.2 Signal Transforms

The BSS/BMI methods presented in this chapter may be directly applied to the observed signals in their above-defined original representation $x_i(t)$ (i.e.,"time-domain" representation, in the above broad sense of the variable t). They may also be applied to a *transformed version* of the observed signals. In particular, we here consider the following two types of transforms.

The first type consists of time-frequency (TF) transforms [10, 21]. It especially includes the Short-Time Fourier Transform (STFT), e.g., defined in [18] for complex-valued signals. Each STFT coefficient $U(t, \omega)$ is the contribution of the considered signal $u(t')$ in the part of the TF plane corresponding to: (i) a short time window centered around t and (ii) the angular frequency ω.

The other considered type of transforms consists of time-scale (TS) transforms and especially includes the Continuous Wavelet Transform (CWT), e.g., detailed in [16, 37, 49]. Each wavelet coefficient $W_u(\tau, d)$ defines the local behavior of the considered signal $u(t)$ around time τ, at scale d (with an associated frequency proportional to $1/d$).

We hereafter use a single notation for the transformed version of a signal $u(t)$, whatever transform is used: we denote it as $U(v)$. Its variable v may be a scalar or a vector, depending on which transform is used. For STFT, the variable v stands for the couple, or corresponding vector, (t, ω). For CWT, v represents the couple (τ, d). The notation $U(v)$ also includes the untransformed signal $u(t)$ itself (which may be considered as a transformed version with identity transform), where v stands for t.

Moreover, we here only consider *linear* transforms, since they do not break the simple structure of the linear mixing model (6.1): applying any such transform to (6.1), we get

$$X_i(v) = \sum_{j=1}^{N} a_{ij} S_j(v) \qquad \forall\, i \in \{1, \ldots, P\}. \tag{6.5}$$

The transformed observations $X_i(v)$ thus remain *linear instantaneous* mixtures of the transformed sources $S_j(v)$, and the mixing matrix involved in the transformed domain is exactly the same as in the original domain. We will use this property in Sect. 6.2.2.1.

6.2.1.3 Sparsity of One Signal

Since the above transforms do not change the nature and parameter values of the mixing model, the reader may wonder what advantage they bring. The answer has to do with the sparsity of the considered signals. As a first step, we here introduce the required concepts concerning the sparsity of a *single* signal (for more details on this topic and examples, the reader may refer to [12, 19]).

Let us consider a signal in a given representation domain, such as the time, TF, or TS domain. In this domain, this signal is defined by the values of its "components", e.g., by the values of the time-domain samples $u(t)$, of the STFT coefficients $U(t, \omega)$, or of the CWT coefficients $W_u(\tau, d)$. The degree of sparsity of this signal in the considered domain is basically measured by the number[3] of its components which are equal to zero (or in practice, which are negligible, i.e., which have a much lower magnitude than the maximum-magnitude component of the considered signal). The higher this number of zero components, the higher the sparsity of the considered signal. A highly sparse signal is therefore a signal which is "most often inactive". For instance, considering the time domain representation of a signal u, the number of its values $u(t)$ which are negligible defines its degree of sparsity. This is illustrated for two signals in Fig. 6.1: in the time domain, both signals are active (i.e., have non-negligible values) almost everywhere. They therefore have a low degree of sparsity in this domain.

Any of the above-defined transforms may then be applied to an original signal $u(t)$ in order to obtain a representation $U(v)$ of this signal which exhibits a higher sparsity. For instance, the spectrograms (i.e., squared moduli of STFT coefficients $U(t, \omega)$) of the above two speech signals are represented in Fig. 6.2. This shows that these signals have a very low magnitude (clear zones of the figure) in most of the TF plane, i.e., that they are much sparser in this plane than in the time domain. This may be explained as follows. A speech signal may be split into a series of time intervals, with a duration around 10 to 30 ms, where it has stationary behavior. In each such interval, a speech signal often has non-negligible components only in some frequency bands. Therefore, whereas continuous speech yields a signal which is active everywhere in the time domain, this signal only contributes in some zones of the TF domain. Applying a transform such as STFT to this time-domain signal thus yields a signal representation with a higher sparsity.

6.2.1.4 Joint Sparsity of Several Signals

The above considerations were based on the analysis of the sparsity of a single signal. However, in the framework of BSS/BMI, we deal with a set of observed signals and each of them contains contributions from several source signals. This requires us

[3] The term "number" is to be understood either as the total number or as the fraction (or percentage), e.g., depending whether the set of components of the considered signal is defined over a continuous/discrete and bounded/unbounded domain.

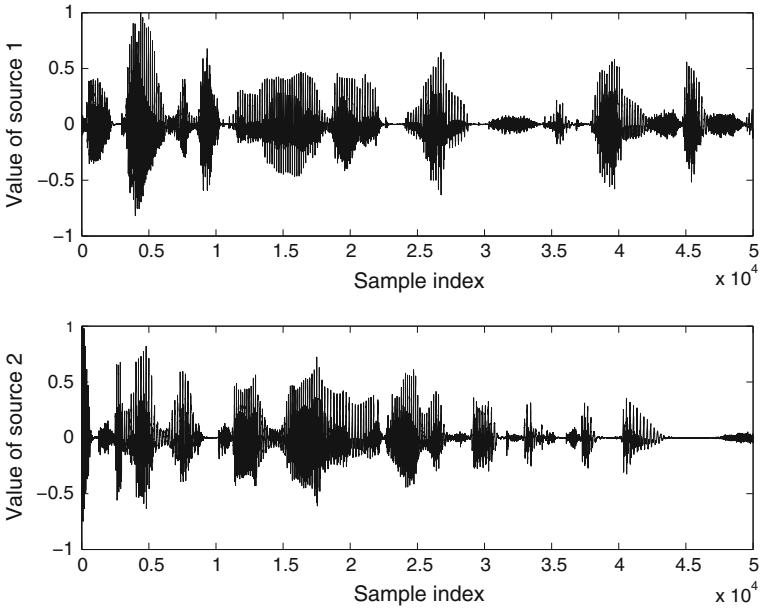


Fig. 6.1 Time-domain representations of two speech signals [18, 19]. The integer-valued variable t associated with the horizontal axis is the sample index of the considered discrete-time signal. The vertical axis corresponds to the signal values $u(t)$

to analyze the *joint* sparsity properties of the considered source signals (this may remind the reader of ICA methods, which are based on the *joint* statistics—e.g., joint moments or cumulants, or joint probabiblity density function—of the source signals, or of the resulting observations or separating system outputs). More precisely, whereas the sparsity of a single signal was analyzed above in terms of the zones of the representation domain where it is inactive, we hereafter consider the zones of that domain where most source signals are inactive. In particular, the BSS/BMI methods defined below focus on zones where a single source signal is active. Before describing these methods, we here define sparsity properties which must be met by the source signals for these methods to be applicable.

The signals are considered in a given representation domain, denoted as $\mathscr{D}$. Inside this domain, the SCA methods described further in this chapter successively use different "analysis zones". Such a zone is denoted as Z and defined as follows. It consists of a set of points of domain $\mathscr{D}$. Hereafter, these points are supposed to be adjacent. For instance, for signals considered in the discrete time domain, whose samples are indexed by integers, an analysis zone is a bounded interval of integers on the time axis. For signals in the TF plane, which depend on discrete time and frequency indices, several options exist, including the following two approaches: (i) for linear instantaneous mixtures [3, 18], we mainly consider analysis zones corresponding to a bounded interval of adjacent time indices and to a single frequency

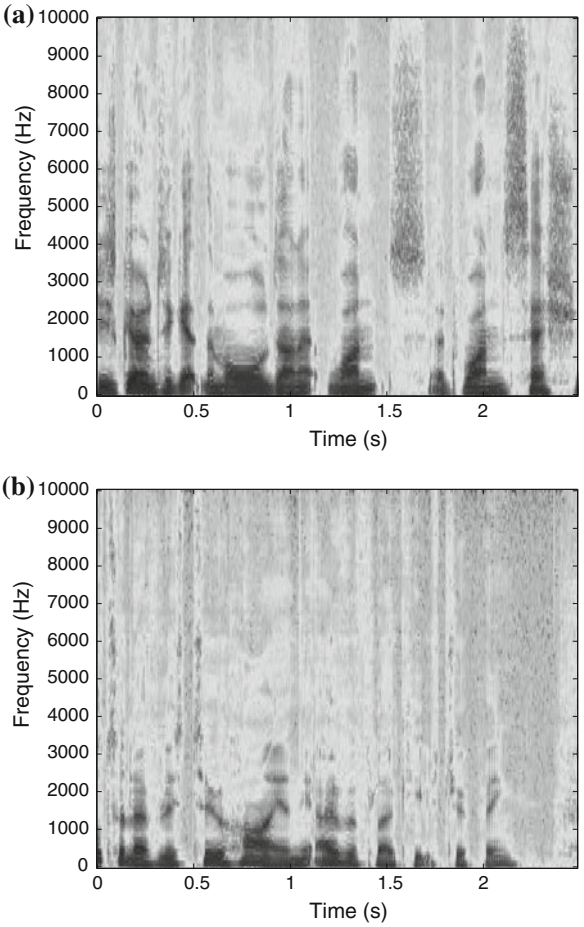


Fig. 6.2 Time-frequency representations (spectrograms) of the two speech signals of Fig. 6.1 [18, 19]. The horizontal axis corresponds to time (in seconds). The vertical axis corresponds to frequency (in Hertz). **a** source 1, **b** source 2

index,[4] (ii) another solution consists in considering a single time index and a bounded interval of adjacent frequency indices (both approaches were used for attenuation-delay mixtures [42]).

The discussion so far was focused on deterministic signals. We here stress that, throughout this chapter, we will also consider the case when the considered signals are defined in a stochastic framework. In the theoretical *definition* of the SCA methods proposed for stochastic signals, each analysis zone Z is restricted to a single value of the signal argument v, e.g., a single time t. The practical *implementation* of these stochastic SCA methods then has a clear link with our deterministic SCA methods, as

[4] This is illustrated in (6.9) further in this chapter.

will now appear, when introducing several definitions and an associated assumption concerning the properties of the source signals. From here on, the concept of an "active signal" is focused on its properties in an analysis zone.

Definition 6.1 A source signal is "active" in an analysis zone if its mean power is non-zero in this zone.

In this chapter, we use the same notation in the deterministic and stochastic cases for the mean power of a complex-valued signal $U(v)$, over an analysis zone Z: we denote it as $\mathbf{M}\{|U(v)|^2\}(Z)$. In this expression, $\mathbf{M}\{.\}(Z)$ is the mean operator,[5] over zone Z in the deterministic and stochastic cases, but it has a different meaning depending on the nature of the signals:

- For deterministic signals, $\mathbf{M}\{.\}(Z)$ represents the arithmetic mean of the set of values available in zone Z, for the considered signal (this signal is $|U(v)|^2$ in the above case).
- For stochastic signals, $\mathbf{M}\{.\}(Z)$ represents the expectation operator, often denoted as $E\{.\}$. As stated above, the considered analysis zone Z is then restricted to a single value of the signal argument v. The averaging performed by operator $E\{.\}$ consists of the statistical mean of the considered random variable over all possible outcomes. Although such a stochastic framework may be used for *defining* the proposed SCA methods and *analyzing* their theoretical properties it should be clear that, when *implementing* these methods, one uses a single realization of the considered stochastic signals, which yields deterministic signals. The single value of the above variable v (e.g., single time t) considered in theoretical investigations is then replaced by a set of values of this variable (e.g., bounded time interval), over which arithmetic averaging is performed, assuming the signals are ergodic. The practical implementation used for stochastic signals thus takes the same form as in the case when the signals are initially defined in a deterministic framework.

Definition 6.2 A source signal is "isolated" in an analysis zone if only this source signal (among all the source signals which are mixed in the observations) is active in this analysis zone. An analysis zone where a source signal is isolated, i.e., where a single source signal is active, is called a "single-source zone". An analysis zone where several source signals are active is called a "multiple-source zone".

The above definition of an isolated source signal corresponds to the theoretical point of view. From a practical point of view, this means that the mean powers of all other source signals are negligible as compared to the mean power of the source signal that is isolated.

---

[5] In notation $\mathbf{M}\{.\}(.)$, the letter $\mathbf{M}$ stands for mean. Besides, in Sect. 6.2 we consider various approaches and we therefore introduce various operators. We favor coherent and readable notations (i.e., no subscripts) throughout this section. We therefore always provide the considered signal(s) as the first operator argument, i.e., inside $\{.\}$, and the considered set of data points as its second argument, i.e., inside $(.)$. In some subsequent sections, we consider a single approach per section and therefore introduce a somewhat simplified and more standard version of the notations used in Sect. 6.2.

Definition 6.3 A source signal is "accessible" in the considered representation domain $\mathscr{D}$ if there exist at least one analysis zone inside this domain where this source signal is isolated.

Assumption 6.2 Each source signal is accessible in the considered representation domain $\mathscr{D}$.

In other words, the only constraint that we set at this stage is that, for each source, there should exist a tiny zone, in the considered representation domain, where only this source is active. In all other zones, we allow several sources to overlap, i.e., to be simultaneously active. The constraint on joint sparsity thus set on the sources is very low and the "sparse component analysis" methods proposed hereafter might therefore be more precisely called "quasi-non-sparse component analysis methods". This should be contrasted with such methods as DUET [24, 55] which require much sparser signals: they typically request the source signals to have no overlap, i.e., for each position v in the considered representation domain, only one source signal is allowed to be active (this is called "W-disjoint orthogonality"). However, the latter methods are thus able to extract sources in underdetermined configurations (i.e., for a number P of observations lower than the number N of sources), whereas we only consider determined mixtures at this stage (underdetermined mixtures are addressed in Sect. 6.2.4.1).

In addition to Assumption 6.2, the source signals should have some other form of "diversity". The required diversity depends on the considered version of our methods and is therefore detailed further in this chapter.

It should also be mentioned that the variant of our approach that we started to define above corresponds to the case when we use the "non-centered version of the signals", i.e., their nonmodified version, as opposed to the "centered version of the signals" obtained by subtracting the mean of each signal to all its available values. Another variant of this approach may be derived by considering the centered version of the signals. Definition 6.1 is then modified by considering signals with nonzero variance, instead of nonzero mean power. Details about this variant may be found in [18]. In the remainder of this chapter, we mainly consider the noncentered version of our approach. We now move to the description of the overall structure of the proposed SCA methods, and then detail their stages.

6.2.2 Proposed SCA Methods

6.2.2.1 Overall Structure

Most versions of our SCA methods for determined linear instantaneous mixtures of 1D sources share the same overall structure, which consists of the following stages:

1. The **"sparsification" stage** is the only optional stage of our methods. It consists in applying a transform, such as those defined in Sect. 6.2.1.2, to the observed

signals. If the signals, in their original representation, do not meet Assumption 6.2 (p. 10), this sparsification must be performed (with an adequate transform, to be defined by the user, depending on the considered class of signals) in order to guarantee that Assumption 6.2 becomes valid for the transformed signals. Even if Assumption 6.2 is likely to be met by the original signals, performing this sparsification may be of interest, because it may yield sparser data, which allow our methods to estimate the mixing matrix and source signals more accurately. For instance, moving from a time-domain representation of speech signals to a TF representation is likely to yield analysis zones where sources are better isolated in practice and thus better extracted by our BSS methods (this is confirmed by the experimental results provided in [18]).

2. In the **detection stage**, our methods detect all single-source zones. Several methods may be used to this end. They are detailed in Sect. 6.2.2.2.
3. In the **mixture-column estimation stage**, our methods use all single-source zones (or the best of them). In each of these zones, these methods derive an estimate of one column of the mixing matrix, up to a scale factor. This may also be performed by using several versions of our methods, which are described in Sect. 6.2.2.3. Note that, for each source, this stage of our methods may yield a whole set of estimates of the column of the mixing matrix corresponding to this source, not a single estimate, since these methods provide one estimate per single-source zone corresponding to this source.
4. In the **mixture-column combination stage**, our methods derive, for each source, a single estimate of the corresponding column of the mixing matrix, based on the above-mentioned complete set of estimates. Again, several methods may be used to this end. They are detailed in Sect. 6.2.2.4.
   The column vectors thus derived, which are obtained in an arbitrary order with respect to source numbering, are then gathered in a matrix. This yields an estimate $\hat{A}$ of the mixing A, up to permutation and scale factors, which are the standard indeterminacies of linear instantaneous BSS. At this stage of the approach, the above-defined BMI task is complete.
5. In the **source estimation stage**, our methods combine each observed vector $x(t)$ with the above estimated matrix $\hat{A}$ in order to derive the corresponding output vector as

$$y(t) = \hat{A}^{-1} x(t). \tag{6.6}$$

   This yields an estimate of the source vector $s(t)$, up to the permutation and scaling indeterminacies of its components, which result from the corresponding indeterminacies in $\hat{A}$.
   When using the sparsification stage in order to first create a transformed observation vector $X(v)$, one might consider performing the source estimation stage by applying $\hat{A}^{-1}$ to $X(v)$ instead of $x(t)$ in (6.6). However, this would yield a *transformed* version $Y(v)$ of the estimated source vector. The inverse transform should then be applied to $Y(v)$, in order to derive the estimated source vector $y(t)$ in the original domain, which is the source representation of interest in most applications. In order to avoid this inverse transformation of $Y(v)$, the inversion

(6.6) is preferably achieved in the original representation domain, even when BMI is performed in a transformed domain. We thus take advantage of the property mentioned in Sect. 6.2.1.2: thanks to the considered transforms, the mixing model is the same in both domains, and its inverse may therefore be applied to either representation of the observations.

We hereafter describe some of the above stages of our methods in more detail.

6.2.2.2 Detection Stage

In this section, we present alternative methods for detecting all single-source zones. These methods typically explore the selected representation domain $\mathscr{D}$. To this end, they consider all analysis zones defined in practice by successive, adjacent or partly overlapping, positions of a sliding window which is moved in all the domain $\mathscr{D}$. The signal properties are analyzed separately in each such analysis zone, by using the alternative parameters defined hereafter.

Detection Based on Correlation Coefficients

Let us consider a single-source zone Z and denote as $S_k(v)$ the (possibly transformed) source signal which is isolated in this zone. At any point v in this zone, (6.5) with $P = N$ shows that the observed signals become restricted to

$$X_i(v) = a_{ik} S_k(v) \quad \forall\, i \in \{1, \ldots, N\}. \tag{6.7}$$

All observed signals are thus proportional in any single-source zone. A simple and appealing approach for detecting these zones therefore consists in checking the cross-correlation coefficients between these observed signals in all analysis zones. More precisely, for any analysis zone Z, we first define the corresponding (zero-lag non-centered[6]) non-normalized correlation parameter of two arbitrary, possibly transformed, complex-valued signals $U_1(v)$ and $U_2(v)$ as

$$\mathbf{R}\{U_1, U_2\}(Z) = \mathbf{M}\{U_1(v) \times U_2^*(v)\}(Z) \tag{6.8}$$

where the superscript * denotes complex conjugation. In the specific case when $U_1 = U_2$, the parameter $\mathbf{R}\{U_1, U_2\}(Z)$ becomes equal to the mean power of signal $U_1(v)$ over analysis zone Z. Expression (6.8) applies to deterministic and stochastic signals, in the same way as in Sect. 6.2.1.4. For instance, when considering deterministic signals expressed in the TF domain as $U_1(t, \omega)$ and $U_2(t, \omega)$, their above-defined correlation parameter explicitly reads as follows over an analysis zone composed

[6] We again stress that we here restrict ourselves to the version of this method based on the noncentered version of the signals and associated parameters, whereas its centered version is described in [18].

of L TF points which correspond to different times t_p and to the same angular frequency ω:

$$\mathbf{R}\{U_1, U_2\}(Z) = \frac{1}{L}\sum_{p=1}^{L} U_1(t_p, \omega) \times U_2^*(t_p, \omega). \tag{6.9}$$

Whatever signal transform and analysis zones are considered, the (zero-lag noncentered) correlation coefficient of two arbitrary signals $U_1(v)$ and $U_2(v)$ is then defined as

$$\rho\{U_1, U_2\}(Z) = \frac{\mathbf{R}\{U_1, U_2\}(Z)}{\sqrt{\mathbf{R}\{U_1, U_1\}(Z) \times \mathbf{R}\{U_2, U_2\}(Z)}}. \tag{6.10}$$

In our SCA methods, the above parameters are used by considering the correlation coefficients $\rho\{X_1, X_i\}(Z)$ between the observed signals $X_1(v)$ and $X_i(v)$, for $2 \leq i \leq N$. Note that these parameters are undefined if all source signals are equal to zero everywhere in the considered analysis zone, because the numerator and denominator of (6.10) are then equal to zero. To avoid this situation, we set the following condition:

Assumption 6.3 On each analysis zone, at least one source is active.

Besides, the parameters $\rho\{X_1, X_i\}(Z)$ are defined only when the considered observed signals have nonzero mean powers (to avoid division by zero in (6.10)). To guarantee that this condition is met for all observations defined by (6.7) and for any isolated source, we set the following condition:

Assumption 6.4 all mixing coefficients a_{ij} are non-zero.

The above Assumptions 6.3 and 6.4 are "technical assumptions", i.e., they are used in a *theorerical* approach to noiseless mixtures, essentially to avoid the above-mentioned observations with mean powers equal to zero. On the contrary, in *practice*, the (possibly transformed) observed signals contain some noise in addition to source contributions, so that they have nonzero mean powers, even in zones where the source signals yield no significant contributions as compared to noise. Assumptions 6.3 and 6.4 are therefore not required in practice to avoid observations with mean powers equal to zero.

The above coefficients $\rho\{X_1, X_i\}(Z)$ are used as follows. It may easily be derived from (6.7) that the moduli of these coefficients are all equal to one in single-source analysis zones. On the contrary, they should not all be equal to one in multiple-source zones, since we want to use these coefficients to discriminate between single-source and multiple-source zones. To this end, we set the following constraint on the source signals:

Assumption 6.5 Over each analysis zone, all active source signals are linearly independent (if there exist at least two active source signals in this zone).

This assumption is expressed in compact form in order to apply to the deterministic and stochastic frameworks. For deterministic signals, each analysis zone in practice

consists of a discrete set of points of the considered representation domain. For each source, we form the vector composed of the values of this source signal in this zone. Assumption 6.5 then refers to the linear independence of these source vectors. For stochastic signals, we consider the corresponding random variables obtained for a single value of the signal argument v. We recall (see Papoulis and Pillai [41], Ed. 2002, p. 251) that the complex-valued random variables w_i are linearly independent if $E\{|c_1 w_1 + \cdots + c_n w_n|^2\} > 0$ for any $C \neq 0$, where $C = [c_1, \ldots, c_n]$.

It may easily be checked that if the active signals in the considered zone are orthogonal, then Assumption 6.5 is met. However, there also exist cases when several source signals are active and such that Assumption 6.5 is still met, although these signals are not orthogonal. The centered version of our methods leads to the same type of results, except that the orthogonality of the active sources is replaced by their uncorrelation (see [16] for deterministic sources or [17] for nonlinear mixtures of stochastic sources). This shows the attractiveness of our approach: there exist signals which cannot be separated by ICA approaches and classical second-order-statistic BSS methods because they are correlated, while our methods apply to them, at the expense of requesting the source signals to mainly meet the sparsity Assumption 6.2 (p. 10) and Assumption 6.5. This applicability to correlated sources is illustrated for speech signals in Sect. 6.2.3.3.

Under the above assumptions, Appendix 1 shows that the following property is met:

Property 6.1 *An analysis zone Z is a single-source zone if and only if*

$$|\rho\{X_1, X_i\}(Z)| = 1 \quad \forall\, i \in \{2, \ldots, N\}. \tag{6.11}$$

This property is used as follows in our practical method for detecting single-source zones, based on the correlation coefficients of observations. For each analysis zone, we combine the above moduli of correlation coefficients, $|\rho\{X_1, X_i\}(Z)|$, by computing their mean (or e.g., their median, ...) over all i, with $2 \leq i \leq N$. This mean over several correlation coefficients is denoted as

$$\overline{|\rho\{X_1, X_i\}(Z)|}. \tag{6.12}$$

The best single-source zones are then considered to be those corresponding to the highest values of $\overline{|\rho\{X_1, X_i\}(Z)|}$, due to Property 6.1. The detection stage therefore consists in keeping, as the single-source zones, all the zones which are such that $\overline{|\rho\{X_1, X_i\}(Z)|}$ is above a threshold, which is real-valued and slightly lower than one. In the simplest version of this approach, the value of this threshold is selected by the user. The automatic selection of various parameters of our methods is discussed in Sect. 6.2.4.4. It should be noted that the above approach not only performs the detection of single-source zones, but also provides an index of the quality of each such zone Z, namely $\overline{|\rho\{X_1, X_i\}(Z)|}$. One may take advantage of this quality index in a subsequent step of our methods, as discussed further in this chapter.

In practice, the analysis zones where only noise significantly contributes to the observations (as opposed to negligible source signals) do not make the proposed detection criterion fail in the usual case when the noise contributions in diffferent observations are orthogonal, or close to orthogonality: these zones yield low values of $\overline{|\rho\{X_1, X_i\}(Z)|}$ and are therefore not considered as single-source zones, hence they are not used in the subsequent stages of our SCA methods which estimate the mixing matrix. The analysis zones where source signals and noise have low levels can also be excluded from the BMI task by using a slightly modified version of the above detection stage, obtained by replacing standard correlation coefficients (6.10) by

$$\rho\{U_1, U_2\}(Z) = \frac{\mathbf{R}\{U_1, U_2\}(Z)}{\sqrt{\mathbf{R}\{U_1, U_1\}(Z) \times \mathbf{R}\{U_2, U_2\}(Z)} + \varepsilon} \tag{6.13}$$

where ε is a small positive constant. When applying (6.13) to observed signals in zones where their mean powers (and therefore their cross-correlations) are low as compared to ε, the mean correlation coefficient $\overline{|\rho\{X_1, X_i\}(Z)|}$ takes low values so that, again, these zones are not considered as single-source zones nor used in the subsequent stages of our SCA methods which estimate the mixing matrix.

Several versions of our methods use this correlation-based detection stage (in its centered or non-centered version). This first includes our LI-TempCorr method, intended for Linear Instantaneous mixtures of 1D signals considered in the TEMPoral domain, and based on the above CORRelation coefficients. This method was introduced in [14] and detailed in [18]. It is also considered further in this chapter (Sect. 6.4.2.1), in the more general framework of linear-quadratic instantaneous mixtures. The above-mentioned papers [14, 18] also describe the LI-TiF Corr version of this method, which first transforms the original signals into the TIme-Frequency domain. Transforming them into the TIme-Scale domain instead, yields our LI-TiSCorr method, described in [16].

Detection Based on Coherence Functions

When using a TF representation of the considered signals, one may derive a detection stage from another signal processing tool available for that representation, namely the time-segmented coherence function. Our corresponding SCA method, intended for Linear Instantaneous mixtures of 1D signals processed in the TIme-Frequency domain by means of Coherence functions, is called LI-TiFCohere. It is detailed in [5], [15]. In these papers, we consider centered uncorrelated random source signals and we assume that each of them is accessible in the TF domain (Assumption 6.2 p. 10). We first segment the observed signals into successive time intervals, indexed by an integer m, assuming that these signals are stationary over these intervals. We then consider the Power Spectral Densities or PSDs, denoted $\mathbf{S}\{x_i, x_i\}(m, \omega)$, and the Cross-PSDs or CPSDs, denoted $\mathbf{S}\{x_i, x_j\}(m, \omega)$, which are associated with the mixed signals $x_i(t)$ and $x_j(t)$ on each time interval. These parameters therefore depend both on the time index m and on the classical argument of PSDs, namely the

angular frequency ω. The time-segmented "complex coherence function" of observed signals $x_1(t)$ and $x_i(t)$ is then defined as

$$\gamma\{x_1, x_i\}(m, \omega) = \frac{\mathbf{S}\{x_1, x_i\}(m, \omega)}{\sqrt{\mathbf{S}\{x_1, x_1\}(m, \omega) \times \mathbf{S}\{x_i, x_i\}(m, \omega)}} \tag{6.14}$$

and the corresponding time-segmented "real coherence function" reads

$$\Gamma\{x_1, x_i\}(m, \omega) = |\gamma\{x_1, x_i\}(m, \omega)|^2 . \tag{6.15}$$

The single-source TF analysis zones are then defined by the TF points (m, ω) where these time-segmented frequency-dependent real coherence functions of observed signals take the highest values.

The above coherence functions are based on PSDs and CPSDs, and therefore defined in a stochastic framework. However, in practice, they are applied to a single realization of the considered random signals, according to the procedure that we defined in Sect. 6.2.1.4. Therefore, the PSDs and CPSDs of the considered signals must be estimated from this realization. A classical solution to this problem consists in using averaged periodograms, which may be defined as follows. The estimate $\hat{\mathbf{S}}\{u_1, u_2\}(m, \omega)$ of the time-segmented CPSD of two arbitrary signals $u_1(t)$ and $u_2(t)$ is obtained by splitting the time interval associated with m into an arbitrary number L of possibly overlapping sub-intervals, corresponding to time positions t_p, with $p = 1 \ldots L$. The STFTs $U_1(t, \omega)$ and $U_2(t, \omega)$ of the considered signals are then computed over these subintervals and $\hat{\mathbf{S}}\{u_1, u_2\}(m, \omega)$ is derived from them as

$$\hat{\mathbf{S}}\{u_1, u_2\}(m, \omega) = \frac{1}{L} \sum_{p=1}^{L} \frac{1}{M} U_1(t_p, \omega) \times U_2^*(t_p, \omega), \tag{6.16}$$

where M is the number of samples in each sub-interval. PDSs are estimated in the same way with $u_1(t) = u_2(t)$. Comparing (6.16) to (6.9), and the associated estimate of (6.14) to (6.10) then shows that the specific implementation of the LI-TiFCohere detection method based on average periodograms is eventually identical to the LI-TiFCorr detection method, although these two methods were initially introduced from different points of view. The general version of the LI-TiFCohere detection method, based on time-segmented PSDs and CPSDs and not considering specific procedures for estimating them, may therefore be considered as a superset of the LI-TiFCorr detection method.

Detection Based on Ratios of Mixtures

We here consider an approach based on the ratios of observed mixtures defined as

$$\alpha_i(v) = \frac{X_i(v)}{X_1(v)} \quad \forall\, i \in \{1, \ldots, N\}. \tag{6.17}$$

These quantities were also used in the DUET approach [24, 55], but assuming different properties ("W-disjoint orthogonality", only two observed mixtures, possibly underdetermined mixing model). These signal ratios $\alpha_i(v)$ are here defined by using the *first* observation, $X_1(v)$, as a reference signal, but it should be clear that any other observation may be used instead. This could be done by rewriting the definition of $\alpha_i(v)$ and subsequent equations: one could replace $X_1(v)$ by another observation in these expressions. Or, more simply, one should keep in mind that the indices assigned to observations are arbitrary, and one may reassign them for his available data so that he calls $X_1(v)$ the signal he wants to use as a reference.[7] Similar comments apply to the reference signal of the correlation-based and coherence-based detection stages presented above.

Let us again consider a single-source zone Z where the isolated source is denoted $S_k(v)$. As shown by (6.7), at any point v in this zone, we have

$$\alpha_i(v) = \frac{a_{ik}}{a_{1k}} \quad \forall\, i \in \{1, \ldots, N\}. \tag{6.18}$$

Each ratio of mixtures $\alpha_i(v)$ therefore remains constant over a single-source zone, although each observation generally varies on this zone, as shown by (6.7). This may be used to discriminate such zones from multiple-source zones, provided we set the following constraint (which here replaces Assumption 6.5 on p. 13, used in our above correlation-based method):

Assumption 6.6 Over each analysis zone, when several sources are active, they vary so that at least one of the ratios of mixtures $\alpha_i(v)$, with $2 \leq i \leq N$, does not take the same value for all the points v situated in this zone.

The ratios of mixtures thus meet the following property:

Property 6.2 *An analysis zone Z is a single-source zone if and only if*

$$\mathbf{V}\{\alpha_i(v)\}(Z) = 0 \qquad \forall\, i \in \{2, \ldots, N\} \tag{6.19}$$

where $\mathbf{V}\{.\}(Z)$ denotes the variance of the considered (deterministic or stochastic) signal in zone Z.

A practical method for detecting single-source zones may then be derived from this property by computing, for each analysis zone, the mean (or median, ...) of $\mathbf{V}\{\alpha_i(v)\}(Z)$ over all i, with $2 \leq i \leq N$. This mean over several variances is denoted as

$$\overline{\mathbf{V}\{\alpha_i(v)\}(Z)}. \tag{6.20}$$

The best single-source zones are then considered to be those corresponding to the lowest values of $\overline{\mathbf{V}\{\alpha_i(v)\}(Z)}$, due to Property 6.2. The detection stage therefore

[7] The selection of this signal has no influence in the ideal configuration considered hereafter, but it may be of importance in real applications, e.g., when avoiding to take a very noisy signal as the reference.

consists in keeping, as the single-source zones, all the zones which are such that $\overline{\mathbf{V}\{\alpha_i(v)\}(Z)}$ is below a threshold, which is real-valued, small and positive. This approach leads to the same type of comments as the correlation-based detection method, concerning the selection of the above threshold and the use of $\overline{\mathbf{V}\{\alpha_i(v)\}(Z)}$ as a quality index for each analysis zone. These two approaches however differ because the ratios of mixtures $\alpha_i(v) = X_i(v)/X_1(v)$ and resulting variance parameters are not symmetrical with respect to the considered two signals, unlike the above moduli of correlation coefficients defined by (6.10). This asymmetry may degrade the performance of this ratio-based approach, as shown in [18]. This problem may be addressed by also considering the inverse ratios $X_1(v)/X_i(v)$, as detailed in [42] (which mainly concerns a more complex mixing model). However, this increases computational complexity. Therefore, the correlation-based approach described in Section "Detection Based on Correlation Coefficients" is somewhat more attractive in the configuration considered in this section.

The Linear Instantaneous TIme-Frequency version of the above method based on Ratios of Mixtures of originally 1D signals is called LI-TiFROM. It was introduced in [1] and then detailed in [3].[8] The temporal version of this method, which might be called LI-TempROM, is only briefly suggested in these papers. Another version, which might be called LI-TiSROM, may be derived from LI-TiFROM by performing sparsification by means of a time-scale transform (as in the above-mentioned LI-TiSCorr method), instead of the time-frequency transform used in LI-TiFROM. This LI-TiSROM method is thus essentially a combination of the principles used in our above LI-TiSCorr and LI-TiFROM methods. This type of LI-TiSROM method was independently reported in [34].

It should be noted that these ratio-based SCA methods are also applicable to dependent (e.g., correlated) source signals, as, e.g., discussed in [3]. This is illustrated for speech and music signals in Sect. 6.2.3.3. Besides, other approaches for the detection stage of our SCA methods may be derived from considerations about the estimation of the number of sources, as discussed in Sect. 6.2.4.3.

6.2.2.3 Mixture-Column Estimation Stage

As explained in Sect. 6.2.2.1, the mixture-column estimation stage is successively applied to each single-source zone found in the detection stage, in order to derive an estimate of one column of the mixing matrix, up to a scale factor. Three methods may be developed to this end, by, respectively, using the same type of signal parameters as in the above-defined three detection methods:

1. Considering correlation parameters as in Section "Detection Based on Correlation Coefficients", Eq. (6.7) easily shows that, in a zone where $S_k(v)$ is isolated, we have

$$\frac{\mathbf{R}\{X_i, X_1\}(Z)}{\mathbf{R}\{X_1, X_1\}(Z)} = \frac{a_{ik}}{a_{1k}} \quad \forall\, i \in \{2, \ldots, N\}. \tag{6.21}$$

[8] The version in [3] only uses a single ratio of mixtures $\alpha_i(v)$. It was then extended to all above ratios, especially in [42].

Therefore, by filling a column vector with a first value equal to one, followed by estimates of all left-hand terms of (6.21) with $i \in \{2, \ldots, N\}$, one gets an estimate of the column of mixing matrix A corresponding to source $S_k(v)$, up to the scale factor $\frac{1}{a_{1k}}$. Of course, in a blind framework, we do not know which source is active in any given analysis zone, and therefore which matrix column is obtained in this zone. This stage therefore provides unlabelled estimated columns.

2. Similarly, using PSDs and CPSDs as in Section "Detection Based on Coherence Functions", Eq. (6.7) yields

$$\frac{\mathbf{S}\{x_i, x_1\}(m, \omega)}{\mathbf{S}\{x_1, x_1\}(m, \omega)} = \frac{a_{ik}}{a_{1k}} \quad \forall \quad i \in \{2, \ldots, N\}. \tag{6.22}$$

The left-hand term of (6.22) therefore allows one to identify the same mixing parameter ratio as the correlation-based approach, namely a_{ik}/a_{1k}.

3. Finally, for the ratio-based approach of Section "Detection Based on Ratios of Mixtures", Eq. (6.18) shows that the above mixing parameter ratio a_{ik}/a_{1k} is directly provided by the ratio $\alpha_i(v)$ of observed values at any point v of the considered analysis zone. In practice, we use the mean (or median) of this ratio $\alpha_i(v)$ over the zone, i.e. $\mathbf{M}\{\alpha_i(v)\}(Z)$, in order to achieve better robustness with respect to the nonideality of this zone (i.e., low but nonzero values of other sources).

It is natural to use the same type of parameters (e.g., correlation) in the detection and mixture-column estimation stages of an overall SCA method. However, one may also develop "hybrid SCA methods", i.e. methods which use different parameters in these two stages. For instance, a correlation-based detection stage may be combined with a ratio-based mixture-column estimation stage.

6.2.2.4 Mixture-Column Combination Stage

The above mixture-column estimation stage provides a complete set of unlabelled estimates of mixing matrix columns, including one or several estimates for each actual column of A, and called the tentative estimates of these actual columns. We now aim at deriving a single estimate for each actual column. Two types of methods may be used to this end. The basic type of methods aims at selecting one of the tentative columns, for each source, in order to then store it as one of the columns of the final estimate $\hat{A}$ of A (again, this estimate of A is obtained up to permutation and scale factors [9]). The extended type of methods aims at deriving an adequate combination of tentative columns corresponding to the considered source. This combination is not necessarily equal to one of the tentative columns but typically situated "between" them (think, e.g., of their center of gravity). Again, the new column vector thus obtained is stored as one of the columns of the final estimate $\hat{A}$ of A. Various methods may be used to implement these two combination principles:

[9] These two types of indeterminacies are more explicitly shown in the notations used in [18] and hereafter in Sect. 6.4.2.1.

1. The overall set of mixing matrix column estimates available from the mixture-column estimation stage may be seen as a set of data points. Moreover, all points which are estimates of the *same* actual column of A may be hoped to be close to one another, as compared with their distances with respect to the points corresponding to another column of A. Identifying these different subsets of unlabelled points and deriving a representative point for each subset, such as its center of gravity, is nothing but a clustering task [25, 54]. Standard clustering methods, such as k-means (also referred to as c-means or Isodata) [25, 54] may therefore be used to this end, as was done in various other SCA methods. To improve robustness with respect to outliers, one may replace the means used in k-means by medians, which yields the k-medians clustering algorithm, that we used in [44] for another mixing model. The number of clusters to be created is equal to the number N of sources, which is known in the determined case considered here, since it is equal to the known number of observations (the extended case when N is unknown is addressed in Sect. 6.2.4.3).
2. Still using a clustering-based approach, we can proceed further for the class of SCA methods considered in this chapter because, in addition, a quality index is available for each data point to be clustered (i.e., each estimate of a mixing matrix column), as explained above. This allows one to use clustering methods which take advantage of this additional information. For instance, the fuzzy k-means algorithm (also referred to as fuzzy c-means) [25, 54] is used in the SCA method that we proposed in [25] (for another type of observations).
3. A more specific approach for taking advantage of one of the above quality indices consists in first ordering the single-source zones from the best one to the worst one, according to the considered quality index. For our correlation-based detection stage, this means ordering them according to decreasing values of $\overline{|\rho\{X_1, X_i\}(Z)|}$. Similarly, when using our coherence-based detection stage, we order single-source zones according to decreasing values of the average (over observations) of real coherence functions of observed signals. For our ratio-based detection stage, we order single-source zones according to increasing values of $\overline{\mathbf{V}\{\alpha_i(\nu)\}(Z)}$.

   Once all single-source zones have been ordered, we aim at selecting a known number N of columns, among all the tentative columns associated with all above zones, with one column selected per source. To this end, the estimated column corresponding to the first (i.e., best) zone in the ordered list is first kept as the first column of $\hat{A}$. Then, the subsequent zones in this list are successively used as follows. For each zone, the corresponding estimated mixing matrix column is available from one of the methods defined in Sect. 6.2.2.3. This column is kept only if its distance[10] with respect to all previously kept columns is above a user-defined threshold, showing that the considered zone does not contain the same

[10] In the standard version of this method, the distance between two vectors containing estimates of mixing matrix columns is just the norm of the difference between these two vectors. Other versions may be derived by using other similarity measures between vectors, such as their angle : see, e.g., [26].

source as previous zones in the ordered list. This combination stage ends when the number of columns of $\hat{A}$ thus kept becomes equal to the specified number N of sources to be separated (this is theoretically guaranteed to occur because all sources are assumed to be accessible in the considered data). In the extended case when N is unknown, this procedure may be modified so as to determine the value of N, as explained in Sect. 6.2.4.3.
This approach was used in several of our investigations: see especially the detailed reports of the LI-TiFCorr and LI-TiFROM methods respectively provided in [18] and [3]. As compared with clustering-based methods, this approach may decrease computational complexity, but the selection of the value of its threshold distance between mixing matrix columns may be cumbersome in practice.
This method may be considered as an off-line (or batch) method, since it requires one to first perform the complete detection and mixing-matrix column estimation stages, then order all single-source zones, and eventually achieve selection from the complete set of column vectors corresponding to all single-source zones.

4. A modified, on-line, version of the above approach may also be developed. In that case, the three stages achieving detection, mixing-matrix column estimation and mixing-matrix column combination are performed jointly, instead of sequentially in the batch version of the method. More precisely, this on-line method performs the overall BMI task as follows. It first possibly achieves signal sparsification. It then explores the resulting transformed domain, by successively considering analysis zones, e.g., again using a sliding window. For each such zone, it directly performs all the following processing tasks:

   - Detect if this zone is single-source, as above. If it is not single-source, go to next zone. Otherwise, continue processing as described hereafter.
   - Estimate the corresponding column of the mixing matrix, as above.
   - Compare this column with all previously kept ones, and only keep it if its distances with respect to all previously kept columns are above a user-defined threshold.

   This on-line version avoids first building the complete ordered list of single-source zones (not considering whether this may yield sub-optimal results) and may decrease computational cost as compared with the batch version. As compared with clustering-based methods, it still has the drawback of requiring the user to select the value of the distance threshold between columns. This type of on-line method was used in some of our investigations, e.g., reported in [26].

6.2.3 Experimental Results

A detailed report of the performance of the LI-TempCorr and LI-TiFCorr methods is available in [18]. Similarly, the performance of LI-TiFROM is presented in [3] (and [18]). To avoid duplication, we do not detail all these results here. However, in Sect. 6.2.3.1, we focus on one of their major aspects, i.e., we show

how the performance of our LI-TiFCorr method compares with that of various ICA methods. Then, in Sects. 6.2.3.2 and 6.2.3.3, we further analyze other aspects of our SCA methods for linear instantaneous mixtures of 1D sources, which were not reported in [3] and [18].

6.2.3.1 Performance of Time-Frequency SCA Methods Versus ICA

Each of the tests considered here was performed with observations consisting of two artificial linear instantaneous mixtures of two independently recorded speech source signals. Six pairs of sources were thus used and performance was averaged over them. The performance for each test is measured by the output Signal/Interference Ratio ($\mathrm{SIR}^{\mathrm{out}}$) [18] averaged over the two outputs of the separating system. The results thus obtained are shown in Table 6.1. The LI-TiFCorr-NC method is the Non-Centered version that we described above, whereas LI-TiFCorr-C is its above-mentioned centered variant. Table 6.1 (top lines) shows that these two versions of our methods yield very similar performance. Two performance figures are provided for each of these methods: (i) mean $\mathrm{SIR}^{\mathrm{out}}$ over all considered values of the parameters of these methods (see first line in Table 6.1 for considered method), (ii) $\mathrm{SIR}^{\mathrm{out}}$ for optimum parameter values (see second line in Table 6.1). These parameters are: (i) the size of the time windows used for computing STFTs, (ii) the overlap between these windows. Table 6.1 shows that our methods are not very sensitive to the selection of their parameter values, and that they achieve mean output SIRs above 60 dB for the considered type of mixtures and sources, which is very satisfactory.

The subsequent lines of Table 6.1 make it possible to compare the performance figures of the above methods with those of various ICA methods, as implemented in the ICAlab software package[11] [8]. The latter methods here yield significantly lower performance than ours, i.e., SIRs at most around 40 dB. This may be due to the nonstationarity of the considered source signals for some of these ICA methods, whereas our SCA methods are very well suited to nonstationary signals. However, even the SONS method intended for nonstationary signals yields an SIR below 40 dB.

6.2.3.2 Performance of Time-Scale SCA Method

The tests reported here were performed with two artificial linear instantaneous mixtures of two independently recorded continuous speech signals. These signals correspond to different sentences uttered by different male speakers and are therefore uncorrelated (the sample zero-lag centered correlation coefficient of these overall time series is -0.0017). These signals were rescaled so that their maximum absolute values are equal to unity. The mixing matrix was set to

[11] The term "ICA" (for Independent Component Analysis) is here to be understood in a broad sense, i.e., the ICAlab software also implements methods which are only based on second-order statistics.

Table 6.1 Output Signal/Interference Ratios (SIR^{out}) of time-frequency SCA methods (LI-TiFCorr-NC and LI-TiFCorr-C) and ICA methods (AMUSE to SYM-WHITE) (adapted from [18])

BSS method		SIR^{out} (dB)
LI-TiFCorr-NC	Mean over parameters	65.0
	Optimum parameters	71.7
LI-TiFCorr-C	Mean over parameters	64.0
	Optimum parameters	69.5
AMUSE		30.5
EVD2		31.5
EVD24		23.6
SOBI		31.5
SOBI-RO		35.7
SOBI-BPF		28.4
SONS		36.2
JADE-op		2.4
JADETD		34.1
FPICA	Hyper tangent	39.5
	Gauss	41.0
	Cubic	41.9
	5th-order cumulant	25.1
	6th-order cumulant	28.4
PEARSON opt.		42.2
SANG		40.5
NG-FICA		35.7
ThinICA		39.0
ERICA		37.3
SIMBEC		38.8
UNICA		37.3
FOBI-E		16.6
SYM-WHITE		20.1

$$A = \begin{bmatrix} 1 & 0.9 \\ 0.8 & 1 \end{bmatrix}. \tag{6.23}$$

We applied the resulting observations to the centered version of our above-defined LI-TiSCorr method. Each analysis zone here consists of adjacent points in the TS plane, which correspond to the same time position τ and to a discrete set of L adjacent scales d_p, with $p = 1 \ldots L$. The reported tests were performed for various mother wavelets $\psi(t)$ and various numbers L of TS points in analysis zones. The CWTs were computed with the Wavelab 802 package available at [56], using default parameter values. The resulting output SIRs are shown in Table 6.2. They are higher than 40 dB whatever the parameter values of our method. This demonstrates the good separation capability of this approach and also shows that it is "automated" in the sense that the above parameters do not require user tuning to achieve good performance.

Table 6.2 Output Signal/Interference Ratio (in dB) of centered LI-TiSCorr method, depending on mother wavelet and number L of time-scale points in analysis zones, for uncorrelated speech signals [16]

Wavelet	L		
	4	8	12
Morlet	66.3	71.4	76.2
Gaussian	48.9	69.0	41.8
Gaussian derivative	41.9	56.8	55.3
Mexican hat	50.1	60.1	73.2

6.2.3.3 Performance for Correlated Sources

One of the advantages that we claimed above for our SCA methods, whatever the considered signal representation domain, is their ability to operate with dependent sources, including correlated ones. We here illustrate this property by means of two examples, which cover different types of SCA methods.

In the first example, we consider the same time-scale correlation-based method as in Sect. 6.2.3.2, i.e., the centered version of LI-TiSCorr. We tested this method with two correlated sources, created as follows. As compared with Sect. 6.2.3.2, we used an additional speech signal, from a female speaker. We rescaled it so that its absolute maximum value is equal to 0.2, and we added it to each of the previous two male speech signals. The two signals thus obtained were considered as the sources in this second series of tests with LI-TiSCorr, and we again mixed them according to (6.23). Unlike ICA approaches, our LI-TiSCorr method is supposed to be applicable to these correlated source signals (the sample zero-lag centered correlation coefficient of these overall time series is 0.090). However, due to the addition of the female speech signal to each male signal, each source considered here may fill the TS plane to a larger extent than in the tests reported in Sect. 6.2.3.2. This may reduce the amount and quality of single-source analysis zones and may therefore somewhat degrade performance as compared with Table 6.2. This analysis is confirmed by our test results (see Table 6.3): the output SIRs achieved here tend to be lower than in Sect. 6.2.3.2, but they are still higher than 40 dB (except in two cases with $L = 4$, so that such small analysis zones should preferably not be used).

Our second example involves two audio source signals, processed by the time-frequency ratio-based version of our methods, i.e., LI-TiFROM. Source s_1 is a guitar playing a D chord, which consists of D, $F\#$, A. Source s_2 is a D from a singer. These sources are strongly correlated, as may be seen in Fig. 6.3: this figure is obtained by splitting the source signals in successive 256-sample time windows and computing, for each such window, the absolute value of the corresponding sample estimate of the zero-lag non-centered correlation coefficient of the source signals, i.e., the estimate of $|E\{s_1(t)s_2(t)\}|/\sqrt{E\{s_1^2(t)\}E\{s_2^2(t)\}}$. The spectrograms provided in Figs. 6.4 and 6.5 show that these source signals are active in different parts of the TF plane. One may therefore hope that they yield single-source zones and can thus be sepa-

Table 6.3 Output Signal/Interference Ratio (in dB) of centered LI-TiSCorr method, depending on mother wavelet and number L of time-scale points in analysis zones, for correlated speech signals [16]

Wavelet	L		
	4	8	12
Morlet	53.9	43.0	65.8
Gaussian	1.4	55.3	43.9
Gaussian derivative	41.7	55.3	45.9
Mexican hat	0.4	55.4	49.5

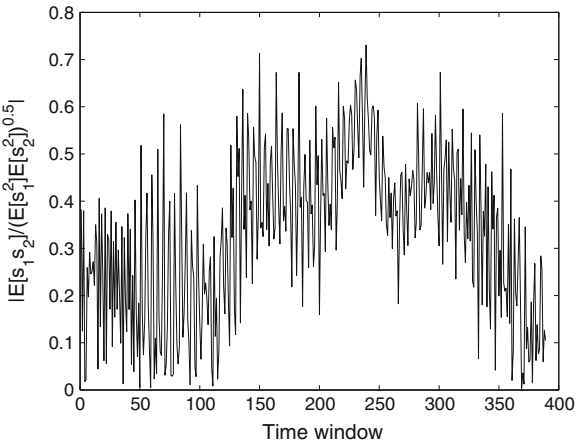


Fig. 6.3 Absolute value of cross-correlation coefficient of source signals versus index of 256-sample time window [2]

rated by our methods, although they are correlated. We checked it by mixing these sources, again by means of matrix (6.23), and processing these observations with our LI-TiFROM method. Each analysis zone here consists of a set of adjacent points in the TF domain, corresponding to: (i) a set of M successive half-overlapping time windows (with indices n_j) which cover an overall time interval here denoted as T_q and (ii) a single angular frequency ω_k. We analyzed the variance of the ratio of mixtures[12] $\alpha(n_j, \omega_k)$ over $M = 8$ TF points. The variations of the inverse of this variance, i.e. $\frac{1}{\mathbf{V}\{\alpha(n,\omega)\}(T_q,\omega_k)}$, with respect to T_q and ω_k are shown in Fig. 6.6. This confirms that some parts of the TF plane yield very low variance (see upwards peaks in Fig. 6.6) and therefore consist of single-source analysis zones. The considered configuration yields output SIRs equal to 34.2 dB for s_1 and 71.3 dB for s_2

[12] For two observations, (6.17) yields a single ratio of mixtures, corresponding to $i = 2$ which is omitted in the notations used here. Moreover, in the investigation [2] reported here, that ratio was defined as the inverse of the right-hand term of (6.17).

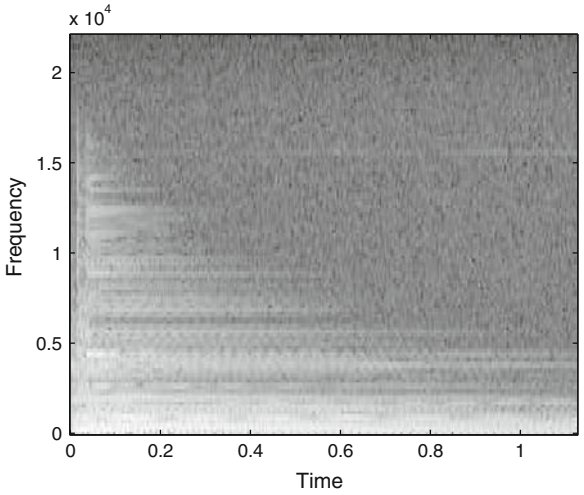


Fig. 6.4 Spectrogram of guitar s_1 (time in seconds, frequency in Hz) [2]

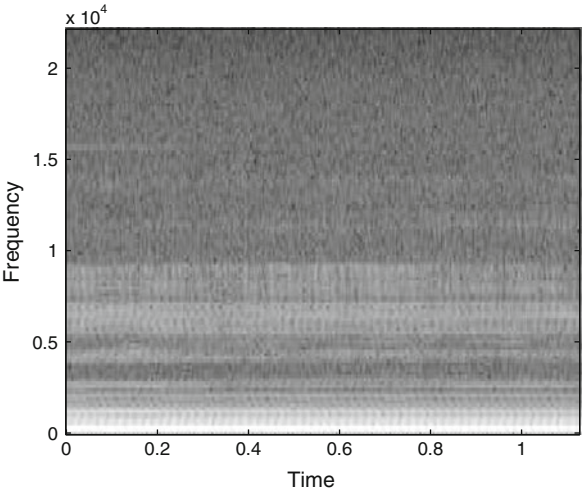


Fig. 6.5 Spectrogram of voice s_2 (time in seconds, frequency in Hz) [2]

which are quite good values for such dependent signals. On the contrary, the classical kurtosis-based FastICA method [22] failed to separate these source signals, due to their dependence.

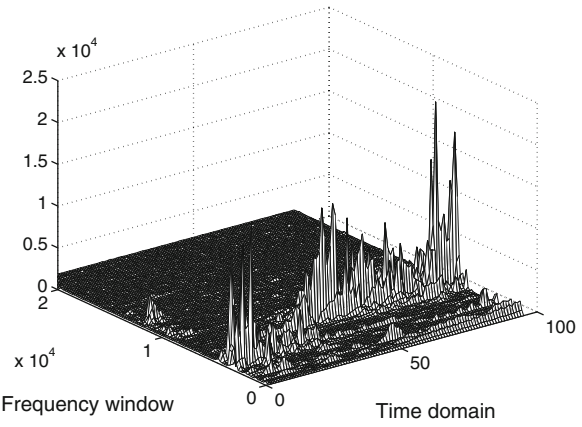


Fig. 6.6 Time-Frequency representation of inverse of variance of ratio of mixtures, i.e. $\frac{1}{\mathbf{V}\{\alpha(n,\omega)\}(T_q,\omega_k)}$ (vertical axis). Horizontal axes variables: index of time interval involved in analysis zone, frequency (in Hz) (adapted from [2])

6.2.4 Extensions and Related Works

6.2.4.1 Underdetermined and Overdetermined Mixtures

Up to this point, we only considered determined mixtures. We now first comment about underdetermined ones, which correspond to the situation when the number P of observed signals is strictly lower than the number N of source signals. For such mixtures, one may consider applying in the same way as in Sect. 6.2.2.1 the first four stages (sparsification to mixture-column combination stages) of the overall structure of our SCA methods, in order to estimate the mixing matrix, i.e., to perform the BMI task. The last stage of our overall SCA methods, i.e., the estimation of the source signals, then yields an issue: the mixing matrix is no more square nor invertible, so that the source signals cannot be estimated by means of (6.6). This is a quite general phenomenon for SCA methods and various solutions to this problem have been reported in the literature (see, e.g., the overview in [12]). As mentioned in Sect. 6.2.1.4, one of them consists in only allowing one source to be active at any point of the considered representation domain ("W-disjoint orthogonality"), as in the DUET method [24, 55]. In [18], we proposed an alternative approach, which sets lower restrictions on the sources (i.e., the same as those defined above in this chapter), at the expense of only performing "partial BSS", defined hereafter. Once the estimate $\hat{A}$ of the (permuted and scaled) rectangular mixing matrix has been obtained, our approach operates as follows. We keep P columns of this matrix $\hat{A}$, which defines the P sources that we select as the sources of interest. We thus obtain a square sub-matrix $\hat{A}'$ of the mixing matrix $\hat{A}$. Moreover, $\hat{A}'$ is invertible thanks to Assumption 6.1 (p. 4). The observed signals may then be considered as mixtures of the P sources of interest associated with the mixing sub-matrix $\hat{A}'$, plus "noise"

composed of contributions from the other $(N - P)$ sources. We then transfer these observed signals through the inverse of this square sub-matrix $\hat{A}'$, as in (6.6). We thus obtain output signals which separate each of the P sources of interest from the other sources of interest, i.e., these output signals each contain a contribution from only one of the P sources of interest, plus again "noise" consisting of contributions from the other $(N - P)$ sources. We thus achieve what we call "partial BSS" for the selected P sources. Note that we may thus choose arbitrarily for which subset of P columns of $\hat{A}$ and associated sources, among the initial N columns and sources, we perform partial BSS.

The third type of linear instantaneous mixtures consists of overdetermined ones, which correspond to the situation when the number P of observed signals is strictly higher than the number N of source signals. Unlike underdetermined mixtures, this case yields no issue. It may e.g. be handled by replacing the inverse of the mixing matrix by its pseudo-inverse, as discussed in [17] (see also Deville and Puigt [18]).

A particular configuration may also be defined by setting the following two constraints. First, the mixture is *globally* underdetermined, i.e., when considering the *complete* set of observed data samples, the number of sources contributing to them is higher than the number of observations, as in the above standard underdetermined case. Second, this mixture is however *locally* determined or overdetermined, i.e., in any analysis zone the number of active sources is at most equal to the number of observations. The version of our BSS methods that we developed in [25] is suited to this configuration.

6.2.4.2 Making Sources Accessible, Deflation-Based Methods

A general issue of various SCA methods is that their constraints on joint source sparsity have lower chance to be met when the number of sources increases. For the methods presented in this chapter, which are focused on single-source zones, this means that the chance to have *only* one source active at a given point of the considered signal representation decreases when the number of sources increases. Some sources may then become unaccessible (in the sense of Definition 6.3, p. 9), so that Assumption 6.2 (p. 10) may not be fulfilled. Fortunately, the following extensions of our methods apply to certain cases when only part of the sources are accessible. The basic version of this type of extension [18] may be defined as follows, with $P = N$ just for the sake of simplicity. Let us assume that at least one of the sources, denoted as $S_k(v)$, is accessible in the considered representation domain. Our above SCA methods then make it possible to estimate the corresponding column of the scaled permuted mixing matrix, say column j. This estimated column consists of elements $\widehat{b_{ij}}$ which are estimates of

$$b_{ij} = \frac{a_{ik}}{a_{1k}} \qquad \forall\, i \in \{1, \ldots, N\}, \tag{6.24}$$

as shown by Sect. 6.2.2.3. We then compute a modified version of each mixed signal $x_i(t)$, with $i = 2 \ldots N$, according to

$$x'_i(t) = x_i(t) - \widehat{b_{ij}} x_1(t) \qquad \forall\, i \in \{2, \ldots, N\}. \tag{6.25}$$

Equations (6.1) and (6.24) show that we thus obtain $N-1$ signals which do not contain any contributions from source $s_k(t)$ (up to errors due to the estimation of b_{ij}). The key point is then that, even if some sources were not accessible from the initial set of N mixed sources, at least one of them may become accessible from the new set of $N-1$ mixed sources involved in the modified mixed signals $x'_i(t)$. This depends on the distributions of all sources in the considered representation domain, and happens if some sources were initially hidden (i.e., they were not isolated in any analysis zone when considering the initial set of N sources), but they are isolated in at least one zone when considering the set of $N-1$ sources which remains after canceling the contributions from source $s_k(t)$ in all mixed signals. If at least one source is accessible from this new set of $N-1$ mixed sources, the same procedure may be applied again. This recursive procedure ends when (no more sources are accessible or when) the number of recombined signals is thus decreased down to one, and this signal contains a single source. This procedure thus succeeds in extracting this source, although not all sources were initially accessible. The same procedure may then again be applied, by selecting other sources $s_k(t)$ at each step of its recursion in order to extract other sources. This approach is reminiscent of deflation-based BSS methods which were used in the framework of ICA, especially in [13], and then became popular with the deflation-based version of the kurtotic FastICA method [22].

More advanced versions of this type of procedure may also be defined in order to cancel, at each intermediate stage of the recursion, the contributions from all the sources which are accessible at that stage. This is detailed in [18]. Besides, deflation-based modified versions of our SCA methods were also proposed by independent research groups: see [52] and, to a lower extent, [53].

6.2.4.3 Estimating the Number of Sources, Other Methods for Detection Stage

Let us consider the case of non-underdetermined mixtures (i.e. $P \geq N$) for the sake of simplicity. If the number of source signals is unknown, one may still apply the first three stages of our SCA methods (sparsification to mixture-column estimation stages) in the same way as above. One may then modify their fourth stage (mixture-column combination stage) so that it estimates the number of sources. For clustering-based combination stages, this corresponds to automatically selecting the number of clusters, and various methods have been developed to this end in the literature (see e.g., references in [25], which uses one of these methods).

One may also change the operation of our combination stages based on an ordered list: for instance, a variant of their batch version still processes the successive columns of the ordered list as explained in Sect. 6.2.2.4, but ends when the considered quality parameter (e.g., $\overline{|\rho\{X_1, X_i\}(Z)|}$) gets beyond a user-defined threshold, which shows that the next zones in the ordered list should not be considered as single-source zones. The number of mixing matrix columns kept at this stage is the estimate of the number of sources provided by this approach.

Whereas the above approach is directly related to the principles of the considered SCA methods, one may instead here estimate the number of sources by using standard methods which have been considered in the general framework of BSS, or in related fields such as array processing for direction of arrival estimation. Various such methods are based on the properties of the correlation or covariance matrix of the observations, especially on the number of its eigenvalues which are above a threshold or on the shape of this series of eigenvalues (see e.g., Luo and Zhang [36]).

All available methods for estimating the number of sources are also of interest for a different aspect of the SCA methods studied in this chapter, i.e., for their detection stage. This results from the fact that the single-source tiny zones to be found by this detection stage consist of the parts of the (possibly transformed) observations where the "local number of sources" is equal to one. Instead of using the procedures described in Sect. 6.2.2.2 for detecting these zones, one may therefore take advantage of any method available from the literature for determining the number of sources involved in mixed signals, by here applying these methods locally, i.e., successively to each analysis zone, in order to only keep the zones where this number of sources is equal to one. In particular, one may use methods based on the number of non-negligible eigenvalues of local correlation or covariance matrices of the observations, or on the magnitude of the first eigenvalue as compared with all subsequent ones. This type of approach was e.g., used in the literature in the method recently described in [6], which has significant relationship with our LI-TiFROM approach, as explained in [6].

6.2.4.4 Automating Selection of Parameter Values

The above SCA methods involve a set of parameters, such as some thresholds or the size of analysis zones and their percentage of overlap. Up to this point, we mainly considered the case when the values of these parameters are selected by the user, which may be cumbersome. A solution to this problem consists in extending these SCA methods by introducing automated procedures for selecting (some of) their parameter values. For instance, we developed an approach which essentially consists in successively running the considered SCA method for various values of its parameters and keeping the parameter values which yield the best performance, according to a performance criterion which may be computed in a blind mode (at the expense of some restrictions on the source signal properties). This approach is detailed in [38]. Another automated approach was proposed by an independent research group in [7] (for our methods and others).

6.2.4.5 Related Works

The methods depicted in this chapter inspired various extensions from other research groups. Besides, our methods were considered in various papers from the literature, to show how our approaches differ from the other methods proposed in those papers.

Some of these aspects were presented above, e.g., in Sections "Detection Based on Ratios of Mixtures", 6.2.4.2, 6.2.4.3 and 6.2.4.4. Moreover, we discussed above the relationship between our methods and DUET. Other related works from the literature especially include different extensions of our methods which were developed by Smith et al. and, e.g., reported in [50–53]. Besides, a method "that can be viewed as an extension of the DUET and the TIFROM methods" is presented in [35]. The relationship between our methods and other approaches proposed in the literature is also, e.g., discussed in [48].

6.3 Convolutive Mixtures of 1D Sources: Major Principles

Beyond the configuration studied in Sect. 6.2, the first natural extension of our investigations concerns *convolutive* mixtures of 1D sources. This case is of major importance, since it is e.g., faced in acoustics or communications. For instance, when using a set of microphones to record simultaneous audio signals, propagation effects generally result in convolutive mixtures. For audio signals, the linear instantaneous mixing model considered in Sect. 6.2 is a restrictive configuration, only applicable to some mixtures, such as those artificially created in studios by combining successive single-track recordings in order to derive the final version of a song record.

The class of SCA methods addressed in this chapter is also applicable to convolutive mixtures, and this would therefore deserve a detailed description. However, we will here restrict ourselves to a presentation of their major principles, because they remain relatively similar to those presented in Sect. 6.2 and we will save space for less similar configurations, namely those based on nonlinear mixtures or 2D sources. We first focus on a specific class of (non-instantaneous) convolutive mixtures in Sect. 6.3.1 and then proceed to general convolutive mixtures in Sect. 6.3.2.

6.3.1 Attenuation-Delay (or Anechoic) Mixtures

Considering a discrete-time representation of signals, with an integer-valued time index denoted as n, determined attenuation-delay mixtures are defined in the time domain as

$$x_i(n) = \sum_{j=1}^{N} a_{ij}\, s_j(n - n_{ij}) \qquad \forall\, i \in \{1, \ldots, N\} \tag{6.26}$$

where we use the same notations as in (6.1) and we introduce the integer-valued[13] time shifts n_{ij}. This model is suited to propagation without reflections (and is therefore also called an "anechoic mixture"): a_{ij} then represents attenuation along the direct path from source i to sensor j, whereas n_{ij} is the propagation delay along that path.

[13] The case of non-integer time shifts is addressed in [45].

Let us take the STFT of (6.26). If the time shifts n_{ij} are negligible as compared with the temporal width of the windowing function used in the STFT transform, this yields

$$X_i(n,\omega) = \sum_{j=1}^{N} a_{ij}\, e^{-j\,\omega n_{ij}}\, S_j(n,\omega) \qquad \forall\, i \in \{1,\ldots,N\}. \tag{6.27}$$

A first SCA method may then be derived by again considering ratios of mixtures, here expressed in the TF domain, i.e.

$$\alpha_i(n,\omega) = \frac{X_i(n,\omega)}{X_1(n,\omega)} \qquad \forall\, i \in \{1,\ldots,N\}. \tag{6.28}$$

If a source $S_k(n,\omega)$ is isolated in an analysis zone composed of TF points (n_p,ω_l), Eq. (6.27) shows that all these points are such that

$$\alpha_i(n_p,\omega_l) = \frac{a_{ik}}{a_{1k}} e^{-j\omega_l(n_{ik}-n_{1k})} \qquad \forall\, i \in \{1,\ldots,N\}. \tag{6.29}$$

Considering real positive coefficients a_{ij} (this condition is met for attenuation without reflection), the modulus of $\alpha_i(n_p,\omega_l)$ in a single-source zone is therefore equal to $\frac{a_{ik}}{a_{1k}}$, i.e., to the same quantity as in Sect. 6.2. The SCA methods described in that section may therefore be easily extended so as to estimate the (ratios of) coefficients a_{ij} of the mixing model considered here, now using the *modulus* of $\alpha_i(n_p,\omega_l)$ in single-source zones. Besides, let us consider a single-source zone corresponding to a single time position n_p and to different frequencies ω_l. Equation (6.29) shows that the unwrapped phase of $\alpha_i(n_p,\omega_l)$ in this zone linearly varies with respect to frequency, and that the slope of this line may be used to estimate (the differences between) the time shifts n_{ij} of the mixing model considered here. This yields an SCA method for Attenuated and Delayed mixtures, operating in the TIme-Frequency domain and based on Ratios Of Mixtures, which is therefore called AD-TiFROM. Many details about its principle, variants and experimental performance are provided in [42].

Using correlation-based parameters instead of the above ratios of mixtures yields the AD-TiFCorr version of this type of methods. Several variants of AD-TiFCorr are described in [43–45]. Using a stochastic framework and coherence functions, one may also develop a corresponding AD-TiFCohere method, as an extension of the LI-TiFCohere method described in Sect. 6.2.

This attenuation-delay mixing model may seem to be quite restrictive, but it has been widely considered in the BSS literature. In particular, the above-mentioned DUET method [24, 55] uses this model. However, a more general, i.e., full convolutive, model should also be considered, in order to handle more complex configurations, e.g., involving multi-path propagation (early reflections and reverberation in acoustics). We hereafter proceed to this case.

6.3.2 General Convolutive Mixtures

Still considering a discrete-time representation of the signals and using the above notations, convolutive mixtures are defined in the time domain as

$$x_i(n) = \sum_{j=1}^{N} a_{ij}(n) * s_j(n) \qquad \forall\, i \in \{1, \ldots, N\} \tag{6.30}$$

where $a_{ij}(n)$ is the impulse response of the filter (e.g., representing overall propagation) from source i to sensor j, and $*$ stands for discrete-time convolution.

Let us take the STFT of (6.30), assuming that the parts of the impulse reponses $a_{ij}(n)$ with non-negligible magnitude are much narrower than the temporal windowing function used in the STFT transform (the impulse responses $a_{ij}(n)$ do not depend on the considered time window of the STFT transform). This yields

$$X_i(n, \omega) = \sum_{j=1}^{N} A_{ij}(\omega)\, S_j(n, \omega) \qquad \forall\, i \in \{1, \ldots, N\}. \tag{6.31}$$

The specific scale factors $a_{ij}\, e^{-j\,\omega n_{ij}}$ of (6.27), encountered in attenuation-delay mixtures, are here replaced by general factors $A_{ij}(\omega)$, but the mixing equations keep their previous linear structure with respect to the transformed sources $S_j(n, \omega)$. Part of the principles that we developed in our above SCA methods therefore extend to general convolutive mixtures. In particular, considering the ratio of mixtures (6.28) in an analysis zone where source $S_k(n, \omega)$ is isolated here yields

$$\alpha_i(n_p, \omega_l) = \frac{A_{ik}(\omega_l)}{A_{1k}(\omega_l)} \qquad \forall\, i \in \{1, \ldots, N\}. \tag{6.32}$$

All these coefficients $\alpha_i(n_p, \omega_l)$ therefore essentially make it possible to identify the column of filters $A_{ik}(\omega)$ associated with the source which is isolated in the considered analysis zone, as in the linear instantaneous case. However, some additional important phenomena should here be taken into account. First, this identification is again achieved up to a "scaling effect", here defined by the denominator of (6.32). This effect is the counterpart of the denominator of (6.18) in the linear instantaneous case. However, instead of a scaling *factor* (i.e., division by a *constant* value) in the latter case, the generalized scaling effect faced here is a division by the frequency response of a filter, i.e., by a frequency-dependent quantity. If using a basic separating system structure, the outputs of that system are therefore equal to the actual sources up to complicated filter frequency responses, as detailed in [4]. The estimated source signals are thus altered by a frequency-dependent gain (frequency distortion), which is a major drawback in various applications, such are speech and/or music separation. Fortunately, this issue may be solved by adding, to the above separating system, post-processing filters which are only expressed with respect to the above-identified

frequency responses (6.32). The resulting signals are equal to each contribution of each source in each observation (6.31). The sources are thus restored with no additional frequency distortion, as compared with the filtering effects obtained when recording each source separately with the same set of sensors, which is quite acceptable. The post-processing filters that may be used to this end are also defined in [4].

Another phenomenon makes this convolutive configuration different from the linear instantaneous and attenuation-delay mixtures. In the latter two configurations, only a limited set of mixing parameters are to be estimated, i.e., the (ratios / differences of) scale factors a_{ij} and possibly time shifts n_{ij}. Mapping the considered signals to the TF domain by means of STFT strongly relaxes the sparsity constraints set on the sources because, essentially, *a tiny single-source zone in the frequency domain*, for a given time interval, is enough for identifying a_{ij} and n_{ij}, and these parameters then apply to the *complete* TF plane to restore the sources. Things are different for convolutive mixtures. Let us consider a single source $S_k(n, \omega)$ as an example. We aim at estimating each corresponding filter response $A_{ik}(\omega)$ (up to the scaling effect). This quantity depends on ω. Therefore, we need single-source zones *for all frequencies* (unless we perform frequency smoothing for nonidentified parts of the frequency response $A_{ik}(\omega)$). The simplest case is when all frequencies where source $S_k(n, \omega)$ is isolated are obtained for the same time interval of the considered STFTs. In other words, in the standard version of the SCA methods that we propose for convolutive mixtures, the sparsity constraint set on the sources is that, for each of them, there should exist a time interval where all other sources are "completely inactive", i.e., equal to zero at *all* frequencies. The source signals then contain silence phases in the time domain. This constraint is therefore more stringent than in the case of linear instantaneous and attenuation-delay mixtures, but this is the price to pay for estimating a much wider set of parameters (i.e. a different complex value at each frequency).

The proposed SCA methods for convolutive mixtures therefore first detect the time intervals where a source is isolated, i.e., where all other sources are completely inactive. This may be achieved by again using the real coherence function of observations, defined in Sect. 6.2. The overall quality, with respect to the single-source property, of a time interval may then, e.g., be measured by the mean of the above coherence function over all frequencies. In a modified version of this approach, one may focus on the frequency bands which are of interest in the considered application, and only compute the mean of the above coherence function over these bands. This approach is used for speech signals in [4], where averaging of the coherence function is performed over the band [0, 800Hz].

The above principles may be used to develop Conv-TiFCorr, Conv-TiFCohere and Conv-TiFROM variants of our methods for Convolutive mixtures of 1D sources. The Conv-TiFCohere method is defined in detail in [4], which also contains various test results.

6.4 Nonlinear Mixtures of 1D Sources

Beyond the above types of linear mixing models, various applications involve *nonlinear* mixtures of 1D sources. This extended case is much tougher than the linear one, and few BSS methods have been proposed to address it (see e.g., the review in [12]): some ICA methods for nonlinear mixtures have been described in the literature, but very few reported BSS methods for such mixtures are based on SCA. In this section, we show that the general framework proposed in this chapter makes it possible to develop SCA-based methods for certain types of nonlinear mixtures. We first focus on a specific type of nonlinear mixtures, namely linear-quadratic instantaneous ones, that we mainly select because they make it relatively easy to show how the proposed SCA methods may be extended beyond linear mixtures. This linear-quadratic instantaneous mixing model is of real practical interest, since it is faced in various application fields, including remote sensing[14] [40]. However, in some of these applications, other BSS approaches than those presented in this chapter should be used, since the signals faced in these applications are not sparse in the considered domain. Beyond linear-quadratic instantaneous mixtures, we also comment about other types of nonlinear mixtures in Sect. 6.4.3.

6.4.1 Problem Statement, Definitions and Assumptions

6.4.1.1 Signal Representation

We here consider the following configuration in the original signal representation domain, again denoting the argument of 1D signals as t. The available P signals $x_i(t)$ are mixtures of N source signals $s_j(t)$, with $P = N(N+1)/2$ in the most general case, as explained below. The source signals are unknown, stochastic and real-valued. The mixing model consists of linear terms, proportional to $s_j(t)$, and quadratic cross-terms, proportional to $\tilde{s}_{jk}(t) = s_j(t)s_k(t)$. Each observed signal then reads

$$x_i(t) = \sum_{j=1,\ldots,N} a_{ij}s_j(t) + \sum_{1\le j<k\le N} q_{ijk}\tilde{s}_{jk}(t) \qquad \forall\, i \in \{1,\ldots,P\} \quad (6.33)$$

where a_{ij} and q_{ijk} are, respectively, linear and quadratic unknown real-valued mixing coefficients. This yields in matrix form

$$x(t) = As(t) + Q\tilde{s}(t) \quad (6.34)$$

with $x(t) = [x_1(t),\ldots,x_P(t)]^T$ and $s(t) = [s_1(t),\ldots,s_N(t)]^T$. The column vector $\tilde{s}(t)$ consists of the signals $\tilde{s}_{jk}(t)$ in a given arbitrary order. Besides, $A = [a_{ij}]$ and

[14] We here refer to the spectral-source approach for remote sensing (i.e., sources depending on wavelength), as opposed to the spatial-source approach.

$Q = [q_{ijk}]$, where i is the row index of Q and the columns of Q are indexed by (j, k) and arranged in the same order as the signals $\tilde{s}_{jk}(t)$ in $\tilde{s}(t)$. We also consider the centered version of the observations, i.e.,

$$x'_i(t) = x_i(t) - E\{x_i(t)\} \qquad \forall\, i \in \{1, \ldots, P\}. \tag{6.35}$$

Eqs. (6.33) and (6.35) then yield

$$x'_i(t) = \sum_{j=1,\ldots,N} a_{ij} s'_j(t) + \sum_{1 \le j < k \le N} q_{ijk} \tilde{s}'_{jk}(t) \qquad \forall\, i \in \{1, \ldots, P\}, \tag{6.36}$$

where $s'_j(t)$ and $\tilde{s}'_{jk}(t)$ are respectively the centered versions of $s_j(t)$ and $\tilde{s}_{jk}(t)$. This yields in matrix form

$$x'(t) = As'(t) + Q\tilde{s}'(t) \tag{6.37}$$

where the vectors $x'(t)$, $s'(t)$ and $\tilde{s}'(t)$ are the centered versions of those involved in (6.34).

We here do not apply any transform to the observed signals, since this is likely to make the mixing model more complex: for instance, the Fourier transform changes each quadratic term of the original model into the convolution of the considered two signals. Since we only consider the original representation of the signals in this section, we keep their original notations for the sake of simplicity.

6.4.1.2 Definitions and Assumptions

We here consider the required definitions and assumptions in the same order as in Sect. 6.2 and we adapt them as follows. Here, the mixing matrix A is not square (see above values of N and P) but the constraint that we set on it can still be expressed as Assumption 6.1 (p. 4). Besides, the proposed method is suited to sources which meet the following condition:

Assumption 6.7 All sources $s_1(t), \ldots, s_N(t)$ are zero-mean at any time t.[15]

Definition 6.1 (p. 8) then becomes, in the time domain:

Definition 6.1′ A signal is "active" at time t if it has non-zero power[16] at that time.[17] It is "inactive" at time t if it has zero power at that time and may then be considered as a deterministic constant.

---

[15] The observations may then be non-zero-mean, due to the nonlinear nature of the mixing model and the possible source correlation. We therefore consider the centered version $x'_i(t)$ of the observations hereafter.

[16] Or, equivalently, nonzero variance.

[17] As explained in Sect. 6.2.1.4, each considered temporal analysis zone is restricted to a single time t in this theoretical statistical framework. However, in practice, all signal moments are estimated over time intervals and each considered temporal analysis zone then consists of such an interval.

Definition 6.2 (p. 9) is here used to define when a signal $s_j(t)$ (not a signal $\tilde{s}_{jk}(t)$ of (6.33)) is isolated. This definition is still expressed as in page 9, taking into account that the analysis zone is here the single time t and that the considered set of sources "mixed in the observations" contains $s_1(t), \ldots, s_N(t)$ but does not contain the signals $\tilde{s}_{jk}(t)$. Definition 6.3 (p. 9) and Assumption 6.2 (p. 10) here remain unchanged, taking into account that they are also applied to $s_j(t)$, not $\tilde{s}_{jk}(t)$, in the time domain, for any analysis zone consisting of a time t.

The considered sources are therefore nonstationary,[18] since their powers are zero at some times and nonzero at others. Moreover, they are only requested to have a slight sparsity in the time domain, in the sense that they are allowed to overlap almost everywhere: for each source, we only request the existence of a time t (i.e., a short time interval for practical estimation) when only this source is active.

We hereafter consider the correlation-based version of our methods. The associated Assumption 6.3 (p. 13) and Assumption 6.4 (p. 13.) then remain unchanged. Assumption 6.5 (p. 13) here becomes:

Assumption 6.5' For any considered time t, the signals which are contained by $s'(t)$ and $\tilde{s}'(t)$ and which are active at that time are linearly independent (if there exist at least two such active signals at that time).

This assumption is again based on the definition of linear independence of random variables that we provided on page 14. It again means that the proposed method also applies to situations where the active signals in $s'(t)$ and $\tilde{s}'(t)$ are correlated, which is an attractive feature as compared with ICA methods.

6.4.2 Proposed SCA Method

Considering the above-defined Linear-Quadratic Instantaneous mixtures, we now present the Temporal SCA method based on Correlation parameters, and therefore here called LQI-TempCorr, that we introduced in [17] for handling this configuration.

6.4.2.1 Identification of Linear Part of Mixture

The first step of our method consists in identifiying the "linear part" of the mixing model, i.e., the matrix A, or more precisely the matrix $B = [b_{ij}]$, where

$$b_{ij} = \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}} \qquad \forall\, i \in \{1, \ldots, P\}, \qquad \forall\, j \in \{1, \ldots, N\} \tag{6.38}$$

[18] More precisely, they are long-term nonstationary, but they should be short-term stationary in practice, in order to make it possible to estimate the above-mentioned signal moments over short time intervals.

and $\sigma(.)$ is a permutation. B is therefore a modified version of A, where the columns are permuted and each column is rescaled with respect to the value in its first row, i.e., with respect to its linear contribution in observation $x_1(t)$.

As shown in [17], despite the presence of the quadratic part of the mixing model, the linear part B of this model may be identified by the same type of procedure as in our LI-TempCorr method, which was defined in Sect. 6.2 (here using the centered version of the signals). This procedure, which is detailed in [17], is therefore skipped in this section, where we directly proceed to the aspects of the BMI and BSS tasks which are specific to the linear-quadratic instantaneous mixing model.

6.4.2.2 Cancelation of Linear Part of Mixture

We then aim at deriving a set of L signals $z_l(t)$ from the observations $x_i(t)$, in such a way that these signals $z_l(t)$ only contain quadratic cross-terms, i.e. terms proportional to $\tilde{s}_{jk}(t)$. To this end, we consider signals defined as

$$z_l(t) = x_1(t) - \sum_{i=2}^{P} c_{li} x_i(t) \qquad \forall\, l \in \{1, \ldots, L\}. \tag{6.39}$$

Combining this expression with (6.33) and (6.38) yields

$$z_l(t) = \sum_{j=1,\ldots,N} a_{1,\sigma(j)} s_{\sigma(j)}(t)[1 - \sum_{i=2}^{P} b_{ij} c_{li}] + \sum_{1 \le j < k \le N} r_{ljk} \tilde{s}_{jk}(t) \qquad \forall l \in \{1, \ldots, L\}. \tag{6.40}$$

To obtain a signal $z_l(t)$ which contains no linear terms associated with any $s_j(t)$, we select the coefficients c_{li} so that

$$\sum_{i=2}^{P} b_{ij} c_{li} = 1 \qquad \forall\, j \in \{1, \ldots, N\}. \tag{6.41}$$

For a given index l, this yields a set of N equations, where the unknowns are the $P-1$ values of c_{li}, while the (estimated) coefficients b_{ij} are available from Sect. 6.4.2.1. If $P - 1 = N$, this set of linear equations has a single solution, i.e. we can only create one such signal $z_l(t)$. More generally speaking, whatever $M \ge 0$, if $P - 1 = N + M$, we can create $M+1$ linearly independent signals $z_l(t)$. Besides, (6.40) then reduces to

$$z_l(t) = \sum_{1 \le j < k \le N} r_{ljk} \tilde{s}_{jk}(t) \qquad \forall\, l \in \{1, \ldots, L\} \tag{6.42}$$

i.e. these signals $z_l(t)$ are then only mixtures of the quadratic signals $\tilde{s}_{jk}(t)$. Moreover, there exist $N(N-1)/2$ signals[19] $\tilde{s}_{jk}(t)$ in the observations (6.33). We want the set of mixtures $z_l(t)$ of the signals $\tilde{s}_{jk}(t)$ to be invertible. We therefore set the numbers L and P of recombined signals $z_l(t)$ and observations $x_i(t)$ to $L = M + 1 = N(N-1)/2$ and therefore $P = N + M + 1 = N(N+1)/2$.

So, we thus obtained the following result: by solving Eq. (6.41) and deriving the resulting signals according to (6.39), we obtain the set of linear instantaneous mixtures $z_l(t)$ of the signals $\tilde{s}_{jk}(t)$ defined by (6.42), which is invertible when $[r_{ljk}]$ is assumed to be invertible. These mixed signals may then be used in various ways, as will now be shown.

6.4.2.3 Remaining BMI and BSS Tasks

One may then proceed in different ways, depending on which parts of the BMI and BSS tasks should be performed in the considered application and which constraints on the sources are acceptable. We now explore these alternatives.

A Method Based on Non-stationarity Conditions

We first again focus on methods for signals which are time-domain sparse, and therefore nonstationary. One may then process the linear instantaneous mixtures $z_l(t)$ of the signals $\tilde{s}_{jk}(t)$, defined in (6.42), by adapting the approach of Sect. 6.4.2.1 to this new context. This achieves both BMI for the mixing matrix in (6.42) (but not yet for the original matrix Q in (6.34)) and BSS for the signals $\tilde{s}_{jk}(t)$ (but not yet for the signals $s_j(t)$). This adaptation of the approach of Sect. 6.4.2.1 requires one to extend the assumptions accordingly. Especially, we then need times when a single signal $\tilde{s}_{jk}(t)$ is active, i.e., essentially times when only the two corresponding sources $s_j(t)$ and $s_k(t)$ are simultaneously active.

It should also be noted that in the basic configuration with $N = 2$ sources, only a single signal $\tilde{s}_{jk}(t)$ exists, namely $s_1(t)s_2(t)$. This signal is then directly provided by the method described in Sect. 6.4.2.2, so that the stage described in the current section then disappears.

A Method also Using Other Correlation Parameters

The method defined in Section "A Method Based on Non-stationarity Conditions" yields scaled permuted versions of the signals $\tilde{s}_{jk}(t)$, i.e., it provides a set of signals

$$y_l(t) = \lambda_{jk}\tilde{s}_{jk}(t) \qquad \forall\, l \in \{1, \ldots, L\}. \tag{6.43}$$

[19] Or less if all coefficients for at least one signal $\tilde{s}_{jk}(t)$ are zero.

We now propose a simple method which may then be applied to these signals when one also wants to identify the matrix Q and/or to separate the signals $s_j(t)$. Considering the signals which are contained by $s'(t)$ and $\tilde{s}'(t)$ at times when they are active, we request them to be uncorrelated, unlike in the previous stages of our approach. Denoting $y'_l(t)$ the centered version of $y_l(t)$, we then have if $\tilde{s}_{jk}(t)$ is active

$$\delta_{il} = \frac{E\{y'_l(t)x'_i(t)\}}{E\{[y'_l(t)]^2\}} = \frac{q_{ijk}}{\lambda_{jk}} \qquad \forall\, i \in \{1, \ldots, P\}, \qquad \forall\, l \in \{1, \ldots, L\}. \tag{6.44}$$

This may be interpreted as in Sect. 6.4.2.1, i.e., one may build the matrix $[\delta_{il}]$, where each column l corresponds to one signal $\tilde{s}_{jk}(t)$. Equation (6.44) then shows that this matrix is equal to Q, up to the scale and permutation indeterminacies. This completes all BMI tasks. Moreover, let us consider the signals

$$u_i(t) = x_i(t) - \sum_{l=1}^{L} \delta_{il} y_l(t) \qquad \forall\, i \in \{1, \ldots, P\}. \tag{6.45}$$

Denoting $u(t)$ the column vector of signals $u_i(t)$, Eqs. (6.33), (6.43), (6.44) and (6.45) then yield in matrix form

$$u(t) = As(t). \tag{6.46}$$

BSS is then straightforwardly achieved for the original sources $s_j(t)$ by computing the vector $B^{\dagger}u(t)$, where † denotes the pseudo-inverse.

A Method Only Using Variance Parameters

Eventually, if one is mainly interested in the BSS of the sources $s_j(t)$, the method of Section "A Method Based on Non-stationarity Conditions" and its constraints may be avoided, again at the expense of requesting the uncorrelation of the signals which are contained by $s'(t)$ and $\tilde{s}'(t)$ (considered at times when they are active). To this end, we introduce the signals

$$v_i(t) = x_i(t) - \sum_{l=1}^{L} d_{il} z_l(t) \qquad \forall\, i \in \{1, \ldots, P\}. \tag{6.47}$$

It may be shown that, by adapting all coefficients d_{il} so as to minimize the variances of all signals $v_i(t)$, the vector $v(t)$ consisting of these signals becomes equal to $As(t)$. BSS is then achieved for the original sources $s_j(t)$ by computing the vector $B^{\dagger}v(t)$.

Experimental results obtained with our overall LQI-TempCorr method are reported in [17] and skipped here, due to space limitations.

6.4.3 Related Works and Extensions

As stated above, we here considered linear-quadratic instantaneous mixtures as a relatively simple application (yet significantly more complex than linear instantaneous mixtures) of the proposed class of SCA methods to nonlinear mixtures. One may then guess how to further extend this approach, e.g., to more general polynomial instantaneous mixtures. To this end, one may also take advantage of the work that we reported in [39] for extending another type of BSS methods (Non-Negative Matrix Factorization) from second-order polynomial mixtures, as above, to third-order ones. The practical applicability of the SCA methods thus derived here for higher-order mixtures may however be limited by the extended sparsity properties that they require.

This type of SCA methods was also independently extended to post-nonlinear mixtures [46] and to a wider class of nonlinear mixtures [47] by a former member of our group.

6.5 Linear Instantaneous Mixtures of 2D Sources

Finally, we briefly discuss the situation when, in their original representation, the considered source "signals" are two-dimensional, i.e., they are functions of two scalar variables, denoted as p_H and p_V. When focusing on the case of image sources as a typical and major example, these variables are, respectively, the Horizontal and Vertical coordinates of the considered pixel. Besides, we here restrict ourselves to the case when these sources are mixed according to the linear instantaneous model. Most of the concepts developed in Sect. 6.2 for the same class of mixtures but for another type of sources can straightforwardly be reused here. This results from the fact that, although the signals considered in Sect. 6.2 were originally 1D, we then intentionally moved to an arbitrary representation domain, which also includes the domains considered here for signals which are originally 2D. We can thus first use the framework of Sect. 6.2 by replacing the temporal variable t by the overall 2D spatial variable (p_H, p_V) which, again, is seen as the argument v of the source and observed signals denoted as $S_j(v)$ and $X_i(v)$. Besides, our framework may be used when a linear sparsifying transform is applied to 2D sources.

The general class of SCA methods that we developed in Sect. 6.2.2 for possibly transformed signals may therefore straightforwardly be extended to image processing. Thus investigating Linear Instantaneous mixtures of originally 2D sources considered in the Spatial domain, and applying the Correlation-based principles of Section "Detection Based on Correlation Coefficients" and item 1 of Sect. 6.2.2.3, e.g., yields the SCA method here called LI-2D-SpaceCorr, which is detailed in [38]. Another version of these methods is obtained by first applying a sparsification stage, which consists in transforming the observed images into the 2D wavelet domain. The resulting method, called LI-2D-WaveCorr, is also detailed in [38]. Other methods, based on ratios of mixtures, may also be developed by using the proposed

framework and applying the principles of Section "Detection Based on Ratios of Mixtures" and item 3 of Sect. 6.2.2.3. This yields two methods, which may be called LI-2D-SpaceROM and LI-2D-WaveROM, since the first of these methods for linear instantaneous mixtures of originally 2D sources uses the original spatial representation of the signals, and the second method first transforms the observations into the 2D wavelet domain.

Moreover, in various image processing applications, the source values and/or mixing coefficients meet specific constraints. In particular, for standard multispectral or hyperspectral reflectance images available in the field of remote sensing (Earth observation),[20] all source values and mixing coefficients are real-valued and non-negative and, when considering spatial unmixing methods, the sum of all source values in each pixel is equal to one. Modified forms of the above-defined SCA methods may be developed for this specific case. Such a modified version of LI-2D-SpaceCorr is described in [25]. We also developed a significantly different method for this case. This method is here called LI-2D-SpaceVM, since it is based on the original Spatial representation of the signals and its detection stage uses Variances of Mixtures. Its stages performing BMI were introduced in [26] and its extension to source estimation will be described in detail in [27]. In all these methods, the non-negativity of the source signals may be used to finally estimate them by means of non-negative least square (NNLS) [30] or Non-negative Matrix Factorization (NMF) [9, 20, 31, 32] algorithms. A comparison of the NNLS-based and NMF-based versions of LI-2D-SpaceCorr and LI-2D-SpaceVM is available in [28].

6.6 Conclusion

Sparse Component Analysis (SCA) is one of the main approaches to Blind Source Separation (BSS) and Blind Mixture Identification (BMI). Using this sparsity concept, we proposed a general framework for developing BSS/BMI methods applicable to different types of sources (one-dimensional signals, images, ...), considered in various domains (original temporal or spatial domain, transformed representation in time-frequency or time-scale/wavelet domain, ...), mixed according to various models (linear instantaneous, anechoic, full convolutive, nonlinear and especially linear-quadratic) and possibly with non-negativity or sum-to-one constraints. These methods essentially require a few tiny single-source zones. They therefore set very limited constraints on source sparsity and could thus be considered as "quasi-non-sparse component analysis" methods. Besides, unlike Independent Component Analysis methods, they are applicable to correlated sources. In this chapter, we provided a unified view and described the latest extensions of our general framework, and we showed that the proposed methods yield attractive experimental performance. Some software and data corresponding to these investigations are currently available

[20] For such data, standard configurations lead to linear instantaneous mixtures [29].

at [57] and will soon be moved to [58]. Further extensions of this general framework for BSS/BMI will be reported in future papers.

Acknowledgments The author would like to thank his colleagues for their participation at some stage in the investigations performed during the last decade concerning the topic presented in this chapter, and/or for helpful discussions, in particular F. Abrard, B. Albouy, D. Benachir, O. Berné, D. Bissessur, A. BOULAIS, H. Carfantan, S. Hosseini, M. S. Karoui, I. Meganem, M. Puigt.

Appendix 1

We here prove the validity of Property 6.1 (see p. 14), used in our correlation-based SCA method for linear instantaneous mixtures. For the sake of simplicity, we express it for a given type of signals, i.e., deterministic signals. The corresponding proof for stochastic signals is similar and is provided for the more general case of linear-quadratic instantaneous mixtures in [17].

For a given analysis zone Z, all values of any given observation $X_i(v)$ are first gathered in a vector $V_{X_i}(Z)$, and all values of any source $S_j(v)$ similarly form a vector $V_{S_j}(Z)$. For determined mixtures, the scalar mixing equations in (6.5) then yield in vector form

$$V_{X_i}(Z) = \sum_{j=1}^{N} a_{ij} V_{S_j}(Z) \qquad \forall\, i \in \{1, \ldots, N\}. \tag{6.48}$$

Besides, the correlation coefficients $\rho\{X_1, X_i\}(Z)$ of observed signals are defined according to (6.10) and (6.9) or its variants for other signal transforms or types of analysis zones. They may therefore be here expressed as

$$\rho\{X_1, X_i\}(Z) = \frac{< V_{X_1}(Z), V_{X_i}(Z) >}{||V_{X_1}(Z)|| \times ||V_{X_i}(Z)||} \tag{6.49}$$

where the notations $< ., . >$ and $||.||$ respectively stand for inner product and vector norm. Applying the Cauchy-Schwarz inequality to (6.49) then shows that

$$|\rho\{X_1, X_i\}(Z)| \leq 1 \qquad \forall\, i \in \{1, \ldots, N\} \tag{6.50}$$

with equality if and only if $V_{X_1}(Z)$ and $V_{X_i}(Z)$ are linearly dependent.

Let us now analyze this condition in a given analysis zone Z, depending on the number of nonzero vectors $V_{S_j}(Z)$, i.e., on the number of sources which are active in this zone. Due to Assumption 6.3 (p. 13), at least one of these vectors $V_{S_j}(Z)$ is not equal to zero. If only one of them is not equal to zero, due to Assumption 6.4 (p. 13), Eq. (6.48) shows that all vectors $V_{X_i}(Z)$, with $1 \leq i \leq N$, are nonzero and collinear. Therefore, equality holds whatever i in (6.50) and the detection condition (6.11) is fulfilled.

The only case that remains to be considered is then the situation when at least two vectors $V_{S_j}(Z)$ are nonzero. It may then easily be shown that if $V_{X_1}(Z)$ and $V_{X_i}(Z)$ were linearly dependent for all i, with $2 \leq i \leq N$, then, due to Assumption 6.5 (p. 13), all the columns of the mixing matrix A with indices equal to the indices j of the nonzero vectors $V_{S_j}(Z)$ would be collinear. This is not true, thanks to Assumption 6.1(p. 4). Therefore, in the considered case, at least one pair of vectors $(V_{X_1}(Z), V_{X_i}(Z))$ does not consist of linearly dependent vectors, so that $|\rho\{X_1, X_i\}(Z)| < 1$ and the detection condition (6.11) is not fulfilled.

As an overall result, condition (6.11) is fulfilled if and only if exactly one of the vectors $V_{S_j}(Z)$ is not equal to zero in the considered analysis zone, i.e., if this is a single-source zone, which completes the proof of Property 6.1.

References

1. Abrard, F., Deville, Y., White, P.: A new source separation approach based on time-frequency analysis for instantaneous mixtures. In: Proceedings of ECM2S'2001, pp. 259–267, Toulouse, France, 30 May–1 June 2001
2. Abrard, F., Deville, Y.: Blind separation of dependent sources using the "TIme-Frequency Ratio Of Mixtures" approach". Proceedings of ISSPA 2003, Paris, France, 1–4 July 2003
3. Abrard, F., Deville, Y.: A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. Signal Process. **85**(7), 1389–1403 (2005)
4. Albouy, B., Deville, Y.: Alternative structures and power spectrum criteria for blind segmentation and separation of convolutive speech mixtures. In: Proceedings of ICA2003, pp. 361–366, Nara, Japan, 1–4 April 2003
5. Albouy, B., Deville, Y.: A time-frequency blind source separation method based on segmented coherence function. In: Proceedings of IWANN 2003, vol. 2, pp. 289–296, Mao, Menorca, Spain, 3–6 June 2003
6. Arberet, S., Gribonval, R., Bimbot, F.: A robust method to count and locate audio sources in a multichannel underdetermined mixture. IEEE Trans. Signal Process. **58**(1), 121–133 (2010)
7. Aziz Sbaï, S.M., Aïssa-El-Bey, A.: Pastor, D.: Contribution of statistical tests to sparseness-based blind source separation. EURASIP J. Adv. Signal Process. **169**, 2012.
8. Cichocki, A., Amari, S., Siwek, K., Tanaka, T. et al.: ICALAB Toolboxes. http://www.bsp.brain.riken.jp/ICALAB
9. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.-I.: Nonnegative matrix and tensor factorizations. Applications to exploratory multi-way data analysis and blind source separation. Wiley, Chichester, UK (2009)
10. Cohen, L.: Time-frequency distributions—a review. Proc. IEEE **77**(7), 941–981 (1989)
11. Comon, P.: Independent component analysis, a new concept? Signal Process. **36**(3), 287–314 (1994)
12. Comon, P., Jutten, C. (eds.) Handbook of Blind Source Separation. Independent Component Analysis and Applications. Academic Press, Oxford, UK (2010)
13. Delfosse, N., Loubaton, P.: Adaptive blind separation of independent sources: a deflation approach. Signal Process. **45**(1), 59–84 (1995)
14. Deville, Y.: Temporal and time-frequency correlation-based blind source separation methods. In: Proceedings of the ICA2003, pp. 1059–1064, Nara, Japan, April 1–4 2003
15. Deville, Y., Puigt, M., Albouy, B.: Time-frequency blind signal separation: extended methods, performance evaluation for speech sources. In: Proceedings of IJCNN 2004, pp. 255–260, Budapest, Hungary, 25–29 July 2004

16. Deville, Y., Bissessur, D., Puigt, M., Hosseini, S., Carfantan, H.: A time-scale correlation-based blind separation method applicable to correlated sources. In: Proceedings of ESANN'2006, Bruges, Belgium, 26–28 April 2006
17. Deville, Y., Hosseini, S.: Blind identification and separation methods for linear-quadratic mixtures and/or linearly independent non-stationary signals. In: Proceedings of ISSPA 2007, Sharjah, United Arab Emirates, 12–15 Feb 2007
18. Deville, Y., Puigt, M.: Temporal and time-frequency correlation-based blind source separation methods. part i: determined and underdetermined linear instantaneous mixtures. Signal Process. **87**(3), 374–407 (2007)
19. Deville, Y.: Traitement du signal : signaux temporels et spatiotemporels - Analyse des signaux, théorie de l'information, traitement d'antenne, séparation aveugle de sources", Ellipses Editions Marketing, Paris, 2011, 312 p. ISBN 978-2-7298-7079-9
20. Donoho, D., Stodden, V.: When does non-negative matrix factorization give a correct decomposition into parts?. In: Proceedings of NIPS 2003, Vancouver and Whistler, Canada, 8–13 Dec 2003
21. Hlawatsch, F., Boudreaux-Bartels, G.F.: Linear and quadratic time-frequency signal representations. IEEE Signal Process. Mag. **9**, 21–67 (1992)
22. Hyvarinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Comput. **9**, 1483–1492 (1997)
23. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
24. Jourjine, A., Rickard, S., Yilmaz, O.: Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures. In: Proceedings of ICASSP 2000, vol. 5, pp. 2985–2988, Istanbul, Turkey, 5–9 June 2000
25. Karoui, M.S., Deville, Y., Hosseini, S., Ouamri, A.: Blind spatial unmixing of multispectral images: new methods combining sparse component analysis, clustering and non-negativity constraints. Pattern Recognit. **45**, 4263–4278 (2012)
26. Karoui, M.S., Deville, Y., Hosseini, S., Ouamri, A.: A new spatial sparsity-based method for extracting endmember spectra from hyperspectral data with some pure pixels. In: Proceedings of IGARSS 2012, pp. 3074–3077, Munich, Germany, 22–27 July 2012
27. Karoui, M.S., Deville, Y., Hosseini, S., Ouamri, A.: Blind unmixing of hyperspectral data with some pure pixels: spatial variance-based methods exploiting sparsity and non-negativity properties. In: Naik, G. (ed.) Signal Processing: New Research. Nova Science Publishers, Hauppauge, NY, USA (2013)
28. Karoui, M.S., Deville, Y., Hosseini, S., Ouamri, A.: Blind unmixing of remote sensing data with some pure pixels: extension and comparison of spatial methods exploiting sparsity and nonnegativity properties. In: Proceedings of WOSSPA 2013, Mazafran, Algiers, Algeria, 12–15 May 2013
29. Keshava, N., Mustard, J.F.: Spectral unmixing. IEEE Signal Process. Mag. **19**, 44–57 (2002)
30. Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems, p. 1995. Prentice-Hall, Englewoods Cliffs, New Jersey, SIAM's Classics in Applied Mathematics (1974)
31. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**, 788–791 (1999)
32. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. Adv. Neural Info. Proc. Syst. **13**, 556–562 (2001)
33. Lee, T.-W., Lewicki, M.S., Girolami, M., Sejnowski, T.J.: Blind source separation of more sources than mixtures using overcomplete representations. IEEE Signal Process. Lett. **6**(4), 87–90 (1999)
34. Li, R., Wang, F.S.: Efficient wavelet based blind source separation algorithm for dependent sources. ICFIE, ASC **40**, 431–441 (2007)
35. Li, Y., Amari, S.-I., Cichocki, A., Ho, D.W.C., Xie, S.: Underdetermined blind source separation based on sparse representation. IEEE Trans. Signal Process. **54**(2), 423–437 (2006)
36. Luo, J., Zhang, Z.: Using eigenvalue grads method to estimate the number of signal source. In: Proceedings of ICSP2000, pp. 223–225, Beijing, China, 21–25 Aug 2000

37. Mallat, S.: A wavelet tour of signal processing. Academic Press, San Diego (1999)
38. Meganem, I., Deville, Y., Puigt, M.: Blind separation methods based on correlation for sparse possibly-correlated images. In: Proceedings of ICASSP 2010, pp. 1334–1337, Dallas, Texas, USA, 14–19 March 2010
39. Meganem, I., Deville, Y., Hosseini, S., Déliot, P., Briottet, X., Duarte, L.T.: Linear-quadratic and polynomial non-negative Matrix Factorization; application to spectral unmixing. In: Proceedings of EUSIPCO 2011, Barcelona, Spain, 29-Aug–2 Sept 2011
40. Meganem, I., Deliot, P., Briottet, X., Deville, Y., Hosseini, S.: Linear-quadratic mixing model for reflectances in urban environments. IEEE Trans. Geosci. Remote Sens. **52**(1), 544–558 (2014)
41. Papoulis, A., Pillai, S.U.: Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York (1965/2002)
42. Puigt, M., Deville, Y.: Time-frequency ratio-based blind separation methods for attenuated and time-delayed sources. Mech. Syst. Signal Process. **19**, 1348–1379 (2005)
43. Puigt, M., Deville, Y.: A time-frequency correlation-based blind source separation method for time-delayed mixtures. In: Proceedings of ICASSP 2006, pp. V-853–V-856, Toulouse, France, 14–19 May 2006
44. M. Puigt, Deville, Y.: A new time-frequency correlation-based source separation method for attenuated and time-shifted mixtures. In: Proceedings of ECMS 2007, pp. 34–39, Liberec, Czech Republic, 21–23 May 2007
45. Puigt, M., Deville, Y.: Iterative-shift cluster-based time-frequency BSS for fractional-time-delay mixtures. In: Proceedings of ICA 2009, pp. 306–313. LNCS, vol. 5441. Springer, Berlin, Paraty, Brazil, 15–18 March 2009
46. Puigt, M., Griffin, A., Mouchtaris, A.: Post-nonlinear speech mixture identification using single-source temporal zones and curve clustering. In: Proceedings of EUSIPCO 2011, pp. 1844–1848, Barcelona, Spain, 29 Aug–2 Sept 2011
47. Puigt, M., Griffin, A., Mouchtaris, A.: Nonlinear blind mixture identification using local source sparsity and functional data clustering. In: Proceedings of SAM 2012, pp. 481–484, Hoboken, NJ, 17–20 June 2012
48. Reju, V.G., Koh, S.N., Soon, I.Y.: An algorithm for mixing matrix estimation in instantaneous blind source separation. Signal Process. **89**, 1762–1773 (2009)
49. Rioul, O., Vetterli, M.: Wavelets and signal processing. IEEE Signal Process. Mag. **8**, 14–38 (1991)
50. Smith, D., Lukasiak, J., Burnett, I.S.: A two channel, block-adaptive audio separation technique based upon time-frequency information. In: Proceedings of EUSIPCO 2004, pp. 393–396, Vienna, Austria, 6–10 Sept 2004
51. Smith, D., Lukasiak, J., Burnett, I.: Two channel, block-adaptive audio separation using the cross correlation of time frequency information. In: Proceedings of ICA 2004, pp. 279–286. LNCS, vol. 3195. Springer, Granada, Spain, 22–24 Sept 2004
52. Smith, D., Lukasiak, J., Burnett, I.: A sequential approach to sparse component analysis. In: Proceedings of the IEEE 7th Workshop on Multimedia Signal Processing, pp. 129–132, Shanghai, 30 Oct–2 Nov 2005
53. Smith, D., Burnett, I.: Blind separation of speech with a switched sparsity and temporal criteria. In: Proceedings of the IEEE 8th Workshop on Multimedia Signal Processing, pp. 136–140, Victoria, BC, 3–6 Oct 2006
54. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Academic Press, San Diego, California, USA (2009)
55. Yilmaz, O., Rickard, S.: Blind separation of speech mixtures via time-frequency masking. IEEE Trans. Signal Process. **52**(7), 1830–1847 (2004)
56. http://www-stat.stanford.edu/~wavelab/
57. http://www.ast.obs-mip.fr/deville
58. http://userpages.irap.omp.eu/~ydeville/

Chapter 7
Underdetermined Audio Source Separation Using Laplacian Mixture Modelling

Nikolaos Mitianoudis

Abstract The problem of underdetermined audio source separation has been explored in the literature for many years. The instantaneous K-sensors, L-sources mixing scenario (where $K < L$) has been tackled by many different approaches, provided the sources remain quite distinct in the virtual positioning space spanned by the sensors. In this case, the source separation problem can be solved as a directional clustering problem along the source position angles in the mixture. The use of Laplacian Mixture Models in order to cluster and thus separate sparse sources in underdetermined mixtures will be explained in detail in this chapter. The novel Generalised Directional Laplacian Density will be derived in order to address the problem of modelling multidimensional angular data. The developed scheme demonstrates robust separation performance along with low processing time.

7.1 Introduction

Let a set of K sensors $\mathbf{x}(n) = [x_1(n), \ldots, x_K(n)]^T$ observe a set of L sound sources $\mathbf{s}(n) = [s_1(n), \ldots, s_L(n)]^T$. We will consider the case of instantaneous mixing, i.e. each sensor captures a scaled version of each signal with no delay in transmission. Moreover, the possible additive noise will be considered negligible. The above instantaneous mixing model can be expressed in mathematical terms, as follows:

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) \tag{7.1}$$

where $\mathbf{A}$ represents the $K \times L$ *mixing matrix* and n the sample index. The blind source separation problem provides an estimate of the source signals $\mathbf{s}(n)$ given

N. Mitianoudis (✉)
Image Processing and Multimedia Lab, Electrical and Computer Engineering Department, Democritus University of Thrace, 67100 Xanthi, Greece
e-mail: nmitiano@ee.duth.gr

G. R. Naik and W. Wang (eds.), *Blind Source Separation*, Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_7,
© Springer-Verlag Berlin Heidelberg 2014

the sensor signals $\mathbf{x}(n)$. Usually, most separation approaches are *semi-blind*, which implies some knowledge of the source signal's general statistical structure. A number of algorithms have been proposed to solve the overdetermined and complete source separation problem ($K \geq L$) with great success. The additional assumption of statistical independence between the sources led to a group of source separation algorithms, summarised under the general term *Independent Component Analysis* (ICA). Starting from different interpretations of statistical independence, most algorithms perform source separation with great accuracy. An overview of current ICA and general blind source separation algorithms can be found in tutorial books on ICA by Hyvärinen et al. [27], Cichocki-Amari [11] and Common et al. [12].

The underdetermined source separation problem ($K \leq L$) is more challenging, since in this case, the estimation of the mixing matrix $\mathbf{A}$ is not sufficient for the estimation of the source signals $\mathbf{s}(n)$. This type of mixtures can be encountered in musical audio mixes. A number of solo instrument recordings are combined linearly in a stereo ($K = 2$) or a multichannel ($K = 5$ or $K = 7$) mixture , in order to form a musical score recording. Assuming Gaussian distributions for the sources and a known mixing matrix $\mathbf{A}$, one could estimate the sources using the pseudo-inverse of matrix $\mathbf{A}$ in a Maximum Likelihood sense [34]. As most speech and audio signals tend to follow heavy-tailed "nonGaussian" distributions, the above linear operation is not sufficient to estimate the sources. Therefore, the underdetermined source separation problem can be divided into two subproblems: (i) estimating the mixing matrix $\mathbf{A}$ and (ii) estimating the source signals $\mathbf{s}(n)$.

The existence of a unique source estimate for the underdetermined source separation problem, even in the case that A is known, is always under question, since it is an ill-conditioned problem that has an infinite number of solutions. Any linear system with less equations than unknown variables has an infinite number of solutions (source estimates) [31]. However, according to Eriksson and Koivunen [21], the linear generative model of (7.1) can have a *unique* and *identifiable* solution for the underdetermined case, provided (i) there are no Gaussian sources present in the mixture, (ii) the mixing matrix $\mathbf{A}$ is of full row rank, i.e. $\text{rank}(\mathbf{A}) = K$ and (iii) none of the source variables has a characteristic function featuring a component in the form $\exp(Q(u))$, where $Q(u)$ is a second-order polynomial or higher. This implies that this intractable problem may have a non-infinite number of solutions, under several constraints and probabilistic criteria for the sources.

One probabilistic profile that satisfies the assumptions set above are *sparse* distributions. *Sparsity* is mainly used to describe signals that are mostly close to a mean value with the exception of several large values. Common models that can be used for approximating sparsity are minimum $\mathscr{L}_0$ or $\mathscr{L}_1$ norms [34], *Mixture of Gaussians* (MoG) [3, 14, 43] or *factorable Laplacian distributions* [25]. The separation quality for the underdetermined case seems to improve with sparsity, as usually the performance of source separation algorithms is closely connected with the "non-Gaussianity" of the source signals [9]. However, in many practical applications, the source data are not sparse. For example, some musical instrument signals tend to be less sparse than speech signals in the time-domain. Speech contains a lot of silent segments that guarantee sparsity (many zero samples), however, this might not be

the case with many instrument signals. Therefore, the assumed sparse models are not accurate enough to describe the statistical properties of the signals in the time-domain. Many natural signals can have sparser representations in other transform domains, including the Fourier transform, the Wavelet transform and the Modified Discrete Cosine Transform (MDCT). Since these transformations are linear, it is equivalent to estimate the generative model and the sources in the transform domain. There are also alternative methods, where one can generate sparse representations for a specific dataset [15]. In the following analysis, the MDCT is employed to provide a sparser representation of audio signals.

The underdetermined source separation problem has been covered extensively in the literature. Lewicki [34] provided a complete Bayesian approach, assuming Laplacian source priors to estimate both the mixing matrix and the sources in the time domain. In [33], Lee et al. applied the previous algorithm to the source separation problem. Girolami [25] employed the factorable Laplacian distribution and variational EM to estimate the mixing matrix and the sources. More complete sparse source models, such as the *Student-t* distribution, were employed by Févotte et al. [22]. The parameters of the model, the mixing matrix and the source signals were estimated using either *Markov Chains Monte Carlo* (MCMC) simulations [22] or a *Variational Expectation Maximisation* (EM) algorithm [10], featuring robust performance, however, being computationally expensive. Clustering solutions were introduced by Hyvärinen [28] and Zibulevsky et al. [62], also featuring good results and lower computational complexity. In this case, the mixing matrix and the source signals are estimated by performing clustering in a sparser representation of the signals in the transform domain. Bofill-Zibulevsky [8] presented a shortest path algorithm based on $\mathscr{L}_1$ minimisation that could estimate the mixing matrix and the sources. O'Grady and Pearlmutter [45] proposed an algorithm to perform separation via Oriented Lines Separation (LOST) using clustering along lines (Hard-Lost) in a similar manner to Hyvärinen [28]. In addition, they proposed a soft-thresholding technique using an EM on a mixture of oriented lines to assign points to more than one source [44]. Davies and Mitianoudis [14] employed two-state Gaussian Mixture Models (GMM) to model the source densities in a sparse representation and also the additive noise. An EM-type algorithm was used to estimate the parameters of the two-state models and perform source coefficients clustering. The latter approach can be considered a joint Bayesian and clustering approach. A two-sensor more-sources setup, modelling also some delays between the sensors, was addressed using the DUET algorithm [61] that can separate the sources, by calculating amplitude differences (AD) and phase differences (PD) between the sensors. An online version of the algorithm was also proposed [48]. Recently, Arberet et al. [2] proposed a method to count and locate sources in underdetermined mixtures. Their approach is based on the hypothesis that in localised neighbourhoods around some time–frequency points (t, f) (in the Short-Time Fourier Transform (STFT) representation) only one source essentially contributes to the mixture. Thus, they estimate the most dominant source (the Estimated Steering Vector) and a local confidence Measure which increases where a single component is only present. A clustering approach merges the above information and estimates the mixing matrix $\mathbf{A}$. In [58],

Vincent et al. used local Gaussian Modelling of minimal constrained variance of the local time–frequency neighbours assuming knowledge of the mixing matrix **A**. The candidate sources' variances are estimated after minimising the Kullback-Leibler (KL) divergence between the empirical and expected mixture covariances, assuming that at most 3 sources contribute to each time–frequency neighbourhood and the sources are derived using Wiener filtering.

The instantaneous mixtures model is rather incomplete in the case of sources recorded in a real acoustic room environment. Assume the case of a sound source and a microphone in a room. Previous research has shown that the signal captured by the microphone can be well represented by a *convolution* of the source signal with a high-order FIR filter, modelling the room acoustics between the source and the sensor [41]. In the case of many sources and sensors, the signal at each sensor can be modelled by the following equation:

$$\mathbf{x}(n) = \begin{bmatrix} \mathbf{h}_{11} & \dots & \mathbf{h}_{1L} \\ \dots & \dots & \dots \\ \mathbf{h}_{K1} & \dots & \mathbf{h}_{KL} \end{bmatrix} * \mathbf{s}(n) \tag{7.2}$$

where $*$ denotes the linear convolution operator and $\mathbf{h}_{ij}$ denotes an FIR filter modelling the room impulse response between the i-th microphone and the j-th source.

Many methods have been proposed to solve the square ($K = L$) convolutional ICA problem. Some of them suggested working directly in the time-domain [32, 57]. Working in the time domain has the disadvantage of being rather computationally expensive, due to calculating many convolutions and the size of the unmixing filters. Other approaches suggested moving to the STFT domain in order to transform the convolution into multiplication and apply ICA methods for instantaneous mixtures (i.e. the natural gradient) for each frequency bin [56]. However, there is an inherent *permutation problem* in all FD-ICA methods, which does not exist in time-domain methods. Mitianoudis and Davies [41] proposed a *time–frequency source model* for a ML-ICA approach, incorporating a *time-varying* parameter, aiming to impose *frequency coupling* between neighbouring frequency bins. In addition, a *likelihood ratio* test was proposed to address the permutation problem. In [38], Mitianoudis and Davies described a mechanism to align permutations using subspace methods at each frequency bin. This idea was refined and was extended for underdetermined convolutive mixtures by Sawada et al. [1, 49, 50]. Winter et al. [60] estimate the mixing matrix based on hierarchical clustering, assuming sparsity of the source signals. Sources are then estimated using L1-norm minimisation of complex numbers, using Laplacian source priors. Duong et al. [20] model the contribution of each source to all mixture channels in the time-frequency domain as a zero-mean Gaussian random variable (r.v.) whose covariance encodes the spatial characteristics of the source. They derive a family of iterative EM algorithms to estimate the parameters of each model and propose suitable procedures adapted from previous convolutive approaches to align the order of the estimated sources across all frequency bins.

A more general source separation case can be introduced by using the following nonlinear mixing setup:

$$\mathbf{x} = f(\mathbf{s}) \tag{7.3}$$

where $f(\cdot)$ is a general nonlinear function, which provides a mapping $f : \mathscr{R}^L \rightarrow \mathscr{R}^L$. The solution for this problem forms a new class of source separation algorithms, termed *nonlinear BSS*. The nonlinear problem has a fundamental characteristic that solutions always exist; however, they are highly non-unique [27]. The general nonlinear BSS problem can be addressed using Kohonen Self-Organising maps [46]. In [30], Jutten and Karhunen state that one can reduce these great indeterminacies by constraining the mapping $f(\cdot)$ to a certain set of transformations. *Smooth nonlinear mappings*, i.e. mappings that preserve independence of the components, such as a rotation matrix, can be unmixed using multilayer perceptron (MLP) networks. In the *Post nonlinear* (PNL) model, the nonlinear mapping has the following structure:

$$x_i(n) = f_i\left(\sum_{j=1}^{L} a_{ij} s_j(n)\right), \; i = 1, \ldots, K \tag{7.4}$$

where the nonlinear functions $f_i(\cdot)$ are assumed to be invertible. Such models may appear in array processing, satellite and microwave communications. A separation method for PNL models generally consists of two stages [27]: (a) a nonlinear stage, where the functions f_i are inverted, (b) a linear stage, where the linearised mixture is separated using an ordinary linear instantaneous ICA algorithm. In addition, there are other special nonlinear mixing cases, that can be linearised using another nonlinear mapping function $g(\cdot)$ (see [30]). Recently, Duarte et al. [19] introduced a blind compensation scheme of the nonlinear distortion introduced in PNL mixtures, by using a semi-blind cost function to estimate the parameters of a known inverting function. Nevertheless, further exploration of nonlinear mixtures separation goes beyond the scope of this chapter.

In this chapter, instantaneous underdetermined source separation is examined in the form of a directional clustering problem. Clustering is performed with the application of density mixture models, which are trained on the directional data using the EM algorithm. We examine three cases of candidate densities based on the Laplacian distribution, which is well-suited to model sparse data. The directionality of the source separation data led to the introduction of a wrapped Laplacian density and finally a generalised directional Laplacian density, a closed-form expression that can model multidimensional directional data.

7.2 Underdetermined Source Separation as a Directional Clustering Problem

Let us assume a two-sensor instantaneous mixing approach. In Fig. 7.1a, one can see the scatter plot of the two sensor signals, in the case of two sensors and four sources. The four sources are 7 s of speech, accordion, piano and violin signals. In the time-domain representation, no directions of the input sources are visible in the mixture. Consequently, the separation problem seems very difficult to solve. To get a sparser representation of the data, the MDCT or the *Short-Time Fourier Transform* (STFT) can be applied on the observed signals. The MDCT is a linear, real transform that has excellent sparsifying properties for most audio and speech signals. The harmonic content of most speech and musical instrument signals can be represented by harmonically related sinusoidals with great accuracy (excluding transient and percussive parts in audio and unvoiced segments in speech). Consequently, using a transformation that projects the audio data on sinusoidal bases will most probably result into a more compact and sparse representation of the original data. The MDCT is more preferable than the STFT, since it is real and retains all the required sinusoidal signal structure. The need for sparser representations in underdetermined source separation and audio analysis in general is discussed more rigorously in [7, 15, 26, 35, 47]. When the sources are sparse, smaller coefficients are more probable, whereas all the signal's energy is concentrated in few large values. Therefore, the density of the data in the mixture space shows a tendency to cluster along the directions of the mixing matrix columns [62]. Observing the scatter plot in Fig. 7.1b, it is clear that the angular difference between the two sensors can be used to identify and separate the sources in the mixture. That is to say, the two-dimensional (2D) problem can be transformed to a one-dimensional (1D) problem, as the main important parameter is the angle θ_n of each point.

$$\theta_n = \text{atan}\frac{x_2(n)}{x_1(n)}. \tag{7.5}$$

Using the directional differences between the two sensors is equivalent to mapping all the observed data points on the unit circle. Extending this to a general multichannel scenario, one can map K-dimensional points $\mathbf{x}(n)$ to the K-D unit sphere, by dividing with the vector's norm $||\mathbf{x}||$. In Fig. 7.2a, we plot the histogram of the observed data angle θ_n in the previous example.[1] The strong "superGaussian" characteristics of the individual components in the MDCT domain are preserved in the angle representation θ_n. Then, the vectors $\mathbf{x}_{\text{norm}}(n)$ contain only directional information in a polar reference system.

[1] A π-periodicity is valid for the observed phenomenon, since data in $(\pi/2, 3\pi/2)$ are symmetrical to the ones in $(-\pi/2, \pi/2)$ (See Fig. 7.1b). Hence, the use of the atan function instead of the extended atan2 function is justified. For the rest of the analysis, we will assume that θ_n takes values between $(0, \pi)$ rather than $(-\pi/2, \pi/2)$. This implies that data in the 4th quadrant $(-\pi/2, 0)$ are mapped with odd symmetry to the 2nd quadrant $(\pi/2, \pi)$. This is performed in order to facilate the derivations of the Generalised Directional Laplacian Distribution and does not alter anything in the actual data.

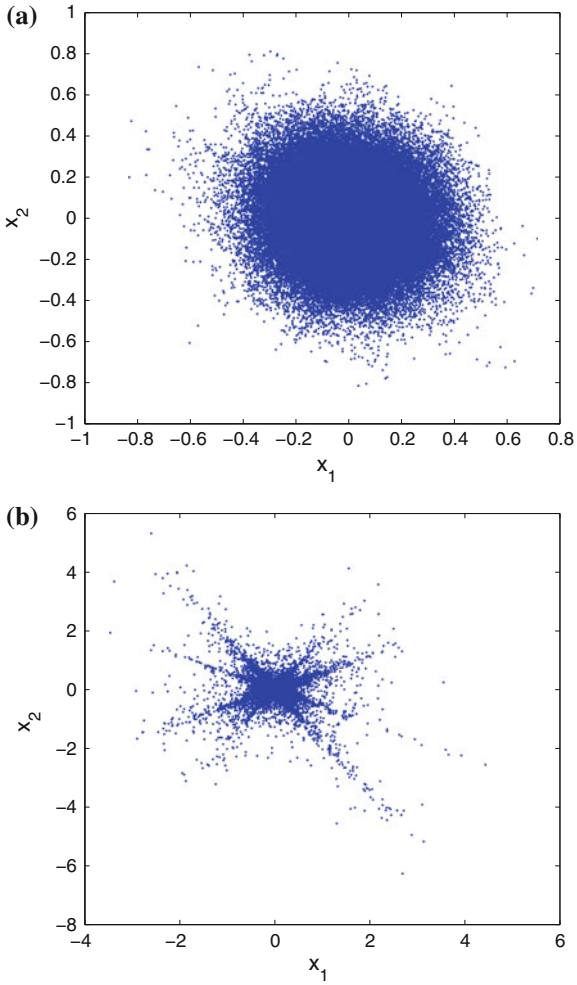


Fig. 7.1 Scatter plot of a two-sensor four-sources mixture in the time domain and in the sparse MDCT domain. The almost Gaussian-like structure of the time-domain representation is enhanced using the MDCT and the four sources can be clearly identified in the mixture. **a** Time domain. **b** MDCT domain

$$\mathbf{x}_{\text{norm}}(n) = \frac{\mathbf{x}(n)}{||\mathbf{x}(n)||}. \tag{7.6}$$

We can also define the magnitude r_n of each point $\mathbf{x}(n)$, as follows:

$$r_n = ||\mathbf{x}(n)|| = \sqrt{x_1(n)^2 + \cdots + x_K(n)^2} \tag{7.7}$$

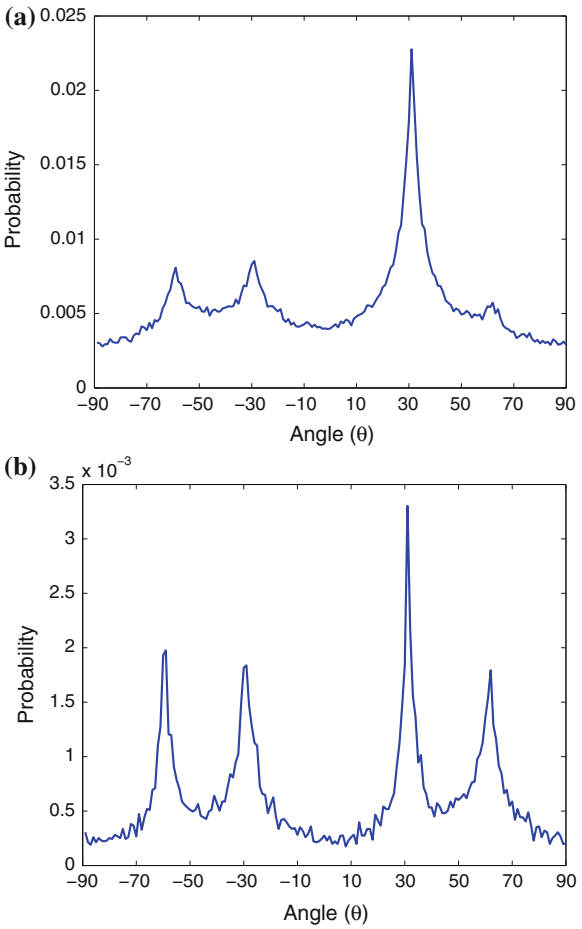


Fig. 7.2 Histograms of angle θ_n in the four sources example of Fig. 7.1. The four sources are identifiable in the original histogram (**a**), however, keeping only the most "superGaussian" components (**b**), we can facilitate the separation process, as the directions of arrival are more clearly identifiable. **a** Original histogram. **b** Modified histogram

We can observe that points that are close to the origin have a more Gaussian structure and thus do not contribute to the desired "superGaussian" profile. Consequently, we can use a "reduced" representation of the original data in order to estimate the columns of the mixing matrix more accurately. In Fig. 7.2b, we can see a histogram of those points n, whose magnitude r_n is above a threshold, e.g. $r_n > 0.1$. Comparing with the original histogram of Fig. 7.2a, the four components are more clearly identifiable in this reduced representation, which can facilitate the estimation of the columns of the mixing matrix, i.e. the directions of arrival for each source. In this representation, we will present three models based on the Laplacian density, that can be applied to cluster and separate the sources.

7.3 Identification Using Laplacian Mixtures Models

A model, that is commonly used in the literature to model sparse data, is the *Laplacian* density function. The definition for the Laplacian probability density function (pdf) is given by the following expression:

$$\mathscr{L}(\theta, k, m) = ke^{-2k|\theta - m|} \tag{7.8}$$

where m defines the mean and $k > 0$ controls the "width" (approximate standard deviation) of the distribution. In a similar fashion to Mixtures of Gaussians (MoG), one can employ *Laplacian Mixture Models* (LMM) in order to model a mixture of "heavy-tailed signals". A *LMM* of K Laplacians can thus be defined, as follows:

$$p(\theta) = \sum_{i=1}^{K} \alpha_i \mathscr{L}(\theta, k_i, m_i) = \sum_{i=1}^{K} \alpha_i k_i e^{-2k_i|\theta - m_i|} \tag{7.9}$$

where α_i, m_i, k_i represent the weight, mean and width of each Laplacian respectively and all weights should sum up to one, i.e. $\sum_{i=1}^{K} \alpha_i = 1$. The EM algorithm is employed to train the parameters of the mixture model. A complete derivation of an EM algorithm was presented by Dempster et al. [16] and has been employed to fit a MoG on a training data set [6]. Assuming N training samples for an 1D r.v. θ_n and Laplacian Mixture densities (7.9), the log-likelihood of these training samples takes the following form:

$$I(\alpha_i, k_i, m_i) = \sum_{n=1}^{N} \log \sum_{i=1}^{K} \alpha_i \mathscr{L}(\theta_n, k_i, m_i). \tag{7.10}$$

Introducing unobserved data items that can identify the components that "generated" each data item, we can simplify the log-likelihood of (7.10) for Laplacian Mixtures, as follows:

$$J(\alpha_i, k_i, m_i) = \sum_{n=1}^{N} \sum_{i=1}^{K} (\log \alpha_i + \log k_i - 2k_i|\theta_n - m_i|) p(i|\theta_n) \tag{7.11}$$

where $p(i|\theta_n)$ represents the probability of sample θ_n belonging to the ith Laplacian of the LMM. In a similar fashion to MoGs, the updates for $p(i|\theta_n)$ and α_i can be given by the following equations:

$$p(i|\theta_n) = \frac{\alpha_i \mathscr{L}(\theta_n, m_i, k_i)}{\sum_{i=1}^{K} \alpha_i \mathscr{L}(\theta_n, m_i, k_i)} \tag{7.12}$$

$$\alpha_i^+ \leftarrow \frac{1}{N}\sum_{n=1}^{N} p(i|\theta_n). \tag{7.13}$$

The updates for m_i and k_i are estimated by setting $\partial J(\alpha_i, k_i, m_i)/\partial m_i = 0$ and $\partial J(\alpha_i, k_i, m_i)/\partial k_i = 0$ respectively. Following some derivation (see [42]), we get the following update rules:

$$m_i^+ \leftarrow \frac{\sum_{n=1}^{N} \frac{\theta_n}{|\theta_n - m_i|} p(i|\theta_n)}{\sum_{n=1}^{N} \frac{1}{|\theta_n - m_i|} p(i|\theta_n)} \tag{7.14}$$

$$k_i^+ \leftarrow \frac{\sum_{n=1}^{N} p(i|\theta_n)}{2\sum_{n=1}^{N} |\theta_n - m_i| p(i|\theta_n)}. \tag{7.15}$$

The four update rules are iterated until convergence. Enhancing the sparsity in the angle representation θ_n will increase EM's convergence speed and will provide more accurate estimates for the sources' angles. Therefore, we train the LMM with a subset of those data points n that satisfy $r_n > B$, where B is a threshold.

Once the LMM is trained, the centre of each Laplacian m_i should represent a column of the mixing matrix $\mathbf{A}$ in the form of $[\cos(m_i)\ \sin(m_i)]^T$. Each wrapped Laplacian should model the statistics of each source in the transform domain and can be used to perform underdetermined source separation.

The main issue with LMMs is that they attempt to model a circular r.v. (angles) using a pdf that has infinite support. The Laplacian density, as described in (7.8), is valid $\forall\, \theta \in (-\infty, +\infty)$. However, the range of θ_n is not only bounded to the $(0, \pi)$ interval but the two boundaries are actually connected. Assume that you have a concentration of points close to π. The EM algorithm will attempt to fit a Laplacian around this cluster, however, assuming a linear support on θ. As a result, the algorithm cannot attribute points that belong to the same cluster, but are close to 0, due to the assumed linear support. Therefore, the algorithm cannot model densities with m_i close to 0 or π with great accuracy. To alleviate the problem, the estimated centres m_i can be rotated, so that the affected boundary (0 or π) is mapped to the middle of the centres m_i that feature the greatest distance (see [42]). This can offer a heuristic but not complete solution to the problem.

7.4 Identification Using Mixtures of Wrapped Laplacian Models

To address this problem in a more elegant manner, we examine the use of an approximate wrapped Laplacian distribution to model the π periodicity that exists in atan$(\cdot)$. The observed angles θ_n of the input data can be modelled, as a Laplacian wrapped around the interval $(0, \pi)$.

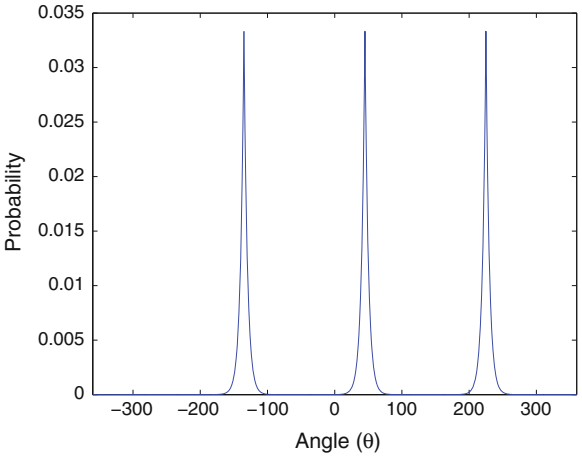


Fig. 7.3 An example of the wrapped Laplacian for $T = [-1, 0, 1]$ $k = 0.01$ and $m = \pi/4$

Definition 1 A wrapped Laplacian can be described by the following additive model

$$\mathscr{L}_w(\theta, k, m) = \frac{1}{2T-1} \sum_{t=-T}^{T} k e^{-2k|\theta - m - \pi t|} = \frac{1}{2T-1} \sum_{t=-T}^{T} \mathscr{L}(\theta - \pi t, k, m) \tag{7.16}$$

where $T \in \mathbf{Z}^+$ denotes the number of ordinary Laplacians with mean m and width k that participate in the wrapped version.

The above expression models the wrapped Laplacian by an ordinary Laplacian and its periodic repetitions by π (see Fig. 7.3). This is an extension of the wrapped Gaussian distribution proposed by Smaragdis and Boufounos [55] for the Laplacian case. The addition of the wrapping of the distribution aims at mirroring the wrapping of the observed angles at $\pm\pi$. In theory, the model should have $T \to \infty$ components, however, it seems that a small range of values for T can successfully approximate the full wrapped probability density function in practice.

In a similar fashion to LMMs, one can introduce *Mixture of wrapped Laplacians* (MoWL) in order to model a mixture of angular or circular sparse signals. A *MoWL* can thus be defined, as follows:

$$p(\theta) = \sum_{i=1}^{K} \alpha_i \mathscr{L}_w(\theta, k_i, m_i) = \sum_{i=1}^{K} \alpha_i \frac{1}{2T-1} \sum_{t=-T}^{T} k_i e^{-2k_i|\theta - m_i - \pi t|} \tag{7.17}$$

where α_i, m_i, k_i represent the weight, mean and width of each Laplacian respectively and all weights should sum up to one, i.e. $\sum_{i=1}^{K} \alpha_i = 1$. We can derive the EM algorithm, based on the previous analysis. Assuming N training samples for θ_n and a MoWL densities (7.17), the log-likelihood of these training samples θ_n takes the

following form:

$$I(\alpha_i, k_i, m_i) = \sum_{n=1}^{N} \log \sum_{i=1}^{K} \alpha_i \mathscr{L}_w(\theta_n, k_i, m_i). \tag{7.18}$$

One can introduce the probability $p(i|\theta_n)$ of sample θ_n belonging to the ith wrapped Laplacian of the MoWL and the probability $p(t|i, \theta_n)$ of sample θ_n belonging to the tth individual Laplacian of the ith wrapped Laplacian $\mathscr{L}_w(k_i, m_i)$. The updates for $p(t|i, \theta_n)$, $p(i|\theta_n)$ and α_i can be then given by the following equations:

$$p(t|i, \theta_n) = \frac{\mathscr{L}(\theta_n - \pi t, m_i, k_i)}{\sum_{t=-T}^{T} \mathscr{L}(\theta_n - \pi t, m_i, k_i)} \tag{7.19}$$

$$p(i|\theta_n) = \frac{\alpha_i \mathscr{L}_w(\theta_n, m_i, k_i)}{\sum_{i=1}^{K} \alpha_i \mathscr{L}_w(\theta_n, m_i, k_i)} \tag{7.20}$$

$$\alpha_i \leftarrow \frac{1}{N} \sum_{n=1}^{N} p(i|\theta_n) \tag{7.21}$$

$$m_i \leftarrow \frac{\sum_{n=1}^{N} \sum_{t=-T}^{T} \frac{\theta_n - \pi t}{|\theta_n - \pi t - m_i|} p(t|i, \theta_n) p(i|\theta_n)}{\sum_{n=1}^{K} \sum_{t=-T}^{T} \frac{1}{|\theta_n - \pi t - m_i|} p(t|i, \theta_n) p(i|\theta_n)} \tag{7.22}$$

$$k_i \leftarrow \frac{\sum_{n=1}^{N} p(i|\theta_n)}{2 \sum_{n=1}^{N} \sum_{t=-T}^{T} |\theta_n - \pi t - m_i| p(t|i, \theta_n) p(i|\theta_n)}. \tag{7.23}$$

Once the MoWL is trained, the centre of each wrapped Laplacian m_i should represent a column of the mixing matrix $\mathbf{A}$ in the form of $[\cos(m_i) \ \sin(m_i)]^T$. Each wrapped Laplacian should model the statistics of each source in the transform domain and can be used to perform underdetermined source separation. This approach addresses the problem of modelling directional data in a more elegant manner, however, the cost of training two EM algorithms makes this approach less attractive.

7.5 A Complete Solution Using the Generalised Directional Laplacian Distribution

The previous two efforts do not offer a closed-form solution to the problem and they can not be easily expanded to more than two sensors. The proposed multidimensional Directional Laplacian model offers a closed-form solution to the modelling of directional sparse data and can also address the general $K \times L$ underdetermined source

separation problem, which is rarely tackled in the literature. There exist distributions that are periodic by definition and can therefore offer closed-form models for circular or directional data.

The *von Mises distribution* (also known as the circular normal distribution) is a continuous probability distribution on the unit circle [24, 29]. It may be considered the circular equivalent of the normal distribution and is defined by:

$$p(\theta) = \frac{e^{k\cos(\theta - m)}}{2\pi I_0(k)}, \quad \forall\, \theta \in [0, 2\pi) \tag{7.24}$$

where $I_0(k)$ is the modified Bessel function of the first kind of order 0, m is the mean and $k > 0$ describes the "width" of the distribution. A generalisation of the previous density is the p-dimensional (p-D) von Mises-Fisher distribution [18, 37]. A p-D unit random vector $\mathbf{x}$ ($||\mathbf{x}|| = 1$) follows a *von Mises-Fisher* distribution, if its probability density function is described by:

$$p(\mathbf{x}) \propto e^{k\mathbf{m}^T\mathbf{x}}, \quad \forall ||\mathbf{x}|| \in \mathscr{S}^{p-1} \tag{7.25}$$

where $||\mathbf{m}|| = 1$ defines the centre, $k \geq 0$ and $\mathscr{S}^{p-1}$ is the p-D unit hypersphere. Since the random vector $\mathbf{x}$ resides on the surface of a p-D unit-sphere, $\mathbf{x}$ essentially describes directional data. In the case of $p = 2$, $\mathbf{x}$ models data that exist on the unit circle and thus can be described only by an angle. In this case, the von Mises-Fisher distribution is reduced to the von Mises distribution of (7.24). The von Mises-Fisher distribution has been extensively studied and many methods have been proposed to fit the distribution or its mixtures to normally distributed circular data [5, 18, 29, 37].

7.5.1 A Generalised Directional Laplacian Model

Assume a r.v. θ modelling directional data with π-periodicity. The periodicity of the density function can be amended to reflect a "fully circular" phenomenon (2π), however, for the rest of the paper we will assume that $\theta \in [0, \pi)$, since it is required by the source separation application. From the definition of the von Mises distribution in (7.24), one can create a Laplacian structure simply by introducing a $|\cdot|$ operator in the superscript of the exponential. This action introduces a large concentration around the mean, which is needed to describe a sparse or Laplacian density. Values far away from the mean are smoothed out by the exponential. Additionally, we have to perform some minor amendments to the phase shift and also invert the distribution in order to impose the desired shape on the derived density.

Definition 2 The following probability density function models directional Laplacian data over $[0, \pi)$ and is termed *Directional Laplacian Density* (DLD):

$$p(\theta) = c(k)e^{-k|\sin(\theta - m)|}, \quad \forall\, \theta \in [0, \pi) \tag{7.26}$$

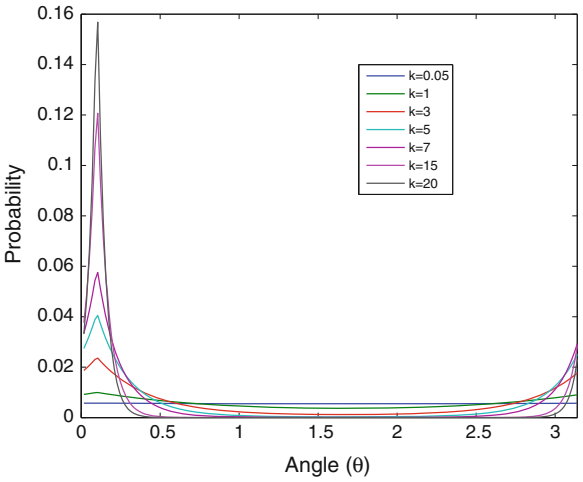


Fig. 7.4 The proposed Directional Laplacian Density (DLD) for various values of k

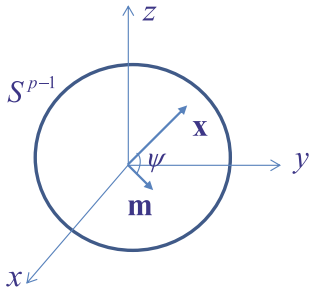


Fig. 7.5 Generalising the Directional Laplacian density in $\mathscr{S}^{p-1}$

where $m \in [0, \pi)$ defines the mean, $k > 0$ defines the width ("approximate variance") of the distribution, $c(k) = \frac{1}{\pi I_0(k)}$ and $I_0(k) = \frac{1}{\pi} \int_0^{\pi} e^{-k \sin\theta} d\theta$.

The normalisation coefficient $c(k) = 1/\pi I_0(k)$ is derived from the fundamental normalisation property of probability density functions [40]. Examples of (7.26) are shown in Fig. 7.4. More details on the special 1D DLD case can be found in [40].

The next step is to derive a generalised definition for the Directional Laplacian model. To generalise the concept of 1D DLD in the p-D space, we will be inspired by the p-D von Mises-Fisher distribution [18, 37]. The von Mises-Fisher distribution is described by $p(\mathbf{x}) \propto e^{k\mathbf{m}^T\mathbf{x}}$ (see (7.25)). Since $||\mathbf{x}|| = ||\mathbf{m}|| = 1$, the inner product $\mathbf{m}^T\mathbf{x} = \cos\psi$, where ψ is the angle between the two vectors $\mathbf{x}$ and $\mathbf{m}$ (see Fig. 7.5). Following a similar methodology to the 1D-DLD, we need to formulate the term $-k|\sin\psi|$ in the superscript of the exponential. We can then derive $|\sin\psi| = \sqrt{1-\cos^2\psi} = \sqrt{1-(\mathbf{m}^T\mathbf{x})^2}$. Thus, the superscript of the generalised DLD can be given by $-k\sqrt{1-(\mathbf{m}^T\mathbf{x})^2}$.

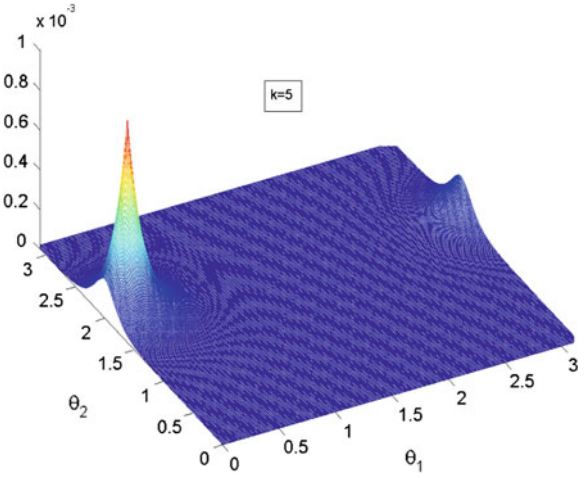


Fig. 7.6 The proposed Generalised Directional Laplacian Distribution for $k = 5$ and $p = 3$

Definition 3 The following probability density function models p-D directional Laplacian data and is termed *Generalised Directional Laplacian Distribution* (DLD):

$$p(\mathbf{x}) = c_p(k) e^{-k\sqrt{1-(\mathbf{m}^T\mathbf{x})^2}}, \quad \forall\, ||\mathbf{x}|| \in \mathscr{S}^{p-1} \tag{7.27}$$

where $\mathbf{m}$ defines the mean, $k \geq 0$ defines the width ("approximate variance") of the distribution, $c_p(k) = \frac{\Gamma\left(\frac{p-1}{2}\right)}{\pi^{\frac{p+1}{2}} I_{p-2}(k)}$, $I_p(k) = \frac{1}{\pi}\int_0^{\pi} e^{-k\sin\theta}\sin^p\theta d\theta$ and $\Gamma(\cdot)$ represents the Gamma function.[2]

The normalisation coefficient $c_p(k)$ is calculated in Appendix 1. In the case of $p = 2$, the generalised DLD is reduced to the one- dimensional DLD of (7.26), verifying the validity of the above model. The generalised DLD density models "directional" data on the half-unit p-D sphere, however, it can be extended to the unit p-D sphere, depending on the specifications of the application. In Fig. 7.6, an example of the generalised DLD is depicted for $p = 3$ and $k = 5$. The centre $\mathbf{m}$ is calculated using spherical coordinates $\mathbf{m} = [\cos\theta_1\cos\theta_2;\ \cos\theta_1\sin\theta_2;\ \sin\theta_1]$ for $\theta_1 = 0.2$ and $\theta_2 = 2$.

[2] Note that for n positive integer, we have that $\Gamma(n) = (n-1)!$

7.5.2 Generalised Directional Laplacian Density Samples Generation

To generate 1D Directional Laplacian data, we employed the inversion of the cumulative distribution method [17]. Inversion methods are based on the observation that continuous cumulative distribution functions (cdf) range uniformly over the interval (0, 1). Since the proposed density is bounded between $[0, \pi)$, we can evaluate the cdf of the DLD with uniform sampling at $[0, \pi)$ and approximate the inverse mapping using spline interpolation. Thus, uniform random data in the interval (0, 1) can be transformed to 1D Directional Laplacian random samples, using the described inverse mapping procedure.

To simulate 2-D Directional Laplacian random data ($p = 3$), we sampled the 2-D density function for specific $\mathbf{m}$, k. The bounded value space ($\theta_1, \theta_2 \in [0, \pi)$) is quantised into small rectangular blocks, where the density is assumed to be uniform. Consequently, we generate a number of uniform random samples for each block. The number of samples generated from each block is different and defined by the overall DL density. The required 3-D unit-norm random vectors are produced using spherical coordinates with unit distance and angles θ_1, θ_2 from the random 2-D Directional data. The above procedure can be extended for the generation of p-D directional data.

7.5.3 Maximum Likelihood Estimation of Parameters m, k

Assume a population of p-D angular data $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_N\}$ that follow a p-D Directional Laplacian Distribution. To estimate the model parameters using Maximum Likelihood Estimation (MLE), one can form the log-likelihood and estimate the parameters $\mathbf{m}$, k that maximise it. For the Generalised DLD density, the log-likelihood function can be expressed, as follows:

$$J(\mathbf{X}, \mathbf{m}, k) = N \log \frac{\Gamma\left(\frac{p-1}{2}\right)}{\pi^{\frac{p+1}{2}} I_{p-2}(k)} - k \sum_{n=1}^{N} \sqrt{1 - (\mathbf{m}^T \mathbf{x}_n)^2}. \tag{7.28}$$

Alternate optimisation is performed to estimate $\mathbf{m}$ and k. The gradients of J along $\mathbf{m}$ and k are calculated in Appendix 2. The update for $\mathbf{m}$ is given by gradient ascent on the log-likelihood via:

$$\mathbf{m}^+ \leftarrow \mathbf{m} + \eta \sum_{n=1}^{N} \frac{\mathbf{m}^T \mathbf{x}_n}{\sqrt{1 - (\mathbf{m}^T \mathbf{x}_n)^2}} \mathbf{x}_n \tag{7.29}$$

$$\mathbf{m}^+ \leftarrow \mathbf{m}^+ / ||\mathbf{m}^+|| \tag{7.30}$$

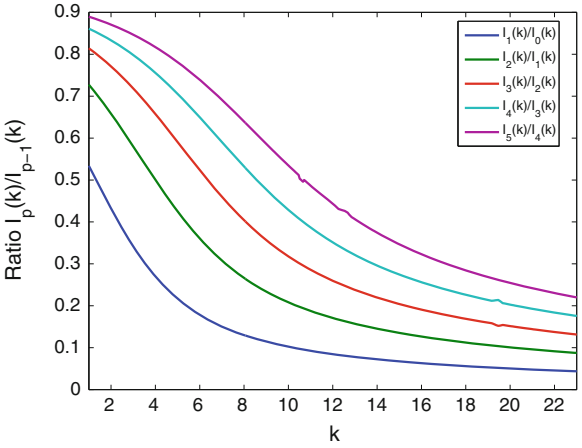


Fig. 7.7 The ratio $I_p(k)/I_{p-1}(k)$ is a monotonic $1-1$ function of k

where η defines the gradient step size. Since the gradient step does not guarantee that the new update for $\mathbf{m}$ will remain on the surface of $\mathscr{S}^{p-1}$, we normalise the new update to unit norm. To estimate k, a numerical solution to the equation $\partial J(\mathbf{X}, \mathbf{m}, k)/\partial k = 0$ is estimated. From the analysis in Appendix 2, we have that

$$\frac{I_{p-1}(k)}{I_{p-2}(k)} = \frac{1}{N}\sum_{n=1}^{N}\sqrt{1-(\mathbf{m}^T\mathbf{x}_n)^2} \tag{7.31}$$

To calculate k analytically from the ratio $I_{p-1}(k)/I_{p-2}(k)$ is not straightforward. However, after numerical evaluation, it can be demonstrated that the ratio $I_{p-1}(k)/I_{p-2}(k)$ is a smooth monotonic $1-1$ function of k. In Fig. 7.7, the ratio $I_{p-1}(k)/I_{p-2}(k)$ is estimated for uniformly sampled values of $k \in [0.01, 30]$ and $p = 2, 3, 4, 5, 6$. Since this ratio is not dependent on data, one can create a look-up table for a variety of k values and use interpolation to estimate k from an arbitrary value of $I_{p-1}(k)/I_{p-2}(k)$. This look-up table solution is more efficient compared to possible iterative estimation approaches of k and generally accelerates the model's training.

7.5.4 Mixtures of Generalised Directional Laplacians

One can employ *Mixtures of Generalised Directional Laplacians* (MDLD) in order to model multiple concentrations of directional generalised "heavy-tailed signals".

Definition 4 Mixtures of Generalised Directional Laplacian Distributions are defined by the following pdf:

$$p(\mathbf{x}) = \sum_{i=1}^{K} a_i c_p(k_i) e^{-k_i \sqrt{1-(\mathbf{m}_i^T \mathbf{x})^2}}, \quad \forall \, ||\mathbf{x}|| \in \mathscr{S}^{p-1} \tag{7.32}$$

where a_i denotes the weight of each distribution in the mixture, K the number of DLDs used in the mixture and $\mathbf{m}_i$, k_i denote the mean and the "width" (approximate variance) of each distribution.

The mixtures of DLD can be trained using the EM algorithm. Following the previous analysis in [6], one can yield the following simplified likelihood function:

$$\mathscr{L}(a_i, \mathbf{m}_i, k_i) = \sum_{n=1}^{N} \sum_{i=1}^{K} \left(\log \frac{a_i \Gamma(\frac{p-1}{2})}{\pi^{\frac{p+1}{2}} I_{p-2}(k)} - k\sqrt{1-(\mathbf{m}^T \mathbf{x})^2} \right) p(i|\mathbf{x}_n) \tag{7.33}$$

where $p(i|\mathbf{x}_n)$ represents the probability of sample $\mathbf{x}_n$ belonging to the ith Directional Laplacian of the mixture. In a similar fashion to other mixture model estimation, the updates for $p(i|\mathbf{x}_n)$ and α_i can be given by the following equations:

$$p(i|\mathbf{x}_n) \leftarrow \frac{a_i c_p(k_i) e^{-k_i \sqrt{1-(\mathbf{m}_i^T \mathbf{x})^2}}}{\sum_{i=1}^{K} a_i c_p(k_i) e^{-k_i \sqrt{1-(\mathbf{m}_i^T \mathbf{x})^2}}} \tag{7.34}$$

$$a_i \leftarrow \frac{1}{N} \sum_{n=1}^{N} p(i|\mathbf{x}_n). \tag{7.35}$$

Based on the derivatives calculated in Appendix 2, it is straightforward to derive the following updates for $\mathbf{m}_i$ and k_i, as follows:

$$\mathbf{m}_i^+ \leftarrow \mathbf{m}_i + \eta \sum_{n=1}^{N} k_i \frac{\mathbf{m}^T \mathbf{x}_n}{\sqrt{1-(\mathbf{m}^T \mathbf{x}_n)^2}} \mathbf{x}_n \, p(i|\mathbf{x}_n) \tag{7.36}$$

$$\mathbf{m}_i^+ \leftarrow \mathbf{m}_i^+ / ||\mathbf{m}_i^+||. \tag{7.37}$$

To estimate k_i, in a similar fashion to the previous MLE, the optimisation yields:

$$\frac{I_{p-1}(k_i)}{I_{p-2}(k_i)} = \frac{\sum_{n=1}^{N} \sqrt{1-(\mathbf{m}_i^T \mathbf{x}_n)^2} \, p(i|\mathbf{x}_n)}{\sum_{n=1}^{N} p(i|\mathbf{x}_n)}. \tag{7.38}$$

The training of this mixture model is also dependent on the initialisation of its parameters, especially the means $\mathbf{m}_i$ [39]. In Appendix 3, the standard K-Means algorithm is reformulated in order to tackle p-D directional data. The proposed p-D *Directional K-Means* is used to initialise the means $\mathbf{m}_i$ of the DLDs in the generalised DLD mixture EM training. A *Directional K-Means* already exists in the literature [4],

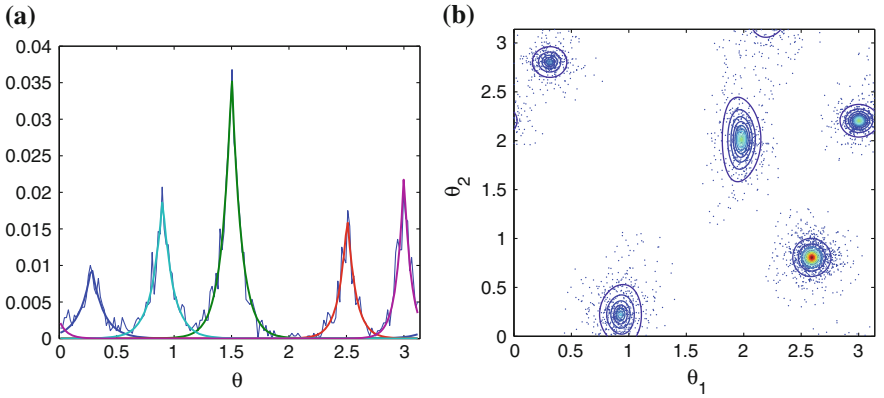


Fig. 7.8 Examples of fitting a Generalised MDLD model on 2,000 randomly generated 1D (*left*) and 2D (*right*) directional Laplacian data. **a** 1D DLD mixture. **b** 2D DLD mixture

however, the proposed p-D *Directional K-Means* in Appendix 3 employs a distance function more relevant to sparse directional data. Examples of trained MDLD are shown in Fig. 7.8.

7.6 Source Separation Using Hard or Soft Thresholding

Once the Mixture Model is trained, optimal detection theory and the estimated individual Laplacians can be employed to provide estimates of the sources. The centre of each Laplacian m_i should represent a column of the mixing matrix $\mathbf{A}$ in the form of $[\cos(m_i)\ \ \sin(m_i)]^T$. Each Laplacian should model the statistics of each source in the transform domain. Thus, using either a *hard* or a *soft decision threshold*, we can perform underdetermined source separation. The same strategy can hold for either of the three proposed LMM.

7.6.1 Hard Thresholding

The hard thresholding ("Winner takes all") strategy attributes each point of the scatter plot of Fig. 7.1b to only one of the sources. This is performed by setting a hard threshold at the intersections between the trained Laplacians. Consequently, the source separation problem becomes an *optimal decision* problem. The decision thresholds θ_{ij}^{opt} between the i-th and the j-th neighbouring Laplacians depend on the type of mixture model (LMM, MoWL or MDLD). Threshold formulas for the LMM and MoWL can be found in [39, 42], respectively. Using these thresholds, the algorithm can attribute the points with $m_{ij}^{\text{opt}} < \theta_n < m_{jk}^{\text{opt}}$ to source j, where i, j, k

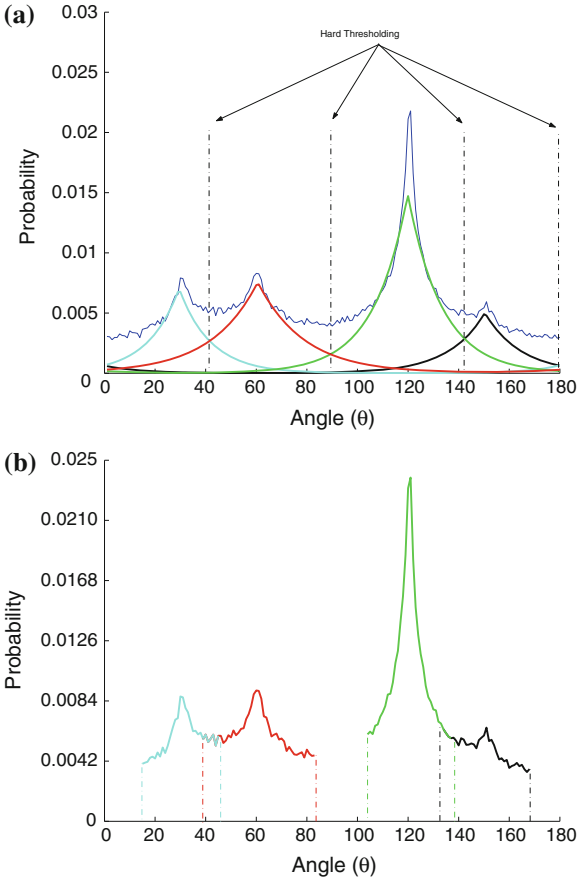


Fig. 7.9 A two-sensors four-sources scenario, separated using LMM. In **a**, the four trained Laplacians are depicted along with the actual density function and the imposed hard thresholds. Applying soft thresholds, the classification shown in (**b**) is achieved, which allows some overlapping between adjacent sources. **a** Hard thresholding. **b** Soft Thresholding

are neighbouring Laplacians (sources). Figure 7.9a depicts the fitted LMM, in a two sensors—four audio sources (voice, piano, accordion and violin) example and the hard thresholds imposed using the above equation. The points that belong to each of the four clusters, shown in Fig. 7.9a, are attributed and are used to reconstruct eachsource.

In the case of the 1D-MDLD, it is possible to derive the thresholds where two neighbouring DLDs intersect and therefore apply a hard thresholding strategy to cluster the audio data. In the case of an p-D MDLD, it is not straightforward to derive the intersecting hyperplanes between two neighbouring DLDs, therefore, in this case we resort to the soft-thresholding technique.

7.6.2 Soft Thresholding

Observing the histograms of Fig. 7.2, we can attribute points that are distant from the centre of the 2D representation to each source with great confidence. In contrast, there exist points that cannot be attributed to either source with great confidence. These points may belong to more than one source. One can then relax the hard threshold strategy, by allowing points belonging to more than one source simultaneously. A "soft-thresholding" strategy can attribute points that constitute a chosen ratio q (i.e. 0.7–0.9) of the density of each Laplacian (any of the three models) to the corresponding source (see Fig. 7.9). Hence, the ith source can be associated with those points θ_n, for which $p(\theta_n) \geq (1-q)\alpha_i k_i$, where $p(\theta_n)$ is given by the corresponding density model. A large value for q allows more points to belong to more than one Laplacian. A small value for q imposes stricter criteria for the points to belong to a Laplacian and essentially becomes a hard thresholding approach. This scheme can be effective, only if the estimated Laplacians are concentrated around each $\mathbf{m}_i$. In the opposite case, there will be components that will dominate the pdf and therefore be attributed with more points than it should and therefore they would contain contamination from other sources. In Fig. 7.9b, we can see the four sources in the previous 1D example, as classified by the soft thresholding strategy. The different colours represent different clusters, i.e. different sources. We can see that several points are attributed to both the first and the second sources and both the third and fourth sources by the soft classification scheme.

7.6.3 Source Reconstruction

Having attributed the points $\mathbf{x}(n)$ to the L sources, using either the "hard" or the "soft" thresholding technique, the next step is to reconstruct the sources. Let $S_i \sqsubseteq N$ represent the point indices that have been attributed to the ith source and $\mathbf{m}_i$ the corresponding mean vector, i.e. the corresponding column of the mixing matrix. We initialise $u_i(n) = 0, \forall\, n = 1, \ldots, N$ and $i = 1, \ldots, L$. The source reconstruction is performed by substituting:

$$u_i(S_i) = \mathbf{m}_i^T \mathbf{x}(S_i) \qquad \forall\, i = 1, \ldots, L. \tag{7.39}$$

In the case that we need to capture the multichannel image of the separated source, the result of the separation is a multichannel output that is initialised to $\mathbf{u}_i(n) = \mathbf{0} \forall$ $n = 1, \ldots, N$. The source image reconstruction is performed by:

$$\mathbf{u}_i(S_i) = \mathbf{x}(S_i) \qquad \forall\, i = 1, \ldots, L. \tag{7.40}$$

7.7 Experiments

In this section, we verify the validity of the above derived MLE algorithms and demonstrate the density's relevance and performance in underdetermined audio source separation. We can see that the proposed MDLD model improves the LMM and MoWL modelling efforts in terms of stability, speed and performance and offers a fast alternative to state-of-the-art algorithms with reasonable separation performance.

We will use Hyvärinen's clustering approach [28], the MoWL algorithm [39] and the "GaussSep" algorithm [58] for comparison. We prefered not to benchmark the LMM model, because the other two models (MoWL and MDLD) tackle data's directionality more efficiently. After fitting the MDLD with the proposed EM algorithm, separation will be performed using hard or soft thresholding, as described earlier. In order to quantify the performance of the algorithms, we estimate the *Signal-to-Distortion Ratio* (SDR), the *Signal-to-Interference Ratio* (SIR) and the *Signal-to-Artefact Ratio* from the BSS_EVAL Toolbox v.3 [23]. The input signals for the MDLD, MoWL and Hyvärinen's approaches are sparsified using the MDCT, as developed by Daudet and Sandler [13]. The frame length for the MDCT analysis is set to 32 ms for the speech signals and 128 ms for the music signals sampled at 16 KHz, and to 46.4 ms for the music signals at 44.1 KHz. We initialise the parameters of the MoWL and MDLD as follows: $\alpha_i = 1/K$ and $c_i = 0.001$, $T = [-1, 0, 1]$ (for MoWL only) and $k_i = 15$ (for the DLD only). The centres m_i were initialised in either case using the Directional *K-means* step, as described in Appendix 3. We used the "GaussSep" algorithm, as publicly available by the authors.[3] For the estimation of the mixing matrix, we used Arberet et al.'s [2] DEMIX algorithm,[4] as suggested in [58]. The number of sources in the mixture was also provided to the DEMIX algorithm, as it was provided to all other algorithms. The "GaussSep" algorithm operates in the STFT domain, where we used the same frame length with the other approaches and a time–frequency neighbourhood size of 5 for speech sources and 15 for music sources.

7.7.1 Two-Microphone Examples

We tested the algorithms with the *Groove*, *Latino1* and *Latino2* datasets, available by BASS-dB [59], and sampled at 44.1 KHz. The "Groove" dataset features four widely spaced sources: bass (far left), distorted guitar (centre left), clean guitar (centre right) and drums (far right). The two "Latino" datasets features four widely spaced sources: bass (far left), drums (centre left), keyboards (centre right) and distorted guitar

[3] MATLAB code for the "GaussSep" algorithm is available from http://www.irisa.fr/metiss/members/evincent/software.

[4] MATLAB code for the "DEMIX" algorithm is available from http://infoscience.epfl.ch/record/165878/files/.

(far right). We also used a variety of test signals from the Signal Separation Evaluation Campaigns SiSEC2008 [52] and SiSEC2010 [53]. We employed two audio instantaneous mixtures from the "dev1" and "dev2" data sets ("Dev2WDrums" and "Dev1WDrums" sets—three instruments at 16 KHz) and two speech instantaneous mixtures from the "dev2" data set ("Dev2Male3" and "Dev2Female3" sets—four closely located sources at 16 KHz). We used the development (dev) datasets instead of the test data sets, in order to have all the source audio files for proper benchmarking.

In Table 7.1, we can see the results for the four methods in terms of SDR, SIR and SAR. For simplicity, we averaged the results for all sources at each experiment. The reader of the paper can visit the following url[5] and listen to the described separation results. The proposed MDLD approach seems to outperform our previous separation effort MoWL and Hyvärinen's algorithm in terms of all the performance indexes. The proposed MDLD approach is not susceptible to bordering effects, since it is circular by definition and avoids shortcomings of our previous offerings. Compared to a state-of-the-art method, such as "GaussSep", our method is better in terms of the SIR index but is falling behind in terms of the SDR and SAR indexes. The SIR index reflects the capability of an algorithm to remove interference from other sources in the mixture. The SAR index refers to the audible artefacts that remain in the separated signals, due to the overlapping of several points in the time–frequency space (even in the MDCT representation) in the underdetermined mixture that are incorrectly attributed to either source. In this sense, our algorithm seems to perform slightly better compared to "GaussSep" in terms of removing "crosstalk" from other sources, but there seem to be more audible artefacts after separation in our approach compared to "GaussSep". This is due to the fact that the "GaussSep" segments the time–frequency representation in small localised neighbourhoods and performs local Gaussian Modelling so as to separate and filter sources from those areas that separation is more achievable. Instead, our approach simply clusters all time–frequency points according to the fitted DLD using hard thresholds (or soft-thresholds in the case $K > 2$).

Another important issue is to compare the processing time of the three best performing algorithms. All experiments were conducted on an Intel Core i5-460M (2.53 GHz) with 4GB DDR3 SDRAM running Windows Professional 64-bit and MATLAB R2012b. Our MATLAB implementations of the MDLD and MoWL algorithms were not optimised in terms of execution speed. In Table 7.2, the typical running time in seconds is summarised for each experiment and method. The first observation is that the MDLD approach is faster compared to the MoWL approach. As it was previously mentioned, employing a mixture of wrapped Laplacians to solve the "circularity" problem entails the running of two EM algorithms: one for the wrapped Laplacians and one for the MoWL. This seems to delay the convergence of the algorithm. Instead, MDLD requires the training of one EM algorithm for the mixture and it seems to converge faster compared to MoWL. The second observation is that there is an important difference between the processing time of the MDLD approach and the "GaussSep" algorithm. As previously mentioned, the "GaussSep" algorithm is more complicated

[5] http://utopia.duth.gr/~nmitiano/mdld.htm

Table 7.1 The proposed MDLD approach is compared for source estimation performance ($K = 2$) in terms of SDR (dB), SIR (dB) and SAR(dB) with GaussSep, WMoL and Hyvärinen's approach

	SDR (dB)				SIR (dB)				SAR (dB)			
	MDLD	GaussSep	MoWL	Hyva	MDLD	GaussSep	MoWL	Hyva	MDLD	GaussSep	MoWL	Hyva
Latino1	6.38	5.51	5.72	0.89	18.63	8.96	18.59	9.61	6.93	9.20	6.26	3.63
Latino2	3.21	4.71	2.10	0.89	11.50	8.87	11.28	9.61	4.95	9.20	3.85	3.63
Groove	0.22	0.39	−0.43	−0.08	9.48	3.62	9.60	8.88	2.12	7.37	1.00	1.83
Dev2Male3	3.04	6.22	2.11	−3.10	13.69	12.14	13.30	4.73	4.10	8.04	3.33	2.72
Dev2Female3	4.68	5.70	3.86	−1.85	15.28	11.45	16.58	5.02	5.41	7.51	4.61	3.13
Dev2WDrums	9.59	16.57	10.16	0.63	19.77	23.83	19.98	7.57	10.55	17.68	10.54	5.54
Dev1WDrums	4.96	16.54	3.81	6.86	13.88	20.94	12.38	16.75	6.37	19.30	5.20	7.73
Average	4.58	**7.96**	3.91	0.6	**14.61**	12.83	13.82	8.88	5.78	**11.19**	4.97	4.03

The measurements are averaged for all sources of each experiment

Table 7.2 Running time comparison with GaussSep and MoWL approaches

	MDLD	Gaussep	MoWL
Groove	2.39	224.21	20.46
Latino1	1.27	122.02	5.48
Latino2	1.28	129.09	3.59
Dev2Male3	2.31	72.64	19.67
Dev2Female3	2.33	75.92	16.09
Dev2WDrums	2.07	56.79	8.55
Dev1WDrums	1.55	54.06	11.88
Average	**1.88**	**104.96**	**12.24**
Dev3Female3	9.56	1021.31	–
Example(3×5)	4.04	1598.7	–
Example(4×8)	9.393	2359.1	–
Average	**7.66**	**1659.70**	–

The measurements are in seconds

in structure thus justifying its long running time. Nevertheless, the proposed MDLD approach offers a very fast underdetermined source separation alternative with high SIR performance that can be used in environments where processing time is important. The third observation is that the processing time for the "GaussSep" algorithm scales significantly with the duration of the signals and the number of sources, i.e. the "Groove", "Latino1", "Latino2" (44.1 KHz—4 sources) require more time than the Dev2Male3 and Dev2Female3 sets (16 KHz—4 sources) and the Dev2WDrums and Dev1WDrums sets (16 KHz—3 sources). Instead, MDLD's running time seems to be closer to the average in most cases, maybe slightly deteriorating with the complexity of the source separation problem.

7.7.2 Underdetermined Source Separation Examples with More Than Two Mixtures

In this section, we employ the described generalised DLD approach to perform separation of $3 \times L$ and $4 \times L$ mixtures. The 2-mixtures setup, that dominates the literature, may also arise from the fact that most audio recordings and CD masters are available as stereo recordings (two channels is equivalent to two mixtures), where we need to separate the instruments that are present. Nowadays, the music industry is moving towards multichannel formats, including the 5.1 and the 7.1 surround sound formats, which implies more than two channels will be available for processing. In this section, we will attempt to perform separation of the Dev3Female3 set from SiSEC2011 [54] and a 3×5 (3 mixtures—5 sources) and a 4×8 (4 mixtures—8 sources) scenario using the male and female voices from Dev3. Our MDLD approach will be compared to the "GaussSep" algorithm that is able to work with multi-channel data. We used the same frame length and time–frequency neighbourhood sizes for

Table 7.3 The sources' position angles that were used in the 3×5 and the 4×8 example

	3×5							
	s_1	s_2	s_3	s_4	s_5			
θ_1	$0°$	$-87°$	$-60°$	$0°$	$45°$			
θ_2	$85°$	$0°$	$-60°$	$0°$	$45°$			
	4×8							
	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8
θ_1	$-75°$	$-30°$	$0°$	$50°$	$10°$	$80°$	$-45°$	$0°$
θ_2	$70°$	$30°$	$-20°$	$50°$	$-70°$	$0°$	$15°$	$-70°$
θ_3	$80°$	$20°$	$10°$	$-50°$	$0°$	$-10°$	$-25°$	$-35°$

both algorithms as previously. The MDLD was initialised as described in the previous section. After fitting the model, we employed the soft-thresholding scheme, as it was described in [42]. Since it was not straightforward to calculate the intersection surfaces between the individual p-D DLDs, we employed a soft-thresholding scheme, as described earlier, using a value of $q = 0.8$.

For the 3×5 example, we centred the five speech sources around the angles shown in Table 7.3 which were mixed using the following mixing matrix:

$$\mathbf{A}_{3\times L} = \begin{bmatrix} \cos\theta_2 \cos\theta_1 \\ \cos\theta_2 \sin\theta_1 \\ \sin\theta_2 \end{bmatrix}. \tag{7.41}$$

For the 4×8 example, we centred eight audio sources around the angles shown in Table 7.3 which were mixed using the mixing matrix:

$$\mathbf{A}_{4\times L} = \begin{bmatrix} \cos\theta_3 \cos\theta_2 \cos\theta_1 \\ \cos\theta_3 \cos\theta_2 \sin\theta_1 \\ \cos\theta_3 \sin\theta_2 \\ \sin\theta_3 \end{bmatrix}. \tag{7.42}$$

The separation results for the three experiments in terms of SDR, SIR and SAR can be summarised in Table 7.4. The reader can listen to the audio results from the following url (See Footnote 5). In the case of $K = 3$ mixtures, both algorithms managed to perform separation in either case. Similarly to the $K = 2$ case, "GaussSep" featured higher SDR and SAR performances, whereas the proposed MDLD algorithm featured higher SIR performance. The image is completely different in the case of $K = 4$ mixtures, where MDLD manages to separate all eight sources in contrast to "GaussSep" that fails to perform separation. This might be due to fact that the sparsest ML solution in the optimisation of [58] is restricted to vectors with $K \leq 3$ entries, i.e. three sources present at each point. In contrast, the proposed MDLD algorithm is designed to operate for any arbitrary number of sensors K, without any constraint.

Table 7.4 The proposed MDLD approach is compared for source estimation performance ($K = 3, 4$) in terms of SDR (dB), SIR (dB) and SAR(dB) with the GaussSep approach

	SDR (dB)		SIR (dB)		SAR (dB)	
	MDLD	GaussSep	MDLD	GaussSep	MDLD	GaussSep
Dev3Female3	6.02	16.93	23.84	22.43	6.17	18.40
Example 3×5	3.91	9.94	17.92	15.21	4.17	11.68
Example 4×8	2.24	−18.63	16.4	−17.58	2.52	9.39

The measurements are averaged for all sources of each experiment

In Table 7.2, we can see the processing times for the two algorithms for the three experiments. The MDLD processing time has increased slightly but still remains relatively fast, requiring an average of 7.66 s to perform separation. This implies that the computational complexity of the proposed MDLD algorithm does not scale considerably with the number of sources L and sensors K. In contrast, the "GaussSep" algorithm's processing has increased considerably with K. The processing time seems to scale up dramatically with increasing K and number of estimated sources L. For $K = 3$, it required an average of 1,310 s and for $K = 4$, it required 2,359 s which is almost the double processing time for $K = 3$. Thus, it appears that the proposed MDLD algorithm is capable of offering a faster and more stable multichannel solution to the underdetermined source separation problem, featuring higher SIR rates, compared to a state-of-the-art approach.

The main aspiration for future work behind these experiments is to combine the speed and stability of the MDLD approach with the low-artefact separation quality, proposed by Vincent et al. [58]. It might be possible to import this time–frequency localised source separation framework, where the source clusters can be modelled by mixtures of MDLDs. A more intelligent fuzzy clustering algorithm may combine the information from the MDLD priors to attribute points to multiple sources, overcoming the artefacts that arise from the partitioning of the time–frequency space.

7.8 Conclusions: Possible Extensions

In this chapter, the problem of underdetermined instantaneous source separation is addressed. Since the data can have sparse representations in a transform domain, it is rational to use mixtures of heavy-tailed distributions, such as the Laplacian distribution, to model each source's distribution in the mixture environment. As the main concentrations of data appear on the directions spanned by the columns of the mixing matrix, the source separation problem is transformed to an angular clustering problem. In other words, data that need to be processed are directional, that the use of Laplacian distributions with infinite support is not efficient for sources near 0, or π directions. The first improvement was to wrap the ordinary Laplacian distribution and create a Wrapped Laplacian distribution. Training mixture of the Wrapped Laplacian Distribution is computationally expensive due to the concurrent estimation of two EM algorithms. The existence of closed-form directional Gaussian models inspired

the introduction of a Laplacian directional model. Building on previous work on directional Gaussian models (i.e. the von-Mises and the vonMises-Fisher densities), we proposed a novel generalised Directional Laplacian model for modelling multidimensional directional sparse data. Maximum Likelihood estimates of the densities' parameters were proposed along with an EM-algorithm that handles the training of DLD mixtures. The proposed algorithms were tested and demonstrated good performance in modelling the directionality of the data. The proposed algorithm can also provide a solution for the general multichannel underdetermined source separation problem ($K \geq 2$), offering fast and stable performance and high SIR compared to state-of-the-art methods [58].

Possible extensions is to adapt this technique for a convolutive mixture scenario, where using the STFT, we transform the convolutive mixtures into multiple complex instantaneous mixtures. Source separation-clustering for each frequency bin can be performed using a modified version of the proposed algorithm for complex numbers and permutation alignment can be performed using Time–Frequency Envelopes or Direction-of-Arrival methods as proposed by Mitianoudis and Davies [38, 41] or Sawada et al. [51]. The speed of the proposed MDLD algorithm can be a very positive feature for FD-BSS, since these methods need to solve many complex instantaneous source separation problems simultaneously.

Another possible direction is to adapt the proposed technique for underdetermined PNL mixtures. Once the mixtures have been linearised by the blind compensation method of Duarte et al. [19], it is always possible to use the proposed technique to unmix the PNL mixtures in the linear stage. The speed of the proposed MDLD algorithm may expedite the blind estimation of the inverse nonlinear function of PNL mixtures.

Appendix 1

Calculation of the Normalisation Parameter for the Generalised DLD

To estimate the normalisation coefficient $c_p(k)$ of (7.27), we need to solve the following equation:

$$\int_{\mathbf{x}\in\mathscr{S}^{p-1}} c_p(k) e^{-k\sqrt{1-(\mathbf{m}^T\mathbf{x})^2}} \mathrm{d}\mathbf{x} = 1. \tag{7.43}$$

Following Eq. (B.8) and in a similar manner to the analysis in Appendix B.2 in [18], we can rewrite the above equation as follows:

$$c_p(k) \int_0^{\pi} \mathrm{d}\theta_{p-1} \int_0^{\pi} e^{-k\sqrt{1-\cos^2\theta_1}} \sin^{p-2}\theta_1 \mathrm{d}\theta_1 \prod_{j=3}^{p-1} \int_0^{\pi} \sin^{p-j}\theta_{j-1} \mathrm{d}\theta_{j-1} = 1. \tag{7.44}$$

Following a similar methodology to Appendix B.2 in [18], the above yields:

$$c_p(k)\pi \int_0^{\pi} e^{-k\sin\theta_1} \sin^{p-2}\theta_1 \mathrm{d}\theta_1 \frac{\pi^{\frac{p-3}{2}}}{\Gamma\left(\frac{p-1}{2}\right)} = 1. \tag{7.45}$$

Using the definition of $I_p(k)$, we can write

$$c_p(k) I_{p-2}(k) \frac{\pi^{\frac{p+1}{2}}}{\Gamma\left(\frac{p-1}{2}\right)} = 1 \Rightarrow c_p(k) = \frac{\Gamma\left(\frac{p-1}{2}\right)}{\pi^{\frac{p+1}{2}} I_{p-2}(k)}. \tag{7.46}$$

Appendix 2

Gradient Updates for m and k for the MDDLD

The first-order derivative of the log-likelihood in (7.28) for the estimation of **m** are calculated below:

$$\begin{aligned} \frac{\partial J(\mathbf{X}, \mathbf{m}, k)}{\partial \mathbf{m}} &= -k \sum_{n-1}^{N} \frac{-2\mathbf{m}^T \mathbf{x}_n}{2\sqrt{1-(\mathbf{m}^T \mathbf{x}_n)^2}} \mathbf{x}_n \\ &= k \sum_{n=1}^{N} \frac{\mathbf{m}^T \mathbf{x}_n}{\sqrt{1-(\mathbf{m}^T \mathbf{x}_n)^2}} \mathbf{x}_n. \end{aligned} \tag{7.47}$$

Before we estimate k from the log-likelihood (7.28), we derive the following property:

$$\frac{\partial}{\partial k} I_0(k) = -\frac{1}{\pi} \int_0^{\pi} e^{-k\sin\theta} \sin\theta \mathrm{d}\theta = -I_1(k). \tag{7.48}$$

The above property can be generalised as follows:

$$\frac{\partial^p}{\partial k^p} I_0(k) = (-1)^p \frac{1}{\pi} \int_0^{\pi} \sin^p\theta e^{-k\sin\theta} \mathrm{d}\theta = (-1)^p I_p(k). \tag{7.49}$$

The first-order derivative of the log-likelihood in (7.28) for the estimation of k is then calculated below:

$$\frac{\partial J(\mathbf{X}, \mathbf{m}, k)}{\partial k} = N \frac{I_{p-1}(k)}{I_{p-2}(k)} - \sum_{n=1}^{N} \sqrt{1-(\mathbf{m}^T \mathbf{x}_n)^2}. \tag{7.50}$$

Appendix 3

A Directional K-Means Algorithm

Assume that K is the number of clusters, $\mathscr{C}_i$, $i = 1, \ldots, K$ are the clusters, $\mathbf{m}_i$ are the cluster centres and $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_N\}$ is a p-D angular dataset lying on the half-unit p-D sphere. The original K-means [36] minimises the following non-directional error function:

$$Q = \sum_{n=1}^{N} \sum_{i=1}^{K} ||\mathbf{x}_n - \mathbf{m}_i||^2 \tag{7.51}$$

where $||\cdot||$ represents the Euclidean distance. Instead of using the squared Euclidean distance for the p-D Directional K-Means, we introduce the following distance function:

$$D_l(\mathbf{x}_n, \mathbf{m}_i) = \sqrt{1 - (\mathbf{m}_i^T \mathbf{x}_n)^2}. \tag{7.52}$$

The novel function D_l is similarly monotonic as the original distance but emphasises more on the contribution of points closer to the cluster centre. In addition, D_l is periodic with period π. The p-D Directional K-Means can thus be described as follows:

1. Randomly initialise K cluster centres $\mathbf{m}_i$, where $||\mathbf{m}_i|| = 1$.
2. Calculate the distance of all points $\mathbf{x}_n$ to the cluster centres $\mathbf{m}_i$, using D_l.
3. The points with minimum distance to the centres $\mathbf{m}_i$ form the new clusters $\mathscr{C}_i$.
4. The clusters $\mathscr{C}_i$ vote for their new centres $\mathbf{m}_i^+$. To avoid averaging mistakes with directional data, vector averaging is employed to ensure the validity of the addition. The resulting average is normalised to the half-unit p-D sphere:

$$\mathbf{m}_i^+ = \frac{1}{k_i} \sum_{\mathbf{x}_n \in k_i} \mathbf{x}_n \tag{7.53}$$

$$\mathbf{m}_i^+ \leftarrow \mathbf{m}_i^+ / ||\mathbf{m}_i^+|| \tag{7.54}$$

5. Repeat steps (2)–(4) until the means $\mathbf{m}_i$ have converged.

References

1. Araki, S., Sawada, H., Mukai, R., Makino, S.: Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors. Signal Process. **87**, 1833–1847 (2007)
2. Arberet, S., Gribonval, R., Bimbot, F.: A robust method to count and locate audio sources in a multichannel underdetermined mixture. IEEE Trans. Signal Process. **58**(1), 121–133 (2010)
3. Attias, H.: Independent factor analysis. Neural Comput. **11**(4), 803–851 (1999)

4. Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S.: Clustering on the unit hypersphere using von Mises-Fisher distributions. J. Mach. Learn. Res. **6**, 1345–1382 (2005)
5. Bentley, J.: Modelling circular data using a mixture of von Mises and uniform distributions. Simon Fraser University, MSc thesis (2006)
6. Bilmes, J.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden mixture models. Technical Report, Department of Electrical Engineering and Computer Science, U.C. Berkeley, California (1998)
7. Blumensath, T., Davies, M.: Sparse and shift-invariant representations of music. IEEE Trans. Audio Speech Lang. Process. **14**(1), 50–57 (2006)
8. Bofill, P., Zibulevsky, M.: Underdetermined blind source separation using sparse representations. Signal Process. **81**(11), 2353–2362 (2001)
9. Cardoso, J.F.: Blind signal separation: statistical principles. Proc. IEEE **9**(10), 2009–2025 (1998)
10. Cemgil, A., Févotte, C., Godsill, S.: Variational and stochastic inference for bayesian source separation. Digit. Signal Process. **17**, 891913 (2007)
11. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing. Wiley, New York (2002)
12. Comon, P., Jutten, C.: Handbook of Blind Source Separation: Independent Component Analysis and Applications, 856 p. Academic Press, Waltham (2010)
13. Daudet, L., Sandler, M.: MDCT analysis of sinusoids: explicit results and applications to coding artifacts reduction. IEEE Trans. Speech Audio Process. **12**(3), 302–312 (2004)
14. Davies, M., Mitianoudis, N.: A simple mixture model for sparse overcomplete ICA. IEE Proc. Vis. Image Signal Process. **151**(1), 35–43 (2004)
15. Davies, M., Daudet, L.: Sparse audio representations using the mclt. Signal Process. **86**(3), 358–368 (2006)
16. Dempster, A.P., Laird, N., Rubin, D.: Maximum likelihood for incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B **39**, 1–38 (1977)
17. Devroye, L.: Non-Uniform Random Variate Generation. Springer, New York (1986)
18. Dhillon, I., Sra, S.: Modeling data using directional distributions. Technical Report TR-03-06, University of Texas at Austin, Austin, TX (2003)
19. Duarte, L., Suyama, R., Rivet, B., Attux, R., Romano, J., Jutten, C.: Blind compensation of nonlinear distortions: application to source separation of post-nonlinear mixtures. IEEE Trans. Signal Process. **60**(11), 5832–5844 (2012)
20. Duong, N., Vincent, E., Gribonval, R.: Under-determined reverberant audio source separation using a full-rank spatial covariance model. IEEE Trans. Audio Speech Lang. Process. **18**(7), 1830–1840 (2010)
21. Eriksson, J., Koivunen, V.: Identifiability, separability, and uniqueness of linear ica models. IEEE Signal Process. Lett. **11**(7), 601–604 (2004)
22. Févotte, C., Godsill, S.: A bayesian approach to blind separation of sparse sources. IEEE Trans. Audio Speech Lang. Process. **14**(6), 2174–2188 (2006)
23. Févotte, C., Gribonval, R., Vincent, E.: BSS EVAL toolbox user guide. Techical Report, IRISA Technical Report 1706, Rennes, France, April 2005, http://www.irisa.fr/metiss/bsseval/ (2005)
24. Fisher, N.: Statistical Analysis of Circular Data. Cambridge University Press, Cambridge (1993)
25. Girolami, M.: A variational method for learning sparse and overcomplete representations. Neural Comput. **13**(11), 2517–2532 (2001)
26. Gribonval, R., Nielsen, M.: Sparse decomposition in unions of bases. IEEE Trans. Inf. Theory **49**(12), 3320–3325 (2003)
27. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York, 481+xxii p. http://www.cis.hut.fi/projects/ica/book/ (2001)
28. Hyvärinen, A.: Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood. Neurocomputing **22**, 49–67 (1998)
29. Jammalamadaka, S., Sengupta, A.: Topics in Circular Statistics. World Scientific, Singapore (2001)
30. Jutten, C., Karhunen, J.: Advances in nonlinear blind source separation. In: Proceedings of 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp. 245–256. Nara, Japan (2003)

31. Kreyszig, E.: Advanced Engineering Mathematics, 1264 p. Wiley (2010)
32. Lee, T.W., Bell, A.J., Lambert, R.: Blind separation of delayed and convolved sources. In: Advances in Neural Information Processing Systems (NIPS), vol. 9, pp. 758–764 (1997)
33. Lee, T.W., Lewicki, M., Girolami, M., Sejnowski, T.: Blind source separation of more sources than mixtures using overcomplete representations. IEEE Signal Process. Lett. **4**(5), 87–90 (1999)
34. Lewicki, M., Sejnowski, T.: Learning overcomplete representations. Neural Comput. **12**, 337–365 (2000)
35. Lewicki, M.: Efficient coding of natural sounds. Nat. Neurosci. **5**(4), 356–363 (2002)
36. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297. Berkeley, California (1967)
37. Mardia, K., Kanti, V., Jupp, P.: Directional Statistics. Wiley, Chichester (1999)
38. Mitianoudis, N., Davies, M.: Permutation alignment for frequency domain ICA using subspace beamforming methods. In: Proceedings of the International Workshop on Independent Component Analysis and Source Separation (ICA2004), pp. 127–132. Granada, Spain (2004)
39. Mitianoudis, N., Stathaki, T.: Underdetermined source separation using mixtures of warped Laplacians. In: International Conference on Independent Component Analysis and Source Separation (ICA). London, UK (2007)
40. Mitianoudis, N.: A directional Laplacian density for underdetermined audio source separation. In: 20th International Conference on Artificial Neural Networks (ICANN). Thessaloniki, Greece (2010)
41. Mitianoudis, N., Davies, M.: Audio source separation of convolutive mixtures. IEEE Trans. Audio Speech Process. **11**(5), 489–497 (2003)
42. Mitianoudis, N., Stathaki, T.: Batch and online underdetermined source separation using Laplacian mixture models. IEEE Trans. Audio Speech Lang. Process. **15**(6), 1818–1832 (2007)
43. Moulines, E., Cardoso, J.F., Gassiat, E.: Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97), pp. 3617–3620. Munich, Germany (1997)
44. O'Grady, P., Pearlmutter, B.: Hard-LOST: modified K-means for oriented lines. In: Proceedings of the Irish Signals and Systems Conference, pp. 247–252. Ireland (2004)
45. O'Grady, P., Pearlmutter, B.: Soft-LOST: EM on a mixture of oriented lines. In: Proceedings of the International Conference on Independent Component Analysis 2004, pp. 428–435. Granada, Spain (2004)
46. Pajunen, P., Hyvärinen, A., Karhunen, J.: Nonlinear blind source separation by self-organizing maps. In: Proceedings of the International Conference on Neural Information Processing, pp. 1207–1210. Hong Kong (1996)
47. Plumbley, M., Abdallah, S., Blumensath, T., Davies, M.: Sparse representations of polyphonic music. Signal Process. **86**(3), 417–431 (2006)
48. Rickard, S., Balan, R., Rosca, J.: Real-time time-frequency based blind source separation. In: Proceedings of the ICA2001, pp. 651–656. San Diego, CA (2001)
49. Sawada, H., Araki, S., Makino, S.: A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2007), pp. 139–142 (2007)
50. Sawada, H., Araki, S., Makino, S.: Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. IEEE Trans. Audio Speech Lang. Process. **19**(3), 516–527 (2011)
51. Sawada, H., Mukai, R., Araki, S., Makino, S.: A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Trans. Speech Audio Process. **12**(5), 75–87 (2004)
52. SiSEC 2008: Signal separation evaluation campaign. http://sisec2008.wiki.irisa.fr/tiki-index.php

53. SiSEC 2010: Signal separation evaluation campaign. http://sisec2010.wiki.irisa.fr/tiki-index.php
54. SiSEC 2011: Signal separation evaluation campaign. http://sisec.wiki.irisa.fr/tiki-index.php
55. Smaragdis, P., Boufounos, P.: Position and trajectory learning for microphone arrays. IEEE Trans. Audio Speech Lang. Process. **15**(1), 358–368 (2007)
56. Smaragdis, P.: Blind separation of convolved mixtures in the frequency domain. Neurocomputing **22**, 21–34 (1998)
57. Torkkola, K.: Blind separation of delayed and convolved sources. In: S. Haykin (ed.) Unsupervised Adaptive Filtering, vol. I, pp. 321–375. Wiley (2000)
58. Vincent, E., Arberet, S., Gribonval, R.: Underdetermined instantaneous audio source separation via local gaussian modeling. In: 8th International Conferences on Independent Component Analysis and Signal Separation (ICA), pp. 775–782. Paraty, Brazil (2009)
59. Vincent, E., Gribonval, R., Fevotte, C., Nesbit, A., Plumbley, M., Davies, M., Daudet, L.: BASS-dB: the blind audio source separation evaluation database. http://bass-db.gforge.inria.fr/BASS-dB/
60. Winter, S., Kellermann, W., Sawada, H., Makino, S.: MAP based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and L1-norm minimization. EURASIP J. Adv. Signal Process. **1**, 12 p (2007)
61. Yilmaz, O., Rickard, S.: Blind separation of speech mixtures via time-frequency masking. IEEE Trans. Signal Process. **52**(7), 1830–1847 (2004)
62. Zibulevsky, M., Kisilev, P., Zeevi, Y., Pearlmutter, B.: Blind source separation via multinode sparse representation. Adv. Neural Inf. Process. Syst. **14**, 1049–1056 (2002)

Chapter 8
Itakura-Saito Nonnegative Matrix Two-Dimensional Factorizations for Blind Single Channel Audio Separation

Bin Gao and Wai Lok Woo

Abstract A new blind single channel source separation method is presented. The proposed method does not require training knowledge and the separation system is based on nonuniform time-frequency (TF) analysis and feature extraction. Unlike conventional researches that concentrate on the use of spectrogram or its variants, we develop our separation algorithms using an alternative TF representation based on the gammatone filterbank. In particular, we show that the monaural mixed audio signal is considerably more separable in this nonuniform TF domain. We also provide the analysis of signal separability to verify this finding. In addition, we derive two new algorithms that extend the recently published Itakura-Saito nonnegative matrix factorization to the case of convolutive model for the nonstationary source signals. These formulations are based on the Quasi-EM framework and the Multiplicative Gradient Descent (MGD) rule, respectively. Experimental tests have been conducted which show that the proposed method is efficient in extracting the sources' spectral–temporal features that are characterized by large dynamic range of energy, and thus lead to significant improvement in source separation performance.

8.1 Introduction

The principal aim of blind source separation (BSS) is to extract the underlying source signals from only a set of observations. Due to the diverse promising and exciting applications, BSS has attracted a substantial amount of attention in both the academic field as well as the industry. During the last decade, tremendous developments have been achieved in the application of BSS, particularly in wireless

B. Gao · W. L. Woo (✉)
School of Electrical and Electronic Engineering, Newcastle University, England, UK
e-mail: lok.woo@ncl.ac.uk

B. Gao
e-mail: bin.gao@ncl.ac.uk

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_8,
© Springer-Verlag Berlin Heidelberg 2014

communication, medical signal processing, geophysical exploration, and image enhancement/recognition. The so-called cocktail-party problem within the BSS context refers to the phenomenon of extracting original voice signals of the speakers from the signals recorded from several microphones. Similar examples in the field of radio communication involve the observations that correspond to the outputs of several antenna elements in response to several transmitters that represent the original signals. In the analysis of medical signals, electroencephalography (EEG), magnetoencephalography (MEG), and electrocardiogram (ECG) data represent the observations and BSS is used as a signal processing tool to assist noninvasive medical diagnosis. BSS has also been applied to the data analysis in other areas such as telecommunication, finance, and seismology. Further evidence of these applications can be found in [1–6]. A review of the current literature shows that there are three main classifications of BSS. These include linear and nonlinear, instantaneous and convolutive, overcomplete and underdetermined. In the first classification, linear algorithms dominate the BSS research field due to its simplicity in analysis and its explicit separability. Linear BSS assumes that the mixture is represented by a linear combination of sources. Extension of BSS for solving nonlinear mixtures has also been introduced. This model takes nonlinear distorted signals into consideration and offers a more accurate representation of a realistic environment. In the second classification, when the observed signals consist of combinations of multiple time-delayed versions of the original sources and/or mixed signals themselves, the system is referred as the convolutive mixture. Otherwise, the absence of time delays results in the instantaneous mixture of observed signals. Finally, when the number of observed signals exceeds the number of sources, this refers to the overcomplete BSS. Conversely, when the number of observed signals is less than the number of sources, this becomes the underdetermined BSS.

In general and for many practical applications, the challenging case for source separation is when only one monaural recording is available. This leads to the single channel blind source separation (SCBSS) where the problem can be stated as one observation mixed with several unknown sources. In this work, we consider the case of two sources, namely

$$y(t) = x_1(t) + x_2(t) \tag{8.1}$$

where $t = 1, 2, \ldots, T$ denotes time index and the goal is to estimate the two sources $x_1(t)$ and $x_2(t)$ given only the observation signal $y(t)$. Unlike conventional assumption used in BSS where the sources are assumed to be statistical independent which is rather too restrictive, in this chapter, the sources are characterized as nonstationary processes with time-varying spectra [7]. This assumption is practically justified since most signals encountered in applications are nonstationary with time-varying spectra. Examples include speech, audio, EEG, stock market index, and seismic trace.

Solutions to SCBSS using nonnegative matrix factorization (NMF) [8] have recently gained popularity. They exploit an appropriate time-frequency (TF) analysis on the mono input recordings, yielding a TF representation that can be decomposed as

$$|\mathbf{Y}|^{.2} \approx \mathbf{DH} \tag{8.2}$$

where $|\mathbf{Y}|^{.2} \in \Re_{+}^{F \times T_s}$ is the power TF representation of the mixture $y(t)$ which is factorized as the product of two nonnegative matrices , $\mathbf{D} \in \Re_{+}^{F \times I}$ and $\mathbf{H} \in \Re_{+}^{I \times T_s}$. The superscript '$\cdot$' represents element-wise operation. F and T_s represent the total frequency units and time slots in the TF domain, respectively. If I is chosen to be $I = T_s$, no benefit is achieved in terms of representation. Thus the idea is to determine $I < T_s$ so the matrix $\mathbf{D}$ can be compressed and reduced to its integral components so that it contains only a set of spectral basis vectors, and $\mathbf{H}$ is an encoding matrix that describes the amplitude of each basis vector at each time point. Because NMF gives a parts-based decomposition [8, 9], it has recently been proposed for separating drums from polyphonic music [10] and automatic transcription of polyphonic music [11]. Commonly used cost functions for NMF are the generalized Kullback-Leibler (KL) divergence and Least Square (LS) distance [8]. A sparseness constraint [12] can be added to these cost functions for optimizing $\mathbf{D}$ and $\mathbf{H}$. Other cost functions for audio spectrograms factorization have also been introduced such as that of [13] that assume multiplicative gamma-distributed noise in power spectrograms, while [14] attempts to incorporate phase into the factorization by using a probabilistic phase model. Notwithstanding the above, families of parameterized cost functions, such as the Beta divergence [15] and Csiszar's divergences [16], have also been presented for the source separation. However, they have some crucial limitations that explicitly use training knowledge of the sources [17]. As a consequence, these methods are only able to deal with a very specific set of signals and situations.

Model-based techniques have also been proposed for SCSS which usually require training a set of isolated recordings. The sources are trained by using a Hidden Markov model (HMM) based on Gaussian Mixture Model (GMM) and they are combined in a factorial HMM to separate the mixture [18]. Good separation requires detailed source models that might use thousands of full spectral states, e.g., in [19] HMMs with 8,000 states were required to accurately represent one person's speech for a source separation task. The large state space is required because it attempts to capture every possible instance of the signal. These model-based techniques, however, consume a long time not only in training the prior parameters but also presenting many difficult challenges during the inference stage.

From the above, it is clear that existing solutions to SCBSS are still practically limited and fall short of the success enjoyed in other areas of source separation. In this chapter, a novel separation system is proposed and the contributions are summarized as follows:

(i) A separability analysis in the TF domain for SCBSS and development a quantitative performance measure to evaluate the degree of "separateness" in the monaural mixed signal.
(ii) A separation framework based on the cochleagram. Unlike the spectrogram that deals only with uniform resolution, the gammatone filterbank produces nonuniform TF domain (termed as the cochleagram) whereby each TF unit has different resolution. We prove that the mixed signal is more separable in the cochleagram than the spectrogram and the log-frequency.

(iii) Development of two-dimensional NMF (NMF2D) signal model optimized under the Itakura-Saito (IS) divergence with Quasi-EM and MGD updates (IS-NMF2D). Two new algorithms have been developed to estimate the spectral and temporal features of the audio source model. The first algorithm is founded on the framework of Quasi-EM (Expectation-Maximization) while the second algorithm is based on the multiplicative gradient decent (MGD) update rule. Both algorithms have the unique property of scale-invariant whereby the lower energy components in the TF domain can be treated with equal importance as the higher energy components. This is to be contrasted with other methods based on LS distance [20] and KL divergence [21], which favor the high-energy components but neglect the low-energy components.

The chapter is organized as follows: Sect. 8.2 introduces the TF matrix representation using the gammatone filterbank. Section 8.3 delves into the separability analysis of the single-channel mixture in the nonuniform TF domain. In Sect. 8.4, the two new algorithms are derived and the proposed separation system is developed. Experimental results and a series of performance comparison with methods are presented in Sect. 8.5. Finally, Sect. 8.6 concludes the chapter.

8.2 Time-Frequency Representation

In the task of audio source separation, one critical decision is to choose a suitable TF domain to represent the time-varying contents of the signals. There are several types of TF representations and the most widely used ones are spectrogram [22] and log-frequency spectrogram (using constant-Q transform) [23]. This is documented over the last few years in the research of audio source separation [10–21]. In this work, however, we develop our separation algorithms using a TF representation based on the gammatone filterbank.

8.2.1 Gammatone Filterbank and Cochleagram

The Gammatone filterbank [24] is a cochlear filtering model which decomposes an input signal into the time-frequency domain using a set of gammatone filters. The specific steps of generate cochleagram are summarized as (Table 8.1).

In [25, 26], it was noted that some crucial differences exist in the TF representation of how sound is analyzed by the ear. In particular, the ear's frequency subbands get wider for higher frequencies, whereas the classical spectrogram as computed by the Short-Time Fourier Transform (STFT) has an equal-spaced bandwidth across all frequency channels. Since speech signals are characterized as highly nonstationary and nonperiodic whereas music changes continuously, therefore, application of the Fourier transform will produce errors when complicated transient phenomena such

Table 8.1 Cochleagram computation

1. Give impulse response of a gammatone filter:
   $g(f,t) = t^{h-1}e^{-2\pi vt}\cos(2\pi ft), \quad t \geq 0$ (8.3)
2. The filter output response $x(c,t)$ can be expressed as:
   $x(c,t) = \int_{-\infty}^{\infty} x(\tau) g_{f_c}(t-\tau)\, d\tau$ (8.4)
3. The output of each filter channel is divided into time frames with 50 % overlap between consecutive frames
4. The time-frequency spectra of all the filter outputs are then constructed to form the cochleagram

as the mixture of speech and music is contained in the analyzed signal. Unlike the spectrogram, the log-frequency spectrogram possesses nonuniform TF resolution. However, it does not exactly match the nonlinear resolution of the cochlear since their center frequencies are distributed logarithmically along the frequency axis and all filters have constant-Q factor [23]. On a separate hand, the gammatone filters used in the cochlear model (3) are approximately *logarithmically* spaced with constant-Q for frequencies from $f_s/10$ to $f_s/2$ (f_s denotes the sampling frequency), and approximately *linearly* spaced for frequencies below $f_s/10$. Hence, this characteristic results in selective *nonuniform* resolution in the TF representation of the analyzed audio signal. Figure 8.1 shows the frequency response of a general gammatone filterbank for $f_s = 16\,\text{kHz}$. It is seen that the higher frequencies correspond to the wider frequency subbands which resemble closely to the human perception of frequencies [27]. Therefore, the cochleagram is developed as an alternative TF analysis tool for source separation to overcome the limitations associated with the Fourier transform approach.

8.3 Single Channel Source Separability Analysis

For separation, one generates the TF mask corresponding to each source and applies the generated mask to the mixture to obtain the estimated source TF representation. In particular, when the sources do not overlap in the TF domain, an optimum mask $M_i^{\text{opt}}(f,t_s)$ exists which allows one to extract the ith original source from the mixture as

$$X_i(f,t_s) = M_i^{\text{opt}}(f,t_s)Y(f,t_s) \tag{8.5}$$

Given any TF mask $M_i(f,t_s)$ such that $0 \leq M_i(f,t_s) \leq 1$ for all (f,t_s), we define the separability for the target source $x_i(t)$ in the presence of the interfering sources $p_i(t) = \sum_{j=1, j\neq i}^{N} x_j(t)$ as

$$S_{M_i}^{Y\to X_i,P_i} = \frac{\|M_i(f,t_s)X_i(f,t_s)\|_F^2}{\|X_i(f,t_s)\|_F^2} - \frac{\|M_i(f,t_s)P_i(f,t_s)\|_F^2}{\|X_i(f,t_s)\|_F^2} \tag{8.6}$$

where $X_i(f, t_s)$ and $P_i(f, t_s)$ are the TF representations of $x_i(t)$ and $p_i(t)$, respectively. $\|\cdot\|_F$ is the Frobenius norm. We also define the separability of the mixture with respect to all the N sources as:

$$S_{M_1,\ldots,M_N}^{Y\to X_1,\ldots,X_N} = \frac{1}{N}\sum_{i=1}^{N} S_{M_i}^{Y\to X_i,P_i} \tag{8.7}$$

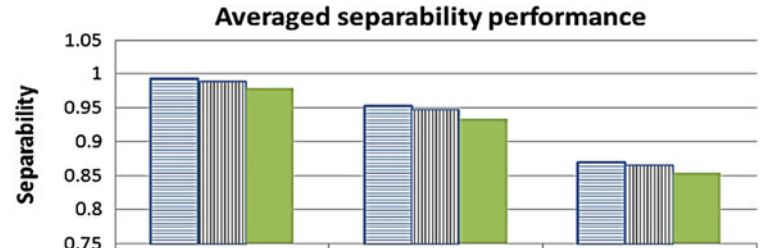
Equation (8.6) is equivalent to measuring the success of extracting the ith source $X_i(f, t_s)$ from the mixture $Y(f, t_s)$ given the TF mask $M_i(f, t_s)$. Similarly, (8.7) measures the success of extracting all the N sources simultaneously from the mixture. To further analyze the separability, we invoke the following: (i) Preserved signal ratio (PSR) that determines how well the mask preserves the source of interest and (ii) Signal-to-interference ratio (SIR) that indicates how well the mask suppresses the interfering sources:

$$\begin{aligned} PSR_{M_i}^{X_i} &= \frac{\|M_i(f,t_s)X_i(f,t_s)\|_F^2}{\|X_i(f,t_s)\|_F^2} \\ SIR_{M_i}^{X_i} &= \frac{\|M_i(f,t_s)X_i(f,t_s)\|_F^2}{\|M_i(f,t_s)P_i(f,t_s)\|_F^2} \end{aligned} \tag{8.8}$$

Using (8.8), it can be shown that (8.7) can be expressed as $S_{M_i}^{Y\to X_i,P_i} = PSR_{M_i}^{X_i} - PSR_{M_i}^{X_i}/SIR_{M_i}^{X_i}$. Analyzing the terms in (8.6), we have

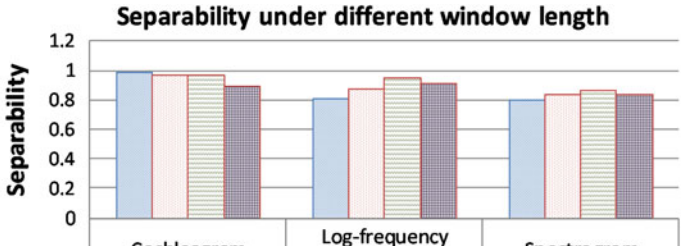
$$\begin{aligned} PSR_{M_i}^{X_i} &:= \begin{cases} 1, & \text{if } \operatorname{sup} pM_i^{\text{opt}} = \operatorname{sup} pM_i \\ <1, & \text{if } \operatorname{sup} pM_i^{\text{opt}} \subset \operatorname{sup} pM_i \end{cases} \\ SIR_{M_i}^{X_i} &:= \begin{cases} \infty, & \text{if } \operatorname{sup} p\,[M_iX_i] \cap \operatorname{sup} pP_i = \oslash \\ \text{finite}, & \text{if } \operatorname{sup} p\,[M_iX_i] \cap \operatorname{sup} pP_i \neq \oslash \end{cases} \end{aligned} \tag{8.9}$$

where 'supp' denotes the support. When $S_{M_i}^{Y\to X_i,P_i} = 1$ (i.e. $PSR_{M_i}^{X_i} = 1$ and $SIR_{M_i}^{X_i} = \infty$), this indicates that the mixture $y(t)$ is separable with respect to the i^{th} source $x_i(t)$. In other words, $X_i(f, t_s)$ does not overlap with $P_i(f, t_s)$ and the TF mask $M_i(f, t_s)$ has perfectly separated the i^{th} source $X_i(f, t_s)$ from the mixture $Y(f, t_s)$. This corresponds to $M_i(f, t_s) = M_i^{opt}(f, t_s)$ in (8.5). Hence, this is the maximum attainable $S_{M_i}^{Y\to X_i,P_i}$ value. For other cases of $PSR_{M_i}^{X_i}$ and $SIR_{M_i}^{X_i}$, we have $S_{M_i}^{Y\to X_i,P_i} < 1$. Using this concept, we can extend the analysis for the case of separating N sources. A mixture is fully separable to all the N sources if and only if $S_{M_1,\ldots,M_N}^{Y\to X_1,\ldots,X_N} = 1$ in (8.7). For the case $S_{M_1,\ldots,M_N}^{Y\to X_1,\ldots,X_N} < 1$, this implies that some of the sources overlap with each other in the TF domain and therefore, they cannot be fully separated. Thus, $S_{M_1,\ldots,M_N}^{Y\to X_1,\ldots,X_N}$ provides the quantitative performance measure for evaluating how separable is the mixture in the TF domain. In our comparison, the following TF representations are used to test the mixture's separability: spectrogram, log-frequency spectrogram, and cochleagram. In the log-frequency spectrogram, the frequency scale is set to logarithmic and grouped into 175 frequency bins in the



	Cochleagram	Log-frequency spectrogram	Spectrogram
M and M	0.993	0.953	0.869
M and S	0.989	0.947	0.865
S and S	0.979	0.934	0.854

Fig. 8.1 Averaged separability performance



	Cochleagram	Log-frequency spectrogram	Spectrogram
20ms	0.985	0.813	0.801
32ms	0.972	0.874	0.834
64ms	0.965	0.948	0.864
128ms	0.892	0.912	0.842

Fig. 8.2 Separability under different window length

range of 50–8 kHz with 24 bins per octave while the bandwidth follows the constant-Q rule [23]. To ensure fair comparison, we generate the ideal binary mask (IBM) [27] directly from the original sources. To reiterate our aim, the separability analysis is undertaken without recourse to any separation algorithms but utilizing only the energy of the sources to ascertain the degree of "separateness" of the mixture in different TF domains. These results have been tabulated in Fig. 8.1. The symbols 'M' and 'S' denotes music and speech, respectively.

In Fig. 8.1, three types of mixture have been used: (i) music mixed with music, (ii) speech mixed with music, and (iii) speech mixed with speech. The speech signals are selected from 10 male and 10 female speeches taken from TIMIT database and are normalized to unit energy. The 10 music sources are selected from the RWC database [28] and also normalized to unit energy. Two sources are randomly chosen from the databases and the mixed signal is generated by adding the sources. All mixed signals

are sampled at 16 kHz sampling rate. TF representation using different window length has also been investigated and the results are tabulated in Fig. 8.2.

Figure 8.2 shows the average separability results for all types of the mixture based on different window length. The bracketed number shows the number of data points corresponding to the particular window length. It is clear that, for both spectrogram and log-frequency spectrogram settings, the STFT with 1024-point window length is the best setting to analyze the separability performance. The results of PSR, SIR, and separability for each TF domain are obtained by averaging over 300 realizations. Following the listening performance test proposed in [29], we conclude that $S_{M_i}^{Y\to X_i,P_i} > 0.8$ leads to acceptable separation performance. Therefore, all TF representations satisfy this condition. While this is true, the spectrogram gives only a mediocre level of separability with averaged $S_{M_1,M_2}^{Y\to X_1,X_2} \approx 0.86$ while the log-frequency spectrogram shows a better result with $S_{M_1,M_2}^{Y\to X_1,X_2} \approx 0.94$. Nevertheless, the cochleagram yields the best separability with $S_{M_1,M_2}^{Y\to X_1,X_2} \approx 0.98$. Notwithstanding this, it is also seen that the average SIR of the cochleagram exhibits a much higher value than those of spectrogram and log-frequency spectrogram. This implies that the amount of interference between any two sources is lesser in the cochleagram.

8.4 The Proposed Method

In this section, two new algorithms are developed, namely the *Quasi-EM IS-NMF2D* and the *MGD IS-NMF2D*. The former algorithm optimizes the parameters of the signal model using the Expectation-Maximization approach, whereas the latter is directly based on the multiplicative gradient descent. To facilitate the derivation of these algorithms, we first consider the signal model in terms of the power TF representation

8.4.1 Signal Models

Since the sources have time-varying spectra, it is befitting to adopt a model whose power spectra can be described separately in terms of time and frequency. Although conventional NMF model can still be used, it will need a large number of spectral components and requires a clustering step to group and assign each spectral component to the appropriate source. As a result, the NMF model may not always yield the optimal results. An alternative model is to use the two-dimensional NMF model (NMF2D) [2, 3, 30, 31]. This model extends the basic NMF to be a two-dimensional convolution of $\mathbf{D}$ and $\mathbf{H}$ i.e. $|\mathbf{Y}|^{\cdot 2} \approx \sum_{\tau,\phi} \overset{\downarrow\phi}{\mathbf{D}}{}^{\tau} \overset{\to\tau}{\mathbf{H}}{}^{\phi}$ where the vertical arrow in $\overset{\downarrow\phi}{\mathbf{D}}{}^{\tau}$ denotes the downward shift that moves each

element in the matrix down by ϕ rows, and the horizontal arrow in $\overrightarrow{\mathbf{H}^{\phi}}^{\tau}$ denotes the right shift operator that moves each element in the matrix to the right by τ columns. In scalar representation, the (f, t_s)th element in $|\mathbf{Y}|^{.2}$ is given by $\left|\mathbf{Y}_{f,t_s}\right|^2 \approx \sum_{i=1}^{I}\sum_{\tau=0}^{\tau_{\max}}\sum_{\phi=0}^{\phi_{\max}}\mathbf{D}_{f-\phi,i}^{\tau}\mathbf{H}_{i,t_s-\tau}^{\phi}$ where $\mathbf{D}_{f',i'}^{\tau'}$ is the $\left(f', \tau', i'\right)$th element of $\mathbf{D}$ and $\mathbf{H}_{i',t_s'}^{\phi'}$ is the $\left(i', \phi', t_s'\right)$th element of $\mathbf{H}$. In source separation, this model compactly represents the characteristics of the nonstationary sources by a time-frequency profile convolved in both time and frequency by a time-frequency weight matrix. $\mathbf{D}_i^{\tau}$ represents the spectral basis of ith source in the TF domain and $\mathbf{H}_i^{\phi}$ represents the corresponding temporal code for each spectral basis.

The TF representation of the mixture in (8.1) is given by $Y(f, t_s) = X_1(f, t_s) + X_2(f, t_s)$ where $Y(f, t_s)$, $X_1(f, t_s)$ and $X_2(f, t_s)$ denote the TF components that are obtained by applying the gammatone filterbank to the mixture. The time slots are given by $t_s = 1, 2, \ldots, T_s$ while frequencies by $f = 1, 2, \ldots, F$. Since each component is a function of t_s and f, we represent this as a $F \times T_s$ matrix $\mathbf{Y} = [Y(f, t_s)]_{t_s=1,2,\ldots,T_s}^{f=1,2,\ldots,F}$ and $\mathbf{X}_i = [X_i(f, t_s)]_{t_s=1,2,\ldots,T_s}^{f=1,2,\ldots,F}$. It is shown in Sect. 8.3 that the sources are almost perfectly separable in the cochleagram. This therefore enable us to express the power TF representation as $|\mathbf{Y}|^{.2} \approx \sum_{i=1}^{I}|\mathbf{X}_i|^{.2}$ which we will model as $\left|\mathbf{Y}_{f,t_s}\right|^2 \approx \sum_{i=1}^{I}\sum_{\tau=0}^{\tau_{\max}}\sum_{\phi=0}^{\phi_{\max}}\mathbf{D}_{f-\phi,i}^{\tau}\mathbf{H}_{i,t_s-\tau}^{\phi}$. The source we seek to determine are $\left\{|X_i(f, t_s)|^{.2}\right\}_{i=1}^{I}$ and this will be obtained by using the matrix factorization as $\left|\tilde{X}_i(f, t_s)\right|^{.2} = \sum_{\tau=0}^{\tau_{\max}}\sum_{\phi=0}^{\phi_{\max}}\mathbf{D}_{f-\phi,i}^{\tau}\mathbf{H}_{i,t_s-\tau}^{\phi}$. In the following, we propose two novel algorithms to estimate $\mathbf{D}_{f,i}^{\tau}$ and $\mathbf{H}_{i,t_s}^{\phi}$ from the mixture signal.

8.4.2 Algorithm 1: Quasi-EM Formulation of IS-NMF2D (Quasi-EM IS-NMF2D)

We consider the following generative model defined as:

$$\mathbf{y}_{t_s} = \sum_{k=1}^{K}\mathbf{c}_{k,t_s}, \forall t_s = 1, \ldots, T_s \mathbf{c}_{k,t_s} = \left[c_{k,1,t_s}, \ldots, c_{F,1,t_s}\right]^{\mathbf{T}}$$

$$c_{k,f,t_s} \sim N_c\left(0\,,\ \sum_{\tau,\phi}\mathbf{H}_{k,t_s-\tau}^{\phi}\mathbf{D}_{f-\phi,k}^{\tau}\right) \tag{8.10}$$

where $\mathbf{y}_{t_s} \in C^{F\times 1}, \mathbf{c}_{k,t_s} \in C^{F\times 1}$ and $N_c\,(u,\,\Sigma)$ denotes the proper complex Gaussian distribution and the components $\mathbf{c}_{1,t_s}, \ldots, \mathbf{c}_{K,t_s}$ are both mutually and individually independent. The Expectation-Maximization (EM) framework is developed for the ML estimation of $\theta = \left\{\mathbf{D}^{\tau}, \mathbf{H}^{\phi}\right\}$. Due to the additive structure of the generative

model (8.10), the parameters describing each component $\mathbf{C}_k = \left[\mathbf{c}_{k,1,} \ldots, \mathbf{c}_{k,T_s}\right]$ can be updated separately. We now consider a partition of the parameter space $\theta = \bigcup_{k=1}^{K} \theta_k$ as $\theta_k = \left\{\mathbf{D}_k^{\tau}, \mathbf{H}_k^{\phi}\right\}$ where $\mathbf{D}_k^{\tau}$ is the kth column of $\mathbf{D}^{\tau}$ and $\mathbf{H}_k^{\phi}$ is the k^{th} row of $\mathbf{H}^{\phi}$. The EM algorithm works by formulating the conditional expectation of the negative log likelihood of $\mathbf{C}_k$ as

$$Q_k^{ML}\left(\theta_k|\theta'\right) = -\int_{\mathbf{C}_k} p\left(\mathbf{C}_k|\mathbf{Y}, \theta'\right) \log p(\mathbf{C}_k|\theta_k)\, d\mathbf{C}_k \tag{8.11}$$

where θ' always contains the most recent parameter values of $\left\{\mathbf{D}^{\tau}, \mathbf{H}^{\phi}\right\}$.

8.4.2.1 Expressions of the E- and M-step

One iteration of the EM algorithm includes computing the E-step and maximizing the M-step $Q_k^{ML}\left(\theta_k|\theta'\right)$ for $k = 1, \ldots, K$. The minus hidden-data log likelihood is defined as

$$\begin{aligned} -\log p\left(\mathbf{C}_k|\theta_k\right) &= -\sum_{t_s=1}^{T_s}\sum_{f=1}^{F} \log N_c\left(c_{k,f,t_s} \,\middle|\, 0,\ \sum_{\tau,\phi} \mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}\right) \\ &\doteq \sum_{t_s=1}^{T_s}\sum_{f=1}^{F} \log\left(\sum_{\tau,\phi} \mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}\right) + \frac{\left|c_{k,f,t_s}\right|^2}{\sum\limits_{\tau,\phi} \mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}} \end{aligned} \tag{8.12}$$

where '$\doteq$' in the second line denotes equality up to constant terms. Then, by virtue of (10), the hidden-data posterior also has a Gaussian form as $p\left(\mathbf{C}_k|\mathbf{Y}, \theta\right) = \prod\limits_{t_s=1}^{T_s}\prod\limits_{f=1}^{F} N_c\left(c_{k,f,t_s} \left| u_{k,f,t_s}^{post}\lambda_{k,f,t_s}^{post}\right.\right)$ where u_{k,f,t_s}^{post} and λ_{k,f,t_s}^{post} are the posterior mean and variance of c_{k,f,t_s} given as:

$$u_{k,f,t_s}^{post} = \frac{\sum\limits_{\tau,\phi} \mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}}{\sum\limits_{\tau,\phi,l} \mathbf{D}_{f-\phi,l}^{\tau}\mathbf{H}_{l,t_s-\tau}^{\phi}}\mathbf{Y}_{f,t_s}$$

$$\lambda_{k,f,t_s}^{post} = \frac{\sum\limits_{\tau,\phi} \mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}}{\sum\limits_{\tau,\phi,l} \mathbf{D}_{f-\phi,l}^{\tau}\mathbf{H}_{l,t_s-\tau}^{\phi}} \sum_{\tau,\phi,l\neq k} \mathbf{D}_{f-\phi,l}^{\tau}\mathbf{H}_{l,t_s-\tau}^{\phi} \tag{8.13}$$

Thus, the E-step merely includes computing the posterior power $\mathbf{V}_k$ of component $\mathbf{C}_k$, defined as $[\mathbf{V}_k]_{f,t_s} = v_{k,f,t_s} = \left|u_{k,f,t_s}^{post}\right|^2 + \lambda_{k,f,t_s}^{post}$. The M-step can be treated as one-component NMF problem:

$$\begin{aligned} Q_k^{ML}\left(\theta_k|\theta'\right) &\doteq \sum_{t_s=1}^{T_s}\sum_{f=1}^{F}\log\left(\sum_{\tau,\phi}\mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}\right) + \frac{\left|u_{k,f,t_s}^{post'}\right|^2 + \lambda_{k,f,t_s}^{post'}}{\sum_{\tau,\phi}\mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}} \\ &\doteq \sum_{t_s=1}^{T_s}\sum_{f=1}^{F} d_{IS}\left(|u_{k,f,t_s}^{post'}|^2 + \lambda_{k,f,t_s}^{post'} \;\middle|\; \sum_{\tau,\phi}\mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}\right) \end{aligned} \tag{8.14}$$

where $d_{IS}(\cdot|\cdot)$ is the IS divergence [32] and is formally defined as $d_{IS}(a|b) = (a/b) - \log(a/b) - 1$. The IS divergence has the property of scale invariant, i.e., $d_{IS}(\kappa\, a|\kappa\, b) = d_{IS}(a|b)$ for any κ. This implies that any low energy components (a, b)will bear the same relative importance as the high energy ones $(\kappa\, a, \kappa\, b)$. This is particularly important in situations where$|\mathbf{Y}|^{\cdot 2}$ is characterized by a large dynamic range such as the audio short-term spectra.

8.4.2.2 Estimation of the Spectral Basis and Temporal Code Using Quasi-EM Method

The spectral basis and temporal code can be obtained from (8.14). The derivative of a given element of $g_{k,f,t_s} = \sum_{\tau,\phi}\mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}$ with respect to $\mathbf{D}_{f,k}^{\tau}$ and $\mathbf{H}_{k,t_s}^{\phi}$ is given by:

$$\begin{aligned} \frac{\partial g_{k,f,t_s}}{\partial \mathbf{D}_{f',k'}^{\tau'}} &= \frac{\partial \sum_{\tau,\phi}\mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}}{\partial \mathbf{D}_{f',k'}^{\tau'}} = \mathbf{H}_{k',t_s-\tau'}^{f-f'} \\ \frac{\partial g_{k,f,t_s}}{\partial \mathbf{H}_{k',t_s'}^{\phi'}} &= \frac{\partial \sum_{\tau,\phi}\mathbf{D}_{f-\phi,k}^{\tau}\mathbf{H}_{k,t_s-\tau}^{\phi}}{\partial \mathbf{H}_{k',t_s'}^{\phi'}} = \mathbf{D}_{f-\phi',k'}^{t_s-t_s'} \end{aligned} \tag{8.15}$$

The derivatives of (8.14) corresponding to $\mathbf{D}_{f,k}^{\tau}$ and $\mathbf{H}_{k,t_s}^{\phi}$ is then obtained as

$$\begin{aligned} \frac{\partial Q_k^{ML}(\theta_k|\theta')}{\partial \mathbf{D}_{f',k'}^{\tau'}} &= \frac{\partial}{\partial \mathbf{D}_{f',k'}^{\tau'}}\sum_{f,t_s}\log\left(g_{k,f,t_s}\right) + \frac{v'_{k,f,t_s}}{g_{k,f,t_s}} \\ &= \sum_{\phi,t_s}\left(\frac{g_{k,f'+\phi,t_s} - v'_{k,f'+\phi,t_s}}{g_{k,f'+\phi,t_s}^2}\right)\mathbf{H}_{k',t_s-\tau'}^{\phi} \\ \frac{\partial Q_k^{ML}(\theta_k|\theta')}{\partial \mathbf{H}_{k',t_s'}^{\phi'}} &= \frac{\partial}{\partial \mathbf{H}_{k',t_s'}^{\phi'}}\sum_{f,t_s}\log\left(g_{k,f,t_s}\right) + \frac{v'_{k,f,t_s}}{g_{k,f,t_s}} \\ &= \sum_{\tau,f}\left(\frac{g_{k,f,t_s'+\tau} - v'_{k,f,t_s'+\tau}}{g_{k,f,t_s'+\tau}^2}\right)\mathbf{D}_{f-\phi',k'}^{\tau} \end{aligned} \tag{8.16}$$

Unlike the conventional EM algorithm, it is not possible to directly set $\partial Q_k^{ML}\left(\theta_k|\theta'\right)/\mathbf{D}_{f',k'}^{\tau'} = 0$ and $\partial Q_k^{ML}\left(\theta_k|\theta'\right)/\mathbf{H}_{k',t_s'}^{\phi'} = 0$ because of the nonlinear coupling between and via v'_{k,f,t_s}. Thus, closed-form expressions for estimating $\mathbf{D}_{f,k}^{\tau}$ and $\mathbf{H}_{k,t_s}^{\phi}$ cannot be accomplished. To overcome this problem, we use the following update rules and unify it as part of the M-step:

$$\theta_k \leftarrow \theta_k \cdot \left(\frac{\left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_-}{\left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_+}\right) \tag{8.17}$$

where $\nabla Q_k^{ML}\left(\theta_k|\theta'\right) = \left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_+ - \left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_-$. For each $\mathbf{D}_k^{\tau}$ and $\mathbf{H}_k^{\phi}$ variables, we have:

$$\begin{aligned}\left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_-^{\mathbf{D}} &= \sum_{\phi,t_s}\left(g_{k,f'+\phi,t_s}\right)^{-2} v'_{k,f'+\phi,t_s}\mathbf{H}_{k',t_s-\tau'}^{\phi} \\ \left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_+^{\mathbf{D}} &= \sum_{\phi,t_s}\left(g_{k,f'+\phi,t_s}\right)^{-1}\mathbf{H}_{k',t_s-\tau'}^{\phi}\end{aligned} \tag{8.18}$$

and

$$\begin{aligned}\left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_-^{\mathbf{H}} &= \sum_{\tau,f}\mathbf{D}_{f-\phi',k'}^{\tau}\left(g_{k,f,t_s'+\tau}\right)^{-2} v'_{k,f,t_s'+\tau} \\ \left[\nabla Q_k^{ML}\left(\theta_k|\theta'\right)\right]_+^{\mathbf{H}} &= \sum_{\tau,f}\mathbf{D}_{f-\phi',k'}^{\tau}\left(g_{k,f,t_s'+\tau}\right)^{-1}\end{aligned} \tag{8.19}$$

Inserting (8.18) and (8.19) into (8.17) leads to

$$\mathbf{D}_{f',k'}^{\tau'} \leftarrow \mathbf{D}_{f',k'}^{\tau'}\frac{\sum_{\phi,t_s}\left(g_{k,f'+\phi,t_s}\right)^{-2} v'_{k,f'+\phi,t_s}\mathbf{H}_{k',t_s-\tau'}^{\phi}}{\sum_{\phi,t_s}\left(g_{k,f'+\phi,t_s}\right)^{-1}\mathbf{H}_{k',t_s-\tau'}^{\phi}} \tag{8.20}$$

Similarly, the update rules in $\mathbf{H}_{k',t_s'}^{\phi'}$ writes

$$\mathbf{H}_{k',t_s'}^{\phi'} \leftarrow \mathbf{H}_{k',t_s'}^{\phi'}\frac{\sum_{\tau,f}\mathbf{D}_{f-\phi',k'}^{\tau}\left(g_{k,f,t_s'+\tau}\right)^{-2} v'_{k,f,t_s'+\tau}}{\sum_{\tau,f}\mathbf{D}_{f-\phi',k'}^{\tau}\left(g_{k,f,t_s'+\tau}\right)^{-1}} \tag{8.21}$$

It can be verified that the above update rules have an advantage of ensuring the nonnegativity constraints of $\mathbf{D}_{f,k}^{\tau}$ and $\mathbf{H}_{k,t_s}^{\phi}$ are always maintained during every iteration.

8.4.3 Algorithm 2: Multiplicative Gradient Descent Formulation of IS-NMF2D (MGD IS-NMF2D)

We consider the following generative model defined as:

$$\left|\mathbf{Y}_{f,t_s}\right|^2 = \left(\sum_{i=1}^{I}\sum_{\tau=0}^{\tau_{\max}}\sum_{\phi=0}^{\phi_{\max}}\mathbf{D}^{\tau}_{f-\phi,i}\mathbf{H}^{\phi}_{i,t_s-\tau}\right)\bullet\mathbf{E}_{f,t_s} \tag{8.22}$$

where $\mathbf{E}_{f,t_s}$ is a scalar of multiplicative independent and identically distributed (i.i.d.) Gamma noise with unit mean, i.e., $p(\mathbf{E}_{f,t_s}) = \xi(\mathbf{E}_{f,t_s}|\alpha,\beta)$ where $\xi(\mathbf{E}_{f,t_s}|\alpha,\beta)$ denotes the Gamma probability density function (pdf) defined as: $\xi(\mathbf{E}_{f,t_s}|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\mathbf{E}_{f,t_s}\right)^{\alpha-1}\exp\left(-\beta\mathbf{E}_{f,t_s}\right),\ \mathbf{E}_{f,t_s}\geq 0$. Next, we define $\mathbf{D} = \left[\mathbf{D}^1\mathbf{D}^2\cdots\mathbf{D}^{\tau_{\max}}\right]$ and $\mathbf{H} = \left[\mathbf{H}^1\mathbf{H}^2\cdots\mathbf{H}^{\phi_{\max}}\right]$. Under the independent and identically distributed (i.i.d.) noise assumption, the term $-\log p(\mathbf{Y}\,|\mathbf{D},\mathbf{H})$ becomes

$$-\log p\left(\mathbf{Y}\,|\mathbf{D},\mathbf{H}\right) = \frac{-\sum_{t_s=1}^{Ts}\sum_{f=1}^{F}\log\xi\left(\frac{|\mathbf{Y}|^{\cdot 2}_{f,t_s}}{\sum_{i=1}^{I}\sum_{\tau=0}^{\tau_{\max}}\sum_{\phi=0}^{\phi_{\max}}\mathbf{D}^{\tau}_{f-\phi,i}\mathbf{H}^{\phi}_{i,t_s-\tau}}\,\middle|\,\alpha,\beta\right)}{\sum_{i=1}^{I}\sum_{\tau=0}^{\tau_{\max}}\sum_{\phi=0}^{\phi_{\max}}\mathbf{D}^{\tau}_{f-\phi,i}\mathbf{H}^{\phi}_{i,t_s-\tau}}$$

$$\doteq d_{IS}\left(|\mathbf{Y}|^{\cdot 2}_{f,t_s}\,\middle|\,\sum_{i=1}^{I}\sum_{\tau=0}^{\tau_{\max}}\sum_{\phi=0}^{\phi_{\max}}\mathbf{D}^{\tau}_{f-\phi,i}\mathbf{H}^{\phi}_{i,t_s-\tau}\right) \tag{8.23}$$

where $\doteq$ in the second line denotes equality up to constant terms. Thus, the cost function is $C^{NMF2D}_{IS} = -\log p\left(\mathbf{Y}\,|\mathbf{D},\mathbf{H}\right)$. The derivatives of (23) corresponding to $\mathbf{D}^{\tau}$ and $\mathbf{H}^{\phi}$are given by

$$\begin{aligned}\frac{\partial C^{NMF2D}_{IS}}{\partial\mathbf{D}^{\tau'}_{f',i'}} &= \frac{\partial}{\partial\mathbf{D}^{\tau'}_{f',i'}}\sum_{f,t_s}\left(\frac{|\mathbf{Y}|^2_{f,t_s}}{\mathbf{Z}_{f,t_s}} - \log\frac{|\mathbf{Y}|^2_{f,t_s}}{\mathbf{Z}_{f,t_s}} - 1\right) \\ &= -\sum_{\phi,t_s}\left(\left(\mathbf{Z}_{f'+\phi,t_s}\right)^{-2}\left(|\mathbf{Y}|^2_{f'+\phi,t_s} - \mathbf{Z}_{f'+\phi,t_s}\right)\right)\mathbf{H}^{\phi}_{i',t_s-\tau'}\end{aligned} \tag{8.24}$$

$$\begin{aligned}\frac{\partial C^{NMF2D}_{IS}}{\partial\mathbf{H}^{\phi'}_{i',t'_s}} &= \sum_{f,t_s}\mathbf{D}^{t_s-t'_s}_{f-\phi',i'}\left(\left(\mathbf{Z}_{f,t_s}\right)^{-2}\left(\mathbf{Z}_{f,t_s} - |\mathbf{Y}|^2_{f,t_s}\right)\right) \\ &= -\sum_{\tau,f}\mathbf{D}^{\tau}_{f-\phi',i'}\left(\left(\mathbf{Z}_{f,t'_s+\tau}\right)^{-2}\left(|\mathbf{Y}|^2_{f,t'_s+\tau} - \mathbf{Z}_{f,t'_s+\tau}\right)\right)\end{aligned} \tag{8.25}$$

where $\mathbf{Z} = \sum_{\tau}\sum_{\phi} \overset{\downarrow\phi}{\mathbf{D}^{\tau}} \overset{\rightarrow\tau}{\mathbf{H}^{\phi}}$. The standard gradient decent approach gives

$$\mathbf{D}^{\tau'}_{f',i'} \leftarrow \mathbf{D}^{\tau'}_{f',i'} - \eta_D \frac{\partial Cost^{NMF2D}_{IS}}{\partial \mathbf{D}^{\tau'}_{f',i'}} \quad \text{and} \quad \mathbf{H}^{\phi'}_{i',t'_s} \leftarrow \mathbf{H}^{\phi'}_{i',t'_s} - \eta_H \frac{\partial Cost^{NMF2D}_{IS}}{\partial \mathbf{H}^{\phi'}_{i',t'_s}} \tag{8.26}$$

where η_D and η_H are positive learning rates and can be obtained as

$$\eta_D = \frac{\mathbf{D}^{\tau'}_{f',i'}}{\sum\limits_{\phi,t_s} \left(\mathbf{Z}_{f'+\phi,t_s}\right)^{-1} \mathbf{H}^{\phi}_{i',t_s-\tau'}} \quad \text{and} \quad \eta_H = \frac{\mathbf{H}^{\phi'}_{i',t'_s}}{\sum\limits_{\tau,f} \mathbf{D}^{\tau}_{f-\phi',i'} \left(\mathbf{Z}_{f,t'_s+\tau}\right)^{-1}} \tag{8.27}$$

Inserting (8.27) into (8.26) gives the multiplicative gradient decent rules

$$\mathbf{D}^{\tau'}_{f',i'} \leftarrow \mathbf{D}^{\tau'}_{f',i'} \frac{\sum\limits_{\phi,t_s} \left(\mathbf{Z}_{f'+\phi,t_s}\right)^{-2} |\mathbf{Y}|^2_{f'+\phi,t_s} \mathbf{H}^{\phi}_{i',t_s-\tau'}}{\sum\limits_{\phi,t_s} \left(\mathbf{Z}_{f'+\phi,t_s}\right)^{-1} \mathbf{H}^{\phi}_{i',t_s-\tau'}} \tag{8.28}$$

and

$$\mathbf{H}^{\phi'}_{i',t'_s} \leftarrow \mathbf{H}^{\phi'}_{i',t'_s} \frac{\sum\limits_{\phi,t_s} \left(\mathbf{Z}_{f,t'_s+\tau}\right)^{-2} |\mathbf{Y}|^2_{f,t'_s+\tau} \mathbf{D}^{\tau}_{f-\phi',i'}}{\sum\limits_{\tau,f} \mathbf{D}^{\tau}_{f-\phi',i'} \left(\mathbf{Z}_{f,t'_s+\tau}\right)^{-1}} \tag{8.29}$$

The key difference between both algorithms is that the Quasi-EM IS-NMF2D algorithm prevents zeros in the factors, i.e., $\mathbf{D}^{\tau}$ and $\mathbf{H}^{\phi}$ cannot take entries equal to zero. On the contrary, this is not a feature shared by the MGD IS-NMF2D algorithm since zero coefficients are invariant under MGD updates. If the MGD IS-NMF2D algorithm attains a fixed point solution with zero entries, then it cannot be determined since the limit point is a stationary point [33]. Consequently, the resulting factorizations rendered by these algorithms are not equivalent. For this reason, the Quasi-EM IS-NMF2D algorithm can be considered more reliable for updating $\mathbf{D}^{\tau}$ and $\mathbf{H}^{\phi}$. We have summarized both proposed algorithms in Table 8.2. Details of the source separation performance between these algorithms will be shown in Sect. 8.5 where $\psi = 10^{-6}$ is the threshold for ascertaining the convergence.

8.4.4 Estimation of Sources

The two matrices that we seek to separate from $\left|\mathbf{Y}_{f,t_s}\right|^2$ are $\left|\tilde{X}_1(f,t_s)\right|^{.2}$ and $\left|\tilde{X}_2(f,t_s)\right|^{.2}$. These matrices are estimated as $\left|\tilde{X}_1(f,t_s)\right|^{.2} = \sum\limits_{\tau=0}^{\tau_{\max}} \sum\limits_{\phi=0}^{\phi_{\max}} \mathbf{D}^{\tau}_{f-\phi,1} \mathbf{H}^{\phi}_{1,t_s-\tau}$

and $\left|\tilde{X}_2(f,t_s)\right|^{.2} = \sum_{\tau=0}^{\tau_{\max}} \sum_{\phi=0}^{\phi_{\max}} \mathbf{D}^{\tau}_{f-\phi,2} \mathbf{H}^{\phi}_{2,t_s-\tau}$ [29] which are then used to generate the binary mask as $\mathbf{mask}_i(f,t_s) = 1$ if $\left|\tilde{X}_i(f,t_s)\right|^{.2} > \left|\tilde{X}_j(f,t_s)\right|^{.2}$ and zero otherwise. Finally, the estimated time-domain sources are obtained as $\tilde{x}_i =$ Resynthesize $(\mathbf{mask}_i \cdot \mathbf{Y})$ for $i = 1, 2$ where $\tilde{x}_i = [\tilde{x}_i(1), \ldots, \tilde{x}_i(T)]^{\mathbf{T}}$ denotes the i^{th} estimated source. The time-domain estimated sources are resynthesized using the approach in [22] by weighting the mixture cochleagram by the mask and correcting phase shifts introduced during the gammatone filtering.

8.5 Experimental Results and Analysis

The proposed separation system is tested on recorded audio signals. All recordings and processing are conducted using a PC with Intel Core 2 CPU 6600 @ 2.4 GHz and 2 GB RAM. For mixture generation, three types of mixtures are used, i.e., mixture of music and speech, mixture of different kinds of music, and mixture of different kinds of speech. The speech sources (male and female) are selected from the TIMIT speech database while the music sources (jazz and piano) from the RWC database [28]. All mixtures are sampled at 16 kHz sampling rate. In all cases, the sources are mixed with equal average power over the duration of the signals. As for our proposed algorithms, the convolutive components are selected as follows:

(i) For jazz and speech mixture, $\tau = \{0, \ldots, 4\}$ and $\phi = \{0, \ldots, 4\}$.
(ii) For jazz and piano mixture, $\tau = \{0, \ldots, 6\}$ and $\phi = \{0, \ldots, 9\}$.
(iii) For piano and speech mixture, $\tau = \{0, \ldots, 6\}$ and $\phi = \{0, \ldots, 9\}$.
(iv) For speech and speech mixture, $\tau = \{0, 1\}$ and $\phi = \{0, 1, 2\}$.

These parameters are selected after conducting Monte Carlo tests over 100 realizations of audio mixture. We have evaluated our separation performance in terms of the Signal-to-Distortion ratio (SDR) that unifies the Signal-to-Interference ratio (SIR) and Signal-to-Artifacts ratio (SAR). MATLAB routines for computing these criteria are obtained from the SiSEC'08 webpage [34].

8.5.1 Separation Performance Under Different TF Representations

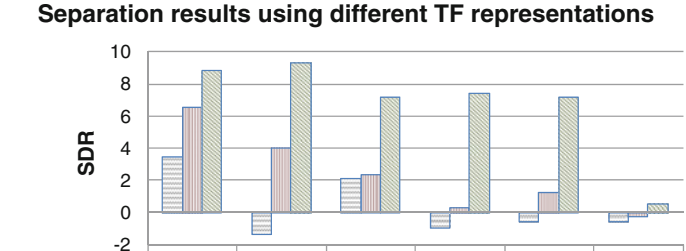
In Sect. 8.2, the separability analysis was undertaken by using the IBM to determine the "separateness" of the mixture without recourse to the separation algorithms. In this section, the impact of separation algorithm is analyzed. Instead of using the IBM, the Quasi-EM IS-NMF2D algorithm is now used to estimate the mask according to Sect. 8.4. In this situation, we are investigating the performance of mixture separation (rather than mixture separability). Speech signals and music are used to generate the

Table 8.2 Pseudo codes for Quasi-EM IS-NMF2D and IS-NMF2D (MGD) algorithms

Quasi-EM IS-NMF2D algorithm	MGD IS-NMF2D algorithm
Input: $\|\mathbf{Y}\|^{\cdot 2}$, random nonnegative matrix $\mathbf{D}^{\tau}$ and $\mathbf{H}^{\phi}$, ϕ, τ Output: $\mathbf{D}^{\tau}$ and $\mathbf{H}^{\phi}$	Input: $\|\mathbf{Y}\|^{\cdot 2}$, random nonnegative matrix $\mathbf{D}^{\tau}$ and $\mathbf{H}^{\phi}$, ϕ, τ Output: $\mathbf{D}^{\tau}$ and $\mathbf{H}^{\phi}$
Procedure: Compute initialize cost value $Cost(1)$ using (8.12)	Procedure: Compute initialize cost value $Cost(1)$ using (8.23)
for n=1: max number of iterations	for n=1: max number of iterations
for k=1:K	Compute $\mathbf{Z} = \sum_{\tau}\sum_{\phi} \mathbf{D}^{\tau}_{f-\phi}\mathbf{H}^{\phi}_{t_s-\tau}$
(E-step): Compute $v_{k,f,t_s} = \left\|u^{post}_{k,f,t_s}\right\|^2 + \lambda^{post}_{k,f,t_s}$ using (8.13)	• Update $\mathbf{D}^{\tau'}_{f',i'}$ using (8.28) for all τ, ϕ
(M-step): Iterate convergence is achieved	Normalize $\mathbf{D}^{\tau'}_{f',i'}$
• Update $\mathbf{D}^{\tau'}_{f',k'}$ using (8.20) for all τ, ϕ	Compute $\mathbf{Z} = \sum_{\tau}\sum_{\phi} \mathbf{D}^{\tau}_{f-\phi}\mathbf{H}^{\phi}_{t_s-\tau}$
Normalize $\mathbf{D}^{\tau'}_{f',k'}$	• Update $\mathbf{H}^{\phi'}_{i',t'_s}$ using (8.29) for all τ, ϕ
• Update $\mathbf{H}^{\phi'}_{k',t'_s}$ using (8.21) for all τ, ϕ	Normalize $\mathbf{H}^{\phi'}_{i',t'_s}$
Normalize $\mathbf{H}^{\phi'}_{k',t'_s}$	Compute cost value using (8.23)
end	end
end	Stopping criterion: $\frac{Cost(n-1)-Cost(n)}{Cost(n)} < \psi$
Stopping criterion: $\frac{Cost(n-1)-Cost(n)}{Cost(n)} < \psi$	

monoaural mixture recording. The separation performance is evaluated by using three types of TF representation: (i) spectrogram (STFT with 1024-point Hamming windowed FFT and 50 % overlap), (ii) log-frequency spectrogram (as described in Sect. 8.3 with 1024-point Hamming windowed FFT), and (iii) cochleagram based on Gammatone filterbank of 128 channels, filter order of 4 (i.e., $h = 4$ in (4)), and each filter output is divided into 20 ms time frames with 50 % overlap. To validate the parameters setting of cochleagram (e.g. h and v), we have constructed an experiment based on three speech sources and tested the result by fixing the parameter h in (3) to unity. The experiment is then repeated by progressively increasing h from 2 to 10. Over this range, the optimal separability is obtained when $h = 4$. The parameter v determines the rate of decay of the impulse response of the gammatone filters. In most audio processing tasks, it is set to $v(f) = 1.019ERB(f)$ where $ERB(f) = 24.7 + 0.108f$ is the equivalent rectangular bandwidth of the filter with the center frequency f. A range of values for v has been tested, i.e., $v(f) = (1.019 + c)ERB(f)$ where c ranges from -0.5 to 0.5 with increment of 0.1. The obtained result indicates that the optimal separability is obtained by setting $c = 0$. As c moves away from 0, the separability result progressively deteriorates. This confirms the validity of setting $v(f) = 1.019ERB(f)$ for the cochleagram.

where 'J', 'M', 'F', 'P', 'S' denote jazz, male speech, female speech, piano, and speech, respectively.



	J and M	J and F	P and M	P and F	J and P	S and S
spectrogram	3.47	-1.41	2.1	-1.01	-0.59	-0.6
log-frequency spectrogram	6.54	3.97	2.31	0.27	1.21	-0.3
cochleagram	8.87	9.34	7.16	7.44	7.21	0.5

Fig. 8.3 Separation results using different TF representations

Figure 8.3 shows the comparison of our proposed algorithm based on the spectrogram, log-frequency spectrogram, and cochleagram under various audio mixtures. The separation results for all mixture types based on the spectrogram gives an average SDR of 0.51 dB while the log-frequency spectrogram an average SDR of 2.8 dB. However, a significantly higher performance is attained by the cochleagram with an average SDR of 8 dB. This leads to a substantial improvement gain of 7.5 dB and 5.2 dB, respectively. The major reason for the large discrepancy is due to the mixing ambiguity between $|\mathbf{X}_1|^{\cdot 2}$ and $|\mathbf{X}_2|^{\cdot 2}$. The larger the mixing ambiguity between $|\mathbf{X}_1|^{\cdot 2}$ and $|\mathbf{X}_2|^{\cdot 2}$, the more TF units will be ambiguous which subsequently decreases the probability of correct assignment of each unit to the sources and eventually results in poorer separation performance. To validate this, Fig. 8.4 shows the spectrogram of the original sources, the mixed signal, and the estimated sources using the proposed Quasi-EM IS-NMF2D algorithm. Both figures indicate that the STFT lacks provision for further low-level information of a TF unit and therefore, the resulting spectrogram fails to infer the dominating source. This leads to high degree of ambiguity in TF domain and causes lack of uniqueness in extracting the spectral-temporal features of the sources

Similar to the above, Fig. 8.5 shows the separation results based on the log-frequency spectrogram. Compared with spectrogram, the separation performance is better since log-frequency spectrogram has the propensity of nonuniform time frequency resolution. However, the transform operation used by the log-frequency spectrogram is still based on the Fourier Transform which may not be an optimal option. On the other hand, the results of separation in the cochleagram have led to significant SDR improvement. The cochleagram enables the mixed signal to be more separable and thus reduces the mixing ambiguity between $|\mathbf{X}_1|^{\cdot 2}$ and $|\mathbf{X}_2|^{\cdot 2}$.

This explains the average performance of separating mixture jazz music and female utterance is the highest among all the mixtures because both sources have very distinguishable TF patterns in the cochleagram. This is evident in Fig. 8.6, which shows the separation results in the cochleagram. The plot clearly shows that the

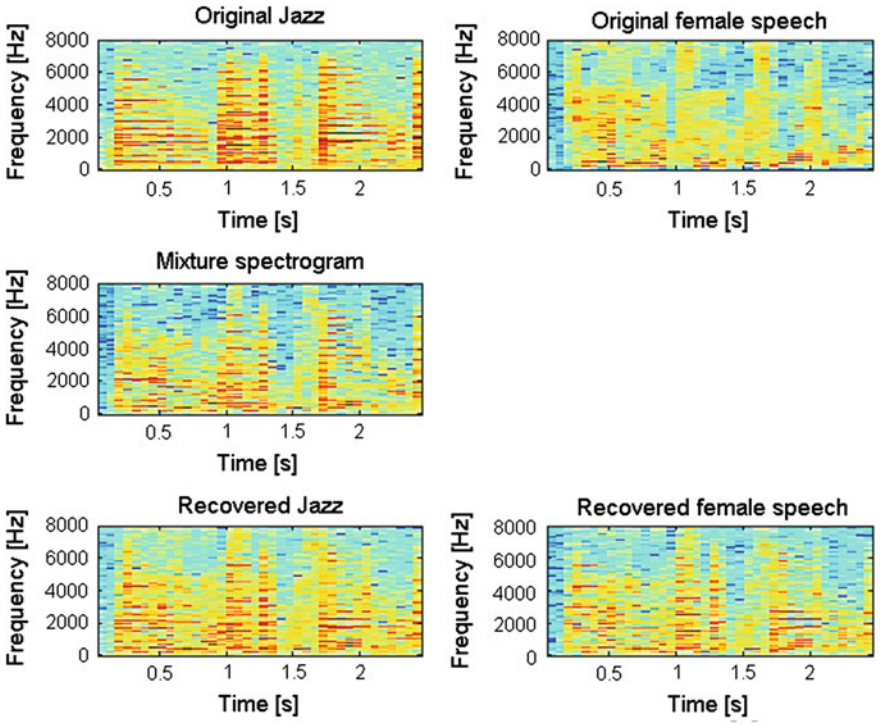


Fig. 8.4 Separation results in spectrogram

spectral energy of the two audio sources has been clustered at different frequencies in the cochleagram due to their different fundamental frequencies. These prominent features have been separated using our proposed Quasi-EM IS-NMF2D algorithm.

The performance of source separation also depends on how accurate the spectral bases are estimated. Given the different types of TF representation, a question arises as to which set of estimated spectral bases have yielded better approximation to the respective original sources' spectral bases. Figure 8.7 shows the results of the original and the estimated spectral basis $\mathbf{D}_i^{\tau}$ for the above mixture when the factorization is performed in the cochleagram. In Fig. 8.7, panels (a and b) refer to the original spectral bases of the jazz music and female utterance, respectively. Panels (c and d) refer to the estimated spectral bases. In comparison, we have also included similar factorization results of the same mixture in the spectrogram and log-frequency spectrogram. These are shown in Figs. 8.8 and 8.9, respectively. In sharp contrast with Fig. 8.7, it is noted that the estimated spectral bases in Figs. 8.8 and 8.9 are quite dissimilar to the original spectral bases. Thus, the construction of the separating mask will inevitably introduce errors in assigning the TF units to the respective sources. Therefore, the recovered sources are very coarse with very low values of SDR in Fig. 8.3.

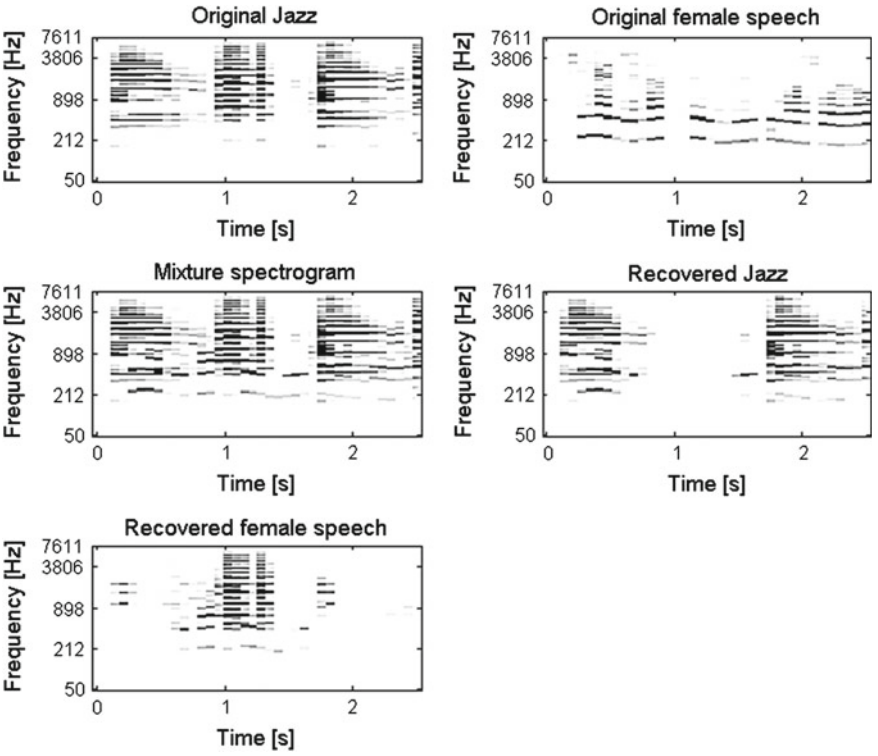


Fig. 8.5 Separation results in log-frequency spectrogram

8.5.2 Comparison Between Different Cost Functions

In the following, experiments are conducted to evaluate the efficiency of the proposed algorithm under different cost functions. Here, we consider the Least Square (LS) distance and Kullback-Leibler (KL) divergence. Figure 8.10 shows the separation results in the cochleagram based on LS, KL, and IS cost functions. In Fig. 8.10, it is noted that Quasi-EM IS-NMF2D algorithm outperforms those of LS distance and KL divergence with an average SDR of 3.1 and 1.8 dB, respectively. This is evidenced by the fact that the IS divergence holds a desirable property of scale invariant so that low energy components can be precisely estimated and they bear the same relative importance as the high energy ones. On the contrary, factorizations obtained with LS distance and KL divergence tend to favor the high energy components at the expense of disregarding the low energy ones. In the cochleagram, the dynamic range of the mixture signal can be considerably large such that the dominating signal at a particular TF unit can manifest either as low or high energy components. In addition, these components tend to exist as clusters. As such, when either LS distance- or KL divergence is used, clusters with low energy tend to be ignored in favor of the

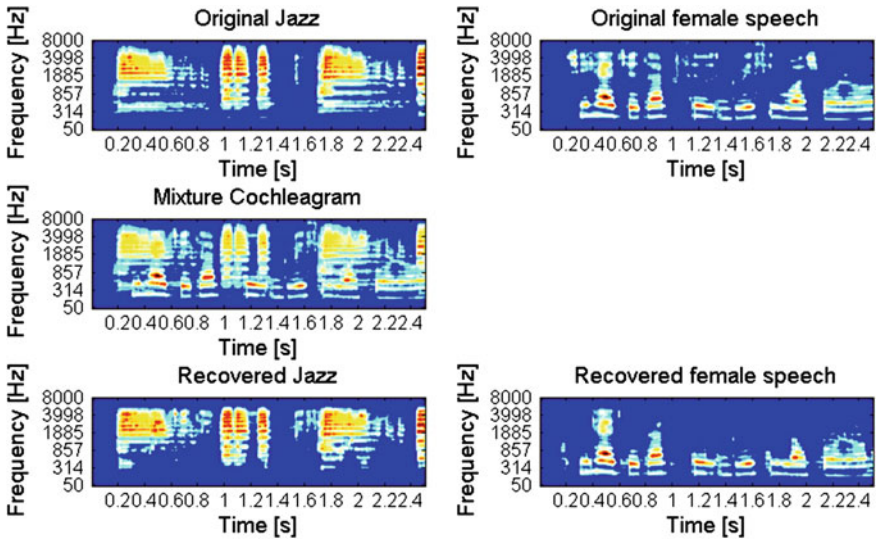


Fig. 8.6 Separation results in cochleagram

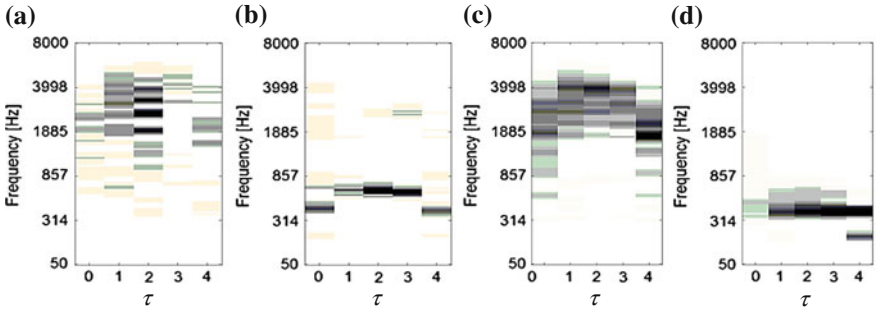


Fig. 8.7 **a–b** Original spectral bases of jazz music and female utterance in the cochleagram. **c–d** The corresponding estimated spectral bases

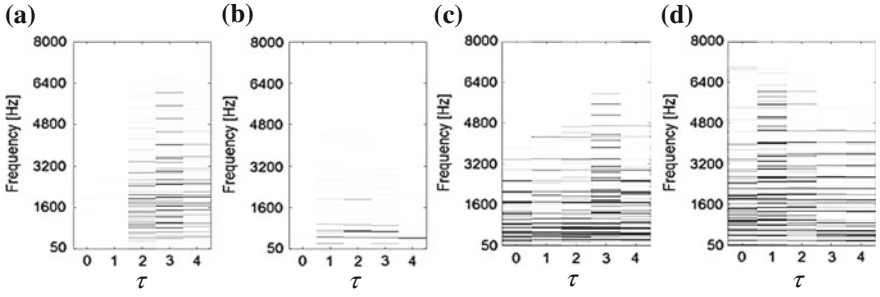


Fig. 8.8 **a–b** Original spectral bases of jazz music and female utterance in the spectrogram. **c–d** The corresponding estimated spectral bases

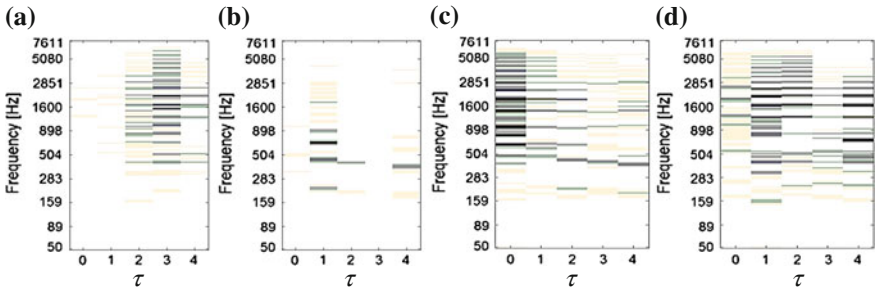
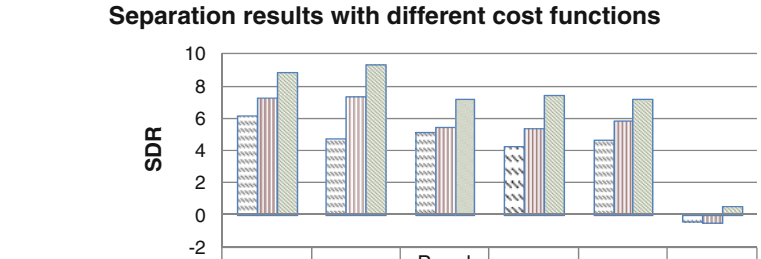


Fig. 8.9 **a**–**b** Original spectral bases of jazz music and female utterance in the log-frequency spectrogram. **c**–**d** The corresponding estimated spectral bases



	J and M	J and F	P and M	P and F	J and P	S and S
LS-NMF2D	6.15	4.69	5.11	4.21	4.61	-0.4
KL-NMF2D	7.24	7.35	5.42	5.38	5.86	-0.5
Quasi-EM IS-NMF2D	8.87	9.34	7.16	7.44	7.21	0.5

Fig. 8.10 Separation results with different cost functions

high energy ones. This leads to mixing ambiguities especially for low energy ones in which case when they are subsumed together leads to significant lost of spectral–temporal information of the sources. Figure 8.11 shows how different cost functions have impacted the separation performance. It is clearly seen that the LS-NMF2D algorithm fails to determine the correct TF components of each source. Panels (a and b) show a considerable level of mixing ambiguities (red box marked area) that have not been accurately resolved by the LS-NMF2D algorithm. The KL-NMF2D exhibits better performance but ignores some low energy TF components in the red box marked area of (c). On the other hand, the proposed algorithm has successfully extracted the low energy components for both female speech and jazz music with high accuracy.

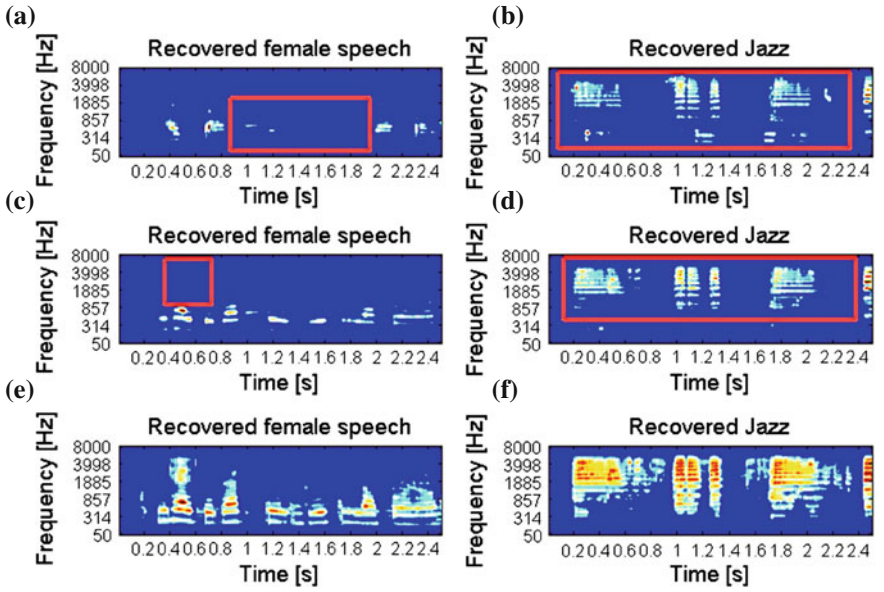


Fig. 8.11 Separation results: **a**–**b**, **c**–**d** and **e**–**f** denote the recovered female speech and jazz music in the cochleagram by using the algorithms with different cost function

8.5.3 Comparing with Different SCBSS Methods

We have made comparison with the recently published EMD SCBSS [35], which first decomposes the given signal into spectrally independent modes using EMD algorithm, and then, ICA is applied to extract statistically independent sources. All the above single channel BSS methods will be tested across all types of mixture and compared in terms of SDR. Table 8.3 summarizes the comparison results. In comparison, the Quasi-EM IS-NMF2D with cochleagram leads to the best separation performance for all types of the mixture. The EMD SCBSS also performs with relative acceptable results compared with Quasi-EM IS-NMF2D. However, it is interesting to point out that the advantage of using Quasi-EM IS-NMF2D with cochleagram is that this method is less complex than the EMD SCBSS and simultaneously it retains a higher level of the separation performance.

8.5.4 Separating More than Two Sources

The proposed method can be extended to the case when $i > 2$ sources. If more than two sources are mixed in a single channel, this requires specifying the number of sources to be separated. Since the method is blind, the separability of the complex

Table 8.3 Separation results using different SCBSS methods

Mixtures	Method	SDR
Jazz and male	EMD SCBSS	6.3
	Quasi-EM IS-NMF2D	8.8
Jazz and female	EMD SCBSS	5.2
	Quasi-EM IS-NMF2D	9.3
Piano and male	EMD SCBSS	5.2
	Quasi-EM IS-NMF2D	7.1
Piano and female	EMD SCBSS	6.6
	Quasi-EM IS-NMF2D	7.4
Jazz and piano	EMD SCBSS	6.6
	Quasi-EM IS-NMF2D	8.5
Speech and speech	EMD SCBSS	0.4
	Quasi-EM IS-NMF2D	0.5

mixture depends highly on how accurate the spectral bases $\mathbf{D}_i^{\tau}$ can be estimated from the TF mixture. Consequently, a set of distinguishable spectral basis of each source for a generic case is a necessary condition to achieve good separation performance. Thus, we adopt three different types of sources, e.g., jazz, piano, and trumpet to generate a complex mixture. The convolutive components in the proposed algorithm are selected as $\tau = \{0, \ldots, 3\}$ and $\phi = \{0, \ldots, 31\}$. Table 8.4 shows the overall separation results. It is seen that mixtures generated by all music sources have been recovered quite successfully. Figure 8.12 shows an example of separating the mixture of Jazz, piano, and trumpet music. It can be seen that three music sources are almost completely separated by using the proposed method. In addition, it is noted that the separation performance has deteriorated when the number of sources increases from two. Increased number of sources will mean that there exists more interference in separating every target source and hence results in higher probability of incurring an error. Comparing the results in the table, mixtures containing speech somehow results in slightly poorer performance than mixtures of music sources only. One reason is the seemingly more overlaps in the TF domain between the speech and music sources. It is observed from Fig. 8.6 that music pitches tend to jump discretely while speech pitches do not. Consequently, this leads to less efficiency in the estimation of the spectral basis from the mixture signal. In addition, we have tested the performance of the proposed method on recordings mixed with $i > 3$ sources. We have found that the proposed method works well for mixtures of music sources that are characterized with distinguishable spectral basis. However, the performance shows degradation when separating mixture contains speech sources.

8.5.5 Separating Real Music Recording

In the final experiment, the proposed method is tested on professionally produced music recordings of the well-known song namely "You raise me up" by Kenny G. The

Table 8.4 Separation results of three sources

Mixtures: $y = x_1 + x_2 + x_3$			SDR of $\hat{x}_1$	SDR of $\hat{x}_2$	SDR of $\hat{x}_3$
x_1	x_2	x_3			
Jazz	Piano	Trumpet	6.51	5.61	5.65
Male	Jazz	Piano	5.23	5.73	4.13
Male	Jazz	Trumpet	5.18	5.65	5.21
Male	Piano	Trumpet	5.20	4.09	4.53
Female	Jazz	Piano	5.36	5.47	4.24
Female	Jazz	Trumpet	5.02	5.51	5.10
Female	Piano	Trumpet	5.02	4.32	4.28
Male	Female	Male	−0.8	1.3	−1.6

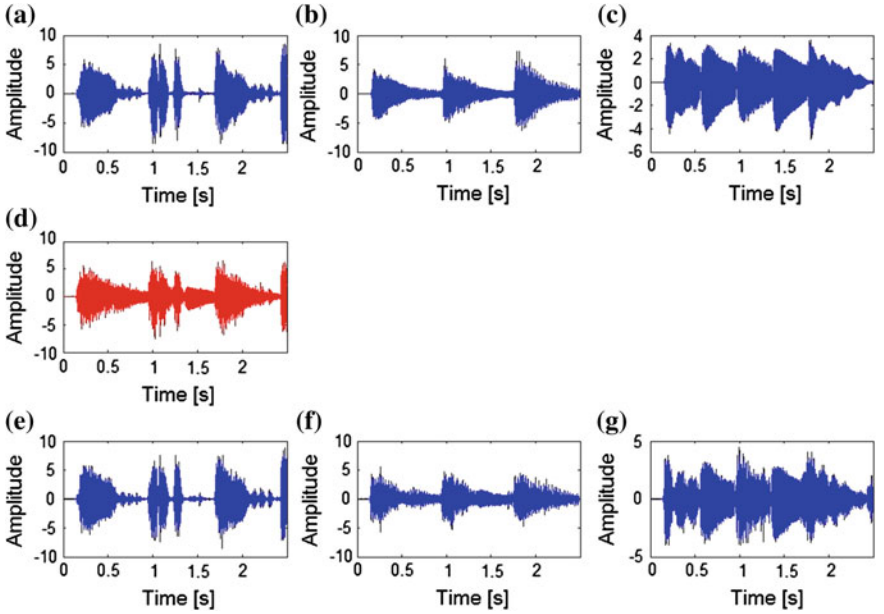


Fig. 8.12 Decomposition results. **a**–**c** denote the original Jazz, piano, and trumpet music, **d** is the mixture and **e**–**g** denote the recovered sources using the proposed method

music consists of two excerpts of length approximately 23 s on mono channel and resampled to 16 kHz. The song is an instrumental music consisting of saxophone and piano sound. The factors of τ and ϕ shifts are set to have $\tau_{\max} = 8$ and $\phi_{\max} = 32$. Since the original source spatial images are not available for this experiment, the separation performance is assessed perceptually and informally by analyzing the log-frequency spectrogram of the estimated source images and listening to the separated sound. This task was a tough task since the instruments play many different notes in the recording. Figure 8.13 shows the separation results of the saxophone and piano

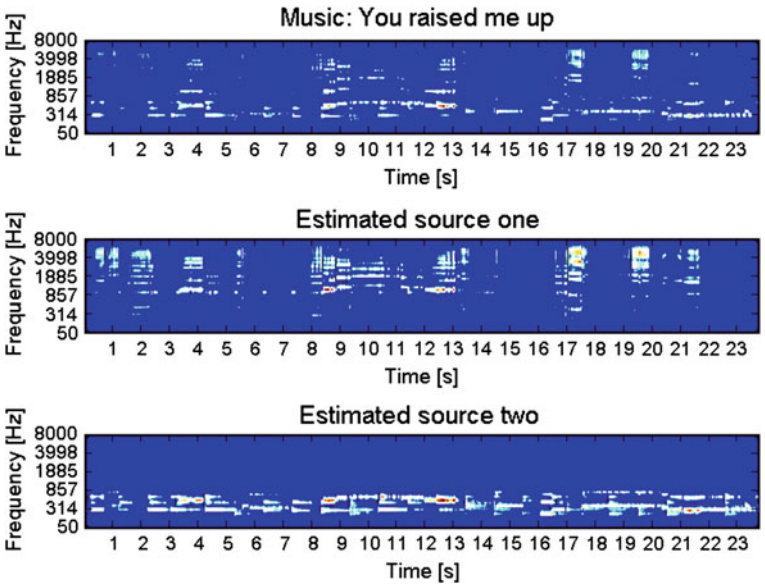


Fig. 8.13 Separation result for song "You raised me up" by Kenny G. *Top* Recorded music. *Middle* Separated saxophone sound. *Bottom* Separated piano sound

sound. The high pitch of continuous saxophone sound is shown in the middle panel of Fig. 8.13 while the notes of the piano are evidently present in Fig. 8.11 bottom panel. Overall, our proposed method successfully separated the professionally produced music recordings and gives a perceptually pleasant listening experience.

8.6 Conclusion

In this chapter, a novel method to solve the single channel audio source separation is proposed. In addition, two algorithms for nonnegative matrix two-dimensional factorization optimized using the Itakura-Saito divergence are presented: Quasi-EM IS-NMF2D and MGD IS-NMF2D. Coupled with the theoretical support of signal separability in the TF domain, the separation system using the gammatone filterbank with these algorithms have shown to yield considerable success. The proposed method enjoys at least three significant advantages: First, it avoids strong constraints of separating sources without training knowledge. Second, the cochleagram rendered by the gammatone filterbank has nonuniform TF resolution which enables the mixed signal to be more separable and thus improves the efficiency of source separation. Finally, the method holds a desirable property of scale invariant which enables low energy components in the cochleagram to bear the same relative importance as the high energy ones. The proposed cochleagram-based IS-NMF2D method in partic-

ular using the Quasi-EM algorithm has yielded significant improvements in source separation compared with other nonnegative matrix factorizations.

References

1. Lee, T.W.: Blind source separation of nonlinear mixing models. Neural Netw. **7**, 121–131 (199**7**)
2. Gao, B., Woo, W.L., Dlay, S.S.: Unsupervised single channel separation of non-stationary signals using gammatone filterbank and Itakura-Saito nonnegative matrix two-dimensional factorizations. IEEE Trans. Circuits Syst. I **60**(3), 662–675 (2013)
3. Gao, B., Woo, W.L., Dlay, S.S.: Variational regularized two-dimensional nonnegative matrix factorization. IEEE Trans. Neural Netw. Learn. Syst. **23**(5), 703–716 (2012)
4. Hyvarinen, A., Karhunen, J., Oja, E.: Independent component analysis and blind source separation, pp. 20–60. Wiley, New York (2001)
5. Cichocki, A., Amari, S.I.: Adaptive Blind Signal and Image Processing—Learning Algorithms and Applications. Wiley (2003)
6. Hyvarinen, A.: Survey on independent component analysis. Neural Comput. Surv. **1**, 94–128 (1999)
7. Taleb, A., Jutten, C.: Source separation in post-nonlinear mixtures. IEEE Trans. Sign. Process. **47**(10), 2807–2820 (1999)
8. Lee, D., Seung, H.: Learning the parts of objects by nonnegative matrix factorisation. Nature **401**(6755), 788–791 (1999)
9. Xie, S., Yang, Z.Y., Fu, Y.L.: Nonnegative matrix factorization applied to nonlinear speech and image Cryptosystems. IEEE Trans. on Circuits Syst. I **55**, 2356–2367 (2008)
10. Helén, M., Virtanen, T.: Separation of drums from polyphonic music using nonnegative matrix factorization and support vector machine. In: Proceedings of 13th European Signal Processing. Turkey (2005)
11. Smaragdis, P., Brown, J.C.: Non-negative matrix factorization for polyphonic music transcription. In: Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 177–180. (2003)
12. Rickard, S., Cichocki, A.: When is non-negative matrix decomposition unique? In: 42nd Annual Conference on Information Sciences and Systems (CISS), pp. 1091–1092. (2008)
13. Abdallah, S.A., Plumbley, M.D.: Polyphonic transcription by non-negative sparse coding of power spectra. In: Proceedings of 5th International Confernence on Music Information Retrieval (ISMIR '04), pp. 318–325. Spain (2004)
14. Parry, R.M., Essa, I.: Incorporating phase information for source separation via spectrogram factorization. In: Proceedings of Conference on Acoustics, Speech and Signal Processing (ICASSP'07), pp. 661–664. Hawaii (2007)
15. Kompass, R.: A generalized divergence measure for nonnegative matrix factorization. Neural Comput. **19**(3), 780–791 (2007)
16. Cichocki, A., Zdunek, R., Amari, S.-I.: Csisz'ar's divergences for non-negative matrix factorization: family of new algorithms. In: Proceedings of 6th International Conference on Independent Component Analysis and Signal Separation (ICA '06), pp. 32–39. Charleston (2006)
17. Virtanen, T.: Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria. IEEE Trans. Audio Speech Lang. Process. **15**(3), 1066–1074 (2007)
18. Radfa, M.H., Dansereau, R.M.: Single-channel speech separation using soft mask filtering. IEEE Trans. Audio Speech Lang. Process. **15**(6) (2007)
19. Roweis, S.: One microphone source separation. In: Proceedings of Neural Information Processing, pp. 793–799 (2000)

20. Morup, M., Schmidt, M.N.: Sparse Non-negative Matrix Factor 2-D Deconvolution. Technical Report, Denmark (2006)
21. Schmidt, M.N., Morup, M.: Nonnegative matrix factor 2-D deconvolution for blind single channel source separation. In: Proceedings 6th Internatinal Conference on Independent Component Analysis and Signal Separation (ICA '06), pp. 700–707. Charleston (2006)
22. Gröchenig, K.: Foundations of Time-Frequency Analysis. Birkhauser, Boston (2001)
23. Brown, Judith C.: Calculation of a constant Q spectral transform. J. Acoust. Soc. Am. **89**(1), 425–434 (1991)
24. Hu, G., Wang, D.L.: Monaural speech segregation based on pitch tracking and amplitude modulation. IEEE Trans. Neural Networks **15**(5), 1135–1150 (2004)
25. Roads, C., et al.: The computer music tutorial. The MIT Press, Cambridge (1996)
26. Schulz, S., Herfet, t.: Binaural source separation in non-ideal reverberant environments. In: Proceedings of 10th International Conference on Digital Audio Effects (DAFx-07), pp. 10–15. Bordeaux, France (2007)
27. Wang, D.L.: On ideal binary mask as the computational goal of auditory scene analysis. In: Divenyi, P. (ed.) Speech Separation by Humans and Machines, pp. 181–197. Norwell, Kluwer (2005)
28. Goto, M., Hashiguchi, H., Nishimura, T., Oka, R.: RWC music database: music genre database and musical instrument sound database. In: Proceedings of International Symposium on Music Information Retrieval (ISMIR), pp. 229–230. Baltimore, Maryland (2003)
29. Yilmaz, O., Rickard, S.: Blind separation of speech mixtures via time-frequency masking. IEEE Trans. Sign. Process. **52**(7), 1830–1847 (2004)
30. Gao, B., Woo, W.L., Dlay, S.S.: Single channel source separation using EMD-subband variable regularized sparse features. IEEE Trans. Audio Speech Lang. Process. **19**, 961–976 (2011)
31. Gao, B., Woo, W.L., Dlay, S.S.: Adaptive sparsity non-negative matrix factorization for single channel source separation. IEEE J. Sel. Top. Sign. Process, **5**, 1932–4553 (2011)
32. Itakura, F., Saito, S.: Analysis synthesis telephony based on the maximum likelihood method. In: Proceedings of 6th International Congress on Acoustics, pp. C-17–C-20. Tokyo, Aug 1968
33. Fevotte, C., Bertin, N., Durrieu, J.L.: Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis. Neural Comput. **21**(3), 793–830 (2009)
34. Signal Separation Evaluation Campaign (SiSEC 2008) (2008) Available http://sisec.wiki.irisa.fr

Chapter 9
Source Localization and Tracking: A Sparsity-Exploiting Maximum a Posteriori Based Approach

Md Mashud Hyder and Kaushik Mahata

Abstract In this work, we explore the potential of sparse recovery algorithms for localization and tracking the direction-of-arrivals (DOA) of multiple targets using a limited number of noisy time samples collected from a small number of sensors. In target tracking problems, the targets are assumed to be moving with a small random angular acceleration. We show that the target tracking problem can be posed as a problem of recursively reconstructing a sequence of sparse signals where the support of the signals changing slowly with time. Here, one can use the support of last signal as a priori information to estimate the behavior of current signal. In particular, we propose a maximum a posteriori (MAP)-based approach to deal with the sparse recovery problem arising in tracking and detection of DOAs. We consider both narrowband and broadband scenarios. Numerical simulations demonstrate the effectiveness of the proposed algorithm. We found that the proposed algorithm can resolve and track closely spaced DOAs with a small number of sensors.

9.1 Introduction

Direction of arrival (DOA) estimation using sensor array has been an active research area [20, 25, 37], playing an important role in smart antennas, next generation mobile communication systems, various type of imaging systems and target tracking applications. Many algorithms have been developed, see [37] and references therein. The algorithms like Capon [6], pose the DOA estimation problem as a beamforming

Research is supported by the Australian Research Council.

Md. M. Hyder (✉) · K. Mahata
The University of Newcastle, Callaghan, NSW 2308, Australia
e-mail: mashud_buet@yahoo.com; md.hyder@uon.edu.au

K. Mahata
e-mail: Kaushik.Mahata@newcastle.edu.au

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_9,
© Springer-Verlag Berlin Heidelberg 2014

problem. Here one designs adaptive filterbanks to obtain nonparametric estimate of the spatial spectrum. The popular alternative to this is the subspace algorithms like MUSIC [35], ESPRIT [33] or weighted subspace fitting [36, 41]. The subspace algorithms which exploit the low-rank structure of the noise-free signal. The maximum-likelihood estimation [25] is another efficient technique, but requires accurate initialization to ensure global convergence. All these methods rely on the statistical properties of the data, and hence requires a large number of time samples.

Conventional DOA estimation techniques cannot exploit the target moving statistics into their formulation and hence their performance degrade when a large number of DOAs moving in a field of interest [45]. Recently, several approaches have been developed for tracking targets [4, 8, 30, 31, 44, 45]. The maximum likelihood (ML) methods [30, 31, 45] have good statistical properties and are robust in target tracking with a relatively small number of samples. The works in [30, 31] incorporate the target motion dynamics in ML estimation and computes the the DOA parameters at each time subinterval and refine the ML estimates through Kalman filtering. The concept of Multiple Target States (MTS) has been introduced in [45] to describe the target motion. The DOA tracking is implemented through updating the MTS by maximizing the likelihood function of the array output. However, ML-based algorithms have high computational cost in general [8]. A recursive expectation and maximization (EM) [12] algorithm has been used in [8] to improve computational efficiency of ML algorithms. Cyclostationarity property of the moving targets has been exploited in [44]. In target tracking, the change in DOA from last time frame to the current time frame is computed by exploiting the difference of the averaged cyclic cross correlation of array output. Target tracking in clutter environment is also addressed in [4].

Sparse signal representation has been applied for spectral analysis [10, 14–16, 26, 34]. In [34], a Cauchy-prior is used to enforce sparsity in a temporal spectrum. A recursive weighted least-squares algorithm called FOCUSS has been developed in [16] for source localization. Fuchs [14, 15] formulates the source localization problem as a sparse recovery problem in the beam-space domain. DOA estimation has been posed into joint-sparse recovery problem in [10, 19, 26]. ℓ_1-SVD [26] combines the singular value decomposition (SVD) step of the subspace algorithms with a sparse recovery method based on ℓ_1-norm minimization. The ℓ_1-SVD algorithm can handle closely spaced correlated sources if the number of sources is known. An alternative strategy called joint $\ell_{2,0}$ approximation of DOA (JLZA-DOA) is proposed in [19]. The algorithm represents the snapshots of sensors as some jointly sparse [18] linear combinations of the columns of a manifold matrix. The resulting optimization problem of JLZA-DOA has been solved using a convex–concave procedure. JLZA is further extended to deal with the joint sparse problem with *multiple* measurement matrices arising in broadband DOA estimation, where the manifold matrices for different frequency bands are different. This allows a sensor spacing larger than the smallest half-wavelength of the broadband signal, which in turn results in a significant improvement in the DOA resolution performance. However, these algorithms have been designed for stationary DOA estimation.

There are two aims of the work: (i) develop an efficient algorithm for stationary DOA estimation and, (ii) adopt the algorithm for multiple target tracking. We pose target tracking as a problem of tracking a sparse signal $\mathbf{x}(t)$. Here the field of interest is discretized into a fine grid consisting of a large number (n) of potential DOAs, and $\mathbf{x}(t)$ is an n dimensional complex valued vector, where the ith component of $\mathbf{x}(t)$ is essentially the signal received from the ith point on the DOA-grid at time t. Since there are only a small number of targets at any time t, the vector $\mathbf{x}(t)$ is sparse, and the support of $\mathbf{x}(t)$ gives the locations of the targets. As the targets move, the support of $\mathbf{x}(t)$ changes with time t. If this change is slow enough, then from the estimate of $\mathbf{x}(t-1)$ we can make a fairly accurate prediction of the support of $\mathbf{x}(t)$. Recent papers [24, 39, 40] in compressive sensing (CS) with partially known support demonstrate that such prior knowledge about the support can be used to significantly lower the number of data samples needed for reconstruction.

While the algorithms for CS with partially known support work well for slowly varying support, these methods cannot be used if the target speed is above a particular threshold. To alleviate this problem we propose a MAP estimation approach. At time t we use the past estimates of $\mathbf{x}(\tau)$, $\tau < t$, to construct a a priori predictive probability density function of $\mathbf{x}(t)$. This prior is such that the components of $\mathbf{x}(t)$ which are close to the predicted future location of DOA, will have large magnitude with very high probability. On the other hand, a component of $\mathbf{x}(t)$ which are far from the predicted location is of large magnitude with a very small probability. Subsequently, we use this prior and the array measurements at time t to derive a MAP estimate of $\mathbf{x}(t)$.

We demonstrate the performance of our method on minimum-redundancy linear arrays [27]. In a minimum-redundancy array, the inter-element spacing is not necessarily required to maintain the half wavelength of the receiving narrowband signal. We can resolve relatively large number of DOAs with small sensors and time samples. Moreover, such array arrangement leads to an increase in the resolution of DOA estimation.

Similar to [19], we enforce joint sparsity in DOA estimation for narrowband and broadband signals. In broadband case, this joint sparsity allows a sensor spacing larger than the smallest half-wavelength of the signal, which in turn results in a significant improvement in the DOA resolution performance. This is possible because the spatial aliasing effect can be suppressed by enforcing the joint sparsity of the recovered spatial spectrum across the whole frequency range under consideration.

The chapter is structured as follows. In Sect. 9.2, narrowband stationary DOA estimation is considered. The DOA estimation problem is set as an underdetermined sparse recovery problem. Some state-of-the-art sparsity-based DOA estimation techniques have been discussed, and a MAP-based DOA estimation framework has been developed. Section 9.3 considers the narrowband DOA tracking problem. We formulate MAP framework in DOA tracking. We also demonstrate a possible way to adopt a conventional sparsity-based DOA estimation technique in tracking problem. The proposed MAP approach has been extended for broadband DOA estimation in Sect. 9.4. Finally, in Sect. 9.5 we present some simulation results.

9.2 Stationary Narrowband DOA Estimation

9.2.1 Background

Consider k narrow-band signals $\{s_j(t)\}_{j=1}^k$ incident on a sensor array, consisting of m omnidirectional sensors. Let

$$\mathbf{y}(t) = [\mathbf{y}_1(t) \cdots \mathbf{y}_m(t)]',$$

where $\mathbf{y}_j(t)$ is the signal recorded after demodulation by the jth sensor and $\mathbf{y}'$ denotes the transpose of $\mathbf{y}$. Defining

$$\mathbf{s}(t) = [s_1(t) \cdots s_k(t)]',$$

and using the narrowband observation model [20, 25], we have

$$\mathbf{y}(t) = A(\theta)\mathbf{s}(t) + e(t). \tag{9.1}$$

Here $A(\theta)$ is the manifold matrix, θ is the DOA vector containing the directions of arrival of individual signals, i.e., the jth component θ_j of θ gives the DOA of the signal $s_j(t)$, and $e(t)$ denotes the measurement noise. The manifold matrix consists of the steering vectors $\{a(\theta_j)\}_{j=1}^k$:

$$A(\theta) = [a(\theta_1) \cdots a(\theta_k)].$$

The mapping $a(\theta)$ depends on the array geometry and the wave velocity, which are assumed to be known for any given θ. The problem is to find θ and k from $\{\mathbf{y}_j(t)\}_{j=1}^m$.

9.2.2 Connection to the Blind Source Separation

Blind source separation (BSS) involves recovering unobserved signals from a set of their mixtures [1, 21, 32]. For instance, the signal received by an antenna is a superposition of signals emitted by all the sources which are in its receptive filed. Let us consider k signals $\{s_j(t)\}_{j=1}^k$ incident on a sensor array, consisting of k sensors. Let

$$\hat{\mathbf{y}}(t) = [\hat{\mathbf{y}}_1(t) \cdots \hat{\mathbf{y}}_k(t)]',$$

where $\hat{\mathbf{y}}_j(t)$ is the signal recorded by the jth sensor. The model of sensor output be [1, 21]:

$$\hat{\mathbf{y}}(t) = \hat{A}\mathbf{s}(t) + \hat{e}(t). \tag{9.2}$$

where $\hat{A} \in \mathbb{R}^{k\times k}$ is an unknown nonsingular mixing matrix. Without knowing the properties of the source signals and the mixing matrix, we want to estimate the source signals from the observations $\hat{\mathbf{y}}(t)$ via some linear transformation of the form [1]

$$\hat{\mathbf{s}}(t) = B\hat{\mathbf{y}}(t) \tag{9.3}$$

where $\hat{\mathbf{s}}(t) = [\hat{s}_1(t) \cdots \hat{s}_k(t)\ ]'$, and $B \in \mathbb{R}^{k\times k}$ is a de-mixing matrix. However, without any information about $\hat{A}$ or $\mathbf{s}(t)$, it is impossible to estimate $\hat{\mathbf{s}}(t)$. In BSS, different assumptions have been made on $\mathbf{s}(t)$. For example, independent component analysis (ICA)-based approach assumes that the sources $\mathbf{s}(t)$ are statistically independent [21]. The goal of ICA is to find the transformation matrix B such that the random variables $\hat{s}_1(t), \ldots, \hat{s}_k(t)$ are as independent as possible [32].

There are interesting similarities between the BSS model in (9.2) and the DOA estimation model in (9.1). In the DOA estimation problem (see (9.1)), we have to estimate the source signals, while the matrix $A(\theta)$ and $\mathbf{s}(t)$ are unknown. Unlike BSS, DOA estimation model assumes that the construction of matrix $A(\theta)$ depends on the sensor array geometry and DOA of source signals. Since, the sensor array geometry is known, if one can estimate the DOA of sources, it is possible to construct $A(\theta)$, and hence estimation of $\mathbf{s}(t)$. It is not necessary the sources to be statistically independent [43]. Hence, DOA estimation can be viewed as a semi-BSS problem. Recently, BSS techniques have been applied for DOA estimation [9, 22, 29].

9.2.3 DOA Estimation as a Joint-Sparse Recovery Problem

We divide the whole area of interest into some discrete set of "potential locations". In this work, we consider the far-field scenario and hence the discrete set be a grid of directions-of-arrival angles. Let the set of all potential DOAs be $\mathbb{G} = \{\bar{\theta}_1, \ldots, \bar{\theta}_n\}$, where typically $n \gg k$. The choice of $\mathbb{G}$ is similar to that used in the Capon or MUSIC algorithms. Collect the steering vectors for each element of $\mathbb{G}$ in

$$\Phi = [a(\bar{\theta}_1) \cdots a(\bar{\theta}_n)].$$

Since $\mathbb{G}$ is known, Φ is known and is independent of θ. Now, represent the signal field at time t by $\mathbf{x}(t) \in \mathbb{C}^n$, where the jth component $\mathbf{x}_j(t)$ of $\mathbf{x}(t)$ is nonzero only if $\bar{\theta}_j = \theta_\ell$ for some ℓ, and in that case $\mathbf{x}_j(t) = \mathbf{s}_\ell(t)$. Then one has a model

$$\mathbf{y}(t) = \Phi\mathbf{x}(t) + \bar{e}(t), \tag{9.4}$$

where $\bar{e}(t)$ is the residual due to measurement noise and model-errors. Since $k \ll n$, $\mathbf{x}(t)$ is sparse. Note that the equality $\bar{\theta}_j = \theta_\ell$ may not hold exactly for any $\ell \in \{1, 2, \ldots, k\}$ in practice. Nevertheless, by making $\mathbb{G}$ dense enough, one can ensure $\bar{\theta}_j \approx \theta_\ell$ closely, and the remaining modeling error is absorbed in the residual term

$\bar{e}(t)$. We model the elements of $\bar{e}(t)$ as mutually independent, and identically complex Gaussian distributed random variables with zero mean and variance $1/\lambda$, where λ is a positive real number.

The model, (9.4) lets us pose the problem of estimating k and θ as that of estimating a sparse $\mathbf{x}(t)$ which can be solved using CS [13] framework. If there is a reliable algorithm to recover the sparse $\mathbf{x}(t)$ from $\mathbf{y}(t)$ using (9.4), then all but a few components of the final solution $\mathbf{x}(t)$ will have very small magnitudes. Thus, if the jth component $x_j(t)$ is a dominant component in the recovered $\mathbf{x}(t)$, then we infer that at time t, there is a source with DOA $\bar{\theta}_j$, with an associated signal $x_j(t)$. Finally, the number of these dominant spikes gives k.

A k sparse $\mathbf{x}(t)$ can be recovered uniquely from (9.4) if $k \leq m/2$, and every m columns of Φ form a basis of $\mathbb{C}^m$. The latter is called the *unique representation property*, and is closely connected to the concept of an *un-ambiguous* array. Apart from the limit on k, the single snapshot setting in (9.4) is sensitive to noise. Since noise is ubiquitous in practical problems, we turn to the so-called joint sparse formulation [18]. In practice, we have several snapshots $\{\mathbf{y}(t)\}_{t=1}^{N}$. Using (9.4), we can write

$$Y := [\mathbf{y}(1) \cdots \mathbf{y}(N)] = \Phi X + E, \tag{9.5}$$

where $X = [\mathbf{x}(1) \cdots \mathbf{x}(N)]$ is a sparse matrix, and $E = [\bar{e}_1 \cdots \bar{e}_N]$. If the DOA vector θ is time-invariant over the period of measurement, then for all t the nonzero dominant peaks in $\mathbf{x}(t)$ occur at the same locations corresponding to the actual DOAs. In other words, only k rows of X are nonzero. Such a matrix is called jointly k-sparse. Hence the DOA estimation problem can be posed as the multiple measurement vectors (MMV) problem [18] of finding a jointly sparse X from Y.

9.2.4 Results on the Joint-Sparse Recovery Problem

Assume that $E = 0$, and Y, X are real-valued.[1] The conditions for the existence of a unique solution to the MMV problem is demonstrated by the following lemma [11].

Lemma 1 *Let rank* $(Y) = r \leq m$, *and every* m *columns of* Φ *form a basis of* $\mathbb{R}^m$. *Then a solution to* (9.5) *with* k *nonzero rows is unique provided that* $k \leq \lceil (m + r)/2 \rceil - 1$, *where* $\lceil . \rceil$ *denotes the ceiling operation.*

We pose DOA estimation as a MMV problem. Assume that $N > m$, and the matrix

$$X = [\mathbf{x}(1) \cdots \mathbf{x}(N)]$$

[1] In the real DOA estimation problem, $E \neq 0$ and X, Y are complex valued. We deal with these issues in the next section.

has rank k. Then according to Lemma 1, the DOAs can be estimated uniquely using the joint sparse framework if $k \leq m - 1$. It is interesting that all the subspace algorithms, e.g., MUSIC or ESPRIT, have the same limitation.

As described in [19], an $\ell_{2,0}$-norm-based minimization approach can be used to solve the MMV problem arising in DOA estimation. However, since zero norm leads to an NP-hard problem, different relaxations used in literature.

9.2.5 DOA Estimation Using ℓ_1 Optimization

ℓ_1-SVD [26] is an efficient algorithm that uses $\ell_{2,1}$-based optimization to solve the DOA estimation problem. In presence of noise, ℓ_1-SVD algorithm considers the following way to solve X given Y in (9.5)

$$\min_X \|Y - \Phi X\|_F^2 + \varsigma \|X\|_{2,1}, \tag{9.6}$$

where $||X||_{2,1}$ is the mixed norm

$$||X||_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{N} |x_i(j)|^2} = \sum_{i=1}^{n} ||X(i,:)||_2, \tag{9.7}$$

and $\varsigma > 0$ is a tuning factor whose value depends on noise present in the signal. Note that we use the Matlab notation $X(i,:)$ to denote the ith row of X.

The computation time needed to optimization in (9.6) increases with increasing N. To reduce both the computational complexity and sensitivity to noise, ℓ_1-SVD uses SVD of the data matrix $Y \in \mathbb{C}^{m \times N}$. Similar to other subspace algorithms (i.e., MUSIC) the ℓ_1-SVD keeps the signal subspace.

9.2.6 MAP Approach for DOA Localization

In this section we develop a maximum a posteriori (MAP) estimation approach for stationary DOA estimation. Recall that individual columns of E are modeled as mutually independent and identically complex Gaussian distributed random vectors with zero mean and covariance matrix $1/\lambda I$. Using (9.5) the conditional density of Y given X is given by

$$p(Y|X) = \left(\frac{\lambda}{2\pi}\right)^{mN} \exp\{-\lambda ||Y - \Phi X||_F^2/2\}. \tag{9.8}$$

Now suppose we know a priori density $p(X)$ of X. Then MAP proposes to estimate X by maximizing the conditional density $p(X|Y)$ given the observed data Y with respect to X. This is same as maximizing the joint log-likelihood function [23]

$$\log p(Y|X) + \log p(X). \tag{9.9}$$

Next we propose a suitable candidate for $p(X)$. Recall that X is row sparse. Furthermore it is reasonable to postulate the rows of X are mutually independent, because the locations of individual targets are independent. In practice, it is common to assume that the elements in a row $X(i,:)$ are independent and identically distributed. The independence follows from the choice of the sampling frequency used in practical arrays. The identical distribution follows because the energy of a target signal remains the same over N snapshots. Now for a given i, we have two possibilities:

- With a high probability $1-q$ there is no target at $\bar{\theta}_i$, and so the elements of $X(i,:)$ are basically of very small energy (contributed by noise and model errors), say μ. We model $X(i,:)$ in this case as a complex Gaussian distributed random vector with mean zero and covariance matrix $\mu^2 I$.
- Otherwise (with a low probability q), there is a target at $\bar{\theta}_i$ so that the elements of $X(i,:)$ have relatively large energy $\rho \gg \mu$ contributed by the target signal. We model $X(i,:)$ in this case as a complex Gaussian distributed random vector with mean zero and covariance matrix $\rho^2 I$.

Consequently, $p(X)$ is a product of Gaussian mixture densities

$$p(X) = \prod_{i=1}^{n} \left\{ \frac{q}{(2\pi\rho)^N} \exp\left(\frac{-||X(i,:)||_2^2}{2\rho^2} \right) + \frac{1-q}{(2\pi\mu)^N} \exp\left(\frac{-||X(i,:)||_2^2}{2\mu^2} \right) \right\}. \tag{9.10}$$

In this work we set $q = k/n$. Such a Gaussian mixture model has been used in simulations [28] and performance analysis [17] in CS literature. Using (9.10) we can write

$$-\ln[p(X)] = \sum_{i=1}^{n} \left(\frac{-||X(i,:)||_2^2}{2\rho^2} - \ln\left[1 + r \exp\left\{ -\frac{||X(i,:)||_2^2}{2\sigma^2} \right\} \right] \right) + \text{constant} \tag{9.11}$$

where the "constant" absorbs the terms independent of X, and

$$r = \frac{(1-q)\rho^N}{q\mu^N}, \qquad \frac{1}{\sigma^2} = \frac{1}{\mu^2} - \frac{1}{\rho^2}. \tag{9.12}$$

Combining (9.11) with (9.8) and (9.9), we can write the criterion function

$$\wp(X) = \sum_{i=1}^{n} \left(\frac{||X(i,:)||_2^2}{2\rho^2} - \ln \left[1 + r_i \exp \left\{ -\frac{||X(i,:)||_2^2}{2\sigma^2} \right\} \right] \right) + \frac{\lambda}{2} \|Y - \Phi X\|_F^2, \tag{9.13}$$

which we need to minimize with respect to X. The reader may have noticed that while moving from (9.11) to (9.13) we have replaced r by r_i. Indeed for stationary DOA estimation case $r_i = r$, $\forall i$. Having different r_i for different values of i will be useful in tracking moving targets, where it will suffice to minimize the same cost function (9.13). Considering the more general case at this stage allows us to use the results in the next section in the tracking problem.

9.2.7 Solution Strategy

Like many optimization problems encountered in CS literature [7, 38], minimizing (9.13) is a nonconvex problem. To deal with the nonconvex optimization problem we use the concept of graduated nonconvexity (GNC) [2, 3]. GNC is a deterministic annealing method for approximating the global solution for nonconvex minimization problems. Here we construct a sequence of functions $\wp_j(X)$, $j = 0, 1, 2, \ldots, w$ such that

- $\wp_0(X)$ is quadratic;
- $|\wp_{j+1}(X) - \wp_j(X)|$ is small in the neighborhood of the minimizer $X^{(j)}$ of $\wp_j(X)$;
- $\wp_w(X) = \wp(X)$ for a user chosen integer w.

Because $\wp_0(X)$ is quadratic, we can compute $X^{(0)}$ using the standard analytical expression. Then as $|\wp_1(X) - \wp_0(X)|$ is small in the neighborhood of $X^{(0)}$, by initializing a numerical algorithm to minimize $\wp_1(X)$ at $X^{(0)}$ one has a high probability of converging to $X^{(1)}$. If we continue this process of initializing the numerical algorithm to optimize $\wp_{j+1}(X)$ at our estimate of $X^{(j)}$ obtained by numerically optimizing $\wp_j(X)$, then one can expect that $X^{(w)}$ is likely to be the minimizer of $\wp_w(X)$.

The sequence of functions $\wp_j(X)$, $j = 0, 1, 2, \ldots, w$ are constructed as follows. We choose an appropriate real number σ_1 (more details on how the choices are made will follow shortly), and define

$$\wp_0(X) = \sum_i \frac{\|X(i,:)\|_2^2}{2\rho^2} + \frac{\lambda}{2} \|Y - \Phi X\|_F^2 \tag{9.14}$$

$$\wp_j(X) = \wp_0(X) - \sum_i \ln \left[1 + r_i \exp \left\{ -\frac{\|X(i,:)\|_2^2}{2\sigma_j^2} \right\} \right] \qquad j = 1, 2, 3, \ldots, w; \tag{9.15}$$

where

$$\sigma_j = (\sigma/\sigma_1)^{j/w}\sigma_1,$$

and w is a user chosen integer. The parameter σ_j controls the degree of nonconvexity in $\wp_j$. As we increase the value of j form 0 to w, we gradually transform $\wp_j$ from a convex function $\wp_0$ to our desired likelihood function $\wp_w$. If w is sufficiently large, then the change from $\wp_{j-1}$ to $\wp_j$ is small, and so is the change from $X^{(j-1)}$ to $X^{(j)}$.

Next we derive an expression for $X^{(0)}$. Let R be a diagonal matrix such that $R_{ii} = \rho^2$. Recall that $X^{(0)}$ is the minimum point of $\wp_0$. Hence, if we differentiate (9.14) with respect to X and evaluate at $X^{(0)}$, we must get zero. Hence

$$X^{(0)} = (R + \lambda\Phi^*\Phi)^{-1}(\lambda\Phi^* Y). \tag{9.16}$$

Note that we denote the conjugate transpose of Φ by Φ^*.

We can reduce the cost of computing $X^{(0)}$ if we use an alternative expression for $X^{(0)}$, which is obtained by applying matrix inversion lemma in the right hand side of (9.16)

$$X^{(0)} = R^{-1}\Phi^*(I/\lambda + \Phi R^{-1}\Phi^*)^{-1}Y. \tag{9.17}$$

where I is a $m \times m$ identity matrix. Computing $X^{(0)}$ via (9.16) requires inverting an $m \times m$ matrix. On the other hand we must invert an $n \times n$ matrix if we compute $X^{(0)}$ via (9.17).

The parameter σ_1 controls the degree of nonconvexity in $\wp_1$. If we take $\sigma_1 \to \infty$, then the logarithmic term in (9.15) tends to $\ln(1+r)$, making $\wp_1$ a quadratic function. In practice, we take

$$\sigma_1 \geq 5 \max_i \|X(i,:)^{(0)}\|_2,$$

This ensures $\exp\{-\|X(i,:)^{(0)}\|_2^2/(2\sigma_1^2)\} \geq 0.99$ for all i. Consequently, $\exp\{-\|X(i,:)^{(0)}\|_2^2/(2\sigma_1^2)\} \approx 1$ for all X satisfying $||X - X^{(0)}||_F < ||X^{(1)} - X^{(0)}||_F$ [28].

9.2.8 Minimizing $\wp_j$

In this section we explore some properties of $X^{(j)}$, and develop a numerical algorithm to compute it. Define

$$\xi_j(X(i,:)) = \frac{1}{\rho^2} + \frac{r_i \exp\left(-\frac{\|X(i,:)\|_2^2}{2\sigma_j^2}\right)}{\sigma_j^2\left[1 + r_i \exp\left(-\frac{\|X(i,:)\|_2^2}{2\sigma_j^2}\right)\right]}, \tag{9.18}$$

and an $n \times n$ diagonal matrix $W_j(X)$ as

$$W_j(X) = \text{diag}\{\, \xi_j(X(1,:)) \;\; \xi_j(X(2,:)) \cdots \xi_j(X(m,:)) \,\}.$$

From (9.18) it is readily verified that $\xi_j\{X(i,:)\} > 0$ for all i. Hence $W_j(X)$ is a positive definite matrix.

Now we can verify that

$$\frac{\partial \wp_j(X)}{\partial X} = W_j(X)X - \lambda \Phi^*(Y - \Phi X). \tag{9.19}$$

Since $X^{(j)}$ is the minimum point of $\wp_j$, setting $X = X^{(j)}$ in (9.19) we get

$$X^{(j)} = g_j(X^{(j)}) \tag{9.20}$$

where

$$g_j(X) := \{W_j(X) + \lambda \Phi^* \Phi\}^{-1}\{\lambda \Phi^* Y\}. \tag{9.21}$$

Also, a calculation similar to (9.17) gives

$$g_j(X) := W_j^{-1}(X)\Phi^* \left[I/\lambda + \Phi W_j^{-1}(X)\Phi^* \right]^{-1} Y. \tag{9.22}$$

The equation $X = g_j(X)$ is nonlinear, and cannot be solved analytically. One possibility is to use a fixed point iteration. However, the convergence of the fixed point iteration is not guaranteed. Nevertheless, using (9.19) and (9.21) we have

$$g_j(X) - X = -\{W_j(X^{(j)}) + \lambda \Phi^* \Phi\}^{-1} \frac{\partial \wp_j(X)}{\partial X}. \tag{9.23}$$

Since W_j is a positive definite matrix, and $\lambda > 0$, the matrix $W_j(X^{(j)}) + \lambda \Phi^* \Phi$ is positive definite. Hence (9.23) implies that $\wp_j(X)$ is decreasing along the vector $g_j(X) - X$. In fact moving to $g_j(X)$ from X is same as taking the Newton step associated with some convex–concave procedure to minimize $\wp_j(X)$ [19].

9.2.9 Numerical Algorithm for Solving MAP Optimization

The MAP optimization strategy is given in Table 9.1. We assume that the values of ρ, and μ are known. In fact, simulation results demonstrate that the accurate values of ρ, and μ are not necessary. Instead, an approximation of these values are sufficient [17]. Using initial $X^{(0)}$ we calculate σ_1, and in Step 3 some parameters including w are set. In each iteration, we find a step-length κ along the decent-direction $g_j(X) - X$ using the standard backtracking strategy (step 4–5) [5]. We set $\beta = 0.5$, which

Table 9.1 MAP for narrowband DOA estimation

1. Set $X = X^{(0)}$
2. Set $\sigma_1 = 5 \max_i \|X(i,:)^{(0)}\|_2$
3. Set $j = 1$, choose $\eta \in (0, 1)$, $\beta = 0.5$ and $w \geq 15$

do {

4. Set $\kappa = 1$
5. **while** $\wp_j(\kappa g_j(X) + (1-\kappa)X]) > \wp_j(X)$ {
   $\kappa = \beta\kappa$
   } **end**
6. $X_o = X$, and $X = \kappa g_j(X) + (1-\kappa)X$
7. If $\frac{\|X - X_o\|_2}{\|X_o\|_2} < \eta$ then $j = j + 1$

} **while** $j \leq w$

is very common [5]. The inner-iteration for updating X for a given j terminates when the relative change in the magnitude of X is below η, see step 7. Hence for a smaller η more accurate solutions are sought in expense of higher computation time. According to our experimental study, having $\eta = 0.02$ makes a good tradeoff. Upon convergence of each inner iteration, we increment j (step 7). Note that choosing a larger w helps the optimization problem in (9.15) to move slowly from convex to its desire nonconvex form. Thus we have lower probability to get trapped in local minimum. However, a larger w increases the number of outer iterations (step 4–7) of the algorithm, and hence the computation time. Our experimental study suggests that choosing $w = 20$ makes a good tradeoff between solution accuracy and computation time. Upon convergence for $j = w$, PMAP stops its iterations. The value of λ depends on noise variance. In our simulation, we set $\lambda = 5$ [19].

9.2.10 Acceleration via QR Factorization

Typically, the matrix $X \in \mathbb{C}^{n \times N}$ in (9.5) is large, as n is a large number (we need $n = 360$ to achieve 0.5° spatial resolution). If the number of data samples N is large, the algorithm may become very slow. To accelerate the algorithm, we use the QR factorization $Y/\sqrt{N} = \bar{R}Q$, where $\bar{R} \in \mathbb{C}^{m \times m}$ is a nonsingular upper triangular matrix, and $Q \in \mathbb{C}^{m \times N}$ is such that $QQ^* = I$. When $E = 0$, then

$$\text{row span}\{X\} \subset \text{row span}\{Q\}. \tag{9.24}$$

Consequently, $\|Y - \Phi X\|_F^2 = \|\bar{R} - \Phi\bar{X}\|_F^2$, where $\bar{X} = XQ^* \in \mathbb{C}^{n \times m}$ must be jointly row-sparse, and is of significantly smaller size than X. Hence, it is more efficient to estimate $\bar{X}$ via minimizing

$$\wp(\bar{X}) = \sum_{i=1}^{n}\left(\frac{\|\bar{X}(i,:)\|_2^2}{2\rho^2} - \ln\left[1 + r_i \exp\left\{-\frac{\|\bar{X}(i,:)\|_2^2}{2\sigma^2}\right\}\right]\right) + \frac{\lambda}{2}\|\bar{R} - \Phi\bar{X}\|_F^2. \tag{9.25}$$

Following (9.10), it is readily verified that $\bar{X}$ and X have identical a priori density function.

9.3 Narrowband Target Tracking

9.3.1 Problem Formulation

Suppose k number of targets are moving in a plane. We wish to

- Detect the targets; and
- Track the DOAs of the targets with a time resolution τ, so that the algorithm will yield estimates DOAs at time instant $0, \tau, 2\tau, 3\tau, \ldots$.

Let f_c be the sampling frequency at the sensors. We assume that over each interval $[\ell\tau, \ell\tau + \frac{N}{f_c})$, the change of $\theta_i(t)$ is negligible, i.e.,

$$\theta(t) \approx \theta(\ell\tau); \quad t \in \left[\ell\tau, \ell\tau + \frac{N}{f_c}\right) \tag{9.26}$$

where N is the number of snapshots used to detect and estimate the DOAs at time $\ell\tau$. The above is a common assumption made by many state-of-the-art target tracking algorithms [4, 30, 31, 44]. Hence under this assumption, the N snapshots of sensor data in (9.1) can be expressed as

$$y(t) \approx A(\theta(\ell\tau))\mathbf{s}(t) + e(t), \quad t \in \left[\ell\tau, \ell\tau + \frac{N}{f_c}\right). \tag{9.27}$$

Just as in (9.5) we can now formulate a stationary target tracking problem at time instant $\ell\tau$, where we need to recover a joint sparse matrix

$$X_\ell = [\mathbf{x}(\ell\tau)\ \ \mathbf{x}(\ell\tau + 1/f_c)\ \ \mathbf{x}\{\ell\tau + (N-1)/f_c\}]$$

given the snapshot matrix

$$Y_\ell = [\mathbf{y}(\ell\tau)\ \ \mathbf{y}(\ell\tau + 1/f_c)\ \ \mathbf{y}\{\ell\tau + (N-1)/f_c\}]$$

such that $Y_\ell = \Phi X_\ell + E_\ell$ holds. To solve this problem we can always use the stationary DOA estimation methods discussed before. However, when we track the targets the algorithm is inherently recursive. This means while estimating X_ℓ, we already

know estimates of $X_{\ell-1}$, $X_{\ell-2}$, ... If we can use this prior knowledge efficiently, we can work with a significantly smaller N making the assumption (9.26) a feasible one. In addition it is possible to work with larger number of targets. In order to exploit the prior information available in the estimates of $X_{\ell-1}$, $X_{\ell-2}$, . . ., we need to consider a dynamic of model for target motion. This is discussed next.

9.3.2 Dynamic Model for the Target Motion

In this contribution, we stick to the most commonly used 'small acceleration' model used in the target tracking literature. For any target i, we assume that its angular acceleration $\ddot{\theta}_i(t)$ is a Wiener process with a small incremental variance. Note that we denote the first order derivative of $\theta(t)$ with respect to t by $\dot{\theta}(t)$, and the second order derivative is denoted by $\ddot{\theta}(t)$.

For a generic DOA $\theta(t)$, it is straightforward to write down the state space equations for $\theta(t)$ in terms of the state

$$\mathbf{s}(t) = [\theta(t) \;\; \dot{\theta}(t)]'. \tag{9.28}$$

We have

$$\dot{\mathbf{s}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{s}(t) + \begin{bmatrix} 0 \\ w(t) \end{bmatrix}, \qquad \theta(t) = [1 \;\; 0]\, \mathbf{s}(t), \tag{9.29}$$

where $w(t)$ denotes the Wiener process with an incremental variance γ. Now we can discretize (9.29) with a time resolution τ, and it is wellknown that the equivalent discrete-time model is given by

$$\mathbf{s}_1(\ell+1) = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} \mathbf{s}_1(\ell) + \mathbf{w}_1(\ell), \qquad \theta(\ell\tau) = [1 \;\; 0]\, \mathbf{s}_1(\ell), \tag{9.30}$$

where we write $\mathbf{s}_1(\ell) := \mathbf{s}(\ell\tau)$ for short; $\mathbf{w}_1(\ell)$ is a discrete-time, zero mean white noise sequence such that

$$\mathsf{E}\{\mathbf{w}_1 \mathbf{w}_1'\} = \gamma \begin{bmatrix} \tau^3/3 & \tau^2/2 \\ \tau^2/2 & \tau \end{bmatrix}. \tag{9.31}$$

Using (9.30), (9.31) and after a few steps of algebra one can show that

$$\theta(\ell\tau) = 2\theta(\ell\tau - \tau) - \theta(\ell\tau - 2\tau) + w_2(\ell), \tag{9.32}$$

where w_2 is a scalar valued discrete-time first order moving average process with zero mean and

$$\mathsf{E}\{w_2\} = 0, \qquad \mathsf{E}\{w_2^2(\ell)\} = \frac{2\gamma\tau^3}{3}, \qquad \mathsf{E}\{w_2(\ell) w_2(\ell-1)\} = \frac{\gamma\tau^3}{6}. \tag{9.33}$$

9.3.3 Extension of ℓ_1-SVD for DOA Tracking

Using (9.32) we can use the estimates of $X_{\ell-1}, X_{\ell-2}, \ldots$ to make predictions about X_ℓ. Then we wish to incorporate this prediction in the MAP framework described above. However, before doing so, we consider how we could naturally extend ℓ_1-SVD method for tracking by using an approach called CS with partially known support [24, 39, 40]. The idea is to use the past estimates $X_{\ell-1}, X_{\ell-2}, \ldots$ to estimate the support of X_ℓ, and use that information to estimate a joint sparse X_ℓ.

Suppose that until time $t = (\ell-1)\tau$ the tracking algorithm has detected k targets. The estimated DOAs for the ith target at time$(\ell - j)\tau$ is denoted by $\hat{\theta}_i(\ell - j)$. From (9.32) we know

$$\mathsf{E}\{\theta_i(\ell\tau)\} = 2\hat{\theta}_i(\ell-1) - \hat{\theta}_i(\ell-2). \tag{9.34}$$

At this stage we neglect the second order statistics of $w_2(\ell)$, because the standard methods for CS with partially known support does not have any provisions to do so. Nevertheless, while discussing MAP approach in the next section, we will use the second order statistics of $w_2(\ell)$.

Define $\mathbb{I} := \{1, 2, \ldots, n\}$. Recall that n is the number of points on the DOA-grid $\mathbb{G} = \{\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_n\}$. CS with partially known support requires us to predict the support $T(\ell)$ of X_ℓ defined as

$$T_\ell := \{i \in \mathbb{I} : ||X_\ell(i,:)||_2 \neq 0\}.$$

We do so as follows. For each i we identify the point of $\mathbb{G}$ which is the nearest to $\mathsf{E}\{\theta_i(\ell\tau)\}$ and denotes the associated index by ι_i:

$$\iota_i = \arg\min_{j\in\mathbb{I}} |\bar{\theta}_j - \mathsf{E}\{\theta_i(\ell\tau)\}|.$$

Then form

$$T_\ell = \{\iota_1, \iota_2, \ldots, \iota_k\}.$$

Different CS-based algorithms have been developed to exploit the support information in sparse recovery process [24, 39, 40]. Least squares CS-residual (LS-CS) [39] is a two step procedure. First, a least squares (LS) estimate $\bar{X}_\ell$ of X_ℓ is computed assuming that the support of X_ℓ is T_ℓ. To explain the details, let

$$\Phi_+ = [\Phi(:,\iota_1) \;\; \Phi(:,\iota_2) \cdots (:,\iota_k)\,]$$

and

$$X_+ = [\Phi_+^*\Phi_+]^{-1}\Phi_+^* Y.$$

Note that X_+ is a $k \times N$ matrix, and we form $\bar{X}_\ell$ as follows:

$$\bar{X}_\ell(i,:) = \begin{cases} X_+(j,:), & \text{if } i = \iota_j \text{ for some } j \in \{1,\ldots,k\}, \\ 0, & \text{otherwise.} \end{cases} \tag{9.35}$$

Next calculate the associated residual

$$\bar{Y} = Y - \Phi_{T_\ell} \bar{X}_\ell.$$

In the subsequent step, LS-CS uses a CS algorithm to find a sparse solution $\hat{X}_\ell$ such that $\bar{Y}_\ell = \Phi \hat{X}_\ell$. The final estimate is $\bar{X}_\ell + \hat{X}_\ell$. Adapting recently proposed modified-CS [40] to our problem, this step requires us to solve

$$\hat{X}_\ell = \arg\min_X \|Y - \Phi X\|_F^2 + \varsigma \sum_{i \in \mathbb{I} \setminus T_\ell} ||X(i,:)||_2. \tag{9.36}$$

We refer the modification of ℓ_1-SVD as ℓ_1-SVD-MCS. It might be worthwhile to note the difference between (9.6) and (9.36), and see how easily ℓ_1-SVD in (9.6) is adapted to the framework of CS with partially known support.

9.3.4 MAP for Tracking

The MAP algorithm can be used for tracking problem with a small modification in the expression for $p(X)$ given in (9.10). Here we have the option to use the estimates of $X_{\ell-1}, X_{\ell-2}, \ldots$ to obtain a better prior density $p(X)$.

As before, suppose that until time $t = (\ell-1)\tau$, the tracking algorithm has detected k targets. Then according to (9.32) the conditional density of $\theta_i(\ell\tau)$ evaluated at θ is proportional to

$$\eta_i(\theta) = \exp\left\{ -\frac{\left[\theta - 2\hat{\theta}_i(\ell-1) + \hat{\theta}_i(\ell-2)\right]^2}{4\gamma\tau^3/3} \right\}.$$

It is natural to use this conditional density as a measure of the probability that $\theta_i(\ell\tau)$ is close to a grid point $\bar{\theta}_j$. In particular, if the ith target was the only target detected, then the probability q_j that we will find that target at the grid point $\bar{\theta}_j$ is evaluated as

$$\frac{\eta_i(\bar{\theta}_j)}{\sum_{j \in \mathbb{I}} \eta_i(\bar{\theta}_j)}.$$

When we have k targets in the field, then the probability q_j of finding a target at grid point $\bar{\theta}_j$ is given by

$$q_j = 1 - \prod_{i=1}^{k} \left\{ 1 - \frac{\eta_i(\bar{\theta}_j)}{\sum_{j \in \mathbb{I}} \eta_i(\bar{\theta}_j).} \right\}, \tag{9.37}$$

The above expression for (9.37) works when no new target can appear in the field, and none of the existing targets can disappear. Nevertheless, we can generalize (9.37) to relax these requirements. Let

- α be the probability that an existing target disappears; and
- β be the probability that a new target appears in the field at a grid point.

Now we modify (9.37) to accommodate the possibility that a new target may appear in the field and an existing target may disappear. The event that "the ith target is not presesent at $\bar{\theta}_j$" is the union of two mutually exclusive events:

1. The target has actually disappeared from the field (with probability α); and
2. The target is still there in the field (with probability $1 - \alpha$), but it is not at $\bar{\theta}_j$. The probability of this event is

$$(1 - \alpha) \left\{ 1 - \frac{\eta_i(\bar{\theta}_j)}{\sum_{j \in \mathbb{I}} \eta_i(\bar{\theta}_j)} \right\}.$$

Now combining the probabilities of (1) and (2), the resultant probability that "the ith target is not present at $\bar{\theta}_j$" is

$$\alpha + (1 - \alpha) \left\{ 1 - \frac{\eta_i(\bar{\theta}_j)}{\sum_{j \in \mathbb{I}} \eta_i(\bar{\theta}_j)} \right\}.$$

Then the probability that none of the existing targets is present at $\bar{\theta}_j$ and no new target appears at $\bar{\theta}_j$ is

$$(1 - \beta) \prod_{i=1}^{k} \left[\alpha + (1 - \alpha) \left\{ 1 - \frac{\eta_i(\bar{\theta}_j)}{\sum_{j \in \mathbb{I}} \eta_i(\bar{\theta}_j)} \right\} \right].$$

Thus, to accommodate the possibility that a new target may appear in the field and an existing target may disappear (9.37) is modified accordingly to

$$q_j = 1 - (1 - \beta) \prod_{i=1}^{k} \left[\alpha + (1 - \alpha) \left\{ 1 - \frac{\eta_i(\bar{\theta}_j)}{\sum_{j \in \mathbb{I}} \eta_i(\bar{\theta}_j)} \right\} \right], \quad j \in \mathbb{I} \tag{9.38}$$

Once we know q_i, $\forall i \in \mathbb{I}$, we can replace q by q_i in (9.10) to get

$$p(X) = \prod_{i=1}^{n} \left\{ \frac{q_i}{(2\pi\rho)^N} \exp\left(\frac{-||X(i,:)||_2^2}{2\rho^2} \right) + \frac{1-q_i}{(2\pi\mu)^N} \exp\left(\frac{-||X(i,:)||_2^2}{(2\mu^2)} \right) \right\}, \tag{9.39}$$

which in turn gives

$$-\ln[p(X)] = \sum_{i=1}^{n} \left(\frac{-||X(i,:)||_2^2}{2\rho^2} - \ln\left[1 + r_i \exp\left\{ -\frac{||X(i,:)||_2^2}{2\sigma^2} \right\} \right] \right) + \text{constant} \tag{9.40}$$

where

$$r_i = \frac{(1-q_i)\rho^N}{q_i \mu^N}, \qquad \frac{1}{\sigma^2} = \frac{1}{\mu^2} - \frac{1}{\rho^2}. \tag{9.41}$$

Combining (9.40) with (9.8) and (9.9), we again arrive at the criterion function (9.13). However, unlike the stationary case, now all r_i are different from each other. Nevertheless, we can follow the procedure in Sect. 9.2.7 to develop a minimization problem and use the algorithm in Table 9.1 for estimation of $X(\ell)$. In this work, we set $\alpha = 0.01, \beta = 0.01$.

9.4 Extension of MAP Framework for Broadband DOA Estimation

We consider the procedure proposed in [19] for broadband DOA estimation. The broadband signal has been splitted into several narrowband signals by using a bank of narrowband filters. Subsequently, the narrowband model (9.5) is applied to each narrowband filter output. Suppose that we have narrowband data at frequencies $\{\omega_i\}_{i=1}^K$, and let Φ_i be the "over-complete" manifold matrix at frequency ω_i. Then, the narrowband model at frequency ω_i is of the form

$$Y_i = \Phi_i X_i + E_i, \quad i \in \{1, 2, \ldots, K\}.$$

Here, E_i is the additive noise at frequency ω_i and X_i is the jointly row-sparse signal matrix at frequency ω_i. Now,

$$\mathbf{X} := [X_1 \;\; X_2 \cdots X_K]$$

is jointly row-sparse. This is because if $X_i(\ell, :)$ is nonzero for some ℓ, then there is source signal at frequency ω_i at direction $\bar{\theta}_\ell$. Therefore, we would expect signals at other frequencies from the direction $\bar{\theta}_\ell$ as well, making $X_i(\ell, :)$ nonzero for all i.

Now to resolve broadband DOA in MAP framework we need to develop a priori density $p(\mathbf{X})$ of $\mathbf{X}$. We assume that the energy of a target at all frequency bands $\{\omega_j\}_{j=1}^K$ is almost similar. Nevertheless, the following approach can be extended easily when signal energy is different at different frequency band. Then as described in Sect. 9.2.6, we have two probabilities for every index i: (i) with very small probability q_i, there is a target at $\bar{\theta}_i$, and hence the elements of $\mathbf{X}(i, :)$ have relatively large energy ρ. We model $\mathbf{X}(i, :)$ as a complex Gaussian distributed random vector with zero mean and covariance matrix $\rho^2 I$; (ii) with probability $1 - q_i$, the elements of $\mathbf{X}(i, :)$ have small energy $\mu \ll \rho$. Hence, $p(\mathbf{X})$ is a product of Gaussian mixture densities

$$p(\mathbf{X}) = \prod_{i=1}^{n} \left\{ \frac{q_i}{(2\pi\rho^2)^{KN}} \exp\left(-\frac{\|\mathbf{X}(i, :)\|_2^2}{2\rho^2} \right) + \frac{1 - q_i}{(2\pi\mu^2)^{KN}} \exp\left(-\frac{\|\mathbf{X}(i, :)\|_2^2}{2\mu^2} \right) \right\}. \tag{9.42}$$

Then following (9.8)–(9.13), we can end up the the criterion function

$$\wp(\mathbf{X}) = \sum_{i=1}^{n} \left(\frac{\|\mathbf{X}(i, :)\|_2^2}{2\rho^2} - \ln\left[1 + r_i \exp\left\{ -\frac{\|\mathbf{X}(i, :)\|_2^2}{2\sigma^2} \right\} \right] \right) + \frac{\lambda}{2} \sum_{i=1}^{K} \|Y_i - \Phi_i X_i\|_F^2 \tag{9.43}$$

where, $r_i = \dfrac{(1 - q_i)\rho^{KN}}{q_i \mu^{KN}}$,

which need to minimize with respect to $\mathbf{X}$, and σ are defined in (9.12).

For minimizing (9.43), we follow the GNC procedure of Sect. 9.2.7 and generate w number of suboptimization problem $\{\wp_j(\mathbf{X})\}_{j=1}^w$. Then following the calculation (9.18)–(9.23), it can be shown that $\wp_j(\mathbf{X})$ is decreasing along $\mathbf{g}_j(\mathbf{X}) - \mathbf{X}$, where

$$\mathbf{g}_j(\mathbf{X}) = [g_j^{(1)}(\mathbf{X})\ g_j^{(2)}(\mathbf{X})\ \cdots\ g_j^{(K)}(\mathbf{X})], \tag{9.44}$$

$$g_j^{(i)}(\mathbf{X}) = W_j^{-1}(\mathbf{X})\Phi_i^* \left[I/\lambda + \Phi_i W_j^{-1}(\mathbf{X})\Phi_i^* \right]^{-1} Y_i. \tag{9.45}$$

Using the direction we can develop a broadband DOA estimation algorithm. The final algorithm is given in Table 9.2.

9.4.1 Jamming Signal Mitigation

The noise term $e(t)$ in (9.1) is the residual noise due to measurement noise and model error. In general, it is assumed that the noise has uniform distribution and smaller

Table 9.2 MAP for broadband DOA estimation

1. Set $X_i^{(0)} = R^{-1}\Phi_i^*(I/\lambda + \Phi_i R^{-1}\Phi_i^*)^{-1}Y_i,\ i = 1, \ldots, K.$
   Form $\mathbf{X}^{(0)} = [\ X_1^{(0)}\ \ X_2^{(0)} \cdots X_K^{(0)}\ ]$
2. Set $\sigma_1 = 5\max_i \|\mathbf{X}(i,:)^{(0)}\|_2$
3. Set $j = 1$, choose $\eta \in (0, 1)$, $\beta = 0.5$ and $w \geq 15$

do {
   4. Set $\kappa = 1$
   5. **while** $\wp_j(\kappa \mathbf{g}_j(\mathbf{X}) + (1-\kappa)\mathbf{X}]) > \wp_j(\mathbf{X})$ **{**
      $\kappa = \beta\kappa$
      } **end**
   6. $\mathbf{X}_o = \mathbf{X}$, and $\mathbf{X} = \kappa \mathbf{g}_j(\mathbf{X}) + (1-\kappa)\mathbf{X}$
   7. If $\frac{\|\mathbf{X}-\mathbf{X}_o\|_F}{\|\mathbf{X}_o\|_F} < \eta$ then $j = j + 1$

} **while** $j \leq w$

magnitude than the source signal. However, there exists some other types of noisy signals; like, jamming signal. Jamming and deception is the intentional emission of radio frequency signals to interfere with the operation of a radar by saturating its receiver with noise or false information. In general, jamming signals come from fixed directions and have magnitude many times larger than actual source signal [42]. Hence, jamming signals hinder actual source. However, due to larger magnitude, it is easier to know the direction of jamming signal in priori. To mitigate from jamming, we will use the crude estimate of the direction of the jamming signal as a 'partially known support' of the sparse signal $\mathbf{X}$. Let the jammer direction be supported on T_j. Then the value of q_i in (9.42) will be very high if $q_i \in T_j$. We then search any target in rest of the support of $\mathbf{X}$. In our experiments we set $q_i = 0.99$ when $q_i \in T_j$.

9.5 Simulation Results

We compare the performance of MAP-based approach with ℓ_1-SVD [26], Capon's method [6], and ℓ_1-SVD-MCS (see (9.36)). We use four sensors and follow the procedure of minimum-redundancy array [27] for the linear array arrangement. The interelement spacings are d, $3d$ and $2d$, respectively. For narrowband DOA estimation, the value of d is equal to the half wavelength of receiving narrowband signal. For broadband signal we consider two types of the value of d for two classes of algorithms. Similar to [19], MAP algorithm can allow a sensor spacing larger than the half wavelength associated with the highest frequency in the broadband signal. Hence, we set the value of d to be 1.5 times the smallest wavelength in the broadband signal. However, ℓ_1-SVD and Capon cannot allow larger d. Hence we set the value of d equals to 0.5 times the smallest wavelength in the broadband signal. The algorithms starts with a uniform grid with 0.5° resolution, i.e., $n = 360$, and $\Phi \in \mathbb{C}^{4\times 360}$.

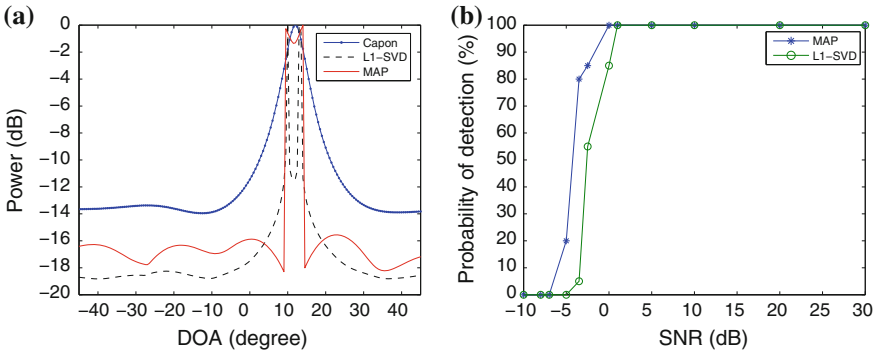


Fig. 9.1 Separating two uncorrelated sources at 10° and 15° by different algorithms. **a** Spatial spectrum obtained by different algorithms. Signal SNR = 2 dB, $N = 50$. **b** Frequency of separating sources versus SNR

In DOA tracking simulation, we select starting DOA locations first. The future DOA locations are generated using (9.30). The starting $\dot{\theta}(t) = 0.5^0$ and $\gamma = 0.05$. The simulations are performed using MATLAB7.

9.5.1 Narrowband DOA Estimation and Tracking

Each narrowband signal is generated from a zero mean Gaussian distribution. The measurements are corrupted by temporally and spatially uncorrelated zero-mean noise sequence. At first we consider stationary DOA estimation. We assume the number of DOAs k is unknown.

Figure 9.1a shows the spatial spectrum plots of the different algorithms when two uncorrelated sources are placed at 10° and 15°. We take $N = 50$ snapshots and SNR = 2 dB. Capon algorithm cannot separate sources. It indicates one source at 12°. ℓ_1-SVD can separate two sources at 10° and 13°. Hence detection bias is 2°. MAP algorithm can separate sources at 9.5° and 15°. Hence bias in only 0.5°. Next, we investigate the impact of the noise power on the performance of algorithms. Here, we simulate two uncorrelated sources at 10° and 15°. We keep the value of N fixed at 50 and vary the noise power. For each SNR, we carry out 100 independent simulations, and the results are shown in Fig. 9.1b, where we plot the frequency at which the different algorithms separate the sources against noise. Note that MAP outperforms the ℓ_1-SVD. The plots for Capon is not shown, as they are unable to resolve the sources when $N < 140$.

Figure 9.2 shows the results when we simulate two strongly correlated sources at 10° and 15° with a correlation coefficient 0.99. SNR = 6 dB and $N = 50$. Note that MAP can locate the sources clearly, while other methods generates single peak and failed separating sources.

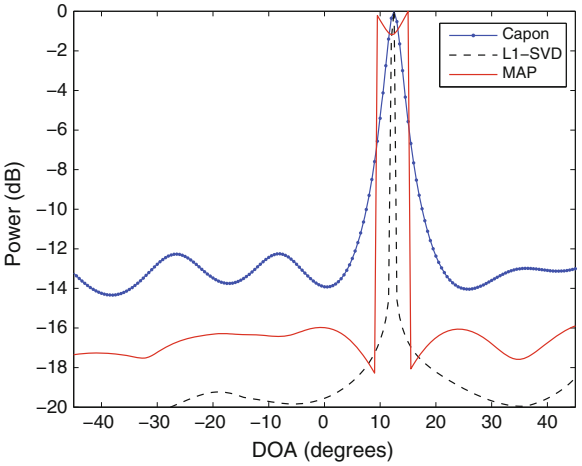


Fig. 9.2 Separating two correlated sources at 10° and 15° by different algorithms. SNR = 6 dB, $N = 50$

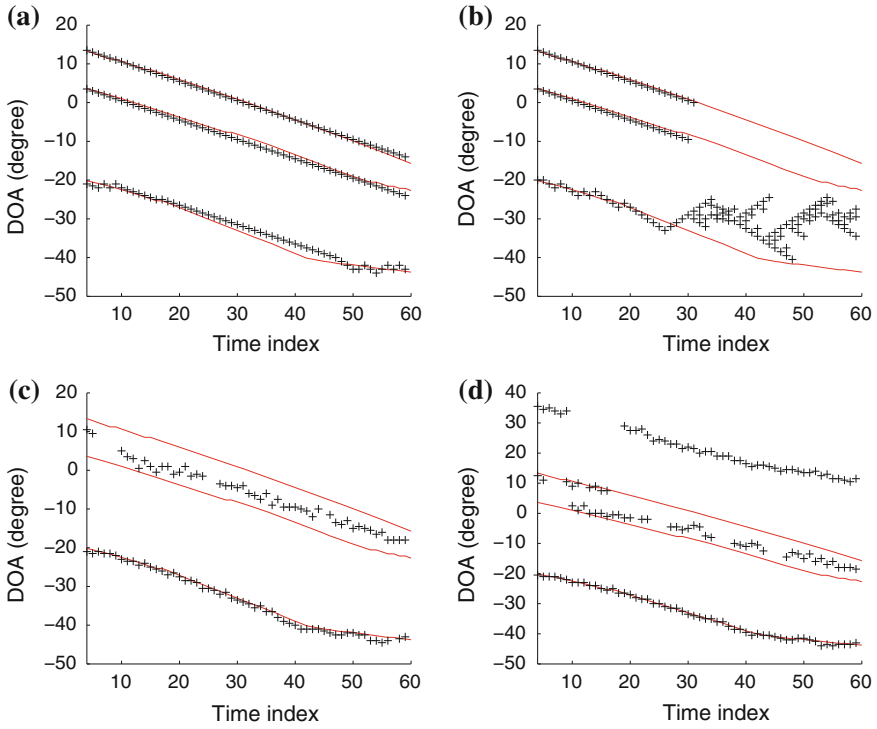


Fig. 9.3 DOA tracking for three targets. SNR = 4 dB, $N = 50$. '−' actual track, '+' estimated track. **a** MAP, **b** ℓ_1-SVD-MCS, **c** ℓ_1-SVD, **d** Capon

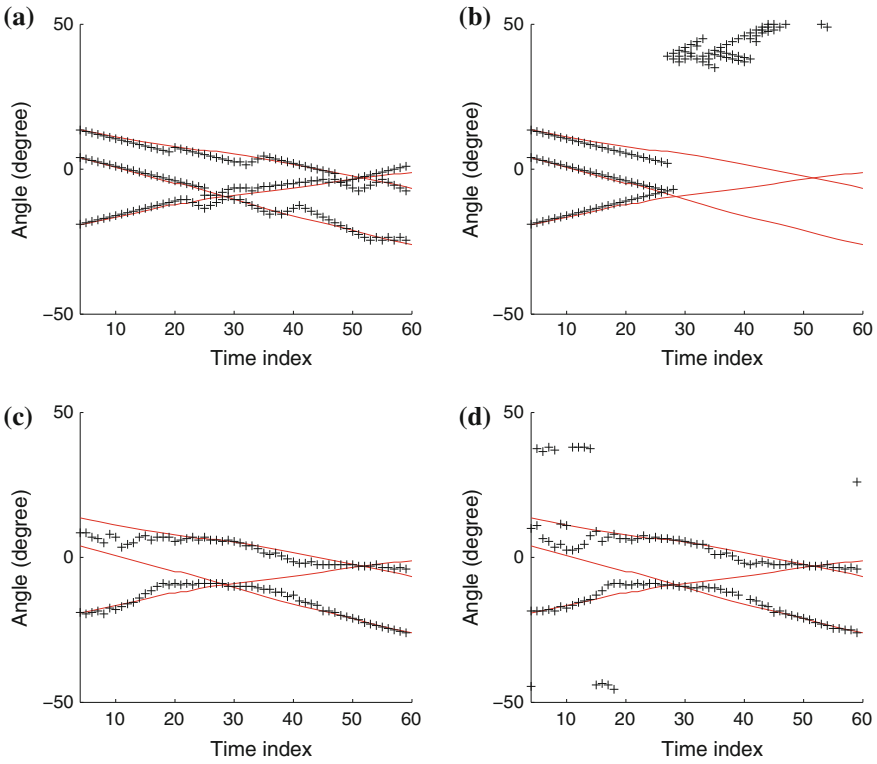


Fig. 9.4 DOA tracking for three targets. SNR $= 4$ dB, $N = 50$. '$-$' actual track, '$+$' estimated track. **a** MAP, **b** ℓ_1-SVD-MCS, **c** ℓ_1-SVD, **d** Capon

Narrowband DOA tracking results are shown in Fig. 9.3. We consider three uncorrelated moving sources. The starting location of sources are $-20°$, $5°$ and $10°$ respectively. The average SNR $= 4$ dB and $N = 50$. As can be seen in Fig. 9.3a MAP can track the sources almost accurately. There is a little error tracking the first source between time index 35 and 45. ℓ_1-SVD-MCS can track sources until time index 25. The interesting observation is that once ℓ_1-SVD-MCS loses track of DOAs, it cannot back to the track again. As illustrated in Fig. 9.3b, after time index 25, ℓ_1-SVD-MCS cannot track second and third DOAs anymore. Instead, it generates some random walk around the track of first source. ℓ_1-SVD can track the first source only. Capon also tracks first source. However, it generates another fictitious path from $40°$ to $20°$.

We consider another route in Fig. 9.4. The starting DOA locations are $-20°$, $5°$ and $15°$, respectively. In this scenario the first source crosses the route of other two sources. It is difficult to keep track of sources when they cross each other. Figure 9.4a illustrates that MAP is able to keep tracking the sources. In some cases, the estimated route of some sources are displaced from the actual route, however, MAP algorithm can back to the actual route of sources immediately. As before, ℓ_1-SVD-MCS can

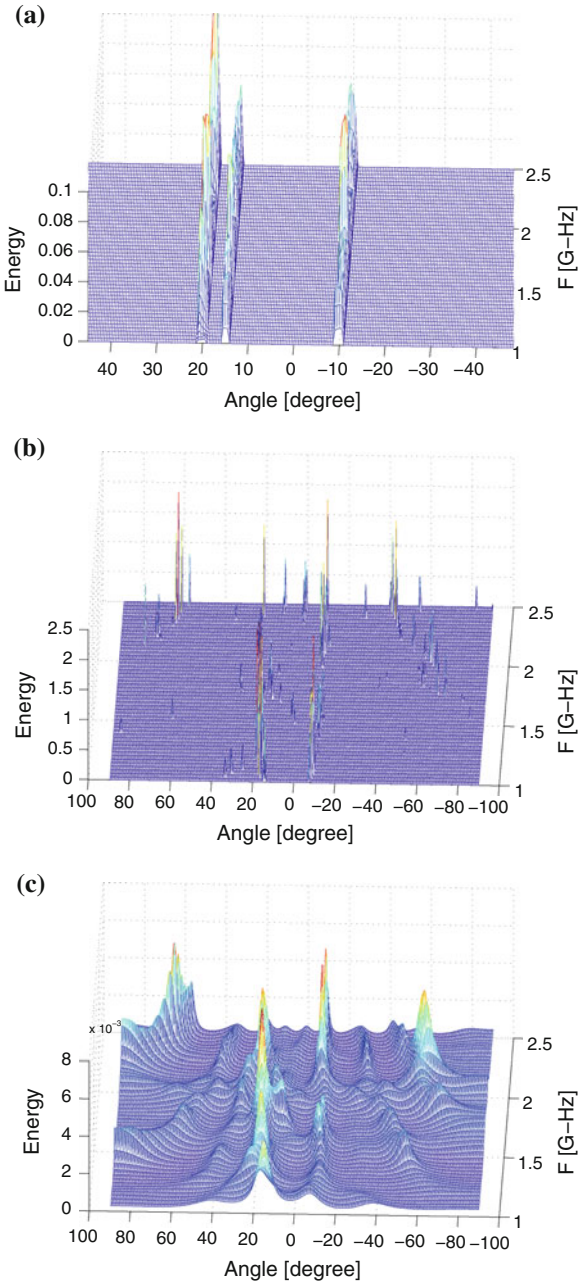


Fig. 9.5 Broadband DOA estimation results for three sources are located at $-10°$, $15°$ and $20°$. SNR $= 6$ dB, $N = 100$. **a** MAP, **b** ℓ_1-SVD, **c** Capon

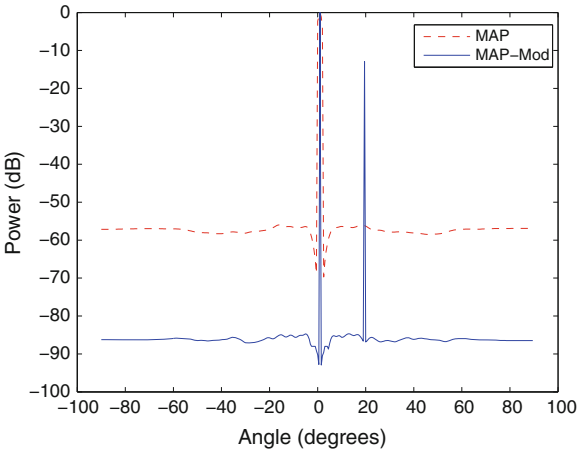


Fig. 9.6 Separating a broadband DOA in presence of jamming signal. The jamming signal comes from 1^0 and actual source is at 20^0. $N = 100$

track the sources until time index 25. The algorithm loses the track of sources when the first and second sources cross each other. ℓ_1-SVD and Capon failed tracking sources. Until time index 25, both algorithms track the first sources, however, they start tracking the second source after time index 25.

9.5.2 Broadband DOA Estimation and Tracking

Broadband sources are generated using the procedure [19]. Each source consists of 10 sinusoids with frequencies randomly chosen from the interval [1, 2.5] GHz. The received signal is sampled at 7.5 GHz. The sampled data is filtered through a bank of first-order bandpass filters of the form

$$H_\omega(z) = \frac{(1-r)\mathrm{e}^{\mathrm{i}\omega}}{z - r\mathrm{e}^{\mathrm{i}\omega}} \tag{9.46}$$

It can be shown that H_ω is a narrow-band filter centered at digital frequency ω (which is related to the analog frequency via the standard relationship). The bandwidth of the filter is controlled by r, where $0 < r < 1$. Taking $r \to 1$ makes the bandwidth smaller, but makes the filter less stable. We take $r = 0.99$. The filterbank consists of 50 filters with center frequencies uniformly distributed over the interval [1, 2.5] GHz.

We simulate three broadband sources at $-10°$, $15°$ and $20°$ in Fig. 9.5. The SNR $= 6$ dB and $N = 100$. As can be seen in Fig. 9.5, MAP algorithm separate three sources fairly accurately. The detected peaks are sharp and clear. ℓ_1-SVD generates

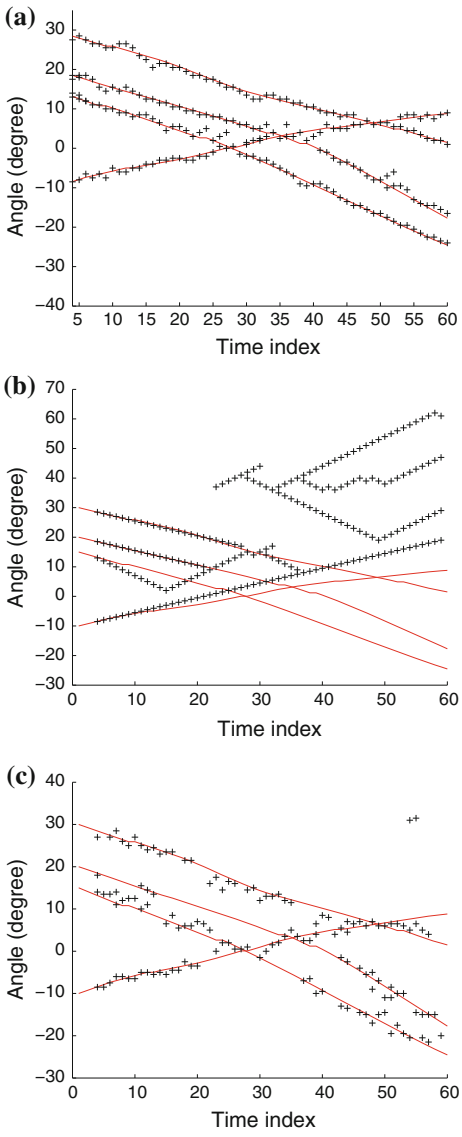


Fig. 9.7 Broadband DOA tracking for four targets. SNR $= 4$ dB, $N = 100$. '$-$' actual track, '$+$' estimated track. **a** MAP, **b** ℓ_1-SVD-MCS, **c** Capon

many spurious peaks and failed to locate the actual DOA locations. Capon roughly generates two peaks at $-9°$ and $19°$. However, It generates many other random peaks.

Figure 9.6 shows the source detection performance in presence of jamming signal. In this setup, a jammer sending signal at an angle 1^0 with a SNR $= 40$ dB. The actual source located at 20^0 transmitting signal with SNR $= 2$ dB. As can be seen, the proposed modification of MAP for jamming signal (MAP-Mod) in Sect. 9.4.1

can detect the actual source. However, when we are applying the conventional MAP, it detects jamming signal only.

Broadband DOA tracking results are shown in Fig. 9.7. We consider four moving sources. The starting location of sources are $-10°$, $15°$, $20°$ and $30°$, respectively. The SNR $= 4$ dB and $N = 50$. As can be seen in Fig. 9.7a MAP can track the sources with reasonable accuracy.

9.6 Conclusion

A sparse signal reconstruction perspective for source localization and tracking has been proposed. We started with a scheme for localizing narrowband sources and developed a tractable MAP-based optimization approach which can exploit the joint-sparsity arises in the source localization problem. The scheme has been extend for wideband source localization. However, the resulting optimization was nonconvex. Hence, we propose an approach similar to the concept of GNC to cope with the issue. We described how to efficiently mitigate the local minima of the nonconvex optimization through an automatic method by choosing the regularization parameter. We then adopt the MAP formulation for narrowband and wideband source tracking scenario. In source tracking formulation, we utilize the information of current location and moving direction of DOA to estimate its future location. We modify the proposed MAP formulation so that it can use the information efficiently. Finally, we examined various aspects of our approach by using numerical simulations. Several advantages over existing source localization methods were identified, including increased resolution, no need for accurate initialization, and improved robustness to noise.

Some of the interesting questions for future research include an investigation of the applicability of GNC-based sparse recovery algorithms, which have a lower computational cost, to blind source localization. A theoretical study for determining the sequence $\wp_j(X)$ in (9.15) so that the algorithm can avoid local minima will be helpful.

References

1. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. In: Advances in Neural Information Processing Systems, pp. 757–763. MIT Press, Cambridge (1996)
2. Blake, A., Zisserman, A.: Visual Reconstruction. MIT Press, Cambridge (1987)
3. Blake, A.: Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. IEEE Trans. Pattern Anal. Mach. Intell. **11**(1), 2–12 (1989)
4. Blanding, W., Willett, P., Bar-Shalom, Y.: Multiple target tracking using maximum likelihood probabilistic data association. In: Aerospace Conference, 2007 IEEE, pp. 1–12 (2007). doi:10.1109/AERO.2007.353035

5. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
6. Capon, J.: High-resolution frequency-wavenumber spectrum analysis. Proc. IEEE **57**(8), 1408–1418 (1969)
7. Chartrand, R.: Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process. Lett. **14**(10), 707–710 (2007)
8. Chung, P.J., Bohme, J.F., Hero, A.O.: Tracking of multiple moving sources using recursive em algorithm. EURASIP J. Adv. Signal Process. **2005**(1), 534, 685 (2005). doi:10.1155/ASP.2005.50. http://asp.eurasipjournals.com/content/2005/1/534685
9. Chun-yu, K., Wen-Tao, F., Xin-hua, Z., Jun, L.: A kind of method for direction of arrival estimation based on blind source separation demixing matrix. In: Eighth International Conference on Natural Computation (ICNC), pp. 134–137 (2012)
10. Cotter, S.: Multiple snapshot matching pursuit for direction of arrival (doa) estimation. In: European Signal Processing Conference, Poznan, Poland, pp. 247–251 (2007)
11. Cotter, S., Rao, B., Engan, K., Kreutz-Delgado, K.: Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans. Signal Process. **53**(7), 2477–2488 (2005)
12. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc B (Methodol.) **39**(1), 1–38 (1977). http://www.jstor.org/stable/2984875
13. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**, 1289–1306 (2006)
14. Fuchs, J.J.: Linear programming in spectral estimation: application to array processing. In: Proceedings of the IEEE International Conference ICASSP-96 on Acoustics, Speech, and Signal Processing 1996, vol. 6, pp. 3161–3164 (1996). doi:10.1109/ICASSP.1996.550547
15. Fuchs, J.J.: On the application of the global matched filter to doa estimation with uniform circular arrays. IEEE Trans. Signal Process. **49**(4), 702–709 (2001)
16. Gorodnitsky, I., Rao, B.: Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm. IEEE Trans. Signal Process. **45**(3), 600–616 (1997)
17. Hyder, M., Mahata, K.: Maximum a posteriori estimation approach to sparse recovery. In: Seventeenth International Conference on Digital Signal Processing (DSP), pp. 1–6 (2011). doi:10.1109/ICDSP.2011.6004892
18. Hyder, M., Mahata, K.: A robust algorithm for joint-sparse recovery. IEEE Signal Process. Lett. **16**(12), 1091–1094 (2009). doi:10.1109/LSP.2009.2028107
19. Hyder, M., Mahata, K.: Direction-of-arrival estimation using a mixed norm approximation. IEEE Trans. Signal Process. **58**(9), 4646–4655 (2010). doi:10.1109/TSP.2010.2050477
20. Johnson, D.H., Dudgeon, D.E.: Array Signal Processing: Concepts and Techniques. Prentice-Hall, Englewood Cliffs (1993)
21. Jutten, C., Herault, J.: Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture. Signal Process. **24**(1), 1–10 (1991)
22. Kang, C., Zhang, X., Han, D.: A new kind of method for DOA estimation based on blind source separation and mvdr beamforming. In: Fifth International Conference on Natural Computation, ICNC '09., vol. 2, pp. 486–490 (2009)
23. Kay, S.M.: Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall Inc, Upper Saddle River (1993)
24. Khajehnejad, M., Xu, W., Avestimehr, A., Hassibi, B.: Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models. IEEE Trans. Signal Process. **59**(5), 1985–2001 (2011)
25. Krim, H., Viberg, M.: Two decades of array signal processing research: the parametric approach. IEEE Signal Process. Mag. **13**(4), 67–94 (1996)
26. Malioutov, D., Cetin, M., Willsky, A.: A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. Signal Process. **53**(8), 3010–3022 (2005)
27. Moffet, A.: Minimum-redundancy linear arrays. IEEE Trans. Antennas Propag. **16**(2), 172–175 (1968). doi:10.1109/TAP.1968.1139138
28. Mohimani, H., Babaie-Zadeh, M., Jutten, C.: A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm. IEEE Trans. Signal Process. **57**(1), 289–301 (2009)

29. Mukai, R., Sawada, H., Araki, S., Makino, S.: Blind source separation and DOA estimation using small 3-D microphone array. In Proc. HSCMA 2005, pp. d.9–10 (2005)
30. Park, S., Ryu, C., Lee, K.: Multiple target angle tracking algorithm using predicted angles. IEEE Trans. Aerosp. Electron. Syst. **30**(2), 643–648 (1994)
31. Rao, C., Sastry, C.R., Zhou, B.: Tracking the direction of arrival of multiple moving targets. IEEE Trans. Signal Process. **42**(5), 1133–1144 (1994). doi:10.1109/78.295205
32. Rojas, F., Puntonet, C., Rodriguez-Alvarez, M., Rojas, I., Martin-Clemente, R.: Blind source separation in post-nonlinear mixtures using competitive learning, simulated annealing, and a genetic algorithm. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **34**(4), 407–416 (2004)
33. Roy, R., Kailath, T.: ESPRIT—estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust. Speech Signal Process. **37**(7), 984–995 (1989)
34. Sacchi, M., Ulrych, T., Walker, C.: Interpolation and extrapolation using a high-resolution discrete fourier transform. IEEE Trans. Signal Process. **46**(1), 31–38 (1998)
35. Schmidt, R.: Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. **34**(3), 276–280 (1986)
36. Stoica, P., Sharman, K.C.: Maximum likelihood methods for direction-of-arrival estimation. IEEE Trans. Acoust. Speech Signal Process. **38**(7), 1132–1143 (1990)
37. Stoica, P., Moses, R.: Introduction to Spectral Analysis, 2nd edn. Prentice-Hall, Upper Saddle River (2004)
38. Trzasko, J., Manduca, A.: Relaxed conditions for sparse signal recovery with general concave priors. IEEE Trans. Signal Process. **57**(11), 4347–4354 (2009)
39. Vaswani, N.: LS-CS-residual (LS-CS): compressive sensing on least squares residual. IEEE Trans. Signal Process. **58**(8), 4108–4120 (2010)
40. Vaswani, N., Lu, W.: Modified-CS: modifying compressive sensing for problems with partially known support. IEEE Trans. Signal Process. **58**(9), 4595–4607 (2010)
41. Viberg, M., Ottersten, B.: Sensor array processing based on subspace fitting. IEEE Trans. Signal Process. **39**(5), 1110–1121 (1991)
42. Xu, L., Li, J., Stoica, P.: Target detection and parameter estimation for mimo radar systems. IEEE Trans. Aerosp. Electron. Syst. **44**(3), 927–939 (2008)
43. Xu, X., Ye, Z., Zhang, Y.: Doa estimation for mixed signals in the presence of mutual coupling. IEEE Trans. Signal Process. **57**(9), 3523–3532 (2009)
44. Yan, H., Fan, H.H.: An algorithm for tracking multiple wideband cyclostationary sources. In: Thirteenth IEEE/SP Workshop on Statistical Signal Processing, pp. 497–502 (2005). doi:10.1109/SSP.2005.1628646
45. Zhou, Y., Yip, P., Leung, H.: Tracking the direction-of-arrival of multiple moving targets by passive arrays: algorithm. IEEE Trans. Signal Process. **47**(10), 2655–2666 (1999). doi:10.1109/78.790648

Part II
Applications

Chapter 10
Statistical Analysis and Evaluation of Blind Speech Extraction Algorithms

Hiroshi Saruwatari and Ryoichi Miyazaki

Abstract In this chapter, a problem of blind source separation for speech applications operated under real acoustic environments is addressed. In particular, we focus on a blind spatial subtraction array (BSSA) consisting of a noise estimator based on independent component analysis (ICA) for efficient speech enhancement. First, it is theoretically and experimentally pointed out that ICA is proficient in noise estimation rather than in speech estimation under a nonpoint-source noise condition. Next, motivated by the above-mentioned fact, we introduce a structure-generalized parametric BSSA, which consists of an ICA-based noise estimator and post-filtering based on generalized spectral subtraction. In addition, we perform its theoretical analysis via higher-order statistics. Comparing a parametric BSSA and a parametric channelwise BSSA, we reveal that a channelwise BSSA structure is recommended for listening but a conventional BSSA is more suitable for speech recognition.

10.1 Introduction

A hands-free speech recognition system [1–3] is essential for the realization of an intuitive, unconstrained, and stress-free human–machine interface, where users can talk naturally because they require no microphone in their hands. In this system, however, since noise and reverberation always degrade speech quality, it is difficult to achieve high recognition performance, compared with the case of using a close-talk microphone such as a headset microphone. Therefore, we must suppress interference sounds to realize a noise-robust hands-free speech recognition system.

H. Saruwatari (✉)
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
e-mail: hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

R. Miyazaki
Nara Institute of Science and Technology, Nara 630-0192, Japan
e-mail: ryoichi-m@is.naist.jp

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_10,
© Springer-Verlag Berlin Heidelberg 2014

Source separation is one approach to removing interference sound source signals. Source separation for acoustic signals involves the estimation of original sound source signals from mixed signals observed in each input channel. There have been various studies on microphone array signal processing; in particular, the delay-and-sum (DS) [4–7] array and adaptive beamformer (ABF) [8–11] are the most conventionally used microphone arrays for source separation and noise reduction. ABF can achieve higher performance than the DS array. However, ABF requires a priori information, e.g., the look direction and speech break interval. These requirements are due to the fact that conventional ABF is based on *supervised* adaptive filtering, which significantly limits its applicability to source separation in practical applications. Indeed, ABF cannot work well when the interfering signal is nonstationary noise.

Recently, alternative approaches have been proposed. Blind source separation (BSS) is an approach to estimating original source signals using only mixed signals observed in each input channel. In particular, BSS based on independent component analysis (ICA) [12], in which the independence among source signals is mainly used for the separation, has recently been studied actively [13–22]. Indeed, the conventional ICA could work, particularly in speech–speech mixing, i.e., all sources can be regarded as point sources, but such a mixing condition is very rare and unrealistic; real noises are often widespread sources. In this chapter, we mainly deal with generalized noise that cannot be regarded as a point source. Moreover, we assume this noise to be nonstationary noise that arises in many acoustical environments; however, ABF could not treat this noise well. Although ICA is not influenced by the nonstationarity of signals unlike ABF, this is still a very challenging task that can hardly be addressed by conventional ICA-based BSS because ICA cannot separate widespread sources.

In this chapter, first, we analyze ICA under a nonpoint-source noise condition and point out that ICA is proficient in noise estimation rather than in speech estimation under such a noise condition. This analysis implies that we can still utilize ICA as an accurate noise estimator. Next, we review blind spatial subtraction array (BSSA) [23], an improved BSS algorithm recently proposed in order to deal with real acoustic sounds. BSSA consists of an ICA-based noise estimator and post-filtering such as spectral subtraction (SS) [24], where noise reduction in BSSA is achieved by subtracting the power spectrum of the estimated noise via ICA from the power spectrum of the noisy observations. This "power-spectrum-domain subtraction" procedure provides better noise reduction than conventional ICA with estimation error robustness. However, BSSA always suffers from artificial distortion, so-called musical noise, owing to nonlinear signal processing. This leads to a serious tradeoff between the noise reduction performance and the amount of signal distortion in speech recognition.

In a recent study, two types of BSSA have been proposed (see Fig. 10.1) [25]. One is the conventional BSSA structure that performs SS after delay-and-sum (DS) (see Fig. 10.1a), and the other involves channelwise SS before DS (chBSSA; see Fig. 10.1b). Also, it has been theoretically clarified that chBSSA is superior to BSSA for the mitigation of the musical noise [26]. Therefore, in this chapter, we generalize

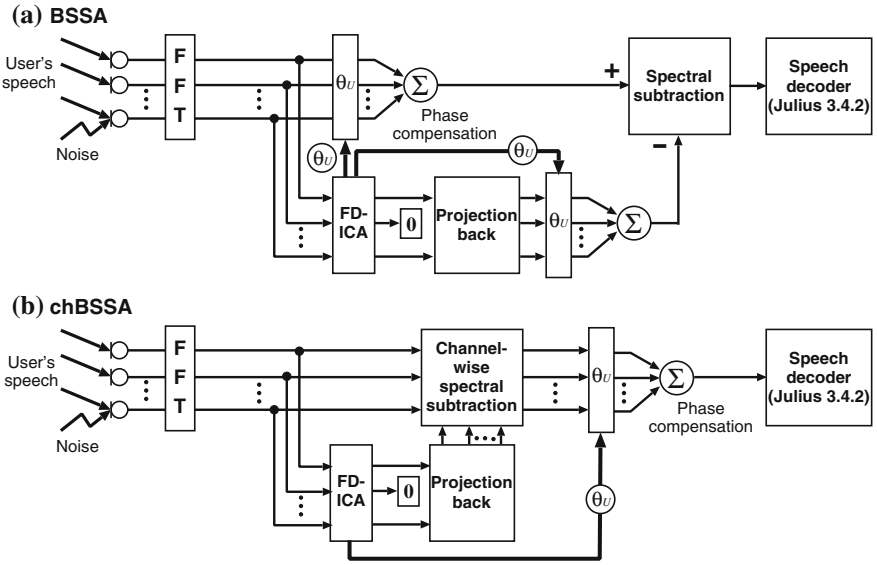


Fig. 10.1 Block diagrams of **a** SS after DS (BSSA) and **b** channelwise SS before DS (chBSSA)

the various types of BSSA as a *structure-generalized parametric BSSA* [27], and we provide a theoretical analysis of the amounts of musical noise and speech distortion generated in several types of methods using the structure-generalized parametric BSSA. From a mathematical analysis based on higher-order statistics, we prove the existence of a tradeoff between the amounts of musical noise and speech distortion in various BSSA structures. From experimental evaluations, it is revealed that the structure should be carefully selected according to the application, i.e., a chBSSA structure is recommended for listening but a conventional BSSA is more suitable for speech recognition.

The outline of this chapter is organized as follows. In Sect. 10.2, we provide a brief review of ICA used for speech applications [28, 29]. In Sect. 10.3, a theoretical analysis of ICA under nonpoint-source noise condition is given, and following this section, we give a review of BSSA and its generalized algorithms [23, 27] in Sect. 10.4. In Sect. 10.5, we describe a musical noise assessment method based on higher-order statistics [30–32]. Using the method, we give a theoretical analysis of musical noise generation and speech distortion for structure-generalized BSSA, where the authors can show that chBSSA is superior to BSSA in terms of less musical noise property, but BSSA is superior to chBSSA in terms of less speech distortion property [27]. In Sect. 10.6, we show results of experimental evaluation [27]. Following a discussion on the theoretical analysis and experimental results, we present our conclusions in Sect. 10.7.

10.2 Data Model and Conventional BSS Method

10.2.1 Sound Mixing Model of Microphone Array

In this chapter, a straight line array is assumed. The coordinates of the elements are designated $d_j (j = 1, \ldots, J)$, and the direction-of-arrivals (DOAs) of multiple sound sources are designated $\theta_k (k = 1, \ldots, K)$ (see Fig. 10.2). Here, we assume the following sound sources: only one target speech signal, some interference signals that can be regarded as point sources, and additive noise. This additive noise represents noises that cannot be regarded as point sources, e.g., spatially uncorrelated noises, background noises, and leakage of reverberation components outside the frame analysis. Multiple mixed signals are observed at microphone array elements, and a short-time analysis of the observed signals is conducted by frame-by-frame discrete Fourier transform (DFT). The observed signals are given by

$$\mathbf{x}(f, \tau) = \mathbf{A}(f) \{\mathbf{s}(f, \tau) + \mathbf{n}(f, \tau)\} + \mathbf{n}_a(f, \tau), \tag{10.1}$$

where f is the frequency bin and τ is the time index of DFT analysis. Also, $\mathbf{x}(f, \tau)$ is the observed signal vector, $\mathbf{A}(f)$ is the mixing matrix, $\mathbf{s}(f, \tau)$ is the target speech signal vector in which only the Uth entry contains the signal component $s_U(f, \tau)$ (U is the target source number), $\mathbf{n}(f, \tau)$ is the interference signal vector that contains the signal components except the Uth component, and $\mathbf{n}_a(f, \tau)$ is the nonstationary additive noise signal term that generally represents nonpoint-source noises. These are defined as

$$\mathbf{x}(f, \tau) = [x_1(f, \tau), \ldots, x_J(f, \tau)]^{\mathrm{T}}, \tag{10.2}$$

$$\mathbf{s}(f, \tau) = [\underbrace{0, \ldots, 0}_{U-1}, s_U(f, \tau), \underbrace{0, \ldots, 0}_{K-U}]^{\mathrm{T}}, \tag{10.3}$$

$$\mathbf{n}(f, \tau) = [n_1(f, \tau), \ldots, n_{U-1}(f, \tau), 0, n_{U+1}(f, \tau), \ldots, n_K(f, \tau)]^{\mathrm{T}}, \tag{10.4}$$

$$\mathbf{n}_a(f, \tau) = [n_1^{(a)}(f, \tau), \ldots, n_J^{(a)}(f, \tau)]^{\mathrm{T}}, \tag{10.5}$$

$$\mathbf{A}(f) = \begin{bmatrix} A_{11}(f) & \cdots & A_{1K}(f) \\ \vdots & & \vdots \\ A_{J1}(f) & \cdots & A_{JK}(f) \end{bmatrix}. \tag{10.6}$$

10.2.2 Conventional Frequency-Domain ICA

Here, we consider a case where the number of sound sources, K, equals the number of microphones, J, i.e., $J = K$. In addition, similarly to that in the case of the conventional ICA contexts, we assume that the additive noise $\mathbf{n}_a(f, \tau)$ is negligible in (10.1). In frequency-domain ICA (FDICA), signal separation is expressed as

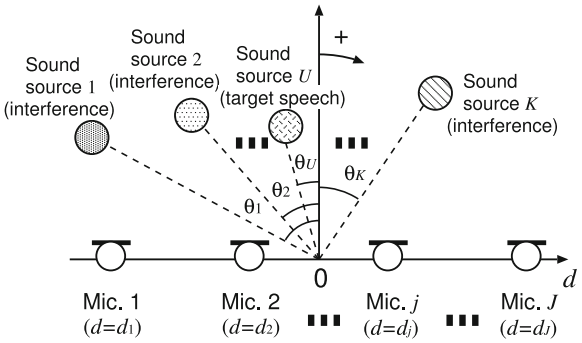


Fig. 10.2 Configurations of microphone array and signals

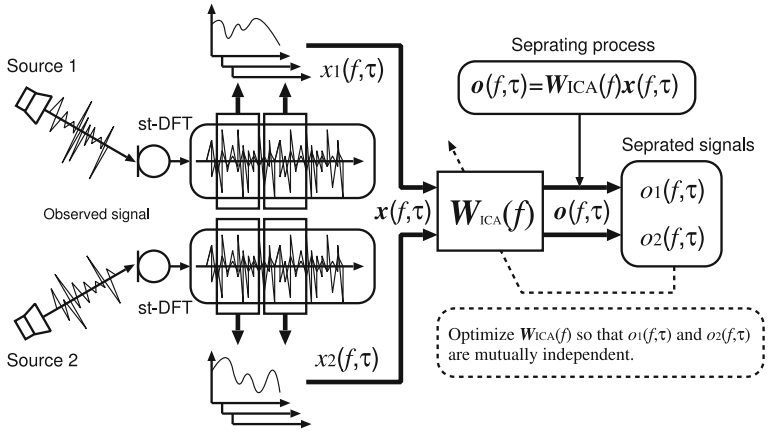


Fig. 10.3 Blind source separation procedure in FDICA in case of $J = K = 2$

$$\mathbf{o}(f,\tau) = [o_1(f,\tau),\ldots,o_K(f,\tau)]^{\mathrm{T}} = \mathbf{W}_{\mathrm{ICA}}(f)\mathbf{x}(f,\tau), \tag{10.7}$$

$$\mathbf{W}_{\mathrm{ICA}}(f) = \begin{bmatrix} W_{11}^{(\mathrm{ICA})}(f) & \cdots & W_{1J}^{(\mathrm{ICA})}(f) \\ \vdots & & \vdots \\ W_{K1}^{(\mathrm{ICA})}(f) & \cdots & W_{KJ}^{(\mathrm{ICA})}(f) \end{bmatrix}, \tag{10.8}$$

where $\mathbf{o}(f,\tau)$ is the resultant output of the separation and $\mathbf{W}_{\mathrm{ICA}}(f)$ is the complex-valued unmixing matrix (see Fig. 10.3).

The unmixing matrix $\mathbf{W}_{\mathrm{ICA}}(f)$ is optimized by ICA so that the output entries of $\mathbf{o}(f,\tau)$ become mutually independent. Indeed, many kinds of ICA algorithms have been proposed. In the second-order ICA (SO-ICA) [17, 19], the separation filter is optimized by the joint diagonalization of co-spectra matrices using the nonstationarity and coloration of the signal. For instance, the following iterative updating equation based on SO-ICA has been proposed by Parra and Spence [17]:

$$\mathbf{W}_{\mathrm{ICA}}^{[p+1]}(f) = -\mu \sum_{\tau_b} \chi(f)\,\mathrm{off\text{-}diag}\left(\mathbf{R}_{oo}(f,\tau_b)\right)\mathbf{W}_{\mathrm{ICA}}^{[p]}(f)\mathbf{R}_{xx}(f,\tau_b) + \mathbf{W}_{\mathrm{ICA}}^{[p]}(f), \tag{10.9}$$

where μ is the step-size parameter, $[p]$ is used to express the value of the pth step in iterations, off-diag[$\mathbf{X}$] is the operation for setting every diagonal element of matrix $\mathbf{X}$ to zero, and $\chi(f) = (\sum_{\tau_b} \|\mathbf{R}_{xx}(f,\tau_b)\|^2)^{-1}$ is a normalization factor ($\|\cdot\|$ represents the Frobenius norm). $\mathbf{R}_{xx}(f,\tau_b)$ and $\mathbf{R}_{oo}(f,\tau_b)$ are the cross-power spectra of the input $\mathbf{x}(f,\tau)$ and output $\mathbf{o}(f,\tau)$, respectively, which are calculated around multiple time blocks τ_b. Also, Pham et al. have proposed the following improved criterion for SO-ICA [19]:

$$\sum_{\tau_b} \left\{ \frac{1}{2} \log\det \mathrm{diag}\left[\mathbf{W}_{\mathrm{ICA}}(f)\mathbf{R}_{oo}(f,\tau_b)\mathbf{W}_{\mathrm{ICA}}(f)^{\mathrm{H}}\right] - \log\det\left[\mathbf{W}_{\mathrm{ICA}}(f)\right] \right\}, \tag{10.10}$$

where the superscript H denotes Hermitian transposition. This criterion is to be minimized with respect to $\mathbf{W}_{\mathrm{ICA}}(f)$. Another possible way to achieve SO-ICA has been proposed as the direct joint diagonalization based on the linear algebraic procedure [33, 34].

On the other hand, a higher-order statistics-based approach exists. In higher-order ICA (HO-ICA), the separation filter is optimized on the basis of the non-Gaussianity of the signal. The optimal $\mathbf{W}_{\mathrm{ICA}}(f)$ in HO-ICA is obtained using the iterative equation

$$\mathbf{W}_{\mathrm{ICA}}^{[p+1]}(f) = \mu[\mathbf{I} - \langle \boldsymbol{\varphi}(\mathbf{o}(f,\tau))\mathbf{o}^{\mathrm{H}}(f,\tau)\rangle_\tau]\mathbf{W}_{\mathrm{ICA}}^{[p]}(f) + \mathbf{W}_{\mathrm{ICA}}^{[p]}(f), \tag{10.11}$$

where $\mathbf{I}$ is the identity matrix, $\langle\cdot\rangle_\tau$ denotes the time-averaging operator, and $\boldsymbol{\varphi}(\cdot)$ is the nonlinear vector function. Many kinds of nonlinear function $\boldsymbol{\varphi}(f,\tau)$ have been proposed. Considering a batch algorithm of ICA, it is well known that $\tanh(\cdot)$ or the sigmoid function is appropriate for super-Gaussian sources such as speech signals [35, 36]. In this study, we define the nonlinear vector function $\boldsymbol{\varphi}(\cdot)$ as

$$\boldsymbol{\varphi}(\mathbf{o}(f,\tau)) \equiv [\varphi(o_1(f,\tau)), \ldots, \varphi(o_K(f,\tau))]^{\mathrm{T}}, \tag{10.12}$$

$$\varphi(o_k(f,\tau)) \equiv \tanh o_k^{(\mathrm{R})}(f,\tau) + i \tanh o_k^{(\mathrm{I})}(f,\tau), \tag{10.13}$$

where the superscripts (R) and (I) denote the real and imaginary parts, respectively. The nonlinear function given by (10.12) indicates that the nonlinearity is applied to the real and imaginary parts of complex-valued signals separately. This type of complex-valued nonlinear function has been introduced by Smaragdis [16] for FDICA, where it can be assumed for speech signals that the real (or imaginary) parts of the time–frequency representations of sources are mutually independent. According to Refs. [21, 37], the source separation performance of HO-ICA is almost the

same as or superior to that of SO-ICA. Thus, in this chapter, HO-ICA is utilized as the basic ICA algorithm hereafter.

FDICA has the inherent problem so-called *permutation problem*, i.e., difficulty in removing the ambiguity of the source order in each frequency subband. In the context of the permutation problem in the ICA study, there exist many methods for solving the permutation problem, such as the source DOA-based method [38], subband correlation-based method [15], and their combination method [39]. The definite way to avoid the permutation problem is to use time-domain ICA (TDICA), which has, however, other problems like relatively slow convergence and complex implementation. Several literatures can be available for understanding the difference and comparison between TDICA and FDICA [40–42].

10.3 Analysis of ICA Under Nonpoint-source Noise Condition

In this section, we investigate the proficiency of ICA under a nonpoint-source noise condition. In relation to the performance analysis of ICA, Araki et al. have reported that ICA-based BSS has equivalence to parallel constructed ABFs [43, 44]. However, this investigation was focused on separation with a nonsingular mixing matrix, and thus was valid for only point sources.

First, we analyze beamformers that are optimized by ICA under a nonpoint-source condition. In the analysis, it is clarified that beamformers optimized by ICA become specific beamformers that maximize the signal-to-noise ratio (SNR) in each output (so-called *SNR-maximize beamformers*). In particular, the beamformer for target speech estimation is optimized to be a DS beamformer, and the beamformer for noise estimation is likely to be a null beamformer (NBF) [18].

Next, a computer simulation is conducted. Its result also indicates that ICA is proficient in noise estimation under a nonpoint-source noise condition. Then, it is concluded that ICA is suitable for noise estimation under such a condition.

10.3.1 Can ICA Separate Any Source Signals?

Many previous studies on BSS provided strong evidence that conventional ICA could perform source separation, particularly in the special case of speech–speech mixing, i.e., all sound sources are point sources. However, such sound mixing is not realistic under common acoustic conditions; indeed the following scenario and problem are likely to arise (see Fig. 10.4):

- The target sound is the user's speech, which can be approximately regarded as a *point source*. In addition, the users themselves locate relatively *near the microphone array* (e.g., 1 m apart), and consequently the accompanying reflection and reverberation components are moderate.

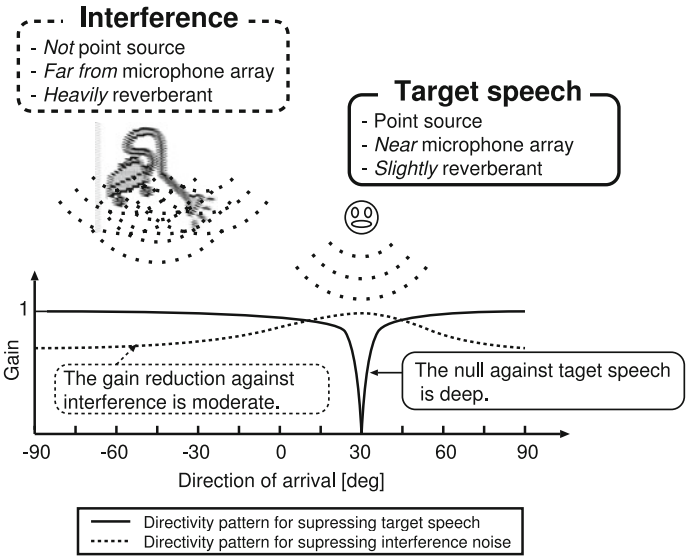


Fig. 10.4 Expected directivity patterns that are shaped by ICA

- For the noise, we are often confronted with interference sound(s) which is *not a point source* but a widespread source. Also, the noise is usually far from the array and is heavily reverberant.

In such an environment, can ICA separate the user's speech signal and a widespread noise signal? The answer is *no*. It is well expected that conventional ICA can suppress the user's speech signal to pick up the noise source, but ICA is very weak in picking up the target speech itself via the suppression of a distant widespread noise. This is due to the fact that ICA with small numbers of sensors and filter taps often provides only directional nulls against undesired source signals. Results of the detailed analysis of ICA for such a case are shown in the following subsections.

10.3.2 SNR-Maximize Beamformers Optimized by ICA

In this subsection, we consider beamformers that are optimized by ICA in the following acoustic scenario: the target signal is the user's speech and the noise is not a point source. Then, the observed signal contains only one target speech signal and an additive noise. In this scenario, the observed signal is defined as

$$\mathbf{x}(f,\tau) = \mathbf{A}(f)\mathbf{s}(f,\tau) + \mathbf{n}_a(f,\tau). \tag{10.14}$$

Note that the additive noise $\mathbf{n}_a(f,\tau)$ cannot be negligible in this scenario. Then, the output of ICA contains two components, i.e., the estimated speech signal $y_s(f,\tau)$ and estimated noise signal $y_n(f,\tau)$; these are given by

$$[y_s(f,\tau), y_n(f,\tau)]^{\mathrm{T}} = \mathbf{W}_{\mathrm{ICA}}(f)\mathbf{x}(f,\tau). \tag{10.15}$$

Therefore, ICA optimizes two beamformers; these can be written as

$$\mathbf{W}_{\mathrm{ICA}}(f) = [\mathbf{g}_s(f), \mathbf{g}_n(f)]^{\mathrm{T}}, \tag{10.16}$$

where $\mathbf{g}_s(f) = \left[g_1^{(s)}(f), \ldots, g_J^{(s)}(f)\right]^{\mathrm{T}}$ is the coefficient vector of the beamformer used to pick up the target speech signal, and $\mathbf{g}_n(f) = \left[g_1^{(n)}(f), \ldots, g_J^{(n)}(f)\right]^{\mathrm{T}}$ is the coefficient vector of the beamformer used to pick up the noise. Therefore, (10.15) can be rewritten as

$$[y_s(f,\tau), y_n(f,\tau)]^{\mathrm{T}} = [\mathbf{g}_s(f), \mathbf{g}_n(f)]^{\mathrm{T}}\mathbf{x}(f,\tau). \tag{10.17}$$

In SO-ICA, the multiple second-order correlation matrices of distinct time block outputs,

$$\langle \mathbf{o}(f,\tau_b)\mathbf{o}^{\mathrm{H}}(f,\tau_b)\rangle_{\tau_b}, \tag{10.18}$$

are diagonalized through joint diagonalization.

On the other hand, in HO-ICA, the higher-order correlation matrix is also diagonalized. Using the Taylor expansion, we can express the factor of the nonlinear vector function of HO-ICA, $\varphi(o_k(f,\tau))$, as

$$\begin{aligned}
\varphi(o_k(f,\tau)) &= \tanh o_k^{(\mathrm{R})}(f,\tau) + i\tanh o_k^{(\mathrm{I})}(f,\tau), \\
&= \left\{ o_k^{(\mathrm{R})}(f,\tau) - \frac{\left(o_k^{(\mathrm{R})}(f,\tau)\right)^3}{3} + \cdots \right\} \\
&\quad + i\left\{ o_k^{(\mathrm{I})}(f,\tau) - \frac{\left(o_k^{(\mathrm{I})}(f,\tau)\right)^3}{3} + \cdots \right\}, \\
&= o_k(f,\tau) - \left(\frac{\left(o_k^{(\mathrm{R})}(f,\tau)\right)^3}{3} + i\frac{\left(o_k^{(\mathrm{I})}(f,\tau)\right)^3}{3} \right) + \cdots .
\end{aligned} \tag{10.19}$$

Thus, the calculation of the higher-order correlation in HO-ICA, $\boldsymbol{\varphi}(\mathbf{o}(f,\tau))\mathbf{o}^{\mathrm{H}}(f,\tau)$, can be decomposed to a second-order correlation matrix and the summation of higher-order correlation matrices of each order. This is shown as

$$\langle \boldsymbol{\varphi}(\mathbf{o}(f,\tau))\mathbf{o}^{\mathrm{H}}(f,\tau)\rangle_\tau = \langle \mathbf{o}(f,\tau)\mathbf{o}^{\mathrm{H}}(f,\tau)\rangle_\tau + \Psi(f), \tag{10.20}$$

where $\Psi(f)$ is a set of higher-order correlation matrices. In HO-ICA, separation filters are optimized so that all orders of correlation matrices become diagonal matrices. Then, at least the second-order correlation matrix is diagonalized by HO-ICA. In both SO-ICA and HO-ICA, at least the second-order correlation matrix is diagonalized. Hence, we prove in the following that ICA optimizes beamformers as SNR-maximize beamformers focusing on only part of the second-order correlation. Then the absolute value of the normalized cross-correlation coefficient (off-diagonal entries) of the second-order correlation, C, is defined by

$$C = \frac{\left|\langle y_s(f,\tau)y_n^*(f,\tau)\rangle_\tau\right|}{\sqrt{\langle |y_s(f,\tau)|^2\rangle_\tau}\sqrt{\langle |y_n(f,\tau)|^2\rangle_\tau}}, \tag{10.21}$$

$$y_s(f,\tau) = \hat{s}(f,\tau) + r_s\hat{n}(f,\tau), \tag{10.22}$$

$$y_n(f,\tau) = \hat{n}(f,\tau) + r_n\hat{s}(f,\tau), \tag{10.23}$$

where $\hat{s}(f,\tau)$ is the target speech component in ICA's output, $\hat{n}(f,\tau)$ is the noise component in ICA's output, r_s is the coefficient of the residual noise component, r_n is the coefficient of the target-leakage component, and the superscript $*$ represents a complex conjugate. Therefore, the SNRs of $y_s(f,\tau)$ and $y_n(f,\tau)$ can be respectively represented by

$$\Gamma_s = \langle |\hat{s}(f,\tau)|^2\rangle_\tau / (|r_s|^2\langle |\hat{n}(f,\tau)|^2\rangle_\tau), \tag{10.24}$$

$$\Gamma_n = \langle |\hat{n}(f,\tau)|^2\rangle_\tau / (|r_n|^2\langle |\hat{s}(f,\tau)|^2\rangle_\tau), \tag{10.25}$$

where Γ_s is the SNR of $y_s(f,\tau)$ and Γ_n is the SNR of $y_n(f,\tau)$. Using (10.22)–(10.25), we can rewrite (10.21) as

$$C = \frac{\left|1/\sqrt{\Gamma_s}\cdot e^{j\arg r_s} + 1/\sqrt{\Gamma_n}\cdot e^{j\arg r_n^*}\right|}{\sqrt{1+1/\Gamma_s}\sqrt{1+1/\Gamma_n}} = \frac{\left|1/\sqrt{\Gamma_s} + 1/\sqrt{\Gamma_n}\cdot e^{j(\arg r_n^* - \arg r_s)}\right|}{\sqrt{1+1/\Gamma_s}\sqrt{1+1/\Gamma_n}}, \tag{10.26}$$

where $\arg r$ represents the argument of r. Thus, C is a function of only Γ_s and Γ_n. Therefore, the cross-correlation between $y_s(f,\tau)$ and $y_n(f,\tau)$ only depends on the SNRs of beamformers $\mathbf{g}_s(f)$ and $\mathbf{g}_n(f)$.

In Ref. [23], the following has been proved.

- The absolute value of cross-correlation only depends on the SNRs of the beamformers spanned by each row of an unmixing matrix.
- The absolute value of cross-correlation is a monotonically decreasing function of SNR.
- Therefore, the diagonalization of a second-order correlation matrix leads to SNR maximization.

Thus, it can be concluded that ICA, in a parallel manner, optimizes multiple beamformers, i.e., $\mathbf{g}_s(f)$ and $\mathbf{g}_n(f)$, so that the SNR of the output of each beamformer becomes maximum.

10.3.3 What Beamformers Are Optimized Under Nonpoint-source Noise Condition?

In the previous subsection, it has been proved that ICA optimizes beamformers as SNR-maximize beamformers. In this subsection, we analyze what beamformers are optimized by ICA, particularly under a nonpoint-source noise condition, where we assume a two-source separation problem. The target speech can be regarded as a point source, and the noise is a nonpoint-source noise. First, we focus on the beamformer $\mathbf{g}_s(f)$ that picks up the target speech signal. The SNR-maximize beamformer for $\mathbf{g}_s(f)$ minimizes the undesired signal's power under the condition that the target signal's gain is kept constant. Thus, the desired beamformer should satisfy

$$\min_{\mathbf{g}_s(f)} \mathbf{g}_s^{\mathrm{T}}(f)\mathbf{R}(f)\mathbf{g}_s(f) \quad \text{subject to } \mathbf{g}_s^{\mathrm{T}}(f)\mathbf{a}(f,\theta_s) = 1, \tag{10.27}$$

$$\mathbf{a}(f,\theta_s(f)) = [\exp(i2\pi(f/M)f_{\mathrm{s}}d_1 \sin\theta_s/c), \ldots, \exp(i2\pi(f/M)f_{\mathrm{s}}d_J \sin\theta_s/c)]^{\mathrm{T}}, \tag{10.28}$$

where $\mathbf{a}(f,\theta_s(f))$ is the steering vector, $\theta_s(f)$ is the direction of the target speech, M is the DFT size, f_{s} is the sampling frequency, c is the sound velocity, and $\mathbf{R}(f) = \langle \mathbf{n}_a(f,\tau)\mathbf{n}_a^{\mathrm{H}}(f,\tau)\rangle_\tau$ is the correlation matrix of $\mathbf{n}_a(f,\tau)$. Note that $\theta_s(f)$ is a function of frequency because the DOA of the source varies in each frequency subband under a reverberant condition. Here, using the Lagrange multiplier, the solution of (10.27) is

$$\mathbf{g}_s(f)^{\mathrm{T}} = \frac{\mathbf{a}(f,\theta_s(f))^{\mathrm{H}}\mathbf{R}^{-1}(f)}{\mathbf{a}(f,\theta_s(f))^{\mathrm{H}}\mathbf{R}^{-1}(f)\mathbf{a}(f,\theta_s(f))}. \tag{10.29}$$

This beamformer is called a minimum variance distortionless response (MVDR) beamformer [45]. Note that the MVDR beamformer requires the true DOA of the target speech and the noise-only time interval. However, we cannot determine the true DOA of the target source signal and the noise-only interval because ICA is an *unsupervised* adaptive technique. Thus, the MVDR beamformer is expected to be the upper limit of ICA in the presence of nonpoint-source noises.

Although the correlation matrix is often not diagonalized in lower frequency subbands [45], e.g., diffuse noise, we approximate that the correlation matrix is almost diagonalized in subbands in the entire frequency. Then, regarding the power of noise signals as approximately $\delta^2(f)$, the correlation matrix results in $\mathbf{R}(f) = \delta^2(f)\cdot\mathbf{I}$. Therefore, the inverse of the correlation matrix $\mathbf{R}^{-1}(f) = \mathbf{I}/\delta^2(f)$ and (10.29) can be rewritten as

$$\mathbf{g}_s(f)^{\mathrm{T}} = \frac{\mathbf{a}(f,\theta_s(f))^{\mathrm{H}}}{\mathbf{a}(f,\theta_s(f))^{\mathrm{H}}\mathbf{a}(f,\theta_s(f))}. \tag{10.30}$$

Since $\mathbf{a}(f,\theta_s(f))^{\mathrm{H}}\mathbf{a}(f,\theta_s(f)) = J$, we finally obtain

$$\begin{aligned}&\mathbf{g}_s(f)\\ &= \frac{1}{J}[\exp(-i2\pi(f/M)f_{\mathrm{s}}d_1\sin\theta_s(f)/c),\ldots,\exp(-i2\pi(f/M)f_{\mathrm{s}}d_J\sin\theta_s(f)/c)]^{\mathrm{T}}.\end{aligned} \tag{10.31}$$

This filter $\mathbf{g}_s(f)$ is approximately equal to a DS beamformer [4]. Note that the filter $\mathbf{g}_s(f)$ is not a simple DS beamformer but a *reverberation-adapted DS beamformer* because it is optimized for a distinct $\theta_s(f)$ in each frequency bin. The resultant noise power is $\delta^2(f)/J$ when the noise is spatially uncorrelated and white Gaussian. Consequently the noise reduction performance of the DS beamformer optimized by ICA under a nonpoint-source noise condition is proportional to $10\log_{10} J$ [dB]; this performance is not particularly good.

Next, we consider the other beamformer $\mathbf{g}_n(f)$, which picks up the noise source. Similar to the noise signal, the beamformer that removes the target signal arriving from $\theta_s(f)$ is the SNR-maximize beamformer. Thus, the beamformer that steers the directional null to $\theta_s(f)$ is the desired one for the noise signal. Such a beamformer is called NBF [18]. This beamformer compensates for the phase of the signal arriving from $\theta_s(f)$, and carries out subtraction. Thus, the signal arriving from $\theta_s(f)$ is removed. For instance, NBF with a two-element array is designed as

$$\begin{aligned}&\mathbf{g}_n(f)\\ &= [\exp(-i2\pi(f/M)f_{\mathrm{s}}d_1\sin\theta_s(f)/c), -\exp(-i2\pi(f/M)f_{\mathrm{s}}d_2\sin\theta_s(f)/c)]^{\mathrm{T}}\cdot\sigma(f),\end{aligned} \tag{10.32}$$

where $\sigma(f)$ is the gain compensation parameter. This beamformer surely satisfies $\mathbf{g}_n^{\mathrm{T}}(f)\cdot\mathbf{a}(f,\theta_s(f)) = 0$. The steering vector $\mathbf{a}(f,\theta_s(f))$ expresses the wavefront of the plane wave arriving from $\theta_s(f)$. Thus, $\mathbf{g}_n(f)$ actually steers the directional null to $\theta_s(f)$. Note that this always occurs regardless of the number of microphones (at least two microphones). Hence, this beamformer achieves a reasonably high, ideally infinite, SNR for the noise signal. Also, note that the filter $\mathbf{g}_n(f)$ is not a simple NBF but a *reverberation-adapted NBF* because it is optimized for a distinct $\theta_s(f)$ in each frequency bin. Overall, the performance of enhancing the target speech is very poor but that of estimating the noise source is good.

10.3.4 Computer Simulations

We conduct computer simulations to confirm the performance of ICA under a nonpoint-source noise condition. Here, we used HO-ICA [16] as the ICA algorithm. We used the following 8 kHz-sampled signals as the ICA's input; the original target

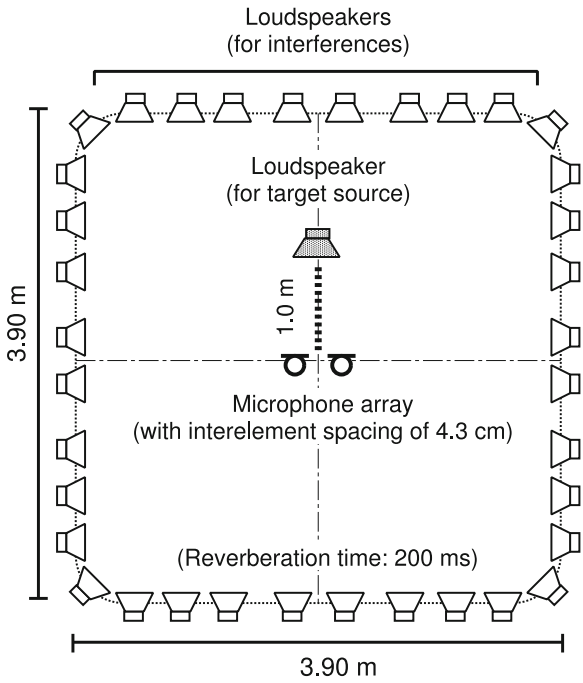


Fig. 10.5 Layout of reverberant room in our simulation

speech (3 s) was convoluted with impulse responses that were recorded in an actual environment, and to which three types of noise from 36 loudspeakers were added. The reverberation time (RT_{60}) is 200 ms; this corresponds to mixing filters with 1,600 taps in 8 kHz sampling. The three types of noise are an independent Gaussian noise, actually recorded railway station noise, and interference speech by 36 people. Figure 10.5 illustrates the reverberant room used in the simulation. We use 12 speakers (6 males and 6 females) as sources of the original target speech, and the input SNR of test data is set to 0 dB. We use a two-, three-, or four-element microphone array with an interelement spacing of 4.3 cm.

The simulation results are shown in Figs. 10.6 and 10.7. Figure 10.6 shows the result for the average noise reduction rate (NRR) [18] of all the target speakers. NRR is defined as the output SNR in dB minus the input SNR in dB. This measure indicates the objective performance of noise reduction. NRR is given by

$$\text{NRR [dB]} = \frac{1}{J}\sum_{j=1}^{J}(\text{OSNR} - \text{ISNR}_j), \tag{10.33}$$

where OSNR is the output SNR and ISNR_j is the input SNR of microphone j.

From this result, we can see an imbalance between the target speech estimation and the noise estimation in every noise case; the performance of the target speech estimation is significantly poor, but that of noise estimation is very high. This result

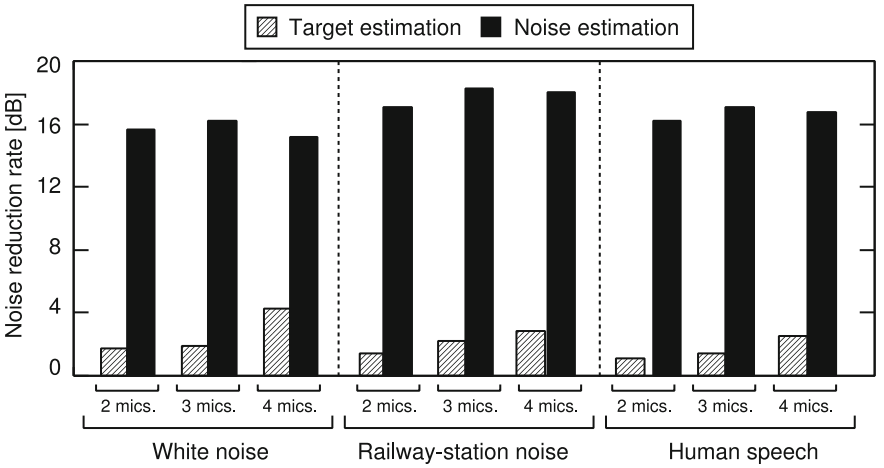
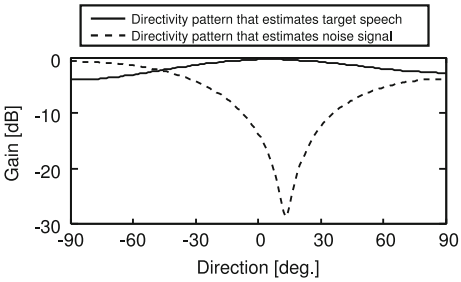


Fig. 10.6 Simulation-based separation results under nonpoint-source noise condition

Fig. 10.7 Typical directivity patterns under nonpoint-source noise condition shaped by ICA at 2 kHz and two-element array for case of white Gaussian noise



is consistent with the previously stated theory. Moreover, Fig. 10.7 shows directivity patterns shaped by the beamformers optimized by ICA in the simulation. It is clearly indicated that beamformer $\mathbf{g}_s(f)$, which picks up the target speech, resembles the DS beamformer, and that beamformer $\mathbf{g}_n(f)$, which picks up the noise, becomes NBF. From these results, it is confirmed that the previously stated theory, i.e., the beamformers optimized by ICA under a nonpoint-source noise condition are DS and NBF, is valid.

10.4 Blind Spectral Subtraction Array

10.4.1 Motivation and Strategy

As clearly shown in Sects. 10.3.3 and 10.3.4, ICA is proficient in noise estimation rather than in target speech estimation under a nonpoint-source noise condition. Thus, we cannot use ICA for direct target estimation under such a condition. However, we can still use ICA as a noise estimator. This motivates us to introduce an improved

speech-enhancement strategy, i.e., BSSA [23]. BSSA consists of a DS-based primary path and a reference path including ICA-based noise estimation (see Fig. 10.1a). The estimated noise component in ICA is efficiently subtracted from the primary path in the power spectrum domain without phase information. This procedure can yield better target speech enhancement than simple ICA, even with the additional benefit of estimation-error robustness in speech recognition applications. The detailed process of signal processing is shown below.

10.4.2 Partial Speech Enhancement in Primary Path

We again consider the generalized form of the observed signal as described in (10.1). The target speech signal is partly enhanced in advance by DS. This procedure can be given as

$$\begin{aligned} y_{\mathrm{DS}}(f,\tau) &= \mathbf{w}_{\mathrm{DS}}^{\mathrm{T}}(f)\mathbf{x}(f,\tau) \\ &= \mathbf{w}_{\mathrm{DS}}^{\mathrm{T}}(f)\mathbf{A}(f)\mathbf{s}(f,\tau) + \mathbf{w}_{\mathrm{DS}}^{\mathrm{T}}(f)\mathbf{A}(f)\mathbf{n}(f,\tau) + \mathbf{w}_{\mathrm{DS}}^{\mathrm{T}}(f)\mathbf{n}_a(f,\tau), \end{aligned} \tag{10.34}$$

$$\mathbf{w}_{\mathrm{DS}} = [w_1^{(\mathrm{DS})}(f), \ldots, w_J^{(\mathrm{DS})}(f)]^{\mathrm{T}}, \tag{10.35}$$

$$w_j^{(\mathrm{DS})}(f) = \frac{1}{J}\exp\left(-i2\pi(f/M)f_s d_j \sin\theta_U/c\right), \tag{10.36}$$

where $y_{\mathrm{DS}}(f,\tau)$ is the primary path output that is a slightly enhanced target speech, $\mathbf{w}_{\mathrm{DS}}(f)$ is the filter coefficient vector of DS, and θ_U is the estimated DOA of the target speech given by the ICA part in Sect. 10.4.3. In (10.34), the second and third terms on the right-hand side express the remaining noise in the output of the primary path.

10.4.3 ICA-Based Noise Estimation in Reference Path

BSSA provides ICA-based noise estimation. First, we separate the observed signal by ICA and obtain the separated signal vector $\mathbf{o}(f,\tau)$ as

$$\mathbf{o}(f,\tau) = \mathbf{W}_{\mathrm{ICA}}(f)\mathbf{x}(f,\tau), \tag{10.37}$$

$$\mathbf{o}(f,\tau) = [o_1(f,\tau), \ldots, o_{K+1}(f,\tau)]^{\mathrm{T}}, \tag{10.38}$$

$$\mathbf{W}_{\mathrm{ICA}}(f) = \begin{bmatrix} W_{11}^{(\mathrm{ICA})}(f) & \cdots & W_{1J}^{(\mathrm{ICA})}(f) \\ \vdots & & \vdots \\ W_{(K+1)1}^{(\mathrm{ICA})}(f) & \cdots & W_{(K+1)J}^{(\mathrm{ICA})}(f) \end{bmatrix}, \tag{10.39}$$

where the unmixing matrix $\mathbf{W}_{\mathrm{ICA}}(f)$ is optimized by (10.11) . Note that the number of ICA outputs becomes $K+1$, and thus the number of sensors, J, is more than $K+1$ because we assume that the additive noise $\mathbf{n}_a(f,\tau)$ is not negligible. We cannot estimate the additive noise perfectly because it is deformed by the filter optimized by ICA. Moreover, other components also cannot be estimated perfectly when the additive noise $\mathbf{n}_a(f,\tau)$ exists. However, we can estimate at least noises (including interference sounds that can be regarded as point sources, and the additive noise) that do not involve the target speech signal, as indicated in Sect. 10.3. Therefore, the estimated noise signal is still beneficial.

Next, we estimate DOAs from the unmixing matrix $\mathbf{W}_{\mathrm{ICA}}(f)$ [18]. This procedure is represented by

$$\theta_u = \sin^{-1} \frac{\arg\left(\frac{[\mathbf{W}_{\mathrm{ICA}}^{-1}(f)]ju}{[\mathbf{W}_{\mathrm{ICA}}^{-1}(f)]j'u}\right)}{2\pi f_s c^{-1}(d_j - d_{j'})}, \tag{10.40}$$

where θ_u is the DOA of the uth sound source. Then, we choose the Uth source signal, which is nearest to the front of the microphone array, and designate the DOA of the chosen source signal as θ_U. This is because almost all users are expected to stand in front of the microphone array in a speech-oriented human–machine interface, e.g., a public guidance system. Other strategies for choosing the target speech signal can be considered as follows.

- If the approximate location of a target speaker is known in advance, we can utilize the location of the target speaker. For instance, we can know the approximate location of the target speaker at a hands-free speech recognition system in a car navigation system in advance. Then, the DOA of the target speech signal is approximately known. For such systems, we can choose the target speech signal, selecting the specific component in which the DOA estimated by ICA is nearest to the known target speech DOA.
- For an interaction robot system [46], we can utilize image information from a camera mounted on a robot. Therefore, we can estimate DOA from this information, and we can choose the target speech signal on the basis of this estimated DOA.
- If the only target signal is speech, i.e., none of the noises are speech, we can choose the target speech signal on the basis of the Gaussian mixture model (GMM), which can classify sound signals into voices and nonvoices [47].

Next, in the reference path, no target speech signal is required because we want to estimate only noise. Therefore, we eliminate the user's signal from the ICA's output signal $\mathbf{o}(f,\tau)$. This can be written as

$$\mathbf{q}(f,\tau) = \left[o_1(f,\tau), ..., o_{U-1}(f,\tau), 0, o_{U+1}(f,\tau), ..., o_{K+1}(f,\tau)\right]^{\mathrm{T}}, \tag{10.41}$$

where $\mathbf{q}(f,\tau)$ is the "noise-only" signal vector that contains only noise components. Next, we apply the projection back (PB) [15] method to remove the ambiguity of amplitude. This procedure can be represented as

$$\hat{\mathbf{q}}(f,\tau) = \left[\hat{q}_1(f,\tau), ..., \hat{q}_J(f,\tau)\right]^{\mathrm{T}} = \mathbf{W}^{+}_{\mathrm{ICA}}(f)\mathbf{q}(f,\tau), \tag{10.42}$$

where $\mathbf{M}^{+}$ denotes the Moore–Penrose pseudoinverse matrix of $\mathbf{M}$. Thus, $\hat{\mathbf{q}}(f,\tau)$ is a good estimate of the noise signals received at the microphone positions, i.e.,

$$\hat{\mathbf{q}}(f,\tau) \simeq \mathbf{A}(f)\mathbf{n}(f,\tau) + \mathbf{W}^{+}_{\mathrm{ICA}}(f)\hat{\mathbf{n}}_a(f,\tau), \tag{10.43}$$

where $\hat{\mathbf{n}}_a(f,\tau)$ contains the deformed additive noise signal and separation error due to an additive noise. Finally, we construct the estimated noise signal $z_{\mathrm{DS}}(f,\tau)$ by applying DS as

$$z_{\mathrm{DS}}(f,\tau) = \mathbf{w}^{\mathrm{T}}_{\mathrm{DS}}(f)\hat{\mathbf{q}}(f,\tau) \simeq \mathbf{w}^{\mathrm{T}}_{\mathrm{DS}}(f)\mathbf{A}(f)\mathbf{n}(f,\tau) + \mathbf{w}^{\mathrm{T}}_{\mathrm{DS}}(f)\mathbf{W}^{+}_{\mathrm{ICA}}(f)\hat{\mathbf{n}}_a(f,\tau). \tag{10.44}$$

This equation means that $z_{\mathrm{DS}}(f,\tau)$ is a good candidate for noise terms of the primary path output $y_{\mathrm{DS}}(f,\tau)$ (see the 2nd and 3rd terms on the right-hand side of (10.34)). Of course this noise estimation is not perfect, but we can still enhance the target speech signal via *oversubtraction* in the amplitude or power spectrum domain, where the overestimated noise component is subtracted from the observed noisy speech component with an allowance of speech distortion, as described in Sect. 10.4.4. Note that $z_{\mathrm{DS}}(f,\tau)$ is a function of the frame index τ, unlike the constant noise prototype in the traditional SS method [24]. Therefore, the proposed BSSA can deal with *nonstationary* noise.

10.4.4 Formulation of Structure-Generalized Parametric BSSA

In a recent study, two types of BSSA have been proposed (see Fig. 10.1). One is the conventional BSSA structure that performs SS after DS (see Fig. 10.1a), and the other involves channelwise SS before DS (chBSSA; see Fig. 10.1b). Also, it has been theoretically clarified that chBSSA is superior to BSSA for the mitigation of the musical noise generation [26]. In this chapter, we generalize the various types of BSSA as a *structure-generalized parametric BSSA* [27].

First, parametric BSSA is described. Using (10.34) and (10.44), we perform generalized SS (GSS) [48] and obtain the enhanced target speech signal as

$$y_{\mathrm{BSSA}}(f,\tau) = \begin{cases} \sqrt[2n]{|y_{\mathrm{DS}}(f,\tau)|^{2n} - \beta\overline{|z_{\mathrm{DS}}(f,\tau)|^{2n}}}\, e^{i\arg(y_{\mathrm{DS}}(f,\tau))} \\ (\text{if} \quad |y_{\mathrm{DS}}(f,\tau)|^{2n} - \beta\overline{|z_{\mathrm{DS}}(f,\tau)|^{2n}} > 0), \\ 0 \quad (\text{otherwise}), \end{cases} \tag{10.45}$$

where $y_{\mathrm{BSSA}}(f,\tau)$ is the final output of the parametric BSSA, β is an oversubtraction parameter, n is an exponent parameter, and $\overline{|z_{\mathrm{DS}}(f,\tau)|^{2n}}$ is the smoothed noise component within a certain time frame window.

Next, in the parametric chBSSA, we first perform GSS independently in each input channel and derive multiple enhanced target speech signals by channelwise GSS using (10.2) and (10.42). This procedure can be given by

$$y_j^{(\mathrm{chGSS})}(f,\tau) = \begin{cases} \sqrt[2n]{|x_j(f,\tau)|^{2n} - \beta\overline{|\hat{q}_j(f,\tau)|^{2n}}}\, e^{i\arg(x_j(f,\tau))} \\ (\text{if} \quad |x_j(f,\tau)|^{2n} - \beta\overline{|\hat{q}_j(f,\tau)|^{2n}} > 0), \\ 0 \quad (\text{otherwise}), \end{cases} \tag{10.46}$$

where $y_j^{(\mathrm{chGSS})}(f,\tau)$ is the enhanced target speech signal obtained by GSS at a specific channel j. Finally, we obtain the resultant-enhanced target speech signal by applying DS to $\mathbf{y}_{\mathrm{chGSS}} = [y_1^{(\mathrm{chGSS})}(f,\tau),\ldots,y_J^{(\mathrm{chGSS})}(f,\tau)]^{\mathrm{T}}$. This procedure can be expressed by

$$y_{\mathrm{chBSSA}}(f,\tau) = \mathbf{w}_{\mathrm{DS}}^{\mathrm{T}}(f)\mathbf{y}_{\mathrm{chGSS}}(f,\tau), \tag{10.47}$$

where $y_{\mathrm{chBSSA}}(f,\tau)$ is the final output of the parametric chBSSA.

10.5 Theoretical Analysis of Structure-Generalized Parametric BSSA

10.5.1 Motivation and Strategy

In general, BSSA can achieve good noise reduction performance but always suffers from artificial distortion, so-called musical noise, owing to its nonlinear signal processing. This leads to a serious tradeoff between the noise reduction performance and the amount of signal distortion in speech recognition. Therefore, in this chapter, we provide a theoretical analysis of the amounts of musical noise and speech distortion generated in several types of methods using the structure-generalized parametric BSSA. From a mathematical analysis based on higher-order statistics, we prove the existence of a tradeoff between the amounts of musical noise and speech distortion in various BSSA structures. From experimental evaluations, we reveal that the structure should be carefully selected according to the application, i.e., a chBSSA structure is recommended for listening but a conventional BSSA is more suitable for speech recognition.

In this chapter, we assume that the input signal x in the power spectral domain can be modeled by the gamma distribution as [49, 50]

$$P_{\mathrm{GM}}(x) = \frac{x^{\alpha-1}\exp(-\frac{x}{\theta})}{\theta^{\alpha}\Gamma(\alpha)}, \tag{10.48}$$

where α is the shape parameter corresponding to the type of the signal, θ is the scale parameter of the gamma distribution. In addition, $\Gamma(\alpha)$ is the *gamma function*, defined as

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}\exp(-t)\mathrm{d}t. \tag{10.49}$$

If the input signal is Gaussian, its complex-valued DFT coefficients also have the Gaussian distributions in the real and imaginary parts. Therefore, the p.d.f. of its power spectra obeys the chi-square distribution with two degrees of freedom, which corresponds to the gamma distribution with $\alpha = 1$. Also, if the input signal is super-Gaussian, the p.d.f. of its power spectra obeys the gamma distribution with $\alpha < 1$.

10.5.2 Analysis of Amount of Musical Noise

10.5.2.1 Metric of Musical Noise Generation: Kurtosis Ratio

We speculate that the amount of musical noise is highly correlated with the number of isolated power spectral components and their level of isolation (see Fig. 10.8). In this chapter, we call these isolated components *tonal components*. Since such tonal components have relatively high power, they are strongly related to the weight of the tail of their probability density function (p.d.f.). Therefore, quantifying the tail of the p.d.f. makes it possible to measure the number of tonal components. Thus, we adopt kurtosis, one of the most commonly used higher-order statistics, to evaluate the percentage of tonal components among all components. A larger kurtosis value indicates a signal with a heavy tail, meaning that the signal has many tonal components. Kurtosis is defined as

$$\mathrm{kurt} = \frac{\mu_4}{\mu_2^2}, \tag{10.50}$$

where "kurt" is the kurtosis and μ_m is the mth-order moment, given by

$$\mu_m = \int_0^{\infty} x^m P(x)\mathrm{d}x, \tag{10.51}$$

where $P(x)$ is the p.d.f. of the random variable X. Note that μ_m is not a central moment but a raw moment. Thus, (10.50) is not kurtosis in the mathematically strict

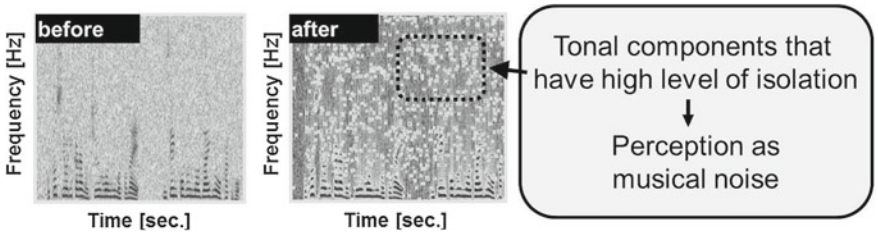


Fig. 10.8 Example of generation of tonal component after signal processing, where input signal is speech with white Gaussian noise and output is processed signal by GSS

definition but a modified version; however, we still refer to (10.50) as kurtosis in this chapter.

In this study, we apply such a kurtosis-based analysis to a time–frequency period of subject signals for the assessment of musical noise. Thus, this analysis should be conducted during, for example, periods of silence in speech when we evaluate the degree of musical noise arising in remaining noise. This is because we aim to quantify the tonal components arising in the noise-only part, which is the main cause of musical noise perception, and not in the target speech-dominant part.

Although kurtosis can be used to measure the number of tonal components, note that the kurtosis itself is not sufficient to measure the amount of musical noise. This is obvious since the kurtosis of some unprocessed noise signals, such as an interfering speech signal, is also high, but we do not recognize speech as musical noise. Hence, we turn our attention to the change in kurtosis between before and after signal processing to identify only the musical noise components. Thus, we adopt the *kurtosis ratio* as a measure to assess musical noise [30–32]. This measure is defined as

$$\text{kurtosis ratio} = \frac{\text{kurt}_{\text{proc}}}{\text{kurt}_{\text{org}}}, \tag{10.52}$$

where $\text{kurt}_{\text{proc}}$ is the kurtosis of the processed signal and kurt_{org} is the kurtosis of the original (unprocessed) signal. This measure increases as the amount of generated musical noise increases. In Ref. [30], it was reported that the kurtosis ratio is strongly correlated with the human perception of musical noise. Figure 10.9 shows an example of the relation between the kurtsis ratio (in log scale) and a human-perceptual score of degree of musical noise generation, where we can confirm the strong correlation.

10.5.2.2 Analysis in the Case of Parametric BSSA

In this section, we analyze the kurtosis ratio in a parametric BSSA. First, using the shape parameter of input noise α_{n}, we express the kurtosis of a gamma distribution, $\text{kurt}_{\text{in}}^{(\text{n})}$, as [51]

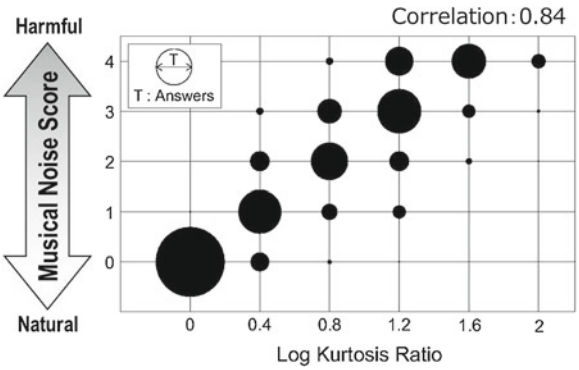


Fig. 10.9 Relation between kurtsis ratio (in log scale) and human-perceptual score of degree of musical noise generation [30]

$$\mathrm{kurt}_{\mathrm{in}}^{(\mathrm{n})} = \frac{\int_0^\infty x^4 P_{\mathrm{GM}}(x)\mathrm{d}x}{\left(\int_0^\infty x^2 P_{\mathrm{GM}}(x)\mathrm{d}x\right)^2} \tag{10.53}$$

$$= \frac{(\alpha_{\mathrm{n}}+2)(\alpha_{\mathrm{n}}+3)}{\alpha_{\mathrm{n}}(\alpha_{\mathrm{n}}+1)}. \tag{10.54}$$

The kurtosis in the power spectral domain after DS is given by [26]

$$\mathrm{kurt}_{\mathrm{DS}}^{(\mathrm{n})} \simeq J^{-0.7} \cdot (\mathrm{kurt}_{\mathrm{in}}^{(\mathrm{n})} - 6) + 6. \tag{10.55}$$

Similarly to (10.53), the shape parameter α_{DS} corresponding to the kurtosis after DS, $\mathrm{kurt}_{\mathrm{DS}}$, is given by solving the following equation in α_{DS}:

$$\mathrm{kurt}_{\mathrm{DS}}^{(\mathrm{n})} = \frac{(\alpha_{\mathrm{DS}}+2)(\alpha_{\mathrm{DS}}+3)}{\alpha_{\mathrm{DS}}(\alpha_{\mathrm{DS}}+1)}. \tag{10.56}$$

This can be expanded as

$$\alpha_{\mathrm{DS}}^2(\mathrm{kurt}_{\mathrm{DS}}^{(\mathrm{n})} - 1) + \alpha_{\mathrm{DS}}(\mathrm{kurt}_{\mathrm{DS}}^{(\mathrm{n})} - 5) - 6 = 0, \tag{10.57}$$

and we have

$$\alpha_{\mathrm{DS}} = \frac{-\mathrm{kurt}_{\mathrm{DS}} + 5 + \sqrt{\mathrm{kurt}_{\mathrm{DS}}^2 + 14\,\mathrm{kurt}_{\mathrm{DS}} + 1}}{2\,\mathrm{kurt}_{\mathrm{DS}} - 2}. \tag{10.58}$$

Then, using (10.53) and (10.55), α_{DS} can be expressed in terms of α_{n} as

$$\alpha_{\mathrm{DS}} = \left[2J^{-0.7} \cdot \left\{\frac{(\alpha_{\mathrm{n}}+2)(\alpha_{\mathrm{n}}+3)}{\alpha_{\mathrm{n}}(\alpha_{\mathrm{n}}+1)} - 6\right\} + 10\right]^{-1} \cdot \left[\left\{\left(J^{-0.7} \cdot \left\{\frac{(\alpha_{\mathrm{n}}+2)(\alpha_{\mathrm{n}}+3)}{\alpha_{\mathrm{n}}(\alpha_{\mathrm{n}}+1)} - 6\right\} + 6\right)^2 + 14J^{-0.7} \cdot \left\{\frac{(\alpha_{\mathrm{n}}+2)(\alpha_{\mathrm{n}}+3)}{\alpha_{\mathrm{n}}(\alpha_{\mathrm{n}}+1)} - 6\right\} + 85\right\}^{0.5} - \left(J^{-0.7} \cdot \left\{\frac{(\alpha_{\mathrm{n}}+2)(\alpha_{\mathrm{n}}+3)}{\alpha_{\mathrm{n}}(\alpha_{\mathrm{n}}+1)} - 6\right\}\right) - 1\right]. \tag{10.59}$$

Next, we calculate the change in kurtosis after parametric BSSA. With the shape parameter after DS, α_{DS}, the resultant kurtosis after the parametric BSSA is represented as

$$\mathrm{kurt}^{(\mathrm{n})}_{\mathrm{BSSA}} = \mathscr{M}(\alpha_{\mathrm{DS}}, \beta, 4, n)/\mathscr{M}^2(\alpha_{\mathrm{DS}}, \beta, 2, n), \tag{10.60}$$

where $\mathscr{M}(\alpha, \beta, m, n)$ is referred to as *normalized moment function* that represents the resultant mth-order moment after GSS in the case that the oversubtraction parameter is β, the exponent parameter is n and the input signal's shape parameter is α. This can be expressed as [52]

$$\mathscr{M}(\alpha, \beta, m, n) = \sum_{l=0}^{m/n} \frac{(-\beta)^l \Gamma^l(\alpha+n)\Gamma(m/n+1)}{\Gamma^{l+1}(\alpha)\Gamma(l+1)\Gamma(m/n-l+1)} \Gamma\left(\alpha + m - nl, \left(\beta \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}\right)^{\frac{1}{n}}\right), \tag{10.61}$$

where $\Gamma(\alpha, z)$ is the upper incomplete gamma function

$$\Gamma(\alpha, z) = \int_z^\infty t^{\alpha-1} \exp(-t)\mathrm{d}t. \tag{10.62}$$

Finally, using (10.52), (10.53), and (10.60),

we can determine the resultant kurtosis ratio through a parametric BSSA as

$$\text{kurtosis ratio}^{(\mathrm{n})}_{\mathrm{BSSA}} = \mathrm{kurt}^{(\mathrm{n})}_{\mathrm{BSSA}}/\mathrm{kurt}^{(\mathrm{n})}_{\mathrm{in}}. \tag{10.63}$$

10.5.2.3 Analysis in the Case of Parametric chBSSA

In this section, we analyze the kurtosis ratio in a parametric chBSSA. First, we calculate the change in kurtosis after channelwise GSS. Using (10.60) with the shape

parameter of input noise α_{n}, we can express the resultant kurtosis after channelwise GSS as

$$\mathrm{kurt}^{(\mathrm{n})}_{\mathrm{chGSS}} = \mathscr{M}(\alpha_{\mathrm{n}}, \beta, 4, n)/\mathscr{M}^2(\alpha_{\mathrm{n}}, \beta, 2, n). \tag{10.64}$$

Next, using (10.55) and (10.64), we can derive the change in kurtosis after a parametric chBSSA as

$$\mathrm{kurt}^{(\mathrm{n})}_{\mathrm{chBSSA}} \simeq J^{-0.7} \cdot (\mathrm{kurt}^{(\mathrm{n})}_{\mathrm{chGSS}} - 6) + 6. \tag{10.65}$$

Finally, we can obtain the resultant kurtosis ratio through a parametric chBSSA as

$$\text{kurtosis ratio}^{(\mathrm{n})}_{\mathrm{chBSSA}} = \mathrm{kurt}^{(\mathrm{n})}_{\mathrm{chBSSA}}/\mathrm{kurt}^{(\mathrm{n})}_{\mathrm{in}}. \tag{10.66}$$

10.5.3 Analysis of Amount of Speech Distortion

10.5.3.1 Analysis in the Case of BSSA

In this section, we analyze the amount of speech distortion on the basis of the kurtosis ratio in speech components. Hereafter, we define $s(f, \tau)$ and $n(f, \tau)$ as the observed speech and noise components at each microphone, respectively. Assuming that speech and noise are disjoint, i.e., there is no overlap in the time–frequency domain, speech distortion is caused by subtracting the average noise from the pure speech component.

Thus, the distorted speech after BSSA is given by

$$\begin{aligned} |s_{\mathrm{BSSA}}(f, \tau)| &= \sqrt[2n]{|s(f, \tau)|^{2n} - \beta\overline{|z_{\mathrm{DS}}(f, \tau)|^{2n}}} \\ &= \sqrt[2n]{|s(f, \tau)|^{2n} - \beta C_{\mathrm{BSSA}}\overline{|s(f, \tau)|^{2n}}}, \end{aligned} \tag{10.67}$$

where $s_{\mathrm{BSSA}}(f, \tau)$ is the output speech component in BSSA. Also, calculating the nth-order moment of the gamma distribution, C_{BSSA} is given by

$$\begin{aligned} C_{\mathrm{BSSA}} &= \overline{|z_{\mathrm{DS}}(f, \tau)|^{2n}}/\overline{|s(f, \tau)|^{2n}} \\ &= J^{-n}\overline{|n(f, \tau)|^{2n}}/\overline{|s(f, \tau)|^{2n}} \\ &= J^{-n}\left(\frac{\alpha_{\mathrm{s}}}{\alpha_{\mathrm{n}}}\right)^n \frac{\Gamma(lpha_{\mathrm{n}} + n)/\Gamma(\alpha_{\mathrm{n}})}{\Gamma(\alpha_{\mathrm{s}} + n)/\Gamma(\alpha_{\mathrm{s}})}\left(\frac{\overline{|n(f, \tau)|^2}}{\overline{|s(f, \tau)|^2}}\right)^n, \end{aligned} \tag{10.68}$$

where α_{s} is the shape parameter of the input speech. Equation (10.68) indicates that the speech distortion increases when the input SNR, $\overline{|s(f, \tau)|^2}/\overline{|n(f, \tau)|^2}$, and/or the number of microphones, J, decreases. Using (10.61) and (10.68) with the input

speech shape parameter α_{s}, we can obtain the speech kurtosis ratio through BSSA as

$$\begin{aligned}&\text{kurtosis ratio}^{(\mathrm{s})}_{\mathrm{BSSA}}\\&= \frac{\mathscr{M}(\alpha_{\mathrm{s}}, \beta C_{\mathrm{BSSA}}, 4, n)}{\mathscr{M}^2(\alpha_{\mathrm{s}}, \beta C_{\mathrm{BSSA}}, 2, n)} \frac{\alpha_{\mathrm{s}}(\alpha_{\mathrm{s}}+1)}{(\alpha_{\mathrm{s}}+2)(\alpha_{\mathrm{s}}+3)}.\end{aligned} \tag{10.69}$$

10.5.3.2 Analysis in the Case of chBSSA

In chBSSA, since channelwise GSS is performed before DS, C_{BSSA} is therefore replaced with

$$\begin{aligned}C_{\mathrm{chBSSA}} &= (\overline{|n(f,\tau)|^{2n}}/\overline{|s(f,\tau)|^{2n}})\\&= \left(\frac{\alpha_{\mathrm{s}}}{\alpha_{\mathrm{n}}}\right)^n \frac{\Gamma(\alpha_n+n)/\Gamma(\alpha_{\mathrm{n}})}{\Gamma(\alpha_{\mathrm{s}}+n)/\Gamma(\alpha_{\mathrm{s}})} \left(\frac{\overline{|n(f,\tau)|^2}}{\overline{|s(f,\tau)|^2}}\right)^n.\end{aligned} \tag{10.70}$$

Equation (10.70) indicates that the speech distortion increases only when the input SNR decreases, regardless of the number of microphones. Thus, the distortion does not change even if we prepare many microphones, unlike the case of a parametric BSSA. Using (10.61) and (10.70) with α_{s}, we can obtain the speech kurtosis ratio through chBSSA as

$$\begin{aligned}&\text{kurtosis ratio}^{(\mathrm{s})}_{\mathrm{chBSSA}}\\&= \frac{\mathscr{M}(\alpha_{\mathrm{s}}, \beta C_{\mathrm{chBSSA}}, 4, n)}{\mathscr{M}^2(\alpha_{\mathrm{s}}, \beta C_{\mathrm{chBSSA}}, 2, n)} \frac{\alpha_{\mathrm{s}}(\alpha_{\mathrm{s}}+1)}{(\alpha_{\mathrm{s}}+2)(\alpha_{\mathrm{s}}+3)}.\end{aligned} \tag{10.71}$$

10.5.4 Comparison of Amounts of Musical Noise and Speech Distortion Under Same Amount of Noise Reduction

According to the previous analysis, we can compare the amounts of musical noise and speech distortion among a parametric BSSA and a parametric chBSSA under the same NRR (output SNR–input SNR in dB). Figure 10.10 shows the theoretical behaviors of the noise kurtosis ratio and speech kurtosis ratio. In Fig. 10.10a, b, the shape parameter of input noise, α_{n}, is set to 0.95 and 0.83, corresponding to almost white Gaussian noise and railway station noise, respectively. Also, in Fig. 10.10c, d, the shape parameter of input speech, α_{s}, is set to 0.1, and the input SNR is set to 10 and 5 dB, respectively. Here, we assume an eight-element array with the interelement spacing of 2.15 cm. The NRR is varied from 11 to 15 dB, and the oversubtraction parameter β is adjusted so that the target speech NRR is achieved. In the parametric

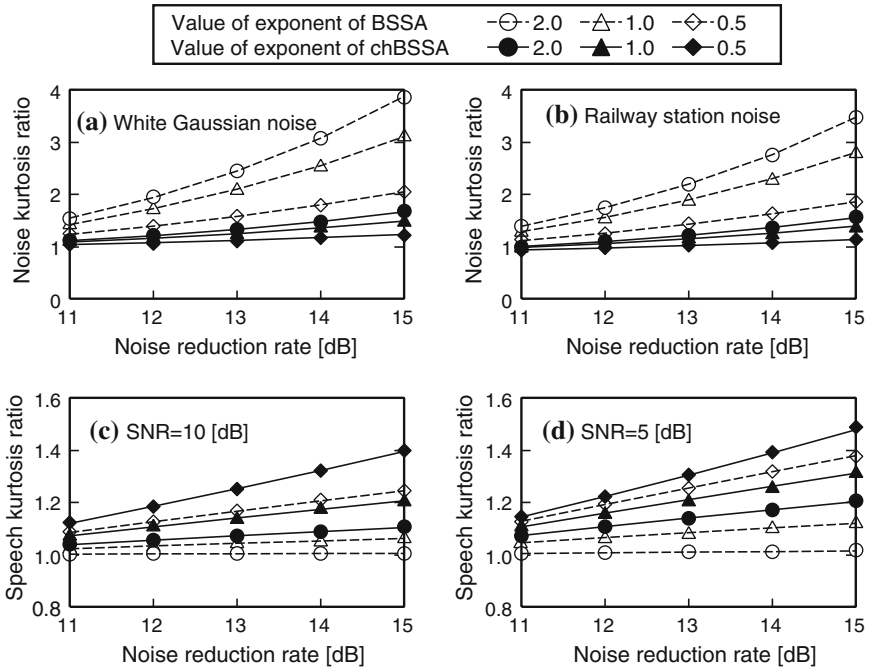


Fig. 10.10 **a** and **b** are theoretical behaviors of noise kurtosis ratio in structure-generalized parametric BSSA. **a** is for white Gaussian noise and **b** is for railway station noise. **c** and **d** are theoretical behaviors of speech kurtosis ratio in structure-generalized parametric BSSA, where the input SNR is set to 10 and 5 dB, respectively

BSSA and parametric chBSSA, the signal exponent parameter $2n$ is set to 2.0, 1.0, and 0.5.

Figure 10.10a, b indicates that the noise kurtosis ratio of chBSSA is smaller than that of BSSA, i.e., less musical noise is generated in a parametric chBSSA than in a parametric BSSA, and a smaller amount of musical noise is generated when a lower exponent parameter is used, regardless of the type of noise and NRR. However, Fig. 10.10c, d shows that speech distortion is lower in a parametric BSSA than in a parametric chBSSA, and a small amount of speech distortion is generated when a higher exponent parameter is used, regardless of the type of noise and NRR. These results theoretically prove the existence of a tradeoff between the amounts of musical noise and speech distortion in BSSA and chBSSA.

10.6 Experiment

10.6.1 Experimental Setup

In this study, we conducted a speech recognition experiment. We used an eight-element microphone array with an interelement spacing of 2.15 cm, and the direction of the target speech was set to be normal to the array. The size of the experimental room is $4.2 \times 3.5 \times 3.0\,\mathrm{m}^3$ and the reverberation time is approximately 200 ms. All the signals used in this experiment are sampled at 16 kHz with 16-bit accuracy. The observed signal consists of the target speech signal of 200 speakers (100 males and 100 females) and two types of diffuse noise (white Gaussian noise and railway station noise) emitted from eight surrounding loudspeakers. The input SNR of the test data is set to 3, 5, and 10 dB. The FFT size is 1,024, and the frame shift length is 256 in BSSA. The speech recognition task is a 20k-word Japanese newspaper dictation, where we used Julius 3.4.2 [53] as the speech decoder. The acoustic model is a phonetic-tied mixture [53], and we use 260 speakers (150 sentences/speaker) to train the acoustic model. In this experiment, the NRR, i.e., the target SNR improvement, is set to 10 dB for white Gaussian noise and 5 dB for railway station noise, the exponent parameter $2n$ is set to 1.0 and 0.5, and the oversubtraction parameter β is adjusted so that the target NRR is achieved.

10.6.2 Evaluation of Speech Recognition Performance and Discussion

Figure 10.11 shows the results of word accuracy in the parametric BSSA and parametric chBSSA, which reveal that better speech recognition performance can be obtained in a parametric BSSA when the input SNR is low (e.g., 3 dB).

This result is of considerable interest because Takahashi et al. [26] reported a contradictory result, i.e., the sound quality of chBSSA is always superior to that of BSSA. Indeed, we conducted a subjective evaluation. We presented 56 pairs of signals processed by a parametric BSSA and a parametric chBSSA, selected from sentences used in the speech recognition experiment, in random order to 10 examinees, who selected which signal they preferred. The result is shown in Fig. 10.12, confirming that chBSSA is preferred by humans, in contrast to the speech recognition results. This is partially true regarding noise distortion, i.e., the amount of musical noise generated, as theoretically shown in Fig. 10.10a, b. Thus, the human evaluation is strongly affected by noise distortion.

However, as shown in Fig. 10.10c, d, the speech distortion in chBSSA is larger than that in BSSA; this leads to the degradation of speech recognition performance. In summary, we should carefully select the structure of signal processing in BSSA, i.e., chBSSA is recommended for listening but BSSA is suitable for speech recognition under a low-input SNR condition.

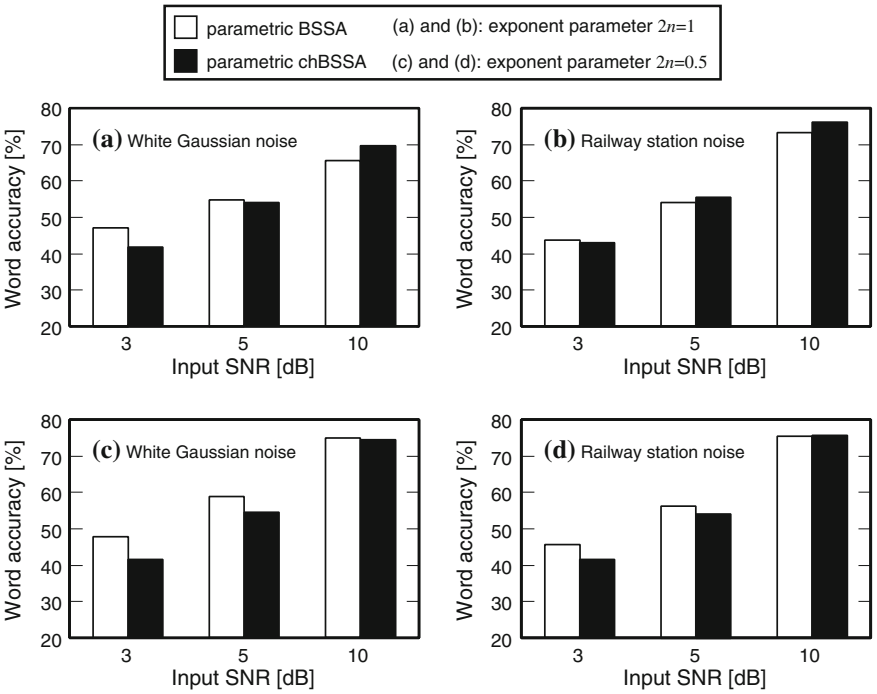


Fig. 10.11 Results of word accuracy in parametric BSSA and parametric chBSSA

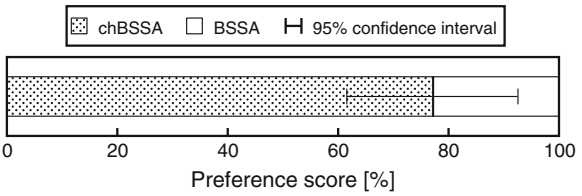


Fig. 10.12 Subjective evaluation results: BSSA versus chBSSA

10.7 Conclusions and Remarks

This chapter addressed the BSS problem for speech applications under real acoustic environments, particularly focusing on BSSA that utilizes ICA as a noise estimator. Under a nonpoint-source noise condition, it was pointed out that beamformers optimized by ICA are a DS beamformer for extracting the target speech signal that can be regarded as a point source and NBF for picking up the noise signal. Thus, ICA is proficient in noise estimation under a nonpoint-source noise condition. Therefore, it is valid to use ICA as a noise estimator.

Motivated by the above-mentioned fact, we introduced a structure-generalized parametric BSSA, which consists of an ICA-based noise estimator and GSS-based

post-filtering. In addition, we performed its theoretical analysis via higher-order statistics. Comparing a parametric BSSA and parametric chBSSA, we revealed that a channelwise BSSA structure is recommended for listening but a conventional BSSA is more suitable for speech recognition.

In this chapter, the SS-based BSSAs, which involve SS-based post-filtering, were mainly addressed. Recent studies have provided the further extended methods that include other types of post-filtering, such as Wiener filtering [54, 55], the minimum mean-square error short-time spectral amplitude (MMSE-STSA) estimator [56, 57], and the combination method with cepstral smoothing for mitigating musical noise [58]. Also, the theoretical analysis based on the higher-order statistics for these methods is available in several literatures [59–63]. In addition, thanks to the same higher-order statistics analysis, *musical-noise-free* post-filtering [64], in which no musical noise is perfectly generated, has been proposed, and successfully introduced into the channelwise BSSA architecture [65, 66].

BSS implementation on a small hardware still receives much attention in industrial applications. Due to the limitation of space, however, the authors skip the discussion on this issue. Instead, several studies [21, 67, 68] have dealt with the issue of real-time implementation of ICA and BSSA, which would be helpful for the readers.

References

1. Juang, B.H., Soong, F.K.: Hands-free telecommunications. In: Proceedings of International Conference on Hands-Free, Speech Communication, pp. 5–10 (2001)
2. Prasad, R., Saruwatari, H., Shikano, K.: Robots that can hear, understand and talk. Adv. Robot. **18**(5), 533–564 (2004)
3. Saruwatari, H., Kawanami, H., Takeuchi, S., Takahashi, Y., Cincarek, T., Shikano, K.: Hands-free speech recognition challenge for real-world speech dialogue systems. In: Proceedings of 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2009), pp. 3729–3782 (2009)
4. Flanagan, J.L., Johnston, J.D., Zahn, R., Elko, G.W.: Computer-steered microphone arrays for sound transduction in large rooms. J. Acoust. Soc. Am. **78**(5), 1508–1518 (1985)
5. Omologo, M., Matassoni, M., Svaizer, P., Giuliani, D.: Microphone array based speech recognition with different talker-array positions. In: Proceedings of ICASSP'97, pp. 227–230 (1997)
6. Silverman, H.F., Patterson, W.R.: Visualizing the performance of large-aperture microphone arrays. In: Proceedings of ICASSP'99, pp. 962–972 (1999)
7. Saruwatari, H., Kajita, S., Takeda, K., Itakura, F.: Speech enhancement using nonlinear microphone array based on complementary beamforming. IEICE Trans. Fundam. **E82-A**(8), 1501–1510 (1999)
8. Frost, O.: An algorithm for linearly constrained adaptive array processing. Proc. IEEE **60**, 926–935 (1972)
9. Griffiths, L.J., Jim, C.W.: An alternative approach to linearly constrained adaptive beamforming. IEEE Trans. Antennas Propag. **30**(1), 27–34 (1982)
10. Kaneda, Y. Ohga, J.: Adaptive microphone-array system for noise reduction. IEEE Trans. Acoust. Speech Signal Process. **34**(6),1391–1400 (1986)
11. Saruwatari, H., Kajita, S., Takeda, K., Itakura, F.: Speech enhancement using nonlinear microphone array based on noise adaptive complementary beamforming. IEICE Trans. Fundam. **E83-A**(5), 866–876 (2000)

12. Comon, P.: Independent component analysis, a new concept? Signal Process. **36**, 287–314 (1994)
13. Cardoso, J.F.: Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem. In: Proceedings of ICASSP'89, pp. 2109–2112 (1989)
14. Jutten, C., Herault, J.: Blind separation of sources Part I: an adaptive algorithm based on neuromimetic architecture. Signal Process. **24**, 1–10 (1991)
15. Ikeda, S., Murata, N.: A method of ICA in the frequency domain. In: Proceedings of International Workshop on Independent Component Analysis and Blind, Signal Separation, pp. 365–371 (1999)
16. Smaragdis, P.: Blind separation of convolved mixtures in the frequency domain. Neurocomputing **22**(1–3), 21–34 (1998)
17. Parra, L., Spence, C.: Convolutive blind separation of non-stationary sources. IEEE Trans. Speech Audio Process. **8**, 320–327 (2000)
18. Saruwatari, H., Kurita, S., Takeda, K., Itakura, F., Nishikawa, T.: Blind source separation combining independent component analysis and beamforming. EURASIP J. Appl. Signal Process. **2003**, 1135–1146 (2003)
19. Pham, D.-T., Serviere, C., Boumaraf, H.: Blind separation of convolutive audio mixtures using nonstationarity. In: International Symposium on Independent Component Analysis and Blind, Signal Separation (ICA2003), pp. 975–980 (2003)
20. Saruwatari, H., Kawamura, T., Nishikawa, T., Lee, A., Shikano, K.: Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. IEEE Trans. Speech Audio Process. **14**(2), 666–678 (2006)
21. Mori, Y., Saruwatari, H., Takatani, T., Ukai, S., Shikano, K., Hiekata, T., Ikeda, Y., Hashimoto, H., Morita, T.: Blind separation of acoustic signals combining SIMO-model-based independent component analysis and binary masking. EURASIP J. Appl. Signal Process. **2006**, ArticleID 34970, 17 (2006)
22. Prasad, R., Saruwatari, H., Shikano, K.: Enhancement of speech signals separated from their convolutive mixture by FDICA algorithm. Digit. Signal Process. **19**(1), 127–133 (2009)
23. Takahashi, Y., Takatani, T., Osako, K., Saruwatari, H., Shikano, K.: Blind spatial subtraction array for speech enhancement in noisy environment. IEEE Trans. Audio Speech Lang. Process. **17**(4), 650–664 (2009)
24. Boll, S.: Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. **ASSP-27**(2), 113–120 (1979)
25. Saruwatari, H., Takahashi, Y., Shikano, K., Kondo, K.: Blind speech extraction combining ICA-based noise estimation and less-musical-noise nonlinear post processing. In: Proceedings of 2010 Asilomar Conference on Signals, Systems, and Computers, pp. 1415–1419 (2010)
26. Takahashi, Y., Saruwatari, H., Shikano, K., Kondo, K.: Musical-noise analysis in methods of integrating microphone array and spectral subtraction based on higher-order statistics. EURASIP J. Adv. Signal Process. **2010**, Article ID 431347, 25 (2010)
27. Miyazaki, R., Saruwatari, H., Shikano, K.: Theoretical analysis of amount of musical noise and speech distortion in structure-generalized parametric blind spatial subtraction array. IEICE Trans. Fundam. **95-A**(2), 586–590 (2011)
28. Saruwatari, H., Takatani, T., Shikano, K.: SIMO-model-based blind source separation -principle and its applications. In: Makino, S., et al. (eds.) Blind Speech Separation, pp. 149–168. Springer, New York (2007). ISBN 978-1-4020-6479-1
29. Saruwatari, H., Takahashi, Y.: Blind source separation for speech application under real acoustic environment. In: Naik, G. (ed.) Independent Component Analysis for Audio and Biosignal Applications, pp. 41–66. InTech Publishing, Rijeka (2012). ISBN 978-953-51-0782-8
30. Uemura, Y., Takahashi, Y., Saruwatari, H., Shikano, K., Kondo, K.: Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics. In: Proceedings of 2008 International Workshop on Acoustic Echo and Noise, Control (IWAENC2008) (2008)

31. Uemura, Y., Takahashi, Y., Saruwatari, H., Shikano, K., Kondo, K.: Musical noise generation analysis for noise reduction methods based on spectral subtraction and MMSE STSA estimation. In: Proceedings of 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2009), pp. 4433–4436 (2009)
32. Takahashi, Y., Miyazaki, R., Saruwatari, H., Kondo, K.: Theoretical analysis of musical noise in nonlinear noise reduction based on higher-order statistics. In: Proceedings of 2012 APSIPA Annual Summit and Conference (APSIPA2012) (2012)
33. Tachibana, K., Saruwatari, H., Mori, Y., Miyabe, S., Shikano, K. Tanaka, A.: Efficient blind source separation combining closed-form second-order ICA and nonclosed-form higher-order ICA. In: Proceedings of 2007 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2007), vol. 1, pp. 45–48 (2007)
34. Saruwatari, H., Takahashi, Y., Tachibana, K., Mori, Y., Miyabe, S., Shikano, K., Tanaka, A.: Fast and versatile blind separation of diverse sounds using closed-form estimation of probability density functions of sources. In: Proceedings of 3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP2009), pp. 249–252 (2009)
35. Lee, T.-W.: Independent Component Analysis. Kluwer Academic, Norwell (1998)
36. Prasad, R., Saruwatari, H., Shikano, K.: Probability distribution of time-series of speech spectral components. IEICE Trans. Fundam. **E87-A**(3), 584–597 (2004)
37. Ukai, S., Takatani, T., Nishikawa, T., Saruwatari, H.: Blind source separation combining SIMO-model-based ICA and adaptive beamforming. In: Proceedings of 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2005), vol. 3, pp. 85–88 (2005)
38. Kurita, S., Saruwatari, H., Kajita, S., Takeda, K., Itakura, F.: Evaluation of blind signal separation method using directivity pattern under reverberant conditions. In: Proceedings of 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2000), no. SAM-P2-5, pp. 3140–3143 (2000)
39. Sawada, H., Mukai, R., Araki, S., Makino, S.: A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Trans. Speech Audio Process. **12**(5), 530–538 (2004)
40. Nishikawa, T., Saruwatari, H., Shikano, K.: Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA. In: IEICE Trans. Fundam. **E86-A**(4), 846–858 (2003)
41. Nishikawa, T., Abe, H., Saruwatari, H., Shikano, K., Kaminuma, A.: Overdetermined blind separation for real convolutive mixtures of speech based on multistage ICA using subarray processing. IEICE Trans. Fundam. **E87-A**(8), 1924–1932 (2004)
42. Araki, S., Makino, S., Aichner, R., Nishikawa, T., Saruwatari, H.: Subband-based blind separation for convolutive mixtures of speech. IEICE Trans. Fundam. **E88-A**(12), 3593–3603 (2005)
43. Araki, S., Mukai, R., Makino, S., Nishikawa, T., Saruwatari, H.: The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. IEEE Trans. Speech Audio Process. **11**(2), 109–116 (2003)
44. Araki, S., Makino, S., Hinamoto, Y., Mukai, R., Nishikawa, T., Saruwatari, H.: Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures. EURASIP J. Appl. Signal Process. **2003**(11), 1157–1166 (2003)
45. Brandstein, M., Ward, D. (eds.): Microphone Arrays: Signal Processing Techniques and Applications. Springer, New York (2001)
46. Saruwatari, H., Hirata, N., Hatta, T., Wakisaka, R., Shikano, K., Takatani, T.: Semi-blind speech extraction for robot using visual information and noise statistics. In: Proceedings of 11th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT2011), pp. 238–243 (2011)
47. Lee, A., Nakamura, K., Nishimura, R., Saruwatari, H., Shikano, K.: Noise robust real world spoken dialogue system using GMM based rejection of unintended inputs. In: Proceedings of 8th International Conference on Spoken Language Processing (ICSLP2004), vol. 1, pp. 173–176 (2004)

48. Sim, B.L., Tong, Y.C., Chang, J.S., Tan, C.T.: A parametric formulation of the generalized spectral subtraction method. IEEE Trans. Speech Audio Process. **6**(4), 328–337 (1998)
49. Stacy, E.W.: A generalization of the gamma distribution. Ann. Math. Stat. **33**(3), 1187–1192 (1962)
50. Shin, J.W., Chang, J.-H., Kim, N.S.: Statistical modeling of speech signal based on generalized gamma distribution. IEEE Signal Process. Lett. **12**(3), 258–261 (2005)
51. Saruwatari, H., Ishikawa, Y., Takahashi, Y., Inoue, T., Shikano, K., Kondo, K.: Musical noise controllable algorithm of channelwise spectral subtraction and adaptive beamforming based on higher-order statistics. IEEE Trans. Audio Speech Lang. Process. **19**(6), 1457–1466 (2011)
52. Inoue, T., Saruwatari, H., Takahashi, Y., Shikano, K., Kondo, K.: Theoretical analysis of musical noise in generalized spectral subtraction based on higher-order statistics. IEEE Trans. Audio Speech Lang. Process. **19**(6), 1770–1779 (2011)
53. Lee, A., Kawahara, T., Shikano, K.: Julius -An open source real-time large vocabulary recognition engine. In: Proceedings of Eurospeech, pp. 1691–1694 (2001)
54. Takahashi, Y., Osako, K., Saruwatari, H., Shikano, K.: Blind source extraction for hands-free speech recognition based on Wiener filtering and ICA-based noise estimation. In: Proceedings of 2008 Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA2008), pp. 164–167 (2008)
55. Even, J., Saruwatari, H., Shikano, K.: Enhanced Wiener post-processing based on partial projection back of the blind signal separation noise estimate. In: Proceedings of 17th European Signal Processing Conference (EUSIPCO2009), pp. 1442–1446 (2009)
56. Okamoto, R., Takahashi, Y., Saruwatari, H., Shikano, K.: MMSE STSA estimator with nonstationary noise estimation based on ICA for high-quality speech enhancement. In: Proceedings of 2010 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2010), pp. 4778–4781 (2010)
57. Saruwatari, H., Go, M., Okamoto, R., Shikano, K.: Binaural hearing aid using sound-localization-preserved MMSE STSA estimator with ICA-based noise estimation. In: Proceedings of 2010 International Workshop on Acoustic Echo and Noise, Control (IWAENC2010) (2010)
58. Jan, T., Wang, W., Wang, D.L.: A multistage approach to blind separation of convolutive speech mixtures. Speech Commun. **53**, 524–539 (2011)
59. Inoue, T., Saruwatari, H., Shikano, K., Kondo, K.: Theoretical analysis of musical noise in Wiener filtering family via higher-order statistics. In: Proceedings of 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2011), pp. 5076–5079 (2011)
60. Yu, H., Fingscheidt, T.: A figure of merit for instrumental optimization of noise reduction algorithms. In: Proceedings of DSP in Vehicles (2011)
61. Kanehara, S., Saruwatari, H., Miyazaki, R., Shikano, K., Kondo, K.: Comparative study on various noise reduction methods with decision-directed a priori SNR estimator via higher-order statistics. In: Proceedings of 2012 APSIPA Annual Summit and Conference (APSIPA2012) (2012)
62. Yu, H., Fingscheidt, T.: Black box measurement of musical tones produced by noise reduction systems. In: Proceedings of ICASSP2012, pp. 4573–4576 (2012)
63. Saruwatari, H., Kanehara, S., Miyazaki, R., Shikano, K., Kondo, K.: Musical noise analysis for Bayesian minimum mean-square error speech amplitude estimators based on higher-order statistics. In: Proceedings of Interspeech 2013 (2013)
64. Miyazaki, R., Saruwatari, H., Inoue, T., Takahashi, Y., Shikano, K., Kondo, K.: Musical-noise-free speech enhancement based on optimized iterative spectral subtraction. IEEE Trans. Audio Speech Lang. Process. **20**(7), 2080–2094 (2012)
65. Miyazaki, R., Saruwatari, H., Shikano, K., Kondo, K.: Musical-noise-free blind speech extraction using ICA-based noise estimation and iterative spectral subtraction. In: Proceedings of 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA2012), pp. 322–327 (2012)

66. Miyazaki, R., Saruwatari, H., Shikano, K., Kondo, K.: Musical-noise-free blind speech extraction using ICA-based noise estimation with channel selection. In: Proceedings of 2012 International Workshop on Acoustic Signal Enhancement (IWAENC2012) (2012)
67. Buchner, H., Aichner, R., Kellermann, W.: A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. IEEE Trans. Speech Audio Process. **13**(1), 120–134 (2005)
68. Hiekata, T., Ikeda, Y., Yamashita, T., Morita, T., Zhang, R., Mori, Y., Saruwatari, H., Shikano, K.: Development and evaluation of pocket-size real-time blind source separation microphone. Acoust. Sci. Technol. **30**(4), 297–304 (2009)

Chapter 11
Speech Separation and Extraction by Combining Superdirective Beamforming and Blind Source Separation

Lin Wang, Heping Ding and Fuliang Yin

Abstract Blind source separation (BSS) and beamforming are two well-known multiple microphone techniques for speech separation and extraction in cocktail-party environments. However, both of them perform limitedly in highly reverberant and dynamic scenarios. Emulating human auditory systems, this chapter proposes a combined method for better separation and extraction performance, which uses superdirective beamforming as a preprocessor of frequency-domain BSS. Based on spatial information only, superdirective beamforming presents abilities of dereverberation and noise reduction and performs robustly in time-varying environments. Using it as a preprocessor can mitigate the inherent "circular convolution approximation problem" of the frequency-domain BSS and enhances its robustness in dynamic environments. Meanwhile, utilizing statistical information only, BSS can further reduce the residual interferences after beamforming efficiently. The combined method can exploit both spatial information and statistical information about microphone signals and hence performs better than using either BSS or beamforming alone. The proposed method is applied to two specific challenging tasks, namely a separation task in highly reverberant environments with the positions of all sources known, and a target speech extraction task in highly dynamic cocktail-party environments with

L. Wang (✉) · F. Yin
School of Electronic and Information Engineering,
Dalian University of Technology, Dalian, China
e-mail: wanglin_2k@sina.com; lin.wang@uni-oldenburg.de

F. Yin
e-mail: flyin@dlut.edu.cn

L. Wang
Institute of Physics - Signal Processing Group, University of Oldenburg, Oldenburg, Germany

H. Ding
Information and Communications Technology,
National Research Council, Ottawa, Canada
e-mail: heping.ding@nrc-cnrc.gc.ca

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_11,
© Springer-Verlag Berlin Heidelberg 2014

only the position of the target known. Experimental results prove the effectiveness of the proposed method.

11.1 Introduction

Extracting one or several desired speech signals from their corrupted observations is essential for many applications of speech processing and communication. One of the hardest situations to handle is the extraction of desired speech signals in a "cocktail party" condition—from mixtures picked up by microphones placed inside a noisy and reverberant enclosure. In this case, the target speech is immersed in ambient noise and interferences, and distorted by reverberation. Furthermore, the environment may be time varying. Generally, there are two well-known techniques that may achieve the objective: blind source separation (BSS) and beamforming.

With a microphone array, beamforming is a well-known technique for directional signal reception [1, 2]. Depending on how the beamformer weights are chosen, it can be implemented as a data-independent fixed beamforming or data-dependent adaptive one [3–7]. Although an adaptive beamformer generally exhibits better noise reduction abilities, a fixed beamformer is more preferred in complicated environments due to its robustness. By coherently summing signals from multiple sensors based on a model of the wavefront from acoustic sources, a fixed beamformer presents a specified directional response. With abilities of enhancing signals from the desired direction while suppressing ones from other directions, it can be used to perform both noise suppression and dereverberation. The most conventional fixed beamformer is a delay-and-sum one, which however requires a large number of microphones to achieve high performance. Another filter-and-sum beamformer has superdirective response with optimized weights [4]. Assuming the directions of the sources are known, speech separation or extraction can be obtained by forming individual beams at the target sources separately. However, fixed beamforming performs limitedly in real cocktail-party scenarios. First, the performance is closely related to the microphone array size—a large array is usually required to obtain a satisfactory result but may not be practically feasible. Second, beamforming cannot suppress the interfering reverberation coming from the desired direction.

BSS is a technique for recovering the source signals from observed signals with the mixing process unknown. By exploiting the statistical independence of the sources, independent component analysis (ICA)-based algorithms are commonly used to solve the BSS problem [8–13]. While time domain ICA-based techniques are well suited for instantaneous mixing problem, they are not efficient in addressing the convolutive mixture problem encountered in reverberant environments [14–17]. By considering the BSS problem in the frequency domain, the convolutive mixing problem can be transformed into an instantaneous mixing problem for each frequency bin, reducing computation complexity significantly. However, the inherent permutation and amplitude scaling ambiguity problem occurs at each frequency bin in frequency-domain BSS, deteriorating signal reconstruction in the time domain significantly

[18, 19]. The permutation ambiguity problem has been investigated intensively and there are generally three strategies to tackle it. The first is to make the separation filters smooth in the frequency domain by limiting the filter length in the time domain [16, 20, 21]. The second strategy is to exploit the interfrequency dependence of the amplitude of separated signals [22–30]. The third strategy is to exploit the position information about sources such as direction of arrival (DOA). By estimating the arriving delay of sources or analyzing the directivity pattern formed by a separation matrix, source direction can be estimated and permutations aligned [31–36].

The relationship between blind source separation and beamforming has been intensively investigated in recent years, and adaptive beamforming is commonly used to explain the physical principle of convolutive BSS [37, 38]. In addition, many approaches have been presented that combine both techniques. Some of these combined approaches are aimed at resolving the permutation ambiguity inherent in frequency-domain BSS [31, 39], whereas other approaches utilize beamforming to provide a good initialization for BSS or to accelerate its convergence [40–43]. For now, there were no systematically studies on combining the two techniques to improve the separation performance in challenging acoustic scenarios.

Compared with beamforming, which extracts desired speech and suppress interference, BSS aims at separating all the involved desired and interfering sources equally. One advantage with blind source separation is that it does not need to know the direction of arrival of any signals and the array geometry can be arbitrary and unknown to the system. Nevertheless, blind source separation also performs limitedly in real cocktail-party scenarios. First, BSS performs poorly in high reverberation with long mixing filters, due to the "circular convolution approximation problem". Second, underdetermined situations can result from the fact that there are only a limited number of microphones. Third, the performance of BSS degrades in dynamic environments.

Due to the reasons above, few methods proposed in recent years show good separation/extraction results in a real cocktail-party environment. In contrast, a human has a remarkable ability to focus on a specific speaker in that case. This selective listening capability is partially attributed to binaural hearing. Two ears work as a beamformer, which enables directive listening [44], then the brain analyzes the received signals to extract sources of interest from the background, just as blind source separation does [45–47]. Stimulating this principle, we propose to do speech separation and extraction by combining beamforming and blind source separation. Specifically, the following two issues will be addressed:

- To improve the separation performance in highly reverberant scenarios, a combined method is proposed which uses beamforming as a preprocessor of blind source separation by forming a number of beams each pointing at a source. With beamforming shortening mixing filters, the inherent "circular convolution approximation" problem in the frequency-domain BSS is mitigated and the performance of proposed method can improve significantly especially in high reverberation.
- The combined method is further extended to a special case of target speech extraction problem in noisy cocktail-party environments, where only one source

is of interest. Instead of focusing on all the sources, the proposed method forms just several fixed beams at an area containing the target source. The proposed scheme can enhance the robustness to time-varying environments and make the target source dominant in the output of the beamformer. Consequently, the subsequent extraction task with blind source separation becomes easier and satisfactory extraction results can be obtained even in challenging scenarios.

The rest of the chapter is organized as follows. The principles of frequency-domain blind source separation and superdirective beamforming are reviewed in Sects. 11.2 and 11.3, respectively. Especially, the inherent "circular convolution approximation" problem with frequency-domain BSS is discussed in detail in Sect. 11.2. BSS and beamforming are combined for a better separation results and the performance of the combined method is experimentally evaluated in Sect. 11.4. The combined method is further extended to a target speech extraction problem with some experimental results shown in Sect. 11.5. Finally, conclusions are drawn in Sect.11.6.

11.2 Frequency-Domain BSS and Its Fundamental Problem

BSS is a powerful tool for solving cocktail-party problems since it aims at recovering the source signals from observed signals with the mixing process unknown. The simplest instantaneous BSS problem can be solved by independent components analysis (ICA), which assumes that all the source signals are independent of each other. One challenge arises when the mixing process is convolutive, i.e., the observations are combinations of filtered versions of sources. The convolutive BSS problem can be solved in the time domain, where the separation network is derived by optimizing a time-domain cost function. However, the task of estimating many parameters simultaneously has to face the challenge of slow convergence and high-computational demand. Alternatively, the convolutive BSS problem can be solved in the frequency domain, where instantaneous BSS is performed at individual frequency bins, reducing the computational complexity significantly. In this section, the principle of the frequency-domain BSS is introduced at first, followed by a discussion of an inherent problem, namely the circular convolution approximation problem, which degrades the performance of the frequency-domain BSS in high reverberation.

11.2.1 Frequency-Domain BSS

Supposing N sources and M sensors in a real-world acoustic scenario, the source vector $\boldsymbol{s}(n) = [s_1(n), \cdots, s_N(n)]^{\mathrm{T}}$, and the observed vector $\boldsymbol{x}(n) = [x_1(n), \cdots, x_M(n)]^{\mathrm{T}}$, the mixing channels can be modeled by FIR filters of length P, the convolutive mixing process is formulated as

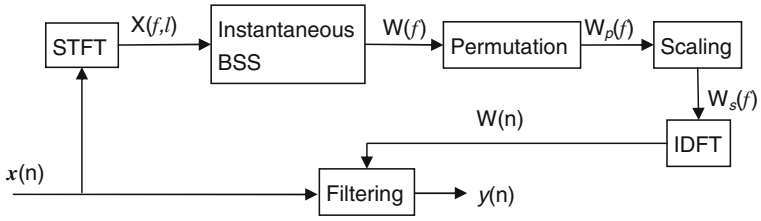


Fig. 11.1 Workflow of frequency-domain blind source separation

$$\boldsymbol{x}(n) = \boldsymbol{H}(n) * \boldsymbol{s}(n) = \sum_{p=0}^{P-1} \boldsymbol{H}(p)\boldsymbol{s}(n-p) \tag{11.1}$$

where $\mathbf{H}(n)$ is a sequence of $M \times N$ matrices containing the impulse responses of the mixing channels, n is the time index, and the operator '*' denotes matrix convolution. For separation, FIR filters of length L can be used to estimate the source signals $\boldsymbol{y}(n) = [y_1(n), \cdots, y_N(n)]^{\mathrm{T}}$ by

$$\boldsymbol{y}(n) = \boldsymbol{W}(n) * \boldsymbol{x}(n) = \sum_{l=0}^{L-1} \boldsymbol{W}(l)\boldsymbol{x}(n-l) \tag{11.2}$$

where $\boldsymbol{W}(n)$ is a sequence of $N \times M$ matrices containing the unmixing filters.

The demixing network $\boldsymbol{W}(n)$ can be estimated in the frequency domain. By using a blockwise Q-point short-time Fourier transform (STFT), the time-domain convolution regarding the mixing process can be converted into frequency-domain multiplications and correspondingly the convolutive BSS problem is converted into multiple instantaneous BSS problem at each frequency bin. This is expressed as

$$\boldsymbol{x}(f,l) = \boldsymbol{H}(f)\boldsymbol{s}(f,l) \tag{11.3}$$

where l is a decimated version of the time index n, f is the frequency index, $\boldsymbol{H}(f)$ is the Fourier transform of $H(n)$, and $\boldsymbol{x}(f,l)$ and $\boldsymbol{s}(f,l)$ are the STFTs of $\boldsymbol{x}(n)$ and $\boldsymbol{s}(n)$, respectively. The remaining task will be to find a demixing matrix $\boldsymbol{W}_{\mathrm{demix}}(f)$ at individual frequency bins, so that the original signals can be recovered. This is expressed as

$$\boldsymbol{y}(f,l) = \boldsymbol{W}_{\mathrm{demix}}(f)\boldsymbol{x}(f,l) \tag{11.4}$$

Based on the discussion above, the workflow of the frequency-domain BSS is shown in Fig. 11.1. The observed time-domain signals are converted into the time–frequency domain by STFT; then instantaneous BSS is applied to each frequency bin; after solving the inherent permutation and scaling ambiguities, the separated signals of all frequency bins are combined and inverse-transformed to the time domain. The procedure of a frequency-domain BSS mainly consists of three blocks: instantaneous BSS, permutation alignment, and scaling correction.

(1) *Instantaneous BSS*
After decomposing time-domain convolutive mixing into frequency-domain instantaneous mixing, it is possible to perform instantaneous separation at each frequency bin with a complex-valued ICA algorithm. The ICA algorithms for instantaneous BSS have been studied for many years and are considered to be quite mature. For instance, the demixing matrix can be estimated iteratively by using the well-known Infomax algorithm [11, 12], i.e.,

$$\begin{cases} \boldsymbol{y}(f,l) = \boldsymbol{W}(f)\boldsymbol{x}(f,l) \\ \boldsymbol{W}(f) = \boldsymbol{W}(f) + \eta(\boldsymbol{I} - \mathrm{E}[\Phi(\boldsymbol{y}(f,l))\boldsymbol{y}^{\mathrm{H}}(f,l)])\boldsymbol{W}(f) \end{cases} \tag{11.5}$$

where $\boldsymbol{I}$ is an identity matrix, $\Phi(\cdot)$ is a nonlinear function and $\mathrm{E}[\cdot]$ is the expectation operator.

(2) *Permutation alignment*
Although satisfactory instantaneous separation may be achieved within all frequency bins, combining them to recover the original sources is a challenge because of the unknown permutations associated with individual frequency bins. This permutation ambiguity problem is the main challenge in the frequency-domain BSS and how to solve this problem has been a hot topic in the research community in recent years.

Two kinds of strategies can be used to solve this problem. The first strategy is to exploit the interfrequency dependence of the amplitude of separated signals [22–25]. The second strategy is to exploit the position information of sources such as direction of arrival or the continuity of the phase of separation matrix [31–33]. By analyzing the directivity pattern formed by a separation matrix, source direction can be estimated and permutations aligned. Since the performance of the second approach is generally limited by the reverberation density of the environment and the source positions, we prefer to use the first approach. In [24], a region-growing permutation alignment approach is proposed with good results, which is based on the interfrequency of separated signals. Bin-wise permutation alignment is applied first across all frequency bins, using the correlation of separated signal powers; then the full frequency band is partitioned into small regions based on the bin-wise permutation alignment result. Finally, region-wise permutation alignment is performed, which can prevent the spreading of the misalignment at isolated frequency bins to others and thus improves permutation alignment results. After permutation alignment, we can assume that the separated frequency components from the same source are grouped together.

(3) *Scaling correction*
The scaling indeterminacy can be resolved relatively easily by using the Minimal Distortion Principle [48]:

$$\boldsymbol{W}_s(f) = \mathrm{diag}(\boldsymbol{W}_p^{-1}(f)) \cdot \boldsymbol{W}_p(f) \tag{11.6}$$

where $\boldsymbol{W}_p(f)$ is $\boldsymbol{W}(f)$ after permutation correction, $(\cdot)^{-1}$ denotes inversion of a square matrix or pseudo inversion of a rectangular matrix; $\mathrm{diag}(\cdot)$ retains only the

main diagonal components of the matrix. $\boldsymbol{W}_s(f)$ is the demixing matrix $\boldsymbol{W}_{\text{demix}}(f)$, which we are looking for.

Finally, the demixing network $\boldsymbol{W}(n)$ is obtained by inverse Fourier transforming $\boldsymbol{W}_s(f)$, and the estimated source $\boldsymbol{y}(n)$ is obtained by filtering $\boldsymbol{x}(n)$ through $\boldsymbol{W}(n)$.

11.2.2 Circular Convolution Approximation Problem

Besides the permutation and scaling ambiguity, another problem also affects the performance of frequency-domain BSS: the STFT circular convolution approximation [24, 37, 43]. The convolutive mixture is decomposed into an instantaneous mixture at each frequency bin as shown in (11.3). Equation (11.3) is only an approximation since it implies a circular convolution but not a linear convolution in the time domain. It is correct only when the STFT analysis frame length L is larger than the mixing filter length P. Thus, a large L is required to ensure a sufficient separation performance. However in that case, the instantaneous separation performance is saturated before reaching a sufficient level, because decreased time resolution for STFT and fewer data available in each frequency bin will violate the independence assumption.

To verify the statement above, a simple example is given below. As is well known, non-Gaussianity is an important measure for the independence of signals while kurtosis is an important measure for non-Gaussianity [8]. The kurtosis of a signal $\boldsymbol{s}$ is defined as

$$\text{kurt}(s) = \text{E}\{s^4\} - 3(\text{E}\{s^2\})^2 \tag{11.7}$$

where the operator $\text{E}\{\cdot\}$ denotes expectation. A high kurtosis value indicates strong non-Gaussianity and independence. We compare kurtosis values of the STFT coefficients of a speech signal when different STFT frame sizes (varying from 128 to 16,384) are used. The kurtosis value is calculated for the real and imaginary parts of the complex-valued coefficients, respectively. Since the kurtosis value, which is calculated for the time sequences at each frequency bin, varies with respect to frequency, a median value is chosen from the set of kurtosis values at all frequencies to represent the independence measure of the signal after STFT analysis. Considering the possible influence of insufficient data points at each frequency bin after a long-frame STFT, three speech signals with lengths of 10s, 40s, and 160s, respectively, are tested. The obtained kurtosis for different test signals and different STFT frame sizes are shown in Fig. 11.2. For reference, the kurtosis of a normalized Gaussian white signal is also given. As can be seen in Fig. 11.2, the real and imaginary parts of the STFT coefficients show similar variation trend with respect to the STFT frame size. Additionally, two phenomena can be observed.

(1) Large kurtosis values can be observed for small STFT frame sizes. The kurtosis value increases slightly with increased STFT frame size and then decreases significantly when the STFT frame size is larger than 1,024. The kurtosis value is close to a Gaussian white signal when the STFT frame size is very large.

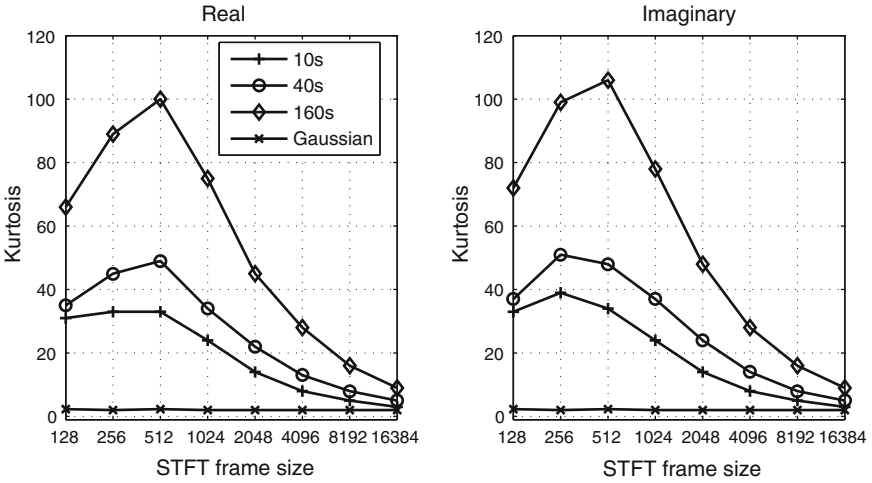


Fig. 11.2 Kurtosis of the STFT coefficients versus STFT frame size (calculated for speech signals of different lengths)

This demonstrates that the STFT coefficients of a speech signal show strong non-Gaussianity for small STFT frame sizes, but tend to be Gaussian for large STFT frame sizes.

(2) Even for a same STFT frame size, the kurtosis of a speech signal with different length is different. Generally, a long signal shows higher kurtosis than short signals. This may be due to insufficient data points available at each frequency bin with short signals.

The results shown in Fig. 11.2 demonstrate that the independence assumption of sources may collapse when a large STFT frame size is used, hence degrading the separation performance significantly. There is a dilemma in determining the STFT frame size: short frames make the conversion to instantaneous mixture incomplete, while long ones disturb the separation. The conflict becomes severer in highly reverberant environments and leads to the degraded performance. Generally, a frequency-domain BSS which works well in low (100–200 ms) reverberation has degraded performance in medium (200–500 ms) and high (>500 ms) reverberation. Since the problem originates from a processing step, which approximates linear convolutions with circular convolutions in frequency-domain BSS, we call it circular convolution approximation problem.

Furthermore, some separation experiments (2×2 and 4×4) are carried out in a reverberant environment (reverberation time 300 ms) using a frequency-domain BSS algorithm proposed in [24] with different STFT frame sizes. The source speech signals are of 8s length and 8 kHz sampling rate.[1] The resultant signal-interference-ratios (SIR) are shown in Fig. 11.3. In both 2×2 and 4×4 cases, the separation

[1] More details of the experiment can be found in Sect. 11.4.2.

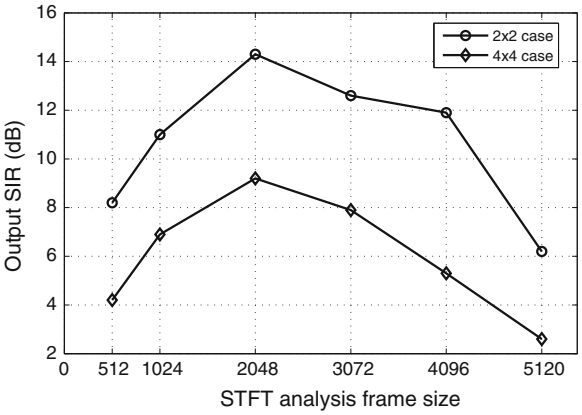


Fig. 11.3 Performance of BSS versus STFT frame size (calculated for speech signals of 8s length, $RT_{60} = 300\,ms$)

performance peaks at the STFT frame size of 2,048, while degrading for both shorter and longer frame sizes. This verifies the discussion about the dilemma in determining the STFT frame size. Obviously, an optimal STFT frame size may exist for a specific reverberation. However, due to complex acoustical environments and varieties of source signals, it is difficult to determine this value precisely. Generally, at sampling frequency of 8,000 Hz, 1,024 or 2,048 can be used as a balanced choice for the frame length.

11.3 Superdirective Beamforming

Beamforming is a technique used in sensor arrays for directional signal reception by enhancing target directions and suppressing unwanted ones. Beamforming can be classified as either fixed or adaptive, depending on how the beamformer weights are chosen. An adaptive beamformer obtains directive response mainly by analyzing the statistical information contained in the array data, not by utilizing the spatial information directly. It generally adapts its weights during breaks in the target signal. The challenge to predict signal breaks when several people are talking concurrently limits the feasibility of adaptive beamforming in cocktail-party environments significantly. In contrast, the weights of a fixed beamformer do not depend on array data and are chosen to present a specified response for all scenarios. The directional response is achieved by coherently summing signals from multiple sensors based on a model of the wavefront from acoustic sources. A filter-and-sum beamformer has super directivity response with optimized weights. The superdirective beamformer can be designed in the frequency-domain.

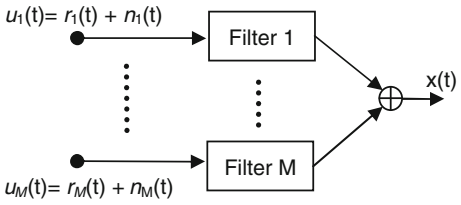


Fig. 11.4 Principle of a filter-and-sum beamformer

The principle of a fixed beamformer is given in Fig. 11.4, where a weighted sum of signals from M sensors is produced to enhance the target direction. Suppose a beamformer model with a target source $r(t)$ and background noise $n(t)$, the components received by the l-th sensor is $u_l(t) = r_l(t) + n_l(t)$ in the time domain. Similarly, in the frequency domain, the l-th sensor output is $u_l(f) = r_l(f) + n_l(f)$. The array output in the frequency domain is

$$\boldsymbol{x}(f) = \boldsymbol{b}^{\mathrm{H}}(f)\boldsymbol{u}(f) \tag{11.8}$$

where $\boldsymbol{b}(f) = [b_1(f), \ldots, b_M(f)]^{\mathrm{T}}$ is the beamforming weight vector composed of beamforming weights from each sensor, and $\boldsymbol{u}(f) = [u_1(f), \ldots, u_M(f)]^{\mathrm{T}}$ is the output vector composed of outputs from each sensor, and $(\cdot)^{\mathrm{H}}$ denotes conjugate transpose. The $\boldsymbol{b}(f)$ depends on the array geometry and source directivity, as well as the array output optimization criterion such as a signal-to-noise ratio (SNR) gain criterion [4, 49, 50].

Suppose $\boldsymbol{r}(f) = [r_1(f), \ldots, r_M(f)]^{\mathrm{T}}$ is the source vector, which is composed of the target source signals from the sensors, and $n(f)$ is the noise vector, which is composed of the spatial diffuse noises from the sensors. The array gain is a measure of the improvement in signal-to-noise ratio. It is defined as the ratio of the SNR at the output of the beamforming array to the SNR at a single reference microphone. For development of the theory, the reference SNR is defined to be the ratio of average signal power spectral densities over the microphone array, $\sigma_r^2(f) = E\{\mathbf{r}^{\mathrm{H}}(f)\mathbf{r}(f)\}/M$, to the average noise power spectral density over the array, $\sigma_n^2(f) = E\{\mathbf{n}^{\mathrm{H}}(f)\mathbf{n}(f)\}/M$. By derivation, the array gain at frequency f is expressed as

$$G(f) = \frac{\mathbf{b}^{\mathrm{H}}(f)\mathbf{R}_{rr}(f)\mathbf{b}(f)}{\mathbf{b}^{\mathrm{H}}(f)\mathbf{R}_{nn}(f)\mathbf{b}(f)} \tag{11.9}$$

where $\boldsymbol{R}_{rr}(f) = \boldsymbol{r}(f)\boldsymbol{r}^{\mathrm{H}}(f)/\sigma_r^2(f)$ is the normalized signal cross-power spectral density matrix, and $\boldsymbol{R}_{nn}(f) = \boldsymbol{n}(f)\boldsymbol{n}^{\mathrm{H}}(f)/\sigma_n^2(f)$ is the normalized noise cross-power spectral density matrix. Provided $\boldsymbol{R}_{nn}(f)$ is nonsingular, the array gain is maximized with the weight vector

$$\boldsymbol{b}_{\mathrm{opt}}(f) = \boldsymbol{R}_{nn}^{-1}(f)\boldsymbol{r}(f) \tag{11.10}$$

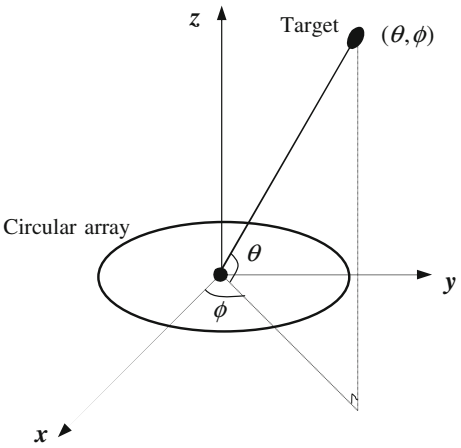


Fig. 11.5 Circular array geometry

The terms $\boldsymbol{R}_{nn}(f)$ and $\boldsymbol{r}(f)$ in (11.10) depend on the array geometry and the target source direction. For instance, given a circular array, $\boldsymbol{R}_{nn}(f)$ and $\boldsymbol{r}(f)$ can be calculated as below [45].

Figure 11.5 shows an M-element circular array with a radius of r and a target source coming from the direction (θ, ϕ). The elements are equally spaced around the circumference, and their positions, which are determined from the layout of array, are given in a matrix form as

$$\boldsymbol{v} = \begin{bmatrix} v_{x_1} & v_{y_1} \\ \vdots & \vdots \\ v_{x_M} & v_{y_M} \end{bmatrix} \tag{11.11}$$

The source vector $\boldsymbol{r}(f)$ can be derived as

$$\boldsymbol{r}(f) = \begin{bmatrix} \exp(-jk(\sin\theta \cdot \cos\phi \cdot v_{x_1} + \sin\theta \cdot \sin\phi \cdot v_{y_1})) \\ \vdots \\ \exp(-jk(\sin\theta \cdot \cos\phi \cdot v_{x_M} + \sin\theta \cdot \sin\phi \cdot v_{y_M})) \end{bmatrix} \tag{11.12}$$

where $k = 2\pi f/c$ is the wave number, and c is the sound velocity. The normalized noise cross-power spectral density matrix $\boldsymbol{R}_{nn}(f)$ is expressed as

$$(\boldsymbol{R}_{nn}(f))_{m_1 m_2} = \begin{cases} \frac{\sin(kd_{m_1 m_2})}{kd_{m_1 m_2}}, & m_1 \neq m_2 \\ 1, & m_1 = m_2 \end{cases} \tag{11.13}$$

where $(\boldsymbol{R}_{nn}(f))_{m_1 m_2}$ is the (m_1, m_2) entry of the matrix $\boldsymbol{R}_{nn}(f)$, $m_1, m_2 = 1, \cdots, M$, k is the wave number, $d_{m_1 m_2}$ is the distance between two microphones m_1 and m_2.

After calculating the beamforming vector by (11.10), (11.12) and (11.13) at each frequency bin, the time-domain beamforming filter $\boldsymbol{b}(n)$ is obtained by inverse Fourier transforming $\boldsymbol{b}_{\text{opt}}(f)$.

The procedure above is to design a beamformer with only one target direction. In a speech separation or extraction system, each source signal may be separately obtained using the directivity of the array, if the directions of sources are known. However, beamforming in principle performs limitedly in highly reverberant conditions because it cannot suppress the interfering reverberation coming from the target direction.

11.4 Enhanced Separation in Reverberant Environments by Combining Beamforming and BSS

Due to the circular convolution approximation problem, the performance of a frequency-domain BSS algorithm degrades seriously when the mixing filters are long, e.g., in high reverberation environments. Thus, the problem may be mitigated if the mixing filters become shorter. With directive response enhancing desired direction and suppressing unwanted ones, a superdirective beamforming can deflate the reflected paths and hence equivalently shorten the mixing filter. It thus may help compensate for the deficiency of blind source separation. If we use beamforming as a preprocessor for blind source separation, at least three advantages can be achieved:

(1) The interfering residuals due to reverberation after beamforming are further reduced by blind source separation;
(2) The poor separation performance of blind source separation in reverberant environments is compensated for by beamforming, which suppresses the reflected paths and shortens the mixing filters;
(3) Beamformer enhances the source in its path and suppresses the ones outside. It thus enhances signal-to-noise ratio and provides a cleaner output for blind source separation to process.

From another point of view, beamforming makes primary use of spatial information while blind source separation utilizes statistical information contained in signals. Integrating both pieces of information should help to get better separation results, just like the way our ears separate audio signals. The details of the combined method are given below, followed by experimental results and analysis.

11.4.1 Workflow of the Combined Method

The illustration of the combined method is shown in Fig. 11.6. For N sources received by an array of M microphones, N beams are formed toward them, respectively,

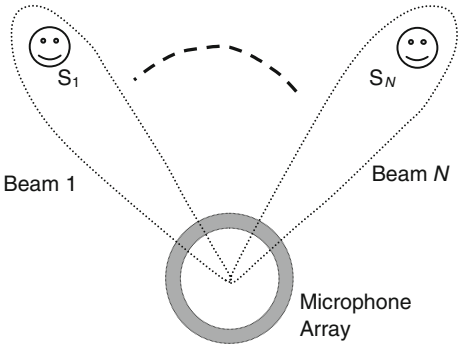


Fig. 11.6 Illustration of the proposed method combining beamforming and blind source separation

assuming the directions of all sources are known. Then the N beamformed outputs are fed to blind separation to recover the N sources. The signal flow of the proposed method is shown in Fig. 11.7, which mainly consists of three stages: acoustic mixing, beamforming, and separation.

The mixing stage results in the observed vector

$$\boldsymbol{u}(n) = \boldsymbol{H}(n) * \boldsymbol{s}(n) \tag{11.14}$$

where $\boldsymbol{u}(n) = [u_1(n), \ldots, u_M(n)]^T$ and $\boldsymbol{s}(n) = [s_1(n), \ldots, s_N(n)]^T$ are the observed and the source vectors, respectively, $\boldsymbol{H}(n)$ is a sequence of $M \times N$ matrices containing the impulse responses of the mixing channels, and the operator '*' denotes matrix convolution.

The beamforming stage is expressed as

$$\boldsymbol{x}(n) = \boldsymbol{B}(n) * \boldsymbol{u}(n) = \boldsymbol{B}(n) * \boldsymbol{H}(n) * \boldsymbol{s}(n) = \boldsymbol{F}(n) * \boldsymbol{s}(n) \tag{11.15}$$

where $x(n) = [x_1(n), \ldots, x_N(n)]^T$ is the beamforming output vector, $B(n)$ is a sequence of $N \times M$ matrices containing the impulse responses of beamformer, $F(n)$ is the global impulse response by combining $H(n)$ and $B(n)$.

The blind source separation stage is expressed as

$$\boldsymbol{y}(n) = \boldsymbol{W}(n) * \boldsymbol{x}(n) = \boldsymbol{W}(n) * \boldsymbol{F}(n) * \boldsymbol{s}(n) \tag{11.16}$$

where $\boldsymbol{y}(n) = [y_1(n), \ldots, y_N(n)]^T$ is the estimated source signal vector, and $\boldsymbol{W}(n)$ is a sequence of $N \times N$ matrices containing the unmixing filters.

It can be seen from (11.14)–(11.16) that with beamforming reducing reverberation and enhancing signal-to-noise ratio, the combined method is able to replace the original mixing network $\boldsymbol{H}(n)$, which results from the room impulse response, with a new mixing network $\boldsymbol{F}(n)$, which is easier to separate.

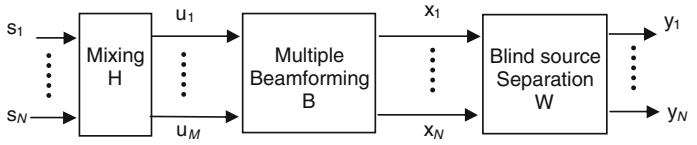


Fig. 11.7 Signal flow of the proposed method combining beamforming and blind source separation

The blind source separation and beamforming algorithms introduced in Sects. 11.2 and 11.3 can be used directly for the combined method. However, the following two issues should be clarified when implementing the combined method.

(1) *The choice of a beamformer*
Beamformer can be implemented as a fixed one or an adaptive one. As mentioned before, comparing to fixed beamforming, an adaptive method is not appropriate for the combined method. First, an adaptive beamformer obtains directive response mainly by analyzing the statistical information contained in the array data, not by utilizing the spatial information directly. Its essence is similar to that of convolutive blind source separation [37, 38]. Cascading them together is equivalent to using the same techniques repeatedly, hence contributing little to performance improvement. Second, An adaptive beamformer generally adapts its weights during breaks in the target signal. However, it is a challenge to predict signal breaks when several people are talking concurrently. This significantly limits the applicability of adaptive beamforming to source separation. In contrast, a fixed beamformer, which relies mainly on the spatial information, does not have such disadvantages. It is data independent and more robust. With its directive response fixed in all acoustic scenarios, a superdirective beamformer is preferred in the combined method.

(2) *The permutation ambiguity problem in BSS*
Permutation ambiguity inherent in frequency-domain BSS is always a challenging problem. Generally, there are two approaches to solve it. One is to exploit the dependence of separated signals across frequencies, and the other is to exploit the position information of sources: the directivity pattern of the mixing/unmixing matrix provides a good reference for permutation alignment. However, in the proposed method, the directivity information contained in the mixing matrix does not exist any longer after beamforming. Even if the source positions are known, they are not much helpful to permutation alignment in the subsequent blind source separation. Consequently, what we can use for permutation is merely the first reference: the interfrequency dependence of separated signals.

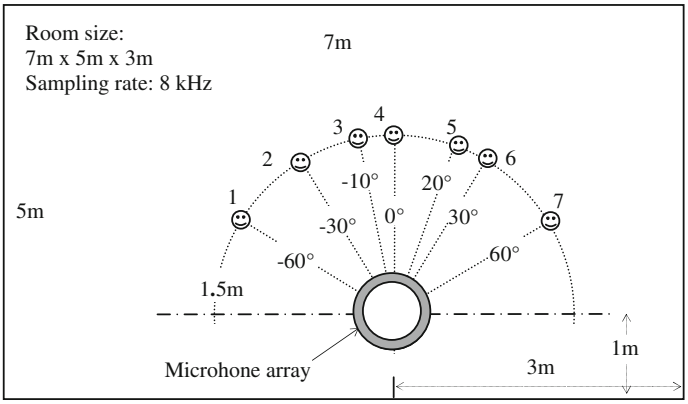


Fig. 11.8 Simulated room environment for speech separation

11.4.2 Experimental Results and Analysis

We evaluate the performance of the proposed method in simulated experiments from two aspects. The first experiment verifies the advantage of the beamforming preprocessing, i.e., dereverberation and noise reduction; the second one investigates the performance of the proposed method in various reverberant conditions, and compares it with a BSS-only method and a beamforming-only one.

The simulation environment is shown in Fig. 11.8, the room size is $7\text{m} \times 5\text{m} \times 3\text{m}$, all sources and microphones are 1.5 m high. The room impulse response was obtained by using the image method [51], and the reverberation time was controlled by varying the absorption coefficient of the wall. The sampling rate is 8 kHz. For BSS, a STFT frame size of 2,048 is used. For beamforming, a circular microphone array is used to design the beamformer with the filter length 2,048. A commonly used objective measure, signal-to-interference ratio (SIR), is employed to evaluate the separation performance [45].

11.4.2.1 Influence of Beamforming Preprocessing

The proposed algorithm is used for separating three sources in the environment shown in Fig. 11.8, using a 16-element circular microphone array with a radius of 0.2 m. The simulated room reverberation time is $\text{RT}_{60} = 300$ ms, where RT_{60} is the time required for the sound level to decrease by 60 dB. This is a medium reverberant condition. Three source locations (2, 4, 6) are used, and the sources are two male speeches and one female speech of 8s each. Three beams are formed by the microphone array pointing at the three sources, respectively. Impulse responses associated with the global transfer function of beamforming is shown in Fig. 11.10, which are calculated from the impulse responses of mixing filters and beamforming filters using

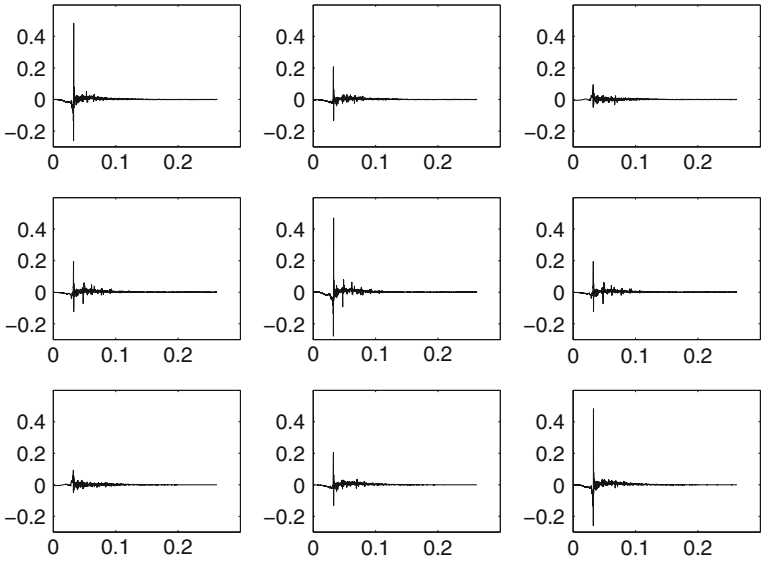


Fig. 11.9 Global impulse responses after beamforming

$$\boldsymbol{F}(n) = \boldsymbol{B}(n) * \boldsymbol{H}(n) \tag{11.17}$$

It can be seen that the diagonal components in Fig. 11.9 are superior to off-diagonal ones. This implies that the target sources are dominant in the outputs. To demonstrate the dereverberation performance of beamforming, the top left panel in Fig. 11.9 is enlarged in Fig. 11.10 and compared with the original impulse response. Obviously, the mixing filter becomes shorter after beamforming, and the reverberation becomes smaller. This indicates that dereverberation is achieved. So far, the two advantages of beamforming, dereverberation and noise reduction, are observed as expected. Thus, the new mixing network (11.17) should be easier to separate than the original mixing network. In this experiment, the average input SIR is -2.8 dB, and the output one, enhanced by beamforming, is 3.3 dB. Applying BSS to the beamformed signals, we get an average output SIR of the combined method of 16.3 dB, a 19.1 dB improvement over the input: 6.1 dB improvement at the beamforming stage, and 13 dB further improvement at the BSS stage.

11.4.2.2 Performance in Reverberant Environments

The performances of the combined method, the BSS-only method and the beamforming-only method, are compared in the simulated environment shown in Fig. 11.8 with different reverberation times. The beamforming-only method is just the first processing stage of the combined method. For the combined method,

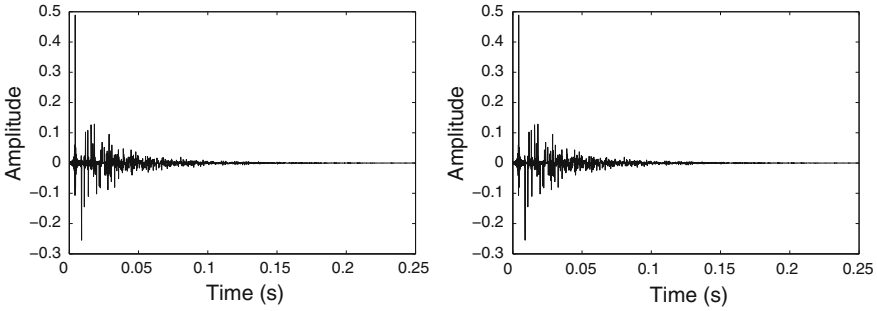


Fig. 11.10 Comparison of the impulse responses before and after beamforming: the left panel is simulated room impulse response for $RT_{60} = 300$ ms; the right panel is the resultant impulse response after beamforming

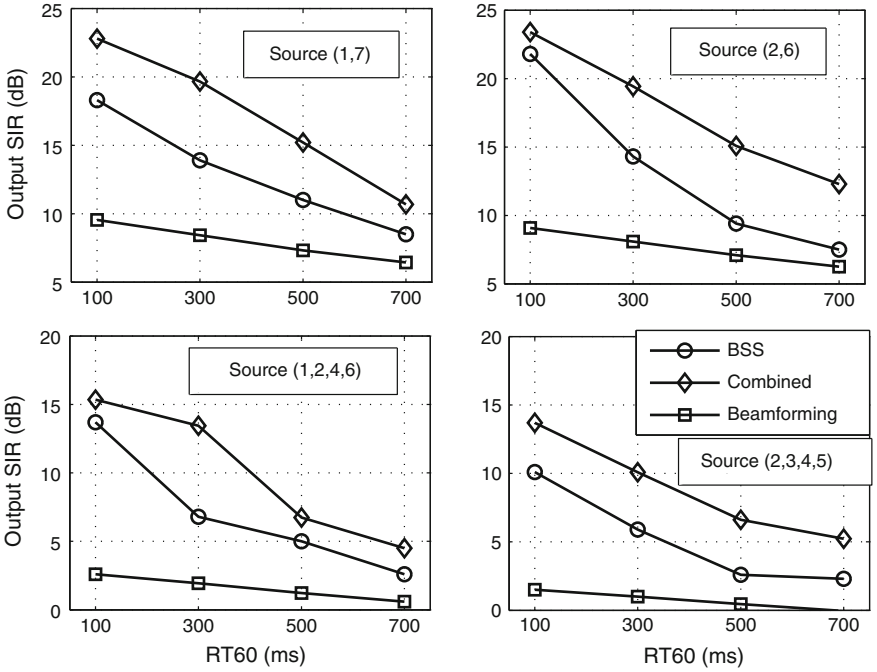


Fig. 11.11 Performance comparison between the combined method, the BSS-only method and the beamforming-only method in different reverberant conditions

a 16-element microphone array with a radius of 0.2 m is used. For the BSS-only method, a linear array consisting of four microphones (inter-space of 6 cm) is used instead of the circular array. Various combinations of source locations are tested (2 sources and 4 sources). The sources are two male speeches and two female speeches of 8s each. RT_{60} ranges from 100 to 700 ms in increments of 200 ms. The average

input SIR does not vary significantly with the reverberation time: it is about 0 dB for two-source cases, and −5 dB for four-source cases. For all three methods, the STFT frame size is set at 2,048. The separation results are shown in Fig. 11.11, with each panel depicting the output SIRs of the three methods for one source combination. It is observed in Fig. 11.11 that for each source configuration, the output SIRs of all methods decrease with increasing reverberation; however, the combined method always outperforms the other two. Beamforming performs worst among the three methods; however, it provides a good preprocessing result, and hence the combined method works better than the BSS-only method.

It is interesting to investigate how big an improvement one can obtain by the use of beamforming preprocessing in different reverberation values. To measure the contribution of this preprocessing, we define the relative improvement of the combined method over the BSS-only method as

$$I_R = \frac{I_c - I_b}{I_b} \times 100\,\% \tag{11.18}$$

where I is the obtained SIR improvement with the subscripts $(\cdot)_b$ and $(\cdot)_c$ standing for the BSS-only method and the combined method, respectively. We calculate the relative performance improvement for the four separation scenarios listed in Fig. 11.11 and show the average result in Fig. 11.12. As discussed previously, the performance is improved by the combined method for all reverberant conditions. However, it is also observed in Fig. 11.12 that the improvement in low reverberation is not as large as in medium and high reverberation. That is, the use of beamforming in low reverberation is not as beneficial as it would be for high reverberation. The reason is that BSS can work well alone when the circular convolution approximation problem is not evident in low reverberation, and thus the contribution of preprocessing is small. On the other hand, when the circular convolution approximation problem becomes severe in high reverberation, the contribution of preprocessing becomes crucial and hence the separation performance is improved significantly.

Based on the two experiments above, a conclusion can be drawn: With beamforming shortening mixing filters and reducing noise before blind source separation, the combined method performs better than using beamforming or blind source separation alone in highly reverberant environments. A disadvantage of the proposed method is that it requires the knowledge of source locations for beamforming. Generally, the source locations may be estimated with an array sound source localization algorithm [52–54].

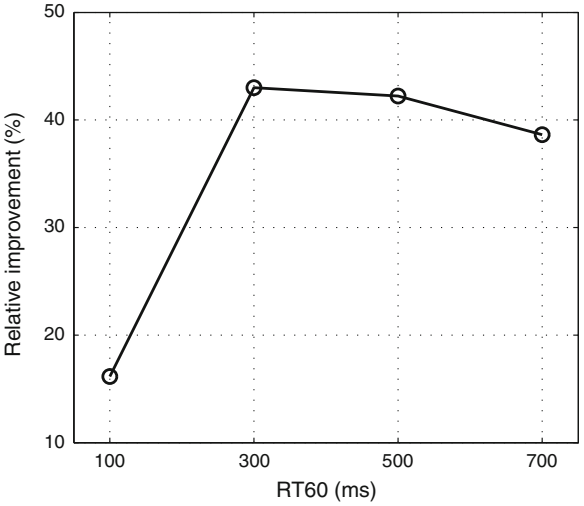


Fig. 11.12 Relative performance improvement of the combined method over the BSS-only method in different reverberant environments

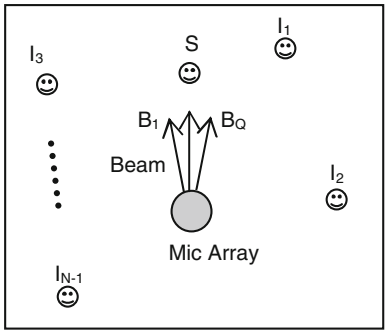


Fig. 11.13 Illustration of the proposed method combining beamforming and blind source separation for target speech extraction

11.5 Target Speech Extraction in a Cocktail-Party Environment

11.5.1 Target Speech Extraction by Combining Beamforming and BSS

In this section, the combined method is extended to a special application of target speech extraction where only the position of the target speaker is known. In real cocktail-party environments, each speaker may move and talk freely. This is very difficult to handle with blind source separation or beamforming alone. Fortunately, it is often in such a case that the target speaker stays in a position or moves slowly and the noisy environment around it is time-varying, *e.g.*, moving interfering speakers and the ambient noise. For this specific situation, a target speech extraction method

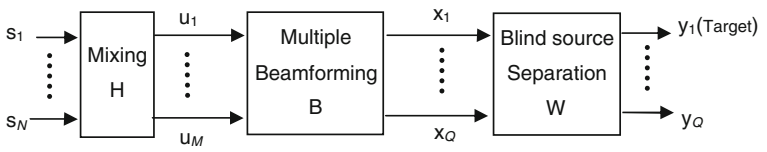


Fig. 11.14 Signal flow of the proposed method combining beamforming and blind source separation for target speech extraction

by combining beamforming and blind source separation is proposed. The principle of the proposed method is illustrated in Fig. 11.13, where the target source S and $N-1$ interfering sources $I_1, \ldots, I_{N-1}$, are convolutively mixed and observed at an array of M microphones. To extract the target, Q beams ($Q \leq N$) are formed at an area containing it, with a small separation angle between adjacent beams; then the Q beamformed outputs are fed to a blind separation scheme. Using beamforming as a preprocessor for BSS, the target signal becomes dominant in the output of the beamformer and is hence easier to extract. Furthermore, as seen in Fig. 11.13, the beams are pointing at an area containing the target, as opposed to the interfering sources. This is very important for operation under a time-varying condition, because of the following reasons:

(1) When the target speaker remains in a constant position while others move, it is impractical to know all speakers' positions and steer a beam at each of them;
(2) There is no need to steer the beams at individual speakers since only the target speaker is of interest;
(3) The target signal is likely to become dominant in at least one of the beamformed output channels if the beams point at an area containing the target speaker. Thus, it is possible to extract it as an independent source even if the number of beams is less than the sources [55]. This feature is very important for the proper operation of the proposed method;
(4) A seamless beam area will be formed by several beams with each covering some beamwidth. It is possible to extract the target signal even if it moves slightly inside this area. This feature may improve the robustness of the proposed method; and
(5) The fact that there are fewer beams than sources reduces the dimensionality of the problem and saves computation.

The signal flow of the proposed method is shown in Fig. 11.14, which is similar to the one shown in Fig. 11.7. The same implementation of beamforming and blind source separation is also employed.

11.5.2 Experimental Results and Analysis

We evaluate the performance of the proposed method in simulated conditions. A typical cocktail-party environment with moving speakers and ambient noises is shown

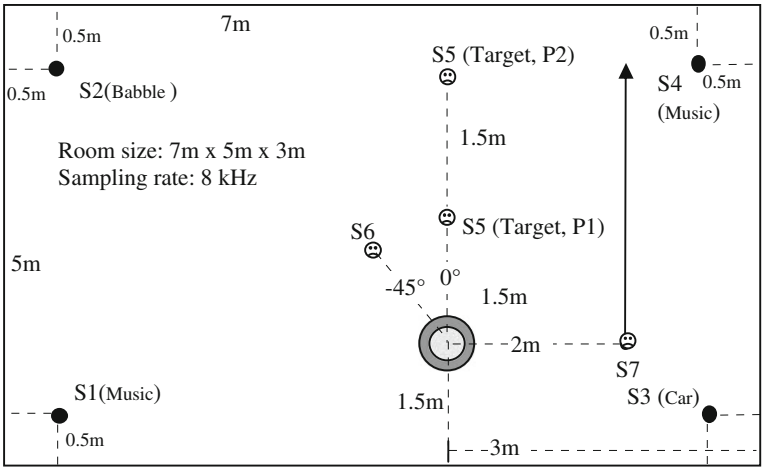


Fig. 11.15 Simulated room environment for target speech extraction

in Fig. 11.15. The room size is 7m × 5m × 3m, and all sources and microphones are 1.5 m high. Four loudspeakers S1–S4 placed near the corners of the room play various interfering sources. Loudspeakers S5, S6, and S7 play speech signals concurrently. S5 and S6 remain in fixed positions, while S7 moves back and forth at a speed of 0.5 m/s. As the target, S5 is placed at either position P1 or P2. S5 simulates a female speaker, while S6 and S7 simulate male speakers. An 8-element circular microphone array with a radius of 0.1 m is placed as shown.

Three beams are formed toward S5, with the separation angle between two adjacent beams being 20°. The room impulse responses are obtained by using the image method, with the reverberation time controlled by varying the absorption coefficient of walls [51]. The test signals last 8s with a sampling rate of 8 kHz. The extraction performance is evaluated in terms of SIR where the signal is the target speech.

With so many speakers present in such a time-varying environment, BSS alone fails to work. Now, we compare the performance of beamforming alone and the proposed method with reverberation RT_{60} of 130 and 300 ms, respectively. The results are given in Table 11.1. As an example, for the close target case (P1) under $RT_{60} = 300$ ms, the input SIR is around -9 dB—the target is almost completely buried in noises and interference. The enhancement by beamforming alone is moderate. On the other hand, the proposed two-stage method improves the SIR by 15.1 dB. In the far target case (P2) of $RT_{60} = 300$ ms, the target signal received at the microphones is much weaker with an input SIR around only -11 dB. The proposed method is still able to extract the target signal with an output SIR of 3.3 dB and a total SIR improvement of 13.5 dB.

For the close target case (P1) under $RT_{60} = 300$ ms, Fig. 11.16 shows the waveforms at various processing stages: sources, microphone signals, beamformer outputs, and finally the BSS outputs. It can be seen that the target signal S5 is totally

Table 11.1 Comparison of beamforming and the proposed method in terms of signal-to-interference ratio (SIR)

Target S5	P1 (close)		P2 (far)	
RT_{60}	130 ms	300 ms	130 ms	300 ms
Input SIR	−8.2 dB	−9.1 dB	−10.7 dB	−10.8 dB
Beamforming	4.6 dB	0.6 dB	2.5 dB	−2.3 dB
Proposed method	11.9 dB	6.0 dB	9.1 dB	3.3 dB
SIR improvement	20.1 dB	15.1 dB	19.8 dB	13.5 dB

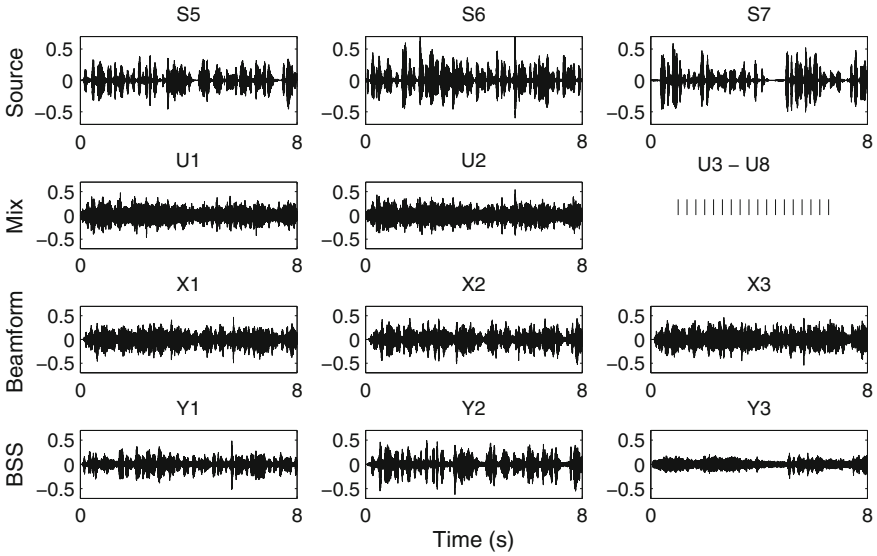


Fig. 11.16 Waveforms at various processing stages

buried in noises and interference in the mixture signals; it is enhanced to a certain degree after beamforming but is still difficult to tell from the background; and after blind source separation, the target signal is clearly exhibited at the channel Y2. In addition, an interference signal (S6) is observed at the output channel Y1, and the noise-like output Y3 is mainly composed of the interfering speech S7 and other noises. The extraction result verifies the validity of the proposed method in noisy cocktail-party environments. Some audio demos can be found at [56].

The good performance of the proposed method in such time-varying environments is due to two reasons. First, fixed beamforming can enhance target signals even in time-varying environments. Second, the spectral components of the target and (moving or static) interfering signals are still independent after beamforming; besides, the target signal becomes dominant in the output of the beamformer. This helps the subsequent blind source separation.

11.6 Conclusions and Prospects

Given the poor performance of blind source separation and beamforming alone in real cocktail-party environments, the chapter proposes a combined method using superdirective beamforming as a preprocessing step of blind source separation. Superdirective beamforming shortens mixing filters and reduces noise for blind source separation, which further reduces the residual interferences. By exploiting both spatial and statistical information, the proposed method can integrate the advantages of beamforming and blind source separation and complement the weakness of them alone. Good results can be obtained when applying the proposed method for speech separation in highly reverberant environments and target speech extraction in dynamic cocktail-party environments.

Although great potentials of the proposed method have been shown, there are still some open problems that need to be addressed. Specifically, beamforming requires the speaker location information to form the beam, but the proposed method in its current form is not capable of identifying the locations, especially with moving speakers. In addition, the separation performance is still limited by the microphone array size, making it a challenge to apply the proposed method to pocket-size applications. These will be investigated in our future research.

Acknowledgments This work is partly supported by the Alexander von Humboldt Foundation.

References

1. Van Veen, B.D., Buckley, K.M.: Beamforming: A versatile approach to spatial filtering. IEEE ASSP Magazine **5**, 4–24 (1988)
2. Van Trees, H.L.: Optimum Array Processing - Part IV of Detection, Estimation, and Modulation Theory, Chapter 4, pp. 231–331, Wiley-Interscience (2002)
3. Griffiths, L.J., Jim, C.W.: An alternative approach to linearly constrained adaptive beamforming. IEEE Trans. Antennas Propag. **30**(1), 27–34 (1982)
4. Cox, H., Zeskind, R.M., Kooij, T.: Practical supergain. IEEE Trans. Speech Audio Process.ing, ASSP-**34**(3), 393–398 (1986)
5. Doclo, S., Moonen, M.: Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. IEEE Trans. Signal Process. **51**(10), 2511–2526 (2003)
6. Doclo, S., Moonen, M.: GSVD-based optimal filtering for single and multimicrophone speech enhancement. IEEE Trans. Signal Process. **50**(9), 2230–2244 (2002)
7. Doclo, S., Spriet, A., Wouters, J., Moonen, M.: Frequency-domain criterion for the speechdistortion weighted multichannel Wiener filter for robust noise reduction. Speech Commun. **49**(7–8), 636–656 (2007)
8. Hyvarien, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
9. Cardoso, J.: Blind signal separation: statistical principles. Proc. IEEE **86**(10), 2009–2025 (1998)
10. Bingham, E., Hyvarien, A.: A fast fixed-point algorithm for independent component analysis of complex valued signals. Int. J. Neural Syst. **10**, 1–8 (2000)
11. Bell, A.J., Sejonwski, T.J.: An information maximization approach to blind separation and blind deconvolution. Neural Comput. **7**(6), 1129–1159 (1995)

12. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. Adv. Neural Inf. Process. Sys. **8**, 757–763 (1996)
13. Wang, W., Sanei, S., Chambers, J.A.: Penalty function based joint diagonalisation approach for convolutive blind separation of nonstationary sources. IEEE Trans. Signal Process. **53**(5), 1654–1669 (2005)
14. Pedersen, M.S., Larsen, J., Kjems, U., Parra, L.C.: A survey of convolutive blind source separation methods. In: Handbook on Speech Processing and Speech Communication, pp. 1–34, Springer (2007)
15. Douglas, S.C., Sun, X.: Convolutive blind separation of speech mixtures using the natural gradient. Speech Commun. **39**, 65–78 (2003)
16. Aichner, R., Buchner, H., Yan, F., Kellermann, W.: A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments. Sig. Process. **86**(6), 1260–1277 (2006)
17. Douglas, S.C., Gupta, M., Sawada, H., Makino, S.: Spatio-temporal FastICA algorithms for the blind separation of convolutive mixtures. IEEE Trans. Audio Speech Lang. Process. **15**(5), 1511–1520 (2007)
18. Sawada, H., Araki, S., Makino, S.: Frequency-domain blind source separation. In: Blind Speech Separation, pp. 47–78, Springer (2007)
19. Smaragdis, P.: Blind separation of convolved mixtures in the frequency domain. Neurocomputing **22**, 21–34 (1998)
20. Parra, L., Spence, C.: Convolutive blind separation of non-stationary sources. IEEE Trans. Speech Audio Process. **8**(3), 320–327 (2000)
21. Mei, T., Mertins, A., Yin, F., Xi, J., Chicharo, J.F.: Blind source separation for convolutive mixtures based on the joint diagonalization of power spectral density matrices. Sig. Process. **88**(8), 1990–2007 (2008)
22. Murata, N., Ikeda, S., Ziehe, A.: An approach to blind source separation based on temporal structure of speech signals. Neurocomputing **41**(1-4), 1–24 (2001)
23. Sawada, H., Araki, S., Makino, S.: Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS. In: 2007 IEEE International Symposium on Circuits and Systems, pp. 3247–3250 (2007)
24. Wang, L., Ding, H., Yin, F.: A region-growing permutation alignment approach in frequency-domain blind source separation of speech mixtures. IEEE Trans. Audio Speech Lang. Process. **19**(3), 549–557 (2011)
25. Wang, L., Ding, H., Yin, F.: An improved method for permutation correction in convolutive blind source separation. Arch. Acoust. **35**(4), 493–504 (2010)
26. Kim, T., Attias, H.T., Lee, S.Y., Lee, T.W.: Blind source separation exploiting higher-order frequency dependencies. IEEE Trans. Audio Speech Lang. Process. **15**(1), 70–79 (2007)
27. Mazur, R., Mertins, A.: An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models. IEEE Trans. Speech Audio Process. **17**(1), 117–126 (2009)
28. Serviere, C., Pham, D.T.: Permutation correction in the frequency domain in blind separation of speech mixtures. EURASIP J. Appl. Sig. Process. **2006**(1), 177–193 (2006)
29. Ono, N.: Stable and fast update rules for independent vector analysis based on auxiliary function technique. In: 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 189–192, New Paltz (2011)
30. Sawada, H., Araki, S., Makino, S.: Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. IEEE Trans. Audio Speech Lang. Process. **19**(3), 516–527 (2011)
31. Saruwatari, H., Kurita, S., Takeda, K.: Blind source separation combining independent component analysis and beamforming. EURASIP J. Appl. Sig. Process. **2003**(11), 1135–1146 (2003)
32. Ikram, M.Z., Morgan, D.R.: Permutation inconsistency in blind speech separation: investigation and solutions. IEEE Trans. Speech Audio Process. **13**(1), 1–13 (2005)
33. Sawada, H., Mukai, R., Araki, S., Makino, S.: A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Trans. Speech Audio Process. **12**(5), 530–538 (2004)

34. Nesta, F., Svaizer, P., Omologo, M.: Convolutive BSS of short mixtures by ICA recursively regularized across frequencies. IEEE Trans. Audio Speech Lang. Process. **19**(3), 624–639 (2011)
35. Nesta, F., Wada, T.S., Juang, B.: Coherent spectral estimation for a robust solution of the permutation problem. In: 2009 IEEE Workshop on Application of Signal Processing to Audio and Acoustics, pp. 1–4, New Paltz, New York (2009)
36. Liu, Q., Wang, W., Jackson, P.: Use of bimodal coherence to resolve the permutation problem in convolutive BSS. Sig. Process. **92**(8), 1916–1927 (2012)
37. Araki, S., Mukai, R., Makino, S., Nishikawa, T., Saruwatari, H.: The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. IEEE Trans. Speech Audio Process. **11**(2), 109–116 (2003)
38. Parra, L., Fancourt, C.: An adaptive beamforming perspective on convolutive blind source separation. In: Davis, G.M. (ed.) Noise Reduction in Speech Applications, pp. 361–376. CRC Press (2002)
39. Ikram, M.Z., Morgan, D.R.: A beamforming approach to permutation alignment for multichannel frequency-domain blind speech separation. In: 2002 IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 881–884 (2002)
40. Parra, L.C., Alvino, C.V.: Geometric source separation: Merging convolutive source separation with geometric beamforming. IEEE Trans. Speech Audio Process. **10**(6), 352–362 (2002)
41. Saruwatari, H., Kawamura, T., Nishikawa, T., Lee, A., Shikano, K.: Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. IEEE Trans. Audio Speech Lang. Process. **14**(2), 666–678 (2006)
42. Gupta, M., Douglas, S.C.: Beamforming initialization and data prewhitening in natural gradient convolutive blind source separation of speech mixtures. In: Independent Component Analysis and Signal Separation, vol. 4666, pp. 512–519, Springer, Berlin (2007)
43. Nishikawa, T.,Saruwatari, H., Shikano, K.: Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **E86-A**(4), 846–858 (2003)
44. Chen, J., Van Veen, B.D., Hecox, K.E.: External ear transfer function modeling: a beamforming approach. J. Acoust. Soc. Am. **92**(4), 1933–1944 (1992)
45. Wang, L., Ding, H., Yin, F.: Combining superdirective beamforming and frequency-domain blind source separation for highly reverberant signals. EURASIP J. Audio Speech Music Process. **2010**, 1–13 (2010). (Article ID 797962)
46. Wang, L., Ding, H., Yin, F.: Target speech extraction in cocktail party by combining beamforming and blind source separation. IEEE Trans. Audio Speech Lang. Process. **39**(2), 64–67 (2011)
47. Pan, Q., Aboulnasr, T.: Combined spatial/beamforming and time/frequency processing for blind source separation. In: European Signal Processing Conference 2005, Antalya, Turkey, pp. 1–4 (2005)
48. Matsuoka, K., Nakashima, S.:Minimal distortion principle for blind source separation. In: 2001 International Workshop on Independent Component, pp. 722–727 (2001)
49. Ryan, J.G., Goubran, R.A.: Array optimization applied in the near field of a microphone array. IEEE Trans. Speech Audio Process. **8**(2), 173–176 (2000)
50. Bouchard, C., Havelock, D.I.: Beamforming with microphone arrays for directional sources. J. Acoust. Soc. Am. **125**(4), 2098–2104 (2008)
51. Allen, J.B., Berkley, D.A.: Image method for efficiently simulating small room acoustics. J. Acoust. Soc. Am. **65**, 943–950 (1979)
52. Silverman, H.F., Yu, Y., Sachar, J.M., Patterson, W.R.: Performance of real-time source-location estimators for a large-aperture microphone array. IEEE Trans. Speech Audio Process. **13**(4) (2005)
53. Madhu, N., Martin, R.: A scalable framework for multiple speaker localisation and tracking. In: 2008 International Workshop on Acoustic Echo and Noise Control, Seatle, Washington, pp. 1–4, (2008)

54. Maazaoui, M., Abed-Meraim, K., Grenier, Y.: Blind source separation for robot audition using fixed HRTF beamforming. EURASIP J. Audio Speech Music Process. **2012**,1–18 (2012)
55. Sawada, H., Araki, S., Mukai, R., Makino, S.: Blind extraction of dominant target sources using ICA and time-frequency masking. IEEE Trans. Audio Speech Lang. Process. **16**(6), 2165–2173 (2006)
56. https://sites.google.com/site/linwangsig/extraction

Chapter 12
On the Ideal Ratio Mask as the Goal of Computational Auditory Scene Analysis

Christopher Hummersone, Toby Stokes and Tim Brookes

Abstract The ideal binary mask (IBM) is widely considered to be the benchmark for time–frequency-based sound source separation techniques such as computational auditory scene analysis (CASA). However, it is well known that binary masking introduces objectionable distortion, especially musical noise. This can make binary masking unsuitable for sound source separation applications where the output is auditioned. It has been suggested that soft masking reduces musical noise and leads to a higher quality output. A previously defined soft mask, the ideal ratio mask (IRM), is found to have similar properties to the IBM, may correspond more closely to auditory processes, and offers additional computational advantages. Consequently, the IRM is proposed as the goal of CASA. To further support this position, a number of studies are reviewed that show soft masks to provide superior performance to the IBM in applications such as automatic speech recognition and speech intelligibility. A brief empirical study provides additional evidence demonstrating the objective and perceptual superiority of the IRM over the IBM.

12.1 Introduction

A natural environment usually consists of a number of sound sources. Some may convey information that is important to the listener (a person speaking for example), whilst others may be less important (a distant vehicle for example). If the important

C. Hummersone (✉) · T. Stokes · T. Brookes
University of Surrey, Guildford, UK
e-mail: c.hummersone@surrey.ac.uk

T. Stokes
e-mail: t.stokes@surrey.ac.uk

T. Brookes
e-mail: t.brookes@surrey.ac.uk

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_12,
© Springer-Verlag Berlin Heidelberg 2014

information is considered to be a target signal and all other sound is considered to be noise/interference, then this situation may be modelled as:

$$z(n) = x(n) * h(n) + d_{\text{st}}(n) + d_{\text{nst}}(n) \tag{12.1}$$

where z is the mixture at sample index n, x is the target signal, h is the acoustic/channel impulse response, and d_{st} and d_{nst} are stationary and non-stationary noise/interference, respectively [33]. In certain situations, these unhelpful *interfering* sound source(s) may prevent the listener from receiving all of the information from the important *target* sound source. A machine listener, such as an automatic speech recognition (ASR) system, may also be impeded by the presence of interfering sounds. But in a natural environment, acoustic interference is often inescapable. Hence reducing the level of acoustic interference may be useful in a number of applications, including: ASR, speaker identification, human–computer interaction, audio information retrieval and hearing prostheses. This broad range of applications has meant that blind source separation (BSS) is an important area of research in signal processing and related fields.

Through the course of research, four main approaches to BSS have emerged: independent component analysis (ICA), spatial filtering, non-negative matrix factorisation (NMF) and computational auditory scene analysis (CASA). ICA seeks to separate components based on statistical independence. The technique aims to find the inverse mixing matrix that provides the most independent separated source signals [7, 23, 34]. Spatial filtering uses microphone array signal processing to enhance sound arriving from a particular direction. NMF [24] aims to factorise a time–frequency (T–F) representation of a mixture in to two matrices: bases and coding. The bases matrix is formed from a set of unique spectral structures; each basis does not represent a source in the mixture but rather each sound that is part of the mixture. For example, the signal from a piano would be divided into each individually occurring note or speech into individual formants. The coding matrix determines the temporal activation of the bases. CASA aims to mimic human auditory scene analysis (ASA) [9], which is the process by which a human makes sense of an auditory scene, a key part of which is the separation of mixtures of sounds. Humans demonstrate a remarkable ability to extract a target sound from a mixture, providing important motivation for research into CASA. It is for this reason that this chapter chooses to focus on CASA.

A typical CASA system broadly consists of two stages [50]. First an analysis of the audio in the T–F domain is used to decide whether a particular T–F unit should be designated as target or interference. Second, this information is used to mask the T–F representation in order to reduce or eliminate the interference. In his seminal treatise on the topic, Wang [48] proposed the ideal binary mask (IBM) as the goal of CASA. The IBM is set to one when the target energy exceeds the interference energy and zero otherwise. Binary masking has also been coupled with other aforementioned BSS techniques including ICA (e.g. [37]) and NMF (e.g. [18]). The proposal of the IBM as the goal of CASA has been supported by a number of studies that have shown the IBM to be advantageous for machine and human listening tasks, including speech

intelligibility (e.g. [11, 27, 40]), and ASR (e.g. [13, 19, 42]). Furthermore, it was shown that under certain constraints the IBM is the optimal binary mask in terms of signal-to-noise ratio (SNR) [28].

However, it is well known that the binary mask separation method introduces audible distortions, especially so-called *musical noise*. The distortion is caused by repeated narrow-frequency-band switching. As will be shown in this chapter, the perceived audio quality of binary-masked audio is poor. This has the potential to limit the applications of binary mask-based techniques such as CASA to domains where the output is not auditioned. This constitutes a significant limitation.

In order to address this limitation and the shortcomings of the IBM, this chapter will propose the ideal ratio mask (IRM) [42] as the goal of CASA. Under certain circumstances and/or for particular applications, the value of the IBM may be great, and this chapter is not intended to refute that. Instead, the chapter will argue that the IRM may be preferable to the IBM as the goal of CASA for a number of theoretical and practical reasons, and across a majority of applications. The chapter will start with a review of the musical noise problem is Sect. 12.2. A number of advantageous features of the IRM will then be reviewed in Sect. 12.3, leading to the proposal of the IRM as the goal of CASA. The IBM and IRM will then be compared in Sect. 12.4 using existing studies and a brief empirical study utilising both purely objective and perceptually informed objective BSS metrics.

12.2 The Problem with the Ideal Binary Mask

Although binary masking has proved to be an effective BSS method, the prevalence of artefacts such as musical noise appears to have a deleterious effect on the audio quality of the separated output. Whilst this might not be problematic for applications where the output is not auditioned (such as ASR or databasing tasks) for other tasks (such as speech enhancement or auditory scene reconstruction) the poor audio quality is likely to prevent adoption of binary mask-based techniques such as CASA. Few studies have compared the audio quality achieved by binary masking to that achieved by other BSS methods, and hence it is difficult to draw meaningful conclusions on the degree to which binary masking is deleterious for audio quality. Some data can be found in Table 12.1 [33]. The paper compares four BSS algorithms against IBM-based separation. The first model, M1, combines a noise tracker based on voice activity detection (VAD) with a minimum mean square error (MMSE) spectral amplitude estimator (STSA) [15]. The second model, M2, combines a VAD-based noise tracker with a log-spectral amplitude estimator (LSA) [16]. M1 and M2 use the 'decision-directed' method [15] to estimate the a priori SNR by weighting the estimated spectral amplitude and noise variance of the previous frame, and the posteriori SNR in the current frame. The third model, M3, uses a magnitude-DFT MMSE estimator under the assumption that the required coefficients have a generalised Gamma distribution (GGD). The fourth model, proposed in the paper, combines a noise estimator designed for highly non-stationary noise (NSNE) [39]

Table 12.1 Performance data [33] comparing the unprocessed noisy mixture with the output of a number of BSS algorithms (M1–M3), a proposed BSS algorithm [33], and the IBM

Noise estimation + method	Metric	Target-to-interference ratio (dB)						Average
		−6	−3	0	3	6	9	
– + Noisy speech [12]	OPS	9.4	8.6	25.9	8.6	9.2	18.2	13.3
	SNR	−8.2	−3.0	−0.6	2.6	−2.8	6.3	−1.0
M1: VAD + LSA [15]	OPS	19.7	15.7	30.6	28.9	34.4	40.9	28.3
	SNR	−6.7	−0.8	1.4	4.3	−1.7	5.4	0.3
M2: VAD + STSA [16]	OPS	20.4	16.2	31.9	29.4	34.4	37.5	28.3
	SNR	−6.6	−0.8	1.4	4.2	−1.7	5.2	0.3
M3: MMSE [20] + GGD [17]	OPS	21.8	19.3	26.3	27.9	31.7	28.1	25.8
	SNR	0.6	0.9	1.2	1.1	0.9	1.2	1.0
NSNE + ML [33]	OPS	27.0	21.8	45.4	34.0	33.4	50.3	35.3
	SNR	2.1	3.1	4.5	4.8	4.7	6.2	4.2
Ideal + IBM	OPS	16.6	12.4	13.3	14.4	14.0	13.9	14.1
	SNR	4.4	5.5	5.1	5.1	3.6	4.3	4.7

The comparisons are in terms of OPS, and SNR (in dB)

Table 12.2 Data from SiSEC2011 [4], for tasks T2 or T3 and instantaneously mixed dataset D1, showing the average OPS and SDR (in dB) of a number of BSS techniques, including the IBM

System	Metric	2 mic 3 speech	2 mic 3 music	2 mic 4 speech	3 mic 4 speech
S1 [35]	OPS	43.9	52.3	42.4	–
	SDR	13.4	16.6	8.9	–
S2 [31]	OPS	43.2	40.0	29.8	39.7
	SDR	7.9	6.9	3.0	11.7
IBM (STFT)	OPS	38.9	33.3	27.1	–
	SDR	10.8	10.4	9.1	–
IBM (GTFB)	OPS	24.0	30.4	22.0	–
	SDR	8.5	9.0	7.5	–

with a maximum likelihood (ML) speech estimator to produce a DFT-based soft mask. The comparison utilises the perceptual evaluation of audio source separation (PEASS) toolbox [14]. The results indicate that in terms of 'overall perceptual score' (OPS) (a metric intended to indicate the 'global quality' of the separated output), the IBM performs poorly compared to the other methods; the difference is as much as 20 points on the 100-point scale.

A similar trend can be observed in Table 12.2 [4], which compares: S1, a generalised expectation–maximisation framework for handling prior information [35]; S2, a k-subspace-based tensor factorization method [31]; the IBM obtained via the short-time Fourier transform (STFT); and the IBM obtained via the gammatone filterbank (GTFB). The table shows some differences of a similar magnitude to Table 12.1, depending on the T–F decomposition and mixture. Note that the SNR and signal-to-distortion ratio (SDR) data presented in the tables show that poor OPS performance is not solely attributable to poor separation performance.

Table 12.3 A comparison using the PEASS metrics of the IBM with three mask postprocessing and/or alternative mask estimation algorithms [43]

Method	APS	IPS	TPS	OPS
IBM	12	76	51	18
DBM	29	62	67	36
NBM	48	66	67	49
CBM	53	61	66	49

Other studies have shown that the IBM is not optimal in terms of audio quality. One study [10] found that although the IBM improves speech intelligibility in noisy conditions and causes the noise to be less annoying, the separated speech is unnatural and consequently listeners do not find it preferable to the unseparated mixture. It is noted that by softening the mask the distortions are reduced and the noise is increased (lowering intelligibility), but that the result is preferred by listeners to the IBM and unprocessed outputs. It should be noted that this study was conducted on normal hearing listeners. There is some evidence to suggest that hearing-impaired listeners are less sensitive to musical noise [1]. Therefore, it may not be advantageous to lower the SNR by softening the mask for applications targeting hearing-impaired listeners. A number of other studies have also shown that musical noise arising from binary masking can be reduced by soft masking, but that this comes at a cost in terms of SNR (e.g. [2, 5, 25]).

Such is the disturbance caused by musical noise that some studies have attempted to improve the perceptual quality of binary-masked audio (e.g. [2, 3, 29]). In one study, summarised in Table 12.3, mask postprocessing and alternative mask estimation algorithms, were compared to the IBM in an attempt to improve the OPS of the separated output [43]. Specifically, the study compared: the IBM; a noisy binary mask (NBM) that had triangular probability density function (TPDF) noise added to the binary values; a dithered binary mask in which the SNR had TPDF dither added prior to mask calculation; and a cepstrally smoothed binary mask [29] that, after optimization, effectively added 0.1 to all zero-valued mask units. Similarly to previous studies noted above, the study concluded that the OPS could be improved, but at the cost of some interference suppression. Although the DBM demonstrated some quality improvement, methods that demonstrated the greatest improvement in OPS allowed the mask values to deviate from zero and one.

These results suggest that a well-defined soft mask may achieve a better audio quality than a binary mask. This has also been suggested by other authors [29, 49]. However, it seems that the choice of soft mask should be made carefully such that it does not introduce an SNR penalty. Several authors (e.g. [5, 6]) have suggested the use of sigmoid functions in order to generate soft masks. One such approach [5] showed that a soft mask defined in this way offers a slight signal-to-interference ratio (SIR) advantage over a binary mask. However, it remains unclear how such sigmoidal masks perform using more common metrics such as SNR. One mask that has received attention in recent years is the IRM [42]. As will be shown in the next

section, the IRM has a number of properties that make it a good alternative to the IBM as the goal of CASA.

12.3 The Ideal Ratio Mask as the Goal of Computational Auditory Scene Analysis

CASA aims to model the human process of ASA [48]. Bregman [9] states that the goal of ASA is 'the recovery of separate descriptions of each separate thing in the environment'. However, this goal is too vague to be transferred directly to CASA. In his important treatise on the goal of CASA, Wang [48] suggests three options, before suggesting that the IBM should be the goal. The first option is to separate out all sound sources in a given mixture. However, this goal is far beyond the capabilities of the human listener who may only be able to separate a handful of concurrent sound sources. The second option is to enhance ASR. Whilst appealing, since this is one of the primary applications of CASA, it is not the only application. Thus in order to retain maximum usefulness across applications, the goal should not be tied to a specific application. The final option is to enhance human listening. However, not all applications involve human listeners (ASR, for example), and thus this goal would also only apply to a subset of applications. Measuring the responses of human listeners may also introduce prohibitive requirements of time, resources and/or expertise that might hinder progress in the field.

Consequently, Wang [48] suggests that the IBM should be the goal of CASA, for a number of reasons discussed in this section. In contrast, this section proposes the IRM as the goal of CASA. The thesis is based on three strands of argumentation: that the IRM matches or exceeds all of the desirable properties laid out by Wang (Sect. 12.3.1); that the IRM provides a closer match to psychophysical and perceptual mechanisms than the IBM (Sect. 12.3.2); and that the IRM provides a number of computational advantages (Sect. 12.3.3).

12.3.1 Properties of the IBM and IRM

First, for the sake of clarity, the IBM $\mathbf{m}_B$ and IRM $\mathbf{m}_R$ are defined in the following way:

$$\mathbf{m}_B(c,m) = \begin{cases} 1 & \dfrac{\mathbf{u}_t(c,m)}{\mathbf{u}_i(c,m)} > 1 \\ 0 & \text{otherwise} \end{cases}, \tag{12.2}$$

$$\mathbf{m}_R(c,m) = \frac{\mathbf{u}_t(c,m)}{\mathbf{u}_t(c,m) + \mathbf{u}_i(c,m)}, \tag{12.3}$$

where $\mathbf{u}_{\{t,i\}}$ is the power of the target and interfering source(s) (t and i respectively) in time frame m and frequency bin/channel c.

In his paper proposing the IBM as the goal of CASA, Wang [48] specified four desirable properties of the IBM. These are:

1. 'flexibility'—for a given mixture, the mask will differ according to which sources are designated target and interference;
2. 'well-definedness'—the ideal mask retains its definition independently of how many sources are present;
3. ceiling performance—the IBM is the optimal binary mask; and
4. psychoacoustical correspondence—the IBM broadly agrees with auditory masking and ASA [9] theories.

Given the similarity of the definitions of the IBM and IRM shown in (12.2) and (12.3), it can be seen that the IRM shares all of these properties. First, the IRM is identically flexible: any source can be designated as the target, and the sum of remaining sources is typically designated as the interference. Second, the IRM is also well-defined, since the interference component may constitute any number of sources. Third, the IRM is the optimal ratio mask and is closely related to the ideal Wiener filter (IWF), which is the optimal linear filter with respect to MMSE [28, 51]. Last, the IRM broadly agrees with psychoacoustic principles. This last point might seem surprising: it might appear counterintuitive that the IBM and IRM can both honour psychoacoustic principles. However, they can and, although they are both approximations, the next section will show how the IRM is perhaps a better approximation of auditory masking and ASA principles than the IBM.

12.3.2 Psychophysical and Perceptual Bases of the IRM

It is argued by Wang [48] that the IBM corresponds closely to auditory masking and ASA theories. However, this section argues that the IRM provides a closer match.

The concept of *binary* masking assumes that auditory masking is dichotomous: that a sound is either masked or it is not. To put it another way, it suggests that a sensory threshold exists. However, it has been known since the 1950s that this is an inadequate description when discussing sensory perception in any modality (see [44] for a review). Like any sensory threshold, auditory masking is only dichotomous in the sense that the experimenter asks the subject a yes/no question, e.g. 'is the sound audible?'. It seems reasonable that under identical circumstances, a listener should always give the same answer. However, this is often not the case. The probability of a consistent answer depends on the relative level of the competing stimulus: the greater the difference, the greater the probability of a consistent answer. Furthermore, individual listeners may give different answers. In auditory masking experiments this *probability of detection* is often plotted as a function of signal magnitude in order to produce a *psychometric function*.

Auditory masking, like many aspects of sensory perception, has therefore been described using signal detection theory [38, 45, 46]. When applied to the auditory domain, the theory defines a decision variable, which often corresponds to physiological or psychological responses to a stimulus, such as the auditory nerve firing rate or sensory impression. Signal detection theory dictates that the *average* value of the decision variable is monotonically related to stimulus magnitude. However, signal detection theory also dictates that the value of the decision variable may fluctuate. There are two causes of fluctuation: external variations such as background noise level, or internal variations such as plausible differences in neural responses or other psychological factors [32].

It can be seen therefore that a ratio mask, where values vary in the range [0, 1], provides a closer match to signal detection theory than a binary mask. A value of 0 or 1 indicates certainty in the absence or presence of a signal, respectively. An intermediate value may be obtained when the signals are of similar magnitude. Although this may or may not agree with an experimentally derived psychometric function, it at least provides a conceptual indication that masking is uncertain.

It should be noted, however, that the concept of a binary mask does agree to some extent with the ASA [9] theory. The theory draws on Gestalt principles of 'exclusive allocation' (sometimes referred to as 'disjoint allocation' or 'belongingness') in visual perception, whereby a sensory element can not be used in the descriptions of more than one object at a time. However, whilst this principle generally holds true, there are a number of examples of violations of this principle in the auditory domain (see [8] for a review). Bregman [8] describes this as 'duplex perception'. Furthermore, the grouping of sensory elements may depend on perspective or attention. Unlike the ratio mask, a binary mask cannot account for these observations because each T–F unit is always assigned to the source with the most energy. In his paper, Wang [48] compares auditory objects to visual objects. Using this analogy, foreground visual objects are assigned a mask value of one, whereas occluded objects are assigned a value of zero. However, Bregman [8] argues that

> … sound is transparent. A sound in the foreground does not 'occlude' a sound in the background in the same way as a visual object occludes our view of objects behind it.

In the visual domain, objects are occluded because light emitted or reflected from the object does not impinge on the retina. In the auditory domain, even if a sound source is visually occluded, its acoustic energy still usually impinges on the ear drums via numerous acoustic pathways. Thereafter sounds are occluded by physiological, psychophysical, or psychological mechanisms, rather than an absence of stimulating input to the auditory system. Furthermore, the human head seldom occludes sound; sound arriving at both ears usually contains information about all sound sources.

Given that auditory objects are transparent, it seems disadvantageous to assign portions of the sensory input to only one object when information about both objects is available. A visual analogy is given in Fig. 12.1. The two images on the left show two objects that are overlapped such that there is now a small common area. Using a disjoint allocation principle—demonstrated in the middle of the figure—the common area must be assigned to one object. The scenario is analogous to binary masking.

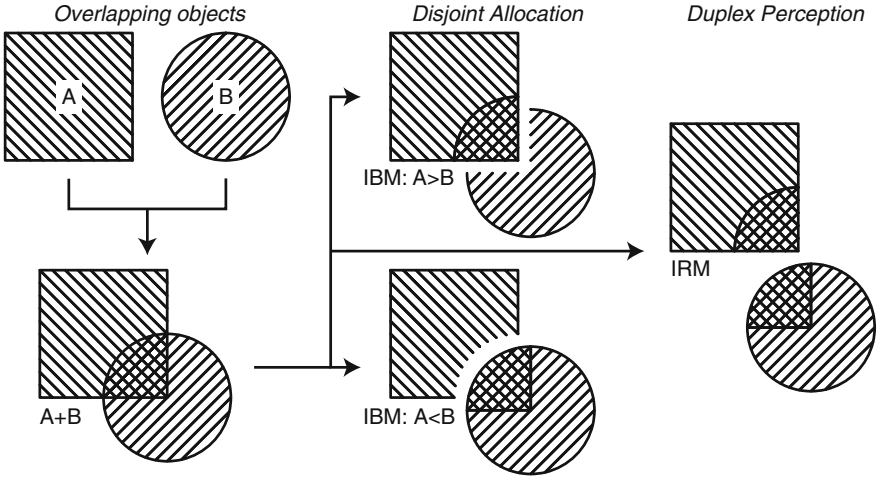


Fig. 12.1 Visual analogies of disjoint allocation and duplex perception when objects overlap (*left*): the disjoint allocation case (*middle*) is analogous to the binary mask; the duplex perception case (*right*) is analogous to the ratio mask

The occluding object is corrupt, whereas the occluded object is incomplete. Using a duplex perception principle—demonstrated in the right of the figure—the common area may be assigned to both objects. The scenario is analogous to ratio masking. The resulting objects are now complete, irrespective of the chosen source, although each is corrupted to some extent by the other. The ratio value indicates the extent of the corruption and hence how meaningful the area is likely to be to the current source. Note that this analogy applies not only to overlapping auditory objects, but also to the BSS problem since algorithms often try to estimate some parameter(s) of both the occluding and occluded signal in order to decide how to assign T–F units ((12.2) and (12.3) assume that some knowledge of both sources is available).

12.3.3 Computational Bases of the IRM

The IRM has a number of theoretical computational advantages, which are discussed in this section.

In their important paper on the relative merits of the IBM and IRM, Li and Wang [28] note that:

> … the IRM achieves higher SNR gains compared to the IBM. However, despite the fact that the IBM is binary and the IRM is not, the SNR gain of the IBM is surprisingly close to that of the IRM. This shows that the IBM is a very reasonable performance metric for sound separation. Indeed, there are reasons to prefer the IBM over the IRM as the computational goal of a separation system. The estimation of the IBM is considerably simpler than that of the IRM: the former requires only binary decisions, whilst the latter requires estimating the

energy ratio of the two signals. Binary estimation is facilitated by the existence of numerous classification and clustering methods.

The SNR achieved by the IRM is shown in the paper to be, on average, 0.1–0.8 dB higher than that achieved by the IBM, depending on the T–F decomposition and constituent signals. Of course, the importance of this gain depends on the overall SNR of the BSS algorithm.

Li and Wang argue above that binary masking facilitates classification and clustering methodologies. Whilst the IRM may preclude these possibilities, it does facilitate alternative probabilistic frameworks; the ratio value may be considered to indicate the probability that a given T–F region is reliable [6] or belongs to a particular source. Additionally, the binary mask can be considered as a special case and subset of the ratio mask. Indeed the equivalent binary mask may always be derived from the ratio mask by rounding or quantising ratio values; the IRM cannot be derived from the IBM. This is important from an application point of view, and suggests that algorithms should always try to estimate the IRM; the algorithm may subsequently choose (or be told) to quantise the mask if the application, for whatever reason, deems it appropriate.

Lastly, Li and Wang suggest that estimating ratio values may be more difficult than making binary decisions. Whilst this may or may not be the case, it should first be considered that the original definitions of the IBM (12.2) and IRM (12.3) contain identical quantities: estimates of the target and total interfering signal energy. Hence, in principle the task does not differ in its complexity, even if practical applications do not intend to estimate these values directly. Furthermore, simplicity is a relativistic concept; undoubtedly as knowledge in this field advances estimates of the IRM will become more accurate. The next section will consider the concept of 'difficulty' in more detail.

12.4 Comparisons of the Ideal Binary and Ratio Masks

Thus far this chapter has outlined theoretical and practical reasons to prefer the IRM over the IBM. The IRM has been shown to offer a small SNR gain over the IBM [28]. It has also been suggested that soft masks may provide superior audio quality to binary masks. A number of other studies have provided reasons to prefer soft masks to binary masks.

In the original paper proposing the IRM [42], summarised in Table 12.4, it was found that, for a small vocabulary digit recognition task, the IBM coupled with a missing-data ASR system marginally outperformed the IRM coupled with a conventional ASR system, by an accuracy of the order of 1 % across all SNRs. However, with a large vocabulary command and control task there was demonstrable improvement from the IRM of as much as 30 % accuracy, with greatest improvement found in higher noise conditions. Similar findings, shown in Table 12.5, were made [6] by employing a soft 'fuzzy' (though non-ideal) mask rather than a binary (non-ideal)

Table 12.4 A comparison of the IBM coupled with a missing-data ASR system and the IRM coupled with a conventional ASR system for two different vocabulary sizes [42]

	Accuracy (%)			
	Small vocabulary		Large vocabulary	
SNR (dB)	IRM	IBM	IRM	IBM
−5	94.8	94.9	96.4	66.5
0	95.7	96.0	97.0	71.2
5	96.4	97.2	97.6	76.5
10	97.7	98.1	97.7	80.1
∞	98.6	97.2	97.7	82.7

Table 12.5 A comparison of fuzzy and binary masks coupled with missing-data ASR performing a digit recognition task under different noise conditions [6]

	Digit recognition accuracy (%)			
	Factory noise		Lynx helicopter noise	
SNR (dB)	Fuzzy	Binary	Fuzzy	Binary
0	60	46	86	77
5	81	73	95	92
10	90	87	98	96
15	95	95	99	98
20	97	97	99	99
200	99	99	99	99

mask. In this case, the fuzzy mask was produced by compressing the difference between the estimated local noise and signal x using a sigmoidal function of the form

$$f(x) = \frac{1}{1 + \mathrm{e}^{-\alpha(x-\beta)}} \tag{12.4}$$

where $\alpha \in [0, \infty)$ and $\beta \in [0, 1]$ are parameters controlling the sigmoid slope and centre, respectively. The binary mask was derived in a similar way; the final values were rounded to 0 or 1. As shown in the table, the fuzzy mask achieves an ASR accuracy gain of a more modest 14 %, again with greatest improvement found in higher noise conditions.

Gains have also been observed for human audition. A recent study [30] found that a soft mask based on the IWF significantly outperformed the IBM (with either a fixed or local threshold, IBM-F and IBM-L, respectively) in terms of speech intelligibility and quality. The speech intelligibility data are summarised in Table 12.6. In this study, the formulation of the IWF was identical to the IRM since the power-spectral density was not smoothed. The table shows that the intelligibility gain can be as much as 100 % in high noise conditions. A similar finding, summarised in Table 12.7, was made in another recent study [22], although the data were yielded from non-ideal masks. In the paper, the authors created a number of mask estimation algorithms based on estimating the MMSE of the spectral magnitude. Specifically, a continuous

Table 12.6 Speech intelligibility of speech separated using different T–F masks and under different interference conditions [30]

	Correct (%)					
	Babble interference			Speech interference		
SNR (dB)	IWF	IBM-L	IBM-F	IWF	IBM-L	IBM-F
−35	96	50	–	87	4	–
−30	100	29	–	92	10	–
−25	98	52	–	98	23	–
−20	96	62	–	96	71	13
−15	100	56	–	100	87	25
−10	–	62	2	–	85	69
−5	–	75	38	–	92	87
0	–	90	75	–	100	100
5	–	90	98	–	–	–
10	–	87	96	–	–	–

Table 12.7 Speech intelligibility of masked speech with speech-shaped-noise interference [22] for a variety of masking algorithms

	Intelligibility (%)			
SNR (dB)	CG-MMSE	BG-HU	Noisy	BG2-MMSE
−8	50.5	21.2	46.2	38.2
−6	69.8	36.3	56.0	55.7
−4	77.2	48.6	68.9	68.9
−2	87.4	64.3	82.8	74.5
0	89.5	79.1	91.1	85.8

gain MMSE mask (CG-MMSE) was compared with two binary gain (BG) estimators (BG-HU [21] and BG2-MMSE proposed in the work). The unprocessed noisy speech was also tested. The table shows that the CG-MMSE system achieved a more modest gain of up to, approximately, 30 % in higher noise conditions. These differences in performance were attributed, in part, to the better preservation of the target envelope by the soft mask. It has been shown that the signal envelope is important for speech intelligibility [41].

These studies have offered compelling evidence that the IRM may provide a superior output to the IBM for a number of applications. It was also shown [30] that the Wiener filtering approach is less sensitive to errors in terms of speech intelligibility. However, it remains unclear how the IBM and IRM trade off in terms of the numerous sources of error and the magnitude of any separation performance gains, and whether a ratio mask retains its error robustness in terms of sound source separation metrics (i.e. whether the supposed difficulty of estimating the IRM [28] incurs a penalty). The rest of this section describes a new study that attempted to address these points. The study compared the separated audio output produced by binary and ratio masks using a number of objective and perceptually informed objective metrics.

As in the previously reported study, [30] it was assumed that task difficulty could be modelled by introducing errors to the target and interfering signal energy. The errors inevitably led to an erroneous mask. Consequently, two additive error components $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_i$ were introduced, resulting in 'estimated' binary and ratio masks ($\hat{\mathbf{m}}_B$ and $\hat{\mathbf{m}}_R$, respectively).

The error components were calculated independently for the target and interfering signal using a previously defined method [30]. Specifically, the error perturbed the spectral coefficient(s) in each T–F unit prior to the calculation of spectral power used to formulate the masks. However, unlike the previous study [30], both the STFT and the GTFB were utilised. Similarly to Li and Wang's study [28], the GTFB was fourth-order and had 64 channels with centre frequencies equally spaced on the ERB-rate scale between 50 and 8,000 Hz, although the time and phase responses were aligned using the method described by Patterson et al. [36]. The STFT was 512-point but the frames were not overlapping (since Li and Wang [28] point out that the IBM may only be optimal when frames do not overlap).

The spectral power for the target and interferer signals, $\hat{\mathbf{u}}_t$ and $\hat{\mathbf{u}}_i$, respectively, were calculated in the following way. For the GTFB

$$\hat{\mathbf{u}}_{\{t,i\}}(c,m) = \sum_{n=mM_{\mathrm{GT}}}^{(m+1)M_{\mathrm{GT}}-1} \left[\mathbf{X}_{\{t,i\}}(c,n) + \theta\boldsymbol{\varepsilon}_{\{t,i\}}(c,m)\right]^2, \tag{12.5}$$

where $\mathbf{X}$ is the output of the GTFB for the target or interfering signals, M_{GT} is the frame length (10 ms in samples), and n is the sample index. For the STFT

$$\hat{\mathbf{u}}_{\{t,i\}}(c,m) = \left| \sum_{n=M_{\mathrm{FFT}}m}^{(m+1)M_{\mathrm{FFT}}-1} x_{\{t,i\}}(n)\mathrm{e}^{-\mathrm{j}2\pi\frac{c}{M_{\mathrm{FFT}}}n} + \theta\boldsymbol{\varepsilon}_{\{t,i\}}(c,m) \right|^2, \tag{12.6}$$

where M_{FFT} is the FFT size (512). In both cases $\boldsymbol{\varepsilon}$ is an error component and $\theta \in [0, 1]$ is the error magnitude. For the GTFB, the error was normally distributed noise with zero mean. For the STFT, the error was complex noise where both real and imaginary parts were normally distributed with zero mean. In each case, the error was scaled in each frequency channel/bin to have equal power to the unperturbed target and interferer signals. The perturbed powers were used to calculate 'estimated' binary and ratio masks, $\hat{\mathbf{m}}_B$ and $\hat{\mathbf{m}}_R$, respectively, such that

$$\hat{\mathbf{m}}_B(c,m) = \begin{cases} 1 & \dfrac{\hat{\mathbf{u}}_t(c,m)}{\hat{\mathbf{u}}_i(c,m)} > 1 \\ 0 & \text{otherwise} \end{cases} \tag{12.7}$$

and

$$\hat{\mathbf{m}}_R(c,m) = \frac{\hat{\mathbf{u}}_t(c,m)}{\hat{\mathbf{u}}_t(c,m) + \hat{\mathbf{u}}_i(c,m)}. \tag{12.8}$$

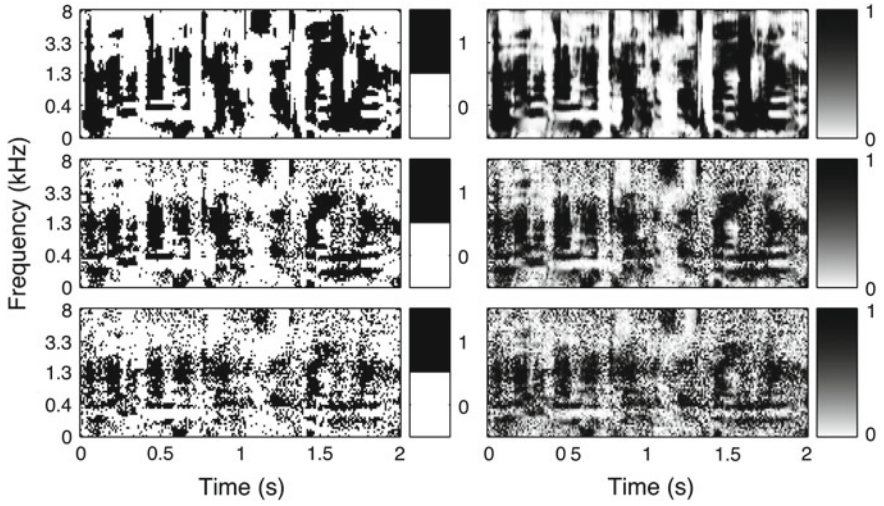


Fig. 12.2 Examples of the ideal and 'estimated' masks using a gammatone filterbank: binary masks (*left column*) and ratio masks (*right column*); ideal masks ($\theta = 0$) (*top row*), $\theta = 0.5$ (*middle row*), and $\theta = 1$ (*bottom row*)

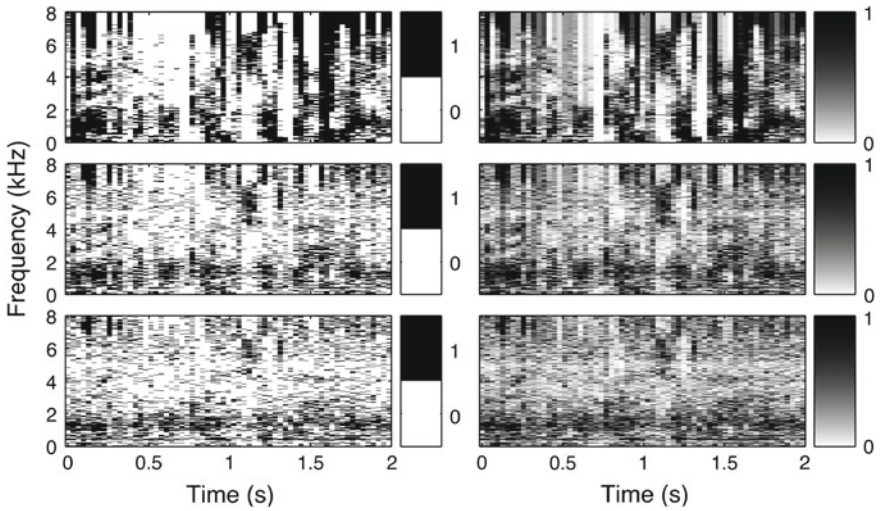


Fig. 12.3 Examples of the ideal and 'estimated' masks using a short-time Fourier transform: binary masks (*left column*) and ratio masks (*right column*); ideal masks ($\theta = 0$) (*top row*), $\theta = 0.5$ (*middle row*), and $\theta = 1$ (*bottom row*)

With $\theta = 0$, the signal power was unperturbed and the masks were ideal, with $\theta = 1$ the error had equal magnitude to the unperturbed signals. The estimated masks were applied to the unperturbed mixture in order to calculate the performance metrics. Examples of the masks are shown in Figs. 12.2 and 12.3.

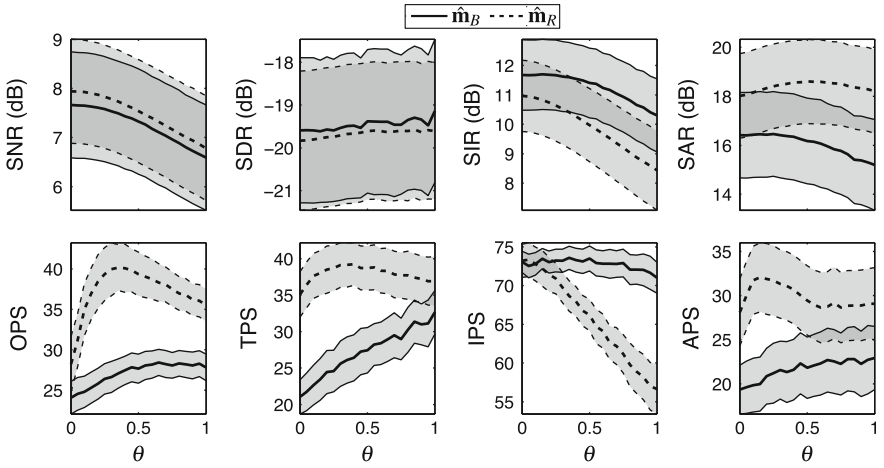


Fig. 12.4 Performance of binary and ratio masks under error conditions, using a gammatone filterbank, measured using a number of objective and perceptually informed objective metrics; *grey regions* show 95 % confidence intervals

The stimuli were taken from the SiSEC2013 corpus.[1] Instantaneous mixtures from the D1 and D2 datasets were used; each source was designated in turn as the target, with the sum of remaining sources designated as the interference (for the purposes of calculating $\hat{\mathbf{u}}_i$ and $\boldsymbol{\varepsilon}_i$). In total eight metrics were employed:

- SNR;
- three metrics from the BSS_eval toolbox [47]: SDR, SIR and signal-to-artefacts ratio (SAR); and
- four metrics from the PEASS toolbox [14]: OPS, target-related perceptual score (TPS), interference-related perceptual score (IPS) and artefact-related perceptual score (APS).

The SNR and OPS metrics are designated here as 'global' metrics, since they produce a single quantity derived from a number of sources of error, whereas the other metrics consider only a subset of error sources.

The results for the GTFB versus error magnitude θ are shown in Fig. 12.4. Differences, calculated as the scores for the ratio masks minus the scores for the binary masks, are given in Fig. 12.5. Note that in some cases the confidence intervals in Fig. 12.4 overlap but the corresponding difference scores in Fig. 12.5 are significant, i.e. the lower bound of the confidence interval is greater than zero. This is because calculating differences eliminates within-group variation, so that only between-group variation is exposed. The results show that the ratio masks retain a small but significant SNR advantage over the binary masks for all θ. The ratio masks are also superior in terms of OPS, TPS, SAR and APS. However, this it at the cost of interference suppression (SIR and IPS). The difference in SDR is negligible. The reduction

[1] http://sisec.wiki.irisa.fr/tiki-index.php

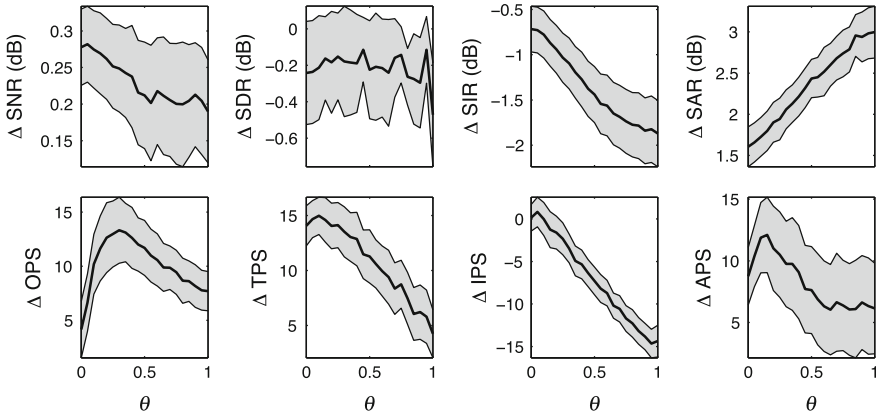


Fig. 12.5 Differences in mask performance for data shown in Fig. 12.4 (scores for the binary masks are subtracted from scores for the ratio masks); *grey regions* show 95 % confidence intervals

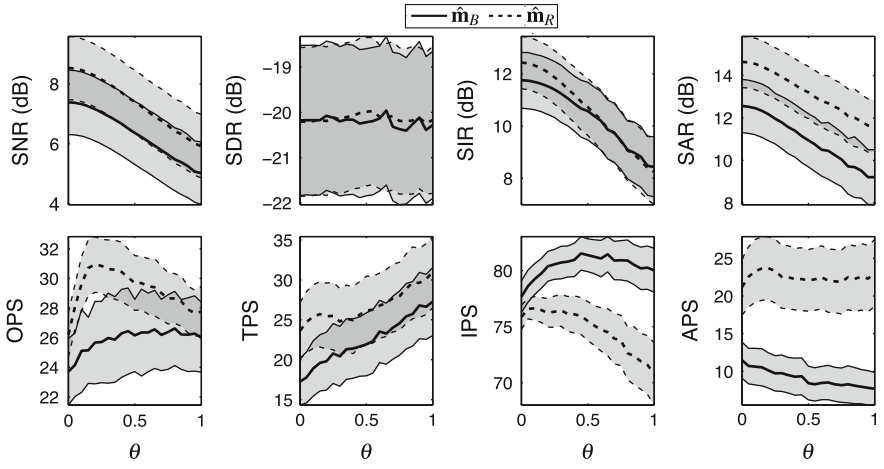


Fig. 12.6 Performance of binary and ratio masks under error conditions, using a short-time Fourier transform, measured using a number of objective and perceptually informed objective metrics; *grey regions* show 95 % confidence intervals

in artefacts is particularly prominent, and seems to have resulted in a substantial improvement in OPS for some values of θ. Informal listening suggested that the increase in OPS is attributable to the reduction in musical noise.

In terms of error resilience, the results show that no significant penalty is incurred when there are errors in estimating source power. The ratio masks retains their SNR, OPS, TPS, SAR and APS advantage for all θ.

Data for the STFT are given in Figs. 12.6 and 12.7. The trends shown in these data appear to mostly align with observations made of the GTFB. Note that the SNR gain is generally larger in this case, whereas the OPS gain is generally smaller.

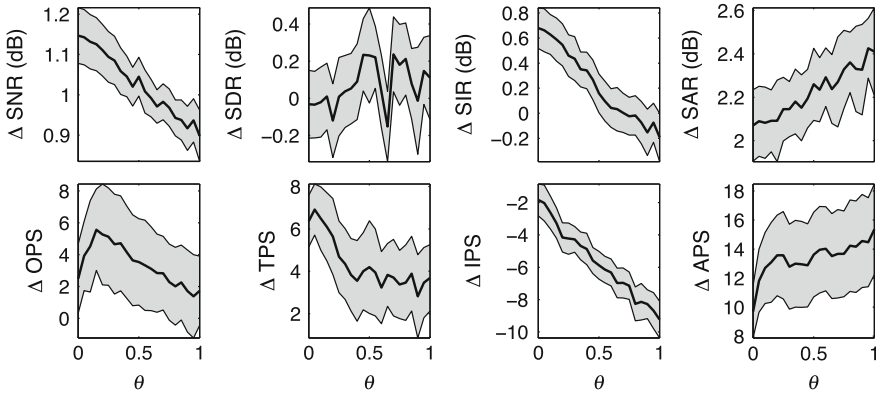


Fig. 12.7 Differences in mask performance for data shown in Fig. 12.6 (scores for the binary masks are subtracted from the scores for the ratio masks); *grey regions* show 95 % confidence intervals

Interestingly, the ratio masks outperform the binary masks in terms of SIR for most values of θ, although this advantage is not reflected in the IPS.

These results suggest that the ratio masks outperform the binary masks in terms of the 'global' metrics (SNR and OPS). As previously found, the IRM's SNR gain over the IBM is small, and this advantage is retained when the masks deviate from ideal. The ratio masks appear to significantly reduce musical noise, outperforming the binary masks in terms of artefact reduction (SAR and APS). They also appear to generally outperform the binary masks in terms of TPS. However, the cost here has been shown to be interference suppression. It remains unclear whether this will be important to any specific applications. Therefore the IRM seems likely to be a more appropriate goal for most applications. It should be noted that these differences may depend upon the T–F resolution, especially in terms of speech intelligibility [26, 30], such that smaller differences may be observed at higher resolutions.

It may be of interest to note that both experiments indicate a maximal OPS for $\theta > 0$, i.e. for a non-ideal mask. It seems that the introduction of random errors into an ideal mask can reduce the severity of objectionable artefacts (as can be seen in the APS plots) due to a lessening of the regularity and, for a ratio mask, the severity of transitions. This phenomenon has been explored in a previous study [43].

12.5 Conclusions

This chapter has argued that the IRM is a more appropriate goal for CASA than the IBM. A number of reasons, summarised in the following sentences, were provided. First, the IRM shares desirable properties with the IBM that make it an appropriate goal for CASA. Second, the IRM seems to correspond more closely to human psychophysical and perceptual mechanisms, such as auditory masking and ASA, than

the IBM. Third, the IRM has some computational advantages, such as facilitating probabilistic frameworks. Fourth, the equivalent binary mask may always be derived from the ratio mask; the reverse is not possible. Fifth, studies have shown the IRM to provide small gains over the IBM for BSS (in terms of SNR), but modest to large gains for ASR and speech intelligibility. Last, hints in the literature that the IRM may lead to a higher audio quality than the IBM by reducing musical noise were confirmed in a brief empirical study. The study showed that although the SNR gain offered by the IRM might be small, and might come at the cost of a slightly reduced interference suppression, a significant gain in OPS can be obtained, together with an improved TPS, APS and SAR. Furthermore, the IRM retains many of its advantages independently of any errors made in estimating source powers. It is acknowledged that the 'ideal' mask might sometimes be beaten in terms of OPS by a slightly less ideal mask. This might suggest that the goal of CASA should perhaps be an 'almost-ideal' ratio mask. However, current algorithms are likely to produce a small degree of error when used for practical BSS and so if the goal is the IRM then it is likely that an 'almost-ideal' ratio mask is what will be actually produced. Thus, adopting the IRM as the goal of CASA is likely to lead to algorithms delivering optimal OPS.

Acknowledgments The authors would like to thank Nicoleta Roman and colleagues for providing the data for Table 12.4, Nilesh Madhu for providing the data for Table 12.6, and Jesper Jensen for providing the data for Table 12.7.

References

1. Anzalone, M., Calandruccio, L., Doherty, K., Carney, L.: Determination of the potential benefit of time-frequency gain manipulation. Ear and hearing **27**(5), 480 (2006)
2. Araki, S., Makino, S., Sawada, H., Mukai, R.: Underdetermined blind separation of convolutive mixtures of speech with directivity pattern based mask and ICA. In: Puntonet, C., Prieto, A. (eds.) Independent Component Analysis and Blind Signal Separation. Lecture Notes in Computer Science, vol. 3195, pp. 898–905. Springer, Berlin (2004)
3. Araki, S., Makino, S., Sawada, H., Mukai, R.: Reducing musical noise by a fine-shift overlap-add method applied to source separation using a time-frequency mask. IEEE Int. Conf. Acoust. Speech Signal Proc. (ICASSP) **III**, 81–84 (2005)
4. Araki, S., Nesta, F., Vincent, E., Koldovsk, Z., Nolte, G., Ziehe, A., Benichoux, A.: The 2011 signal separation evaluation campaign (SiSEC2011): audio source separation. In: Theis, F., Cichocki, A., Yeredor, A., Zibulevsky, M. (eds.) Latent Variable Analysis and Signal Separation. Lecture Notes in Computer Science, vol. 7191, pp. 414–422. Springer, Berlin, Heidelberg (2012)
5. Araki, S., Sawada, H., Mukai, R., Makino, S.: Blind sparse source separation with spatially smoothed time-frequency masking. In: International Workshop on Acoustic, Echo and Noise Control. Paris (2006)
6. Barker, J., Josifovski, L., Cooke, M.P., Green, P.D.: Soft decisions in missing data techniques for robust automatic speech recognition. In: Proceedings of International Conference on Spoken Language Processing, pp. 373–376 (2000)
7. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural computation **7**(6), 1129–1159 (1995)

8. Bregman, A.: The meaning of duplex perception: sounds as transparent objects. In: Schouten, M.E.H. (ed.) The Psychophysics of Speech Perception, pp. 95–111. Martinus Nijhoff, Dordrecht (1987)
9. Bregman, A.S.: Auditory Scene Analysis. MIT Press, Cambridge (1990)
10. Brons, I., Houben, R., Dreschler, W.A.: Perceptual effects of noise reduction by time-frequency masking of noisy speech. J. Acoust. Soc. Am. **132**(4), 2690–2699 (2012)
11. Brungart, D.S., Chang, P.S., Simpson, B.D., Wang, D.: Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J. Acoust. Soc. Am. **120**(6), 4007–4018 (2006)
12. Christensen, H., Barker, J., Ma, N., Green, P.: The chime corpus: a resource and a challenge for computational hearing in multisource environments. In: Proceedings of Interspeech (2010)
13. Coy, A., Barker, J.: An automatic speech recognition system based on the scene analysis account of auditory perception. Speech Commun. **49**(5), 384–401 (2007)
14. Emiya, V., Vincent, E., Harlander, N., Hohmann, V.: Subjective and objective quality assessment of audio source separation. IEEE Trans. Audio Speech Lang. Proc. **19**(7), 2046–2057 (2011)
15. Ephraim, Y., Malah, D.: Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Proc. **32**(6), 1109–1121 (1984)
16. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Proc. **33**(2), 443–445 (1985)
17. Erkelens, J., Hendriks, R., Heusdens, R., Jensen, J.: Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors. IEEE Trans. Audio Speech Lang. Proc. **15**(6), 1741–1752 (2007)
18. Grais, E., Erdogan, H.: Single channel speech music separation using nonnegative matrix factorization and spectral masks. In: The 17th International Conference on Digital Signal Processing, pp. 1–6 (2011)
19. Hartmann, W., Fosler-Lussier, E.: Investigations into the incorporation of the ideal binary mask in ASR. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 4804–4807 (2011)
20. Hendriks, R., Heusdens, R., Jensen, J.: MMSE based noise PSD tracking with low complexity. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 4266–4269 (2010)
21. Hu, Y., Loizou, P.C.: Techniques for estimating the ideal binary mask. In: Proceedings 11th International Workshop on Acoustic Echo and Noise Control (2008)
22. Jensen, J., Hendriks, R.: Spectral magnitude minimum mean-square error estimation using binary and continuous gain functions. IEEE Trans. Audio Speech Lang. Proc. **20**(1), 92–102 (2012)
23. Jutten, C., Hérault, J.: Independent component analysis (inca) versus principal component analysis. In: Signal Processing IV: Theories and applications—Proceedings of EUSIPCO, pp. 643–646. North-Holland, Grenoble (1988)
24. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788–791 (1999)
25. Li, M., McAllister, H., Black, N., De Perez, T.: Perceptual time-frequency subtraction algorithm for noise reduction in hearing aids. IEEE Trans. Biomed. Eng. **48**(9), 979–988 (2001)
26. Li, N., Loizou, P.C.: Effect of spectral resolution on the intelligibility of ideal binary masked speech. J. Acoust. Soc. Am. **123**(4), 59–64 (2008)
27. Li, N., Loizou, P.C.: Factors influencing intelligibility of ideal binary-masked speech: implications for noise reduction. J. Acoust. Soc. Am. **123**(3), 1673–1682 (2008)
28. Li, Y., Wang, D.: On the optimality of ideal binary time-frequency masks. Speech Commun. **51**(3), 230–239 (2009)
29. Madhu, N., Breithaupt, C., Martin, R.: Temporal smoothing of spectral masks in the cepstral domain for speech separation. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 45–48 (2008)

30. Madhu, N., Spriet, A., Jansen, S., Koning, R., Wouters, J.: The potential for speech intelligibility improvement using the ideal binary mask and the ideal Wiener filter in single channel noise reduction systems: Application to auditory prostheses. IEEE Trans. Audio Speech Lang. Proc. **21**(1), 63–72 (2013)
31. Makkiabadi, B., Sanei, S., Marshall, D.: A k-subspace based tensor factorization approach for under-determined blind identification. In: Forty Fourth Asilomar Conference on Signals, Systems and Computers, pp. 18–22 (2010)
32. Moore, B.C.J.: An Introduction to the Psychology of Hearing, 5th edn. Academic Press, London (2004)
33. Mowlaee, P., Saeidi, R., Martin, R.: Model-driven speech enhancement for multisource reverberant environment (signal separation evaluation campaign (SiSEC) 2011). In: Theis, F., Cichocki, A., Yeredor, A., Zibulevsky, M. (eds.) Latent Variable Analysis and Signal Separation. Lecture Notes in Computer Science, vol. 7191, pp. 454–461. Springer, Berlin, Heidelberg (2012)
34. Naik, G.R., Kumar, D.K.: An overview of independent component analysis and its applications. Informatica **35**, 63–81 (2011)
35. Ozerov, A., Vincent, E., Bimbot, F.: A general flexible framework for the handling of prior information in audio source separation. IEEE Trans. Audio Speech Lang. Proc. **20**(4), 1118–1133 (2012)
36. Patterson, R., Nimmo-Smith, I., Holdsworth, J., Rice, P.: An efficient auditory filterbank based on the gammatone function. Technical report, MRC Applied Psychology Unit, Cambridge (1987)
37. Pedersen, M., Wang, D., Larsen, J., Kjems, U.: Overcomplete blind source separation by combining ICA and binary time-frequency masking. In: IEEE Workshop Machine Learning Signal Processing, pp. 15–20 (2005)
38. Peterson, W., Birdsall, T.G., Fox, W.C.: The theory of signal detectability. In: Proceedings of the IRE Professional Group on Information Theory 4, pp. 171–212 (1954)
39. Rangachari, S., Loizou, P.C.: A noise-estimation algorithm for highly non-stationary environments. Speech Commun. **48**(2), 220–231 (2006)
40. Roman, N., Wang, D.: Pitch-based monaural segregation of reverberant speech. J. Acoust. Soc. Am. **120**(1), 458–469 (2006)
41. Shannon, R., Zeng, F., Kamath, V., Wygonski, J., Ekelid, M.: Speech recognition with primarily temporal cues. Science **270**, 303–303 (1995)
42. Srinivasan, S., Roman, N., Wang, D.: Binary and ratio time-frequency masks for robust speech recognition. Speech Commun. **48**(11), 1486–1501 (2006)
43. Stokes, T., Hummersone, C., Brookes, T.: Reducing binary masking artefacts in blind audio source separation. In: Proceedings of 134th Engineering Society Convention Rome (2013)
44. Swets, J.A.: Is there a sensory threshold? Science **134**(3473), 168–177 (1961)
45. Swets, J.A.: Signal Detection and Recognition by Human Observers. Wiley, New York (1964)
46. Tanner Jr, W.P., Swets, J.A.: A decision-making theory of visual detection. Psychol. Rev. **61**(6), 401–409 (1954)
47. Vincent, E., Gribonval, R., Févotte, C.: Performance measurement in blind audio source separation. IEEE Trans. Audio Speech Lang. Proc. **14**(4), 1462–1469 (2006)
48. Wang, D.: On ideal binary mask as the computational goal of auditory scene analysis. In: Divenyi, P. (ed.) Speech Separation by Humans and Machines, pp. 181–197. Kluwer Academic, Norwell (2005)
49. Wang, D.: Time-frequency masking for speech separation and its potential for hearing aid design. Trends Amplif. **12**(4), 332–353 (2008)
50. Wang, D., Brown, G.J.: Fundamentals of computational auditory scene analysis. In: Wang, D., Brown, G.J. (eds.) Computational Auditory Scene Analysis: Principles, Algorithms and Applications, pp. 1–44. Wiley, Hoboken (2006)
51. Wiener, N.: Extrapolation, Interpolation, and Smoothing of Stationary Time Series: with Engineering Applications. MIT Press, Cambridge (1950)

Chapter 13
Monaural Speech Enhancement Based on Multi-threshold Masking

Masoud Geravanchizadeh and Reza Ahmadnia

Abstract The ideal binary mask (IBM) has been assigned as a computational goal in computational auditory scene analysis (CASA) algorithms. Only time–frequency (T-F) units with local signal-to-noise ratio (SNR) exceeding a local criterion (LC) are assigned the binary value 1 in the binary mask. However, there are two problems with employing IBM in source separation applications. First, an optimum LC for a certain SNR may not be appropriate for other SNRs. Second, binary weighting may cause some parts or regions of the synthesized speech to be discarded at the output. If one employs variable weights, as opposed to the hard limiting weights (i.e., 0 or 1) taken in IBM, the above-mentioned problems can be solved considerably. In this chapter, a novel auditory-based mask, called ideal multi-threshold mask (IMM) is proposed which can be used in source separation applications. To show the potential capabilities of the new mask, a minimum mean-square error (MMSE)-based method is proposed to estimate IMM in the framework of monaural speech enhancement system. Various objective and subjective evaluation criteria show the superior performance of the new speech enhancement system as compared to a recently introduced enhancement technique.

13.1 Introduction

In a natural environment, a target sound, such as speech, is usually mixed with acoustic interference. A sound separation system that removes or attenuates acoustic interference has many important applications, such as automatic speech recognition (ASR), speaker identification in real acoustic environments, audio informa-

M. Geravanchizadeh (✉) · R. Ahmadnia
Faculty of Electrical and Computer Engineering, University of Tabriz, 5166615813 Tabriz, Iran
e-mail: geravanchizadeh@tabrizu.ac.ir; mgeravan@yahoo.com

R. Ahmadnia
e-mail: r_ahmadnia89@ms.tabrizu.ac.ir

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_13,
© Springer-Verlag Berlin Heidelberg 2014

tion retrieval, sound-based human–computer interaction, and intelligent hearing aids design. Because of its importance, the sound separation problem has been extensively studied in signal processing and related fields. Three main approaches in this context are speech enhancement, spatial filtering with a microphone array, and blind source separation (BSS) using independent component analysis (ICA). Speech enhancement typically assumes certain prior knowledge of interference. For example, the standard spectral subtraction method is easy to apply and works well when the background noise is stationary. However, the enhancement approach has difficulty in dealing with some nonstationary aspects of the environment where a variety of intrusions, such as competing talkers may occur. While machine separation remains a challenge, the auditory system shows a remarkable capability for sound separation, even monaurally. According to Bregman [1] the auditory system processes the acoustic input in two stages: an analysis and segmentation stage where the sound is decomposed into distinct time–frequency (T–F) segments, followed by a grouping stage. The grouping stage is divided into primitive grouping and schema-driven grouping that represents bottom-up and top-down processes, respectively.

The ideal binary mask (IBM) has been assigned as a computational goal in computational auditory scene analysis (CASA) algorithms [2, 3]. A binary mask is defined in the T–F domain as a matrix of binary numbers. We refer to the basic elements of the T–F representation of a signal as T–F units. Frequency decomposition similar to the human ear can be achieved using a bank of gammatone filters [4], and signal energies are computed in time frames. There have been some studies toward implementing an IBM with a fixed local criterion (LC). The study made in [5] has shown large intelligibility benefits by employing an IBM-based signal segregation method. This study has reported positive results on hearing impaired subjects. Multiplying IBM with the noise-masked signal can yield large gains in intelligibility, even at extremely low SNR levels (-5 dB, -10 dB) [6]. However, there are two problems with employing IBM in source separation problems. First, an optimum LC for a certain SNR may not be appropriate for other SNRs. Second, binary weighting may cause some regions of the synthesized speech to be discarded at the output.

Some other notable work has been done in the realm of soft mask estimation which is closely related to the source separation problem. The method proposed in [7] is a two-stage frequency-domain procedure for underdetermined convolutive BSS. Here, in the first stage, frequency bin-wise samples along the time axis are classified based on Gaussian mixture model fitting. In the second stage, the permutation ambiguities of the bin-wise classified signals are aligned by clustering the posterior probability sequences along the frequency axis. After calculating the posterior probabilities for all sources and for all observation vectors, a probabilistic T–F masking is performed to separate source signals in the frequency domain. In another study, a source separation method using probabilistic models of sources and an expectation–maximization parameter estimation procedure is presented [8]. The proposed system, which is referred to as model-based expectation–maximization source separation and localization (MESSL), clusters individual spectrogram points based on their interaural phase difference (IPD) and interaural level difference (ILD). In this way, MESSL creates probabilistic masks that can be used to separate sound sources from an under-

determined reverberant mixture. In [9], based on weighted combination of various features extracted from binaural recordings, a kind of soft mask is generated and applied to the mixture signal to improve source separation algorithms in reverberant conditions. Here, first, four different features are extracted at each T–F unit of a binaural recording, namely, ILD, IPD, the observation vector, and the interaural coherence (IC). Then, the probability of each T–F unit belonging to each source is obtained from the occupation likelihood and applied to the mixture as a probabilistic soft mask to extract the source signals.

In this chapter, a novel approach in designing a soft mask is proposed which can be used in the source separation problem. Here, an ideal multi-threshold mask (IMM) with variable threshold criteria is designed and employed to improve the quality and intelligibility of the separated speech. The motivation behind using such a mask with variable weights is to overcome the shortcomings of IBM which were mentioned above.

The rest of the chapter is organized in the following manner. The next section provides an overview of IBM and discusses its shortcomings. Section 13.3 presents our proposed ideal multi-threshold mask (IMM). As a special case of source separation problem, Sect. 13.4 describes the monaural speech enhancement system which is based on ideal multi-threshold masking. Systematic evaluations and comparisons are provided in Sect. 13.5. Finally, Sect. 13.6 summarizes the chapter and gives some remarks as to the future work.

13.2 The Ideal Binary Mask

The IBM is defined by comparing the signal-to-noise power ratios within each T–F unit against a local criterion (LC) or threshold measured in units of decibels. Only the T–F units with local signal-to-noise ratio (SNR) exceeding LC are assigned the binary value 1 in the binary mask. Let $T(t, f)$ and $M(t, f)$ denote target and masker signal powers measured in decibels, at time t and frequency f, respectively. The IBM is defined as

$$\mathrm{IBM}(t, f) = \begin{cases} 1, & \text{if } T(t, f) - M(t, f) > \mathrm{LC}, \\ 0, & \text{otherwise.} \end{cases} \tag{13.1}$$

An IBM-based segregated signal can be synthesized from a mixture by deriving a gain from the binary mask, and applying it to the mixture before reconstruction in a synthesis filterbank. Studies in [10] have shown nearly perfect intelligibility of IBM-processed mixture when the value of LC is varied from -12 to 0 dB. Meanwhile, the IBM with LC of 0 dB is considered to be theoretically the optimal mask out of all possible binary masks in terms of SNR gain. In practice, often LC is set as the middle of this interval (i.e., -5 or -6 dB). Along with reporting similar results, other studies have also analyzed the effect of the spectral resolution among other factors influencing intelligibility of ideal binary-masked speech [6]. However, there are two

problems with employing IBM in source separation, namely the SNR-dependence of LC and hard-masking effect of IBM which are discussed in detail below.

13.2.1 SNR Dependence of LC

Optimum LC for a certain SNR may not be appropriate for other SNRs. In other words, the studies show that the intelligibility of speech treated with IBM decreases significantly if LC is not selected properly. Brungart et al. [5] has measured the impact of the threshold value on speech intelligibility and has found that optimal thresholds are dependent on the global signal-to-noise ratio (SNR) values. Specifically, if the SNR value is increased by Δ dB, this leads to an increase of the threshold by an amount of approximately Δ dB. In this way, a high intelligibility can be maintained if the LC value is changed in accordance with changing global SNRs.

13.2.2 Hard-Masking Effect of IBM

In the context of an ideal binary mask, T–F units assigned with the binary value 1 are retained, while those assigned with the value 0 are discarded. Such binary weighting may cause some parts or regions of the synthesized speech to be discarded at the output. We interpret these regions as deep artificial gaps. To solve this problem, the study in [11] has proposed to fill the above-mentioned speech gaps with unmodulated broadband noise. This study reports that adding background noise shallows the areas of silence in the time–frequency domain of the IBM-processed target–speech/masker mixture. In this way, the abruption of transient changes in the mixture is smoothed and the perceived continuity of target–speech components is enhanced, leading to improved target–speech intelligibility. The amount of noise added depends on the input noise type and SNR level.

13.3 Proposed Ideal Multi-threshold Mask (IMM)

As mentioned previously, optimum LC for a certain SNR may not be optimum for other SNR levels. This means that the intelligibility of the IBM-processed speech may decrease significantly if LC is not set to a proper value. Also, the hard masking effect of IBM may cause deep artificial gaps. This is specifically the case in high frequency regions where noise has much energy that causes the loss of the weak unvoiced parts of speech. The resulting artificial gaps sound unnatural at the output. Therefore, we seek a mask that solves the above-mentioned IBM shortcomings simultaneously. Employing a kind of soft threshold masking is probably the strategy which is practically applied to the incoming mixture in the human auditory system.

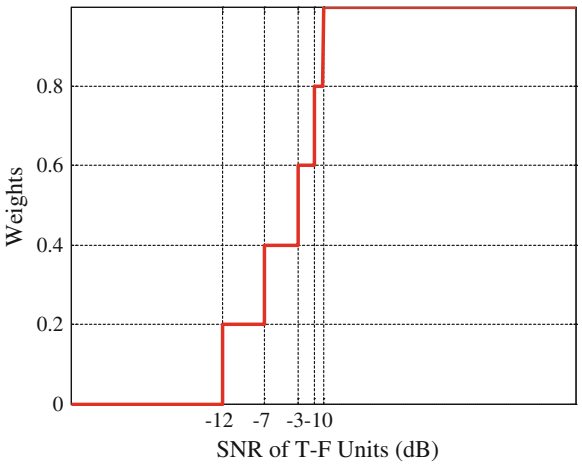


Fig. 13.1 The ideal multi-threshold mask (IMM). The IMM keeps the units with local SNR greater than 0 dB, discards the units with local SNR less than −12 dB, and weights the T-F units in this interval increasingly

The studies made on the design and application of IBM show that, the local SNR value of T-F units in the range of −12 to 0 dB has an important impact on selecting the local criterion value LC. If one employs variable weights in this region, as opposed to the hard limiting weights (i.e., 0 or 1) taken in IBM, the aforementioned problems can be solved considerably. The ideal multi-threshold mask (IMM) which is found empirically is depicted in Fig. 13.1 and can be stated mathematically as follows:

$$W(t,f) = \begin{cases} 0, & \text{if } \mathrm{SNR_{TF}} < -12\,\mathrm{dB}, \\ 0.2, & \text{if } -12\,\mathrm{dB} \le \mathrm{SNR_{TF}} < -7\,\mathrm{dB}, \\ 0.4, & \text{if } -7\,\mathrm{dB} \le \mathrm{SNR_{TF}} < -3\,\mathrm{dB}, \\ 0.6, & \text{if } -3\,\mathrm{dB} \le \mathrm{SNR_{TF}} < -1\,\mathrm{dB}, \\ 0.8, & \text{if } -1\,\mathrm{dB} \le \mathrm{SNR_{TF}} < 0\,\mathrm{dB}, \\ 1, & \text{if } 0\,\mathrm{dB} \le \mathrm{SNR_{TF}}. \end{cases} \tag{13.2}$$

where $\mathrm{SNR_{TF}}$ denotes the SNR value in T-F units, and $W(t, f)$ is the assigned weight to a T-F unit according to its SNR level. Since the T-F units with SNRs close to 0 dB have larger impact on intelligibility than those with SNRs near −12 dB, the SNR range near 0 dB is divided into smaller partitions to increase the resolution of IMM in this area.

The use of ideal multi-threshold mask (IMM) has two advantages over a conventional IBM. First, in this type of thresholding, especially in low SNRs (i.e., smaller than −5 dB), the small amount of noise energy which remains in the segregated speech, does not have a destructive role but it is beneficial in filling the speech gaps. The second advantage concerns the extraction of low-energy parts of speech signal, especially the unvoiced regions of the target. In spite of its low rate of frequency

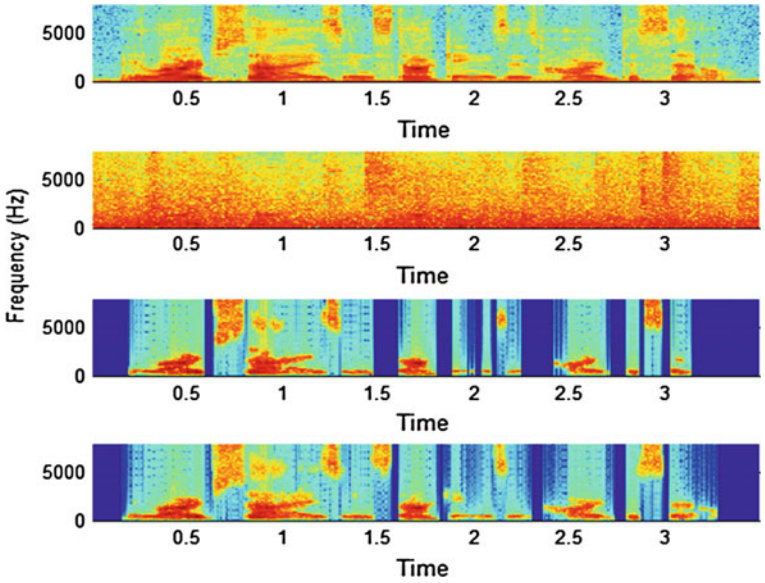


Fig. 13.2 Comparison of the performances of ideal binary mask (IBM) and ideal multi-threshold mask (IMM) in preserving the unvoiced regions of speech. The first panel, from the *top*, shows the cochleagram of a clean speech signal [14]. The second panel is the cochleagram of the noisy signal corrupted with Factory noise [15] at SNR $= -10$ dB. The third panel is the signal cochleagram processed by IBM with LC $= -5$ dB. The last panel is the signal cochleagram processed by IMM. Comparing the third and fourth cochleagrams shows that the unvoiced portions of the speech are retained and the gaps are filled in the case of the cochleagram obtained by IMM

(approx. % 24 in the conversational speech [12, 13]), the unvoiced speech has an important impact on speech intelligibility. By using IMM in the source separation procedure, one expects to increase the intelligibility value by retaining the unvoiced parts of speech. Figure 13.2 shows the cochleagrams of a clean signal taken from the IEEE database [14], mixture signal corrupted with Factory noise taken from the Noisex-92 database [15] at SNR $= -10$ dB, ideal binary masked signal, and ideal multi-threshold masked signal, respectively. It can be seen that the time continuity and the preservation of the unvoiced regions in the IMM-processed signal is more remarkable than those obtained by using IBM.

Figure 13.3 shows the cochleagrams of a clean signal [14], mixture signal corrupted with Factory noise [15] at SNR $= 0$ dB, ideal multi-threshold masked signal, and ideal binary masked signal with LC $= 0, -5$, and -12 dB, respectively. The SNR dependence and hard-masking problems of IBM can be seen in these representations.

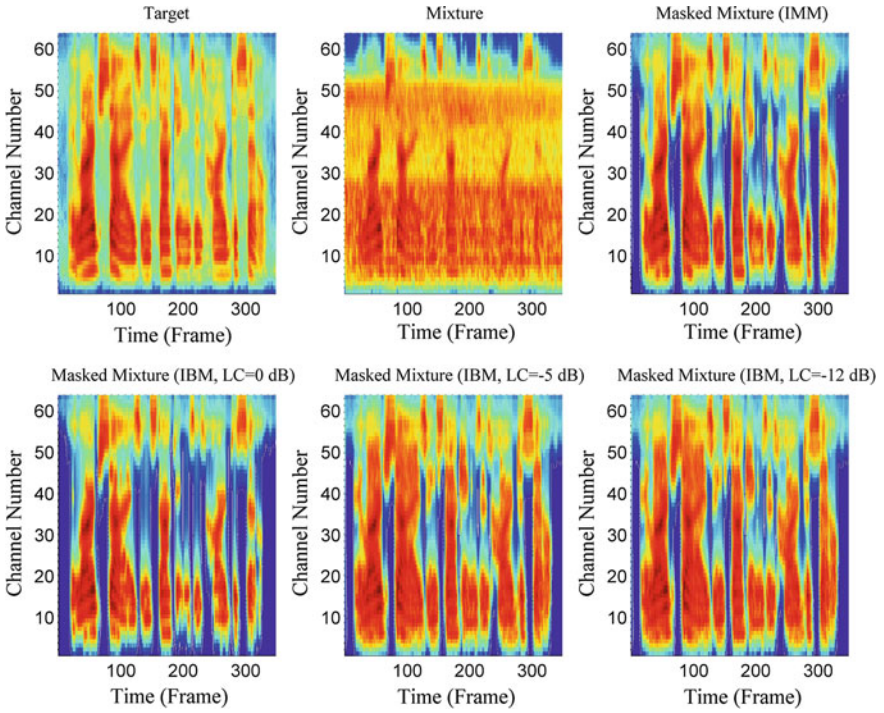


Fig. 13.3 Comparison of the performances of IMM and IBM with different masking thresholds. The *top-left panel* shows the cochleagram of the utterance "A large size in stockings is hard to sell" spoken by a male talker [14]. The *top-middle panel* shows the cochleagram of a mixture of speech and factory noise [15], with SNR $= 0$ dB. The *top-right* panel shows the masked mixture using IMM; note the similarity with the cochleagram of the original target speech (*top-left panel*). The result of applying IBM with LC $= 0, -5$, and -12 dB to the cochleagram of the mixture is shown in the *bottom panels*, respectively

13.4 Speech Enhancement Based on Ideal Multi-threshold Masking

The discussions made above assume the ideal case, where the clean and noise signals are separately available, and so we have access to the ground-truth SNRs of all T-F units. However, in real-world monaural conditions, only the noisy signal is available. Therefore, to use the potential benefits of IMM in speech enhancement applications, this mask should be estimated from noisy signals. For this purpose, a minimum mean-square error (MMSE)-based approach is proposed.

One reason for the incomplete performance of traditional speech enhancement methods such as MMSE-based approach is the inaccurate prediction of the noise power spectrum in nonstationary noisy environments. This does not mean avoiding

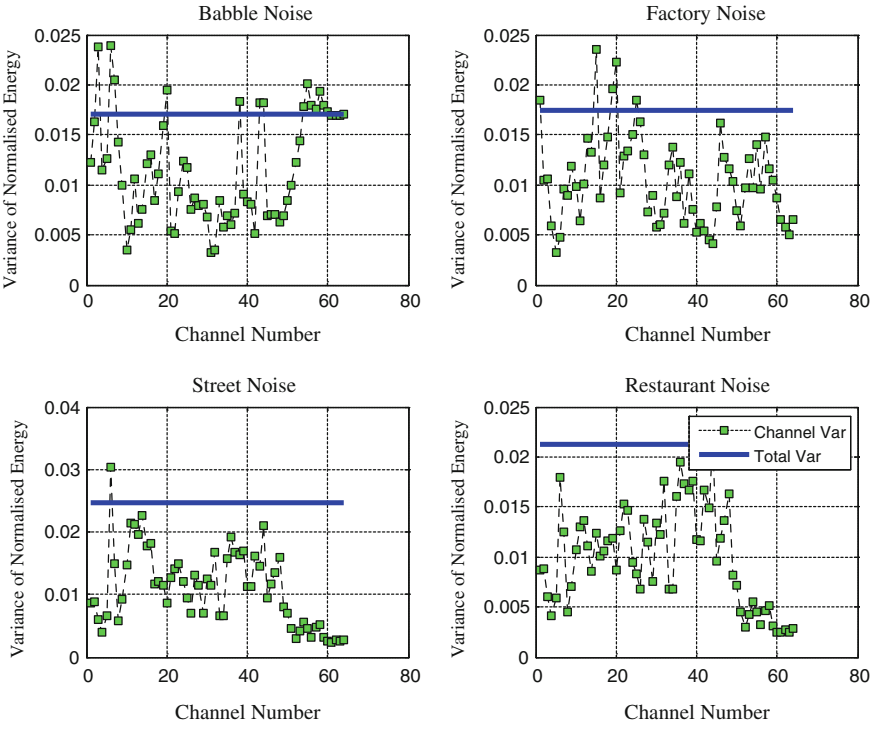


Fig. 13.4 Illustrations of variances of normalized energy for the babble, factory, car, and restaurant noises [15] after decomposing the original noise signal into 64 subband signals by a gammatone filterbank. In each panel, the *solid blue curves*, labeled as "Total Var", indicate the variance of normalized energy of the original noise signals, whereas the *small green squares*, labeled as "Channel Var", indicate the normalized energy variances in each channel. Almost in all cases, the variances of normalized energies in channels have lower values than those of the original noise signals

the use of these algorithms, but it means that if the noise behavior nears to stationary or quasi-stationary, their performance would be better [16].

Our empirical studies show that many types of noises if decomposed into subband signals become more stationary as compared to the original noise signal. The process is made up of two stages. In the first stage, the noise signal is broken down into 64 subband signals using a gammatone filterbank. Signals of all individual channels are divided into 20 ms rectangular frames with 50 % overlap. This creates time-frequency (T-F) units, in which noise energies can be computed. In the second stage, the variances of energies of all T-F units in each channel are computed after a normalization procedure. The normalized variance of the original undecomposed noise energy is obtained subsequently. Figure 13.4 shows the normalized energy variances for the Babble, Factory, Car, and Restaurant noises in gammatone filterbank channels along with variance of normalized energy for the original noise signal. Here, the solid blue curves, labeled as "Total Var", depict the normalized energy variances of the original

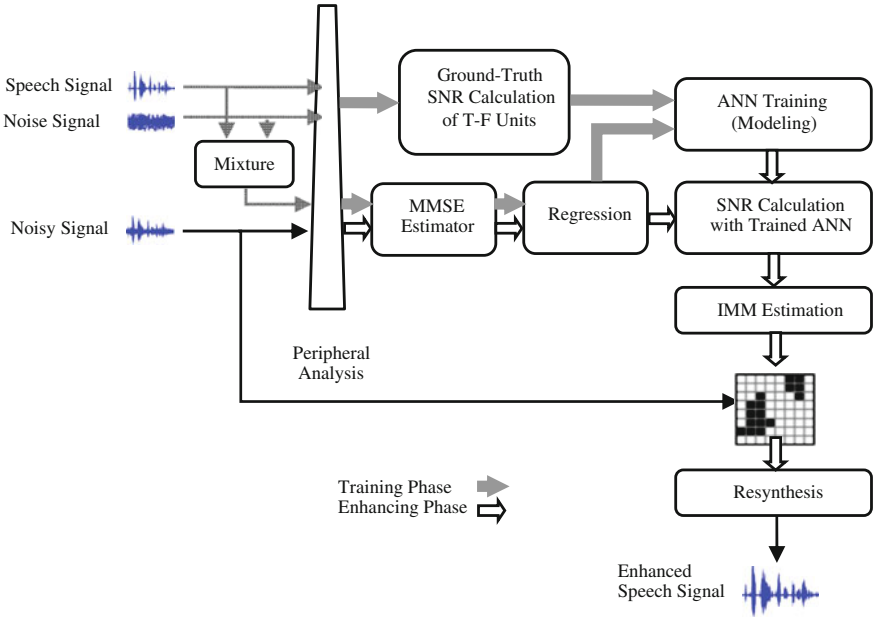


Fig. 13.5 Block diagram of the monaural speech enhancement system using the MMSE-based estimation of ideal multi-threshold mask (IMM)

noise signals, whereas the small green squares in each panel, labeled as "Channel Var", indicate the normalized energy variances in each channel. It can be readily seen that, almost in all cases, the variances of normalized energies for channels are lower than those of the original noise signals themselves. This observation confirms the fact that the noises decomposed in subbands have a stationary or quasi-stationary behavior than the original noise signal. As a consequence of this finding, we are motivated to employ the channel-wise MMSE method to calculate the noise power spectrum, which is then used in the estimation procedure of IMM.

Figure 13.5 illustrates the block diagram of the monaural speech enhancement system based on multi-threshold masking. Generally, the estimation of IMM can be described as a two-stage process, namely the training stage and the enhancement stage. In the first stage, an artificial neural network (ANN) is trained to find the SNR of each T-F unit using the training data. In the next stage, the estimated IMM is applied to the noisy signal using the calculated SNR at the output of ANN. The details of the speech enhancement procedure are described in detail below.

13.4.1 Peripheral Analysis

The input signal is first decomposed in the frequency domain by a bank of 64 gammatone filters [4], with their center frequencies equally distributed on the equivalent rectangular bandwidth (ERB) rate scale from 50 to 8,000 Hz. This filterbank is an empirical one that simulates the cochlear organ of the ear. The impulse response of the gammatone filter is given as

$$g_{f_c}(t) = t^{N-1} \exp\left[-2\pi b\left(f_c\right)\right] \cos(2\pi f_c t + \phi)\, u(t). \tag{13.3}$$

where $N = 4$ is the order of the filter, b is the equivalent rectangular bandwidth, f_c is the center frequency of the filter, ϕ is the phase, and $u(t)$ is the step function.

The outputs of the filterbank are then transformed into neural firing rate by hair cell model [12]. In each filter channel, the output is divided into 20 ms time frames with 10 ms overlapping between consecutive frames. As a result of this process, the input is decomposed into a two-dimensional time-frequency representation, or a collection of T-F units. The resulting T-F representation is known as a cochleagram [2].

13.4.2 Calculation of Ground-Truth SNRs of T-F Units

In this stage, the ground-truth SNRs of all T-F units are calculated for the next stage, namely the artificial neural network which should be trained using the training data. For this purpose, first, pairs of training data, including clean and noise signals are prepared manually at specified SNR levels. The clean and noise signals in each pair are then decomposed separately by the peripheral analysis unit into time-frequency units. By having access to the clean and noise signals in each T-F unit, the ground-truth SNRs of T-F units are calculated as follows:

$$\mathrm{SNR}_{G-T}^{\mathrm{TF}} = 10 \log\left(\frac{\sum_n \left(s_{\mathrm{TF}}(n)\right)^2}{\sum_n \left(n_{\mathrm{TF}}(n)\right)^2}\right). \tag{13.4}$$

Here, $s_{\mathrm{TF}}(n)$ and $n_{\mathrm{TF}}(n)$ represent, respectively, the clean and noise signals in each T-F unit and n is the time index.

13.4.3 MMSE Estimation of the A priori SNR ξ_k

The unit has the task of estimating the *a priori* SNR in the frequency domain for each T-F unit. To this aim, first, a mixture of noisy signal is generated at a known SNR level. Then, the mixture signal is split into subband signals by the gammatone

filterbank in the peripheral analysis stage, yielding a representation of noisy signal in the form of T-F units.

Let $y(n) = x(n) + d(n)$ be the sampled noisy speech signal of a T-F unit in channel c consisting of the clean signal $x(n)$ and the noise signal $d(n)$. In the frequency domain, we have:

$$Y(\omega_k) = X(\omega_k) + D(\omega_k) \tag{13.5}$$

for $\omega_k = 2\pi k/N$ and $k = 0, 1, 2, \ldots, N-1$, where N is the number of samples in each frame or T-F unit of channel c.

The log-MMSE algorithm for the estimation of the *a priori* SNR ξ_k can be implemented recursively using the following steps. For each windowed frame m (i.e., T-F unit) of the mixture signal in channel c do:

Step 1 Compute the DFT of the noisy speech signal: $Y(\omega_k) = Y_k \exp\left(j\theta_y(k)\right)$.

Step 2 Estimate the *a posteriori* SNR γ_k as $\gamma_k = Y_k^2/\lambda_d(k)$, where $\lambda_d(k)$ is the power spectrum of the noise signal computed during non-speech activity (e.g., during speech pauses). Then, estimate $\hat{\xi}_k$ using the decision-directed approach [17]:

$$\hat{\xi}_k(m) = \alpha \frac{\hat{X}_k^2(m-1)}{\lambda_d(k, m-1)} + (1-\alpha)\max[\gamma_k(m) - 1, 0], \tag{13.6}$$

where $0 < \alpha < 1$ is the weighting factor, and $\hat{X}_k^2(m-1)$ is the amplitude estimator obtained in the past analysis frame (i.e., past T-F unit in channel c). The above equation needs initial conditions for the first frame, i.e., for $m = 0$. The following initial conditions are recommended [17] for $\hat{\xi}_k(m)$:

$$\hat{\xi}_k(0) = \alpha + (1-\alpha)\max\left[\gamma_k(0) - 1, 0\right].$$

where good results are obtained with $\alpha = 0.98$.

Step 3 Based on the estimated value of $\hat{\xi}_k$ in the previous step, estimate the enhanced signal magnitude $\hat{X}_k$ using the log-MMSE estimator [18]:

$$\begin{aligned}\hat{X}_k &= \frac{\xi_k}{1+\xi_k} \exp\left\{\frac{1}{2}\int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\right\} Y_k \\ &= G_{LSA}(\xi_k, v_k) Y_k,\end{aligned} \tag{13.7}$$

where $G_{LSA}(\xi_k, v_k)$ is the gain function of the log-MMSE estimator, ξ_k is the *a priori* SNR, and v_k is defined as

$$v_k = \frac{\xi_k}{1+\xi_k}\gamma_k.$$

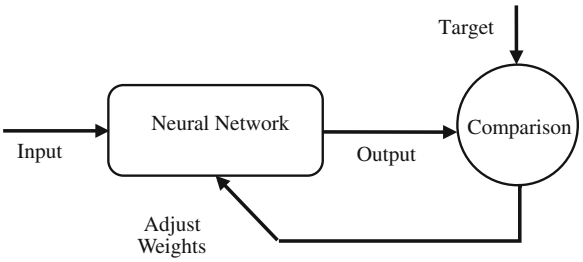


Fig. 13.6 Recursive learning process in a typical artificial neural network system

Step 4 Return to the Step 1, and repeat the procedure until a convergence criterion is satisfied.

The above procedure results in a large-dimensional vector quantity $\hat{\xi}_k$ (here, a vector with $N = 256$ dimensions) for each T-F unit, which will be further processed by the next unit.

13.4.4 Regression

The estimated *a priori* SNRs in T-F units are vectors of large dimension. In order to drive the ANN classifier with a low-dimensional vector, it is necessary to reduce the dimension of the vectors generated by the MMSE estimation unit described above. This task is achieved by the unit of regression.

Without losing the generality, a linear regression is used for the process of dimension reduction [19]. It is known that polynomial models are a special case of the linear regression models. Polynomial models have the advantages of being simple and familiar in their properties. Therefore, one way of reducing the dimensions of the *a priori* SNR vectors is to fit them with linear polynomials of fixed order and then extract their coefficients as the desired feature vector. Assuming that each element of the *a priori* SNR vector represents an observation variable, the process of linear regression by a polynomial of order 4 transforms the large-dimensional vectors of the *a priori* SNRs (here with a dimension of 256) into feature vectors of dimension 5.

13.4.5 Training Artificial Neural Network

Typically, neural networks are adjusted or trained, such that a particular input leads to a specific target output. Figure 13.6 illustrates such a situation. Here, the network is adjusted, based on a comparison of output and target, until the network output matches target.

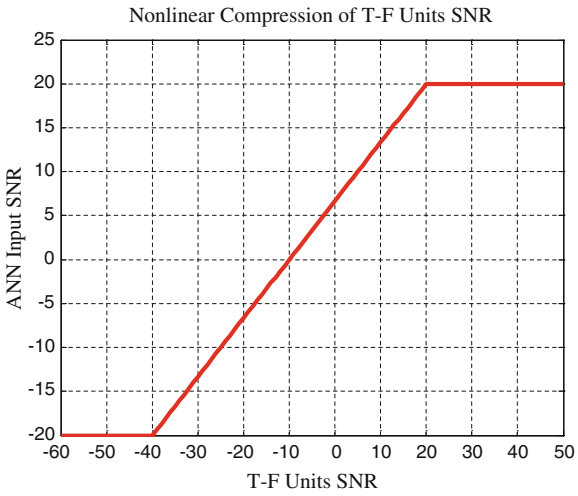


Fig. 13.7 The nonlinear compression to reduce SNR variances of T-F units. The function maps SNR values of units to the range [−20 dB 20 dB]

In the process of estimating an ideal multi-threshold mask, a neural network consisting of one hidden layer with 50 neurons is used. The estimation procedure of IMM consists of two steps, namely the training phase and the testing phase.

The training is performed in a supervised manner in which the low-dimensional *a priori* SNR feature vectors and the corresponding ground-truth SNRs of all T-F units are used as input and target to ANN, respectively. However, since T-F units may have very large or very small values of ground-truth SNRs, it is important to reduce first the impact of SNR variances on the training process of ANN. To this aim, a simple nonlinear transformation is applied on the extracted ground-truth SNRs of T-F units to compress (or map) the SNR values to a specified range. This mapping function is shown in Fig. 13.7, and is described mathematically as

$$\mathrm{SNR}_{\mathrm{TF}}^{C} = \frac{1}{3}\big(|\mathrm{SNR}_{\mathrm{TF}} + 40| - |\mathrm{SNR}_{\mathrm{TF}} - 20|\big), \tag{13.8}$$

where $\mathrm{SNR}_{\mathrm{TF}}$ and $\mathrm{SNR}_{\mathrm{TF}}^{C}$ are, respectively, the true SNR and the compressed SNR of the unit TF. The compressed values of units SNR are given as input to ANN.

13.4.6 SNR Calculation with Trained ANN

In the testing or evaluation phase of the proposed method for estimating IMM, the reduced feature vector of each unit, i.e., the *a priori* SNR vector of T-F unit generated by the regression algorithm, is extracted and given as input to the trained ANN. The output given by ANN is the SNR value for the corresponding unit. However, before using the calculated SNR of the unit in the estimation of IMM, it must be processed

by a decompressing function. This is the inverse of the compressing function as used in the training phase:

$$\hat{\mathrm{SNR}}_{\mathrm{TF}} = \left(3 \times \hat{\mathrm{SNR}}^{C}_{\mathrm{TF}} - 20\right)/2, \tag{13.9}$$

where $\hat{\mathrm{SNR}}^{C}_{\mathrm{TF}}$ is the SNR value obtained by ANN and $\hat{\mathrm{SNR}}_{\mathrm{TF}}$ is the decompressed value of SNR for the T-F unit denoted as TF.

13.4.7 Estimation of IMM

The SNR value, as obtained by Eq. (13.9), is used to estimate the ideal multi-threshold mask (IMM). In other words, by having access to SNRs of all T-F units and considering IMM criterion as given by Eq. (13.2), different weights are assigned to different T-F units of the input signal according to their calculated SNR values. The constructed mask is then applied to the input noisy signal in the next stage to produce the enhanced speech.

13.4.8 Resynthesis Procedure

Using the estimated IMM, it is straightforward to resynthesize the enhanced speech signal from the output of the gammatone filterbank. This can be achieved by employing an approach introduced by Weintraub [20] which is described in the following stages. In the first stage, first, the response of each filter is reversed in time. Then, the reversed response is passed back through the filter. Finally, the filtered response is time-reversed again. In this way, across-channel phase shifts in the filterbank output are removed. In the second stage, the phase-corrected output from each channel is divided into time frames by windowing with a raised cosine, with a frame size equal to the one used in the decomposition of input mixture into T-F units. The energy in each T-F unit is then weighted by the corresponding T-F mask value obtained from the estimated IMM. In the last stage, the weighted responses are summed across all frequency channels to yield an enhanced speech waveform. The process is depicted in Fig. 13.8.

13.5 Evaluations and Experimental Results

For simulations, clean speech signals are selected randomly from the IEEE corpus [14] and noise signals are taken from the Noisex-92 database [15]. We have assessed the performance of our proposed ideal multi-threshold mask (IMM) both in comparison with IBM and in the framework of the monaural speech enhancement system by investigating the amount of SNR improvement, PESQ [21] value, and listening tests.

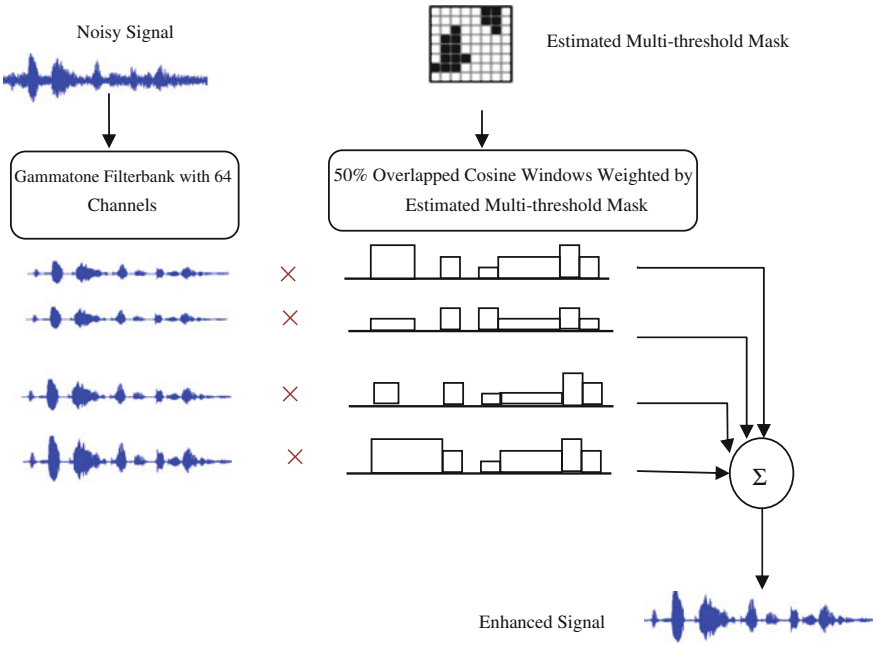


Fig. 13.8 The process of applying the estimated multi-threshold mask to generate the enhanced speech signal from the noisy input mixture

To generate noisy signals, clean signals are mixed with the Babble, Car, Factory, and Restaurant noises at different SNR levels. Table 13.1 shows the experimental conditions used in the evaluation procedure.

13.5.1 Evaluation Criteria

In the following, different objective and subjective measures are introduced to assess the performance and capability of the proposed ideal multi-threshold mask (IMM). These include examining the results of SNR improvement, PESQ test, and listening tests.

13.5.1.1 SNR Improvement

The SNR criterion is an objective measure which is calculated as follows:

$$\mathrm{SNR} = 10 \log \left(\frac{\sum_n S_{\mathrm{AllOne}}(n)^2}{\sum_n \left(S_{\mathrm{AllOne}}(n) - \hat{S}_O(n) \right)^2} \right), \tag{13.10}$$

Table 13.1 Experimental conditions

Number of training data	50 sentences
Number of evaluation data	10 sentences
Sampling frequency (f_s)	16 kHz
Window type	Hamming
Frame length (N)	320 samples (20 ms)
Overlap of frames ($N/2$)	160 samples (10 ms)
Number of channels in gammatone filterbank	64

where $S_{\text{AllOne}}(n)$ is the clean speech signal that is masked with an all-one mask and resynthesized again to cancel the delay and nonlinear modifications which occur in the synthesis process, and $\hat{S}_O(n)$ is the enhanced signal.

Higher values of SNR improvement is an indication of higher speech qualities.

13.5.1.2 PESQ Evaluation

The perceptual evaluation of speech quality (PESQ) is an objective evaluation of speech quality. PESQ is the ITU-T P.862 recommendation [22], which has been found to have a good correlation with the mean opinion score (MOS) test. This score ranges from 4.5 (the highest quality of speech) down to −0.5 (the lowest quality of speech). Among all objective measures considered, the PESQ measure is the most complex to compute and is the one recommended by ITU-T for speech quality assessment of 3.2 kHz (narrowband) handset telephony and narrowband speech codecs [22].

13.5.1.3 Listening Tests

In order to assess the proposed multi-threshold mask (IMM) subjectively, the MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA) is used, which is an ITU-R Recommendation BS.1534-1 [23] as implemented in [24].

The experiments are performed in a sound-proof room. Stimuli are played to listeners through headphone. The subjects (i.e., human listeners) are provided with test utterances plus one reference and one hidden anchor, and are asked to rate different signals (i.e., to give scores for the masked signals) obtained for each noise type and SNR level on a scale of 0–100, where 100 represents the best score. The listeners are permitted to listen to each sentence several times and always have access to clean signal reference. The test signals are the same as those used for the objective evaluation. Four types of noises (i.e., Babble, Car, Factory, and Restaurant) are used during the listening tests. A total of 10 listeners (three females and seven males between the ages of 18 and 30) have participated in these tests.

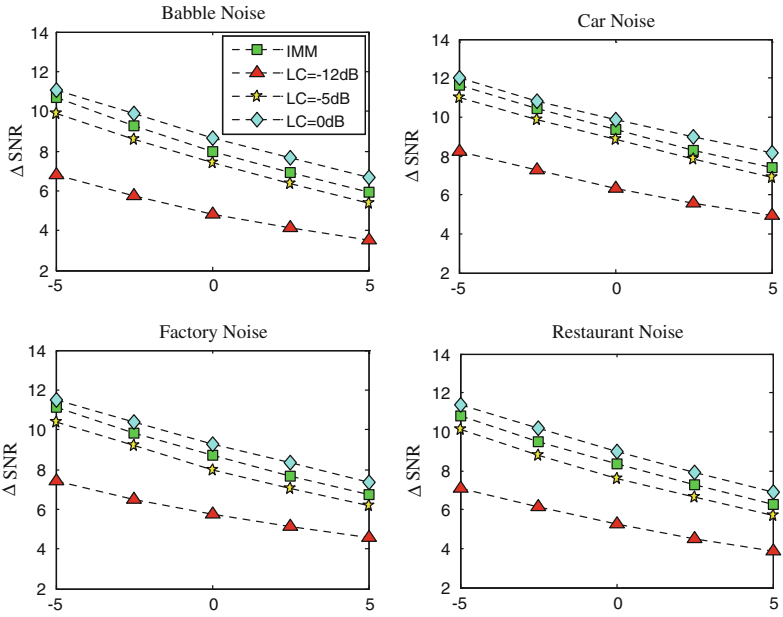


Fig. 13.9 Average SNR-Improvement obtained by processing mixture signals with the proposed IMM and IBMs with LC = 0, −5, and −12 dB, at different input SNRs and noise types

13.5.2 Evaluation of Ideal Masks (IMM Versus IBM)

To evaluate the performance of the ideal multi-threshold mask (IMM) against the ideal binary mask (IBM), these masks are applied directly to the input noisy mixture to generate the enhanced speech signal. The process is as follows. First, the clean and noise signals are manually mixed at different SNRs. Then, IBM and the proposed IMM are applied to the mixture signals. The weighted responses are finally processed by the resynthesis module to yield a reconstructed ideally masked mixture waveform.

Figure 13.9 shows the average SNR-improvement obtained by processing mixture signals with the proposed IMM and IBMs with three different LCs, at different input SNRs, and for the Babble, Factory, Car, and Restaurant noises, respectively. As seen from the figure, the values of SNR-improvement resulted from IBM with LC = −12 dB lies far below those obtained from other IBMs and IMM in all noise types. This implies the fact that the processed signal with this mask has more residual noise compared with other masks. The improvements from IMM and other IBMs are close together, and the amount of improvement from IMM is slightly less than that obtained from IBM with LC = 0 dB.

Figures 13.10 and 13.11 show, respectively, the SNR-improvement at channel 50 and over the top 20 high frequency channels of the gammatone filterbank resulted from different IBMs and IMM, under the same experimental conditions as given in Fig. 13.9. The band-wise evaluations reveal that IMM has in general higher perfor-

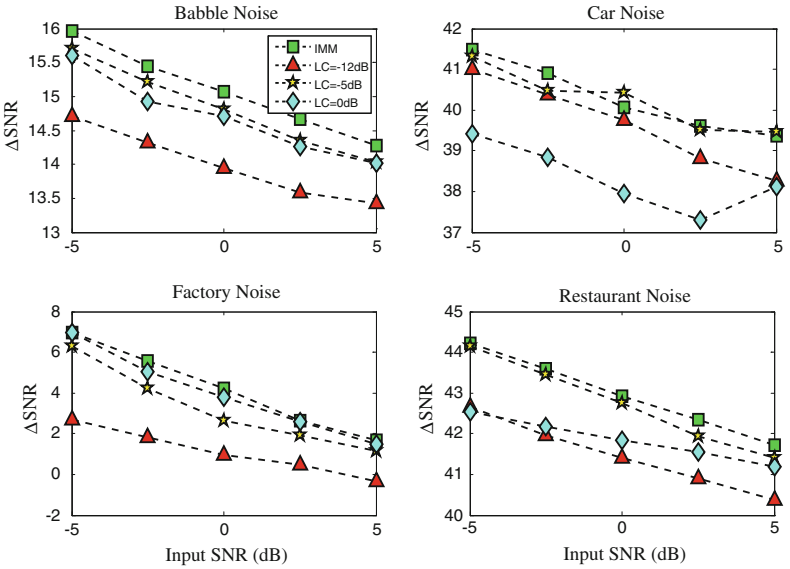


Fig. 13.10 Average SNR-Improvement in the 50th frequency channel obtained by processing mixture signals with the proposed IMM and IBMs with LC $= 0, -5$, and -12 dB, at different input SNRs and noise types

mance than IBM at high frequency channels in the sense of SNR-improvement. Also, it is seen that the amount of improvement resulting from IBM depends highly on the noise type and input SNR value, whereas IMM shows relatively stable behavior in these conditions. It is known that most of the weak unvoiced parts of speech are located at high frequencies. As stated before, the high energy of noise in these frequency regions may cause the loss of unvoiced parts of speech which results in the degradation of speech intelligibility. It can be concluded, therefore, that the IMM-processed mixture, because of its multiple thresholds, can lead to an improvement in speech intelligibility at high frequency range of input mixture.

Figure 13.12 shows the PESQ values for the proposed IMM and IBMs with three different thresholds obtained at different noisy conditions and input SNR values. As it can be seen from the figure, the IMM-processed signal has the best quality, compared with those signals which use IBMs as the processing mask.

The results of subjective listening tests for evaluating the performance of IMM and IBMs with different thresholds in the processing of enhancing the input noisy mixture are depicted in Fig. 13.13. By examining the results, it is observed that the proposed IMM produces the highest speech quality, compared with IBMs. The superior performance of IMM is again in agreement with the results obtained during the objective evaluations tests.

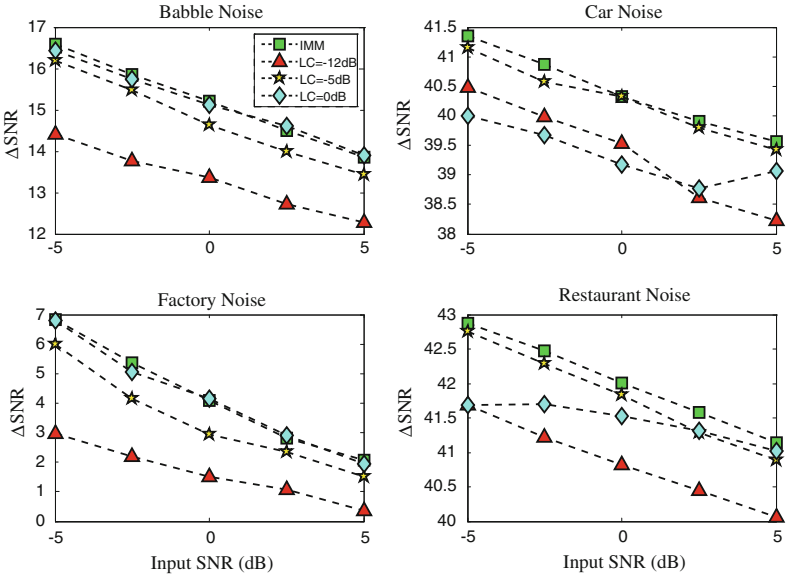


Fig. 13.11 Average SNR-Improvement over 20 high frequency channels obtained by processing mixture signals with the proposed IMM and IBMs with LC $= 0$, -5, and -12 dB, at different input SNRs and noise types

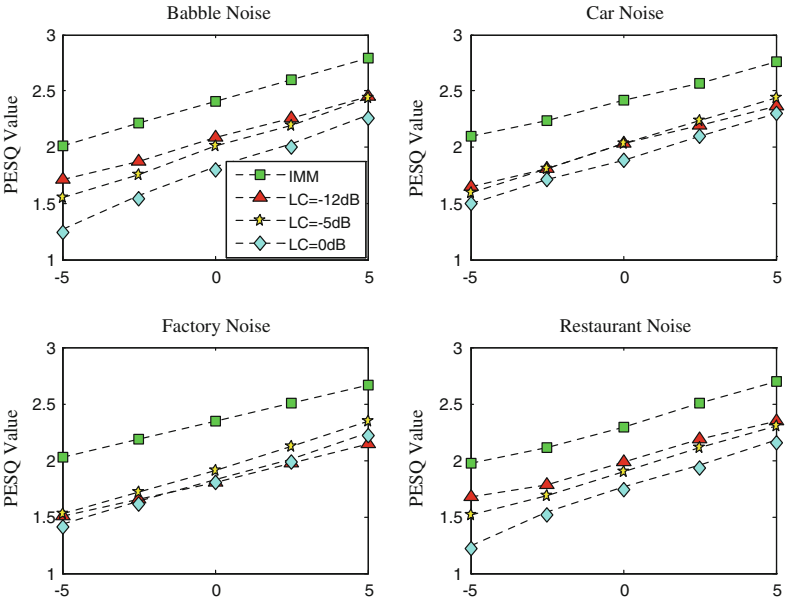


Fig. 13.12 Average PESQ values obtained by processing mixture signals with the proposed IMM and IBMs with LC $= 0$, -5, and -12 dB, at different input SNRs and noise types

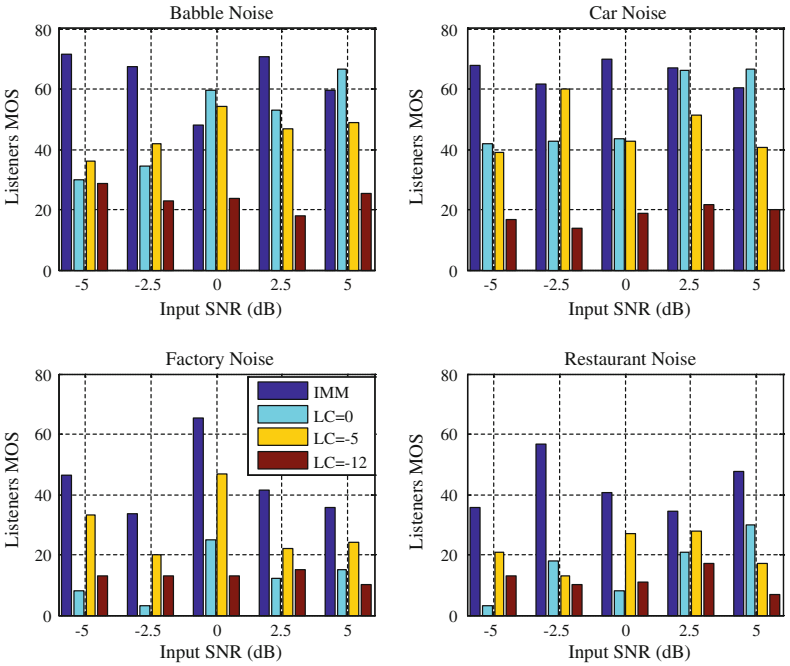


Fig. 13.13 The MUSHRA listening test results obtained by processing mixture signals with the proposed IMM and IBMs with LC $= 0$, -5, and -12 dB, at different input SNRs and noise types

13.5.3 Evaluation of Estimation Methods for IMM and IBM

In this section, the performance of our proposed MMSE-based multi-threshold mask estimation system is compared with that of a classification-based system for estimating binary mask in the framework of speech enhancement application. To this aim and motivated by a recent classification-based approach for unvoiced–voiced speech separation [25, 26], we first implement an SVM-based classifier for estimating binary mask which uses the gammatone frequency cepstral coefficients (GFCC) as classification features. Then, using the previously mentioned objective and subjective criteria, the performances of both systems are evaluated in the sense of improving the quality of input noisy mixture.

Here, the proposed multi-threshold mask estimation system and the SVM-based classifier (called SVM-GFCC) are trained with noisy signals obtained by mixing clean signals with Factory noise at SNR $= 0$ dB. In the evaluation phase, the performances of both systems are examined for all noise types, including the Babble, Car, and Restaurant noises, and at different SNR values of input mixture signals. Figures 13.14, 13.15, and 13.16 show the results of SNR improvement, PESQ measures, and listening tests, respectively. Although the results of Fig. 13.15 show that the proposed mask estimation method performs slightly poorer than the SVM-GFCC approach in

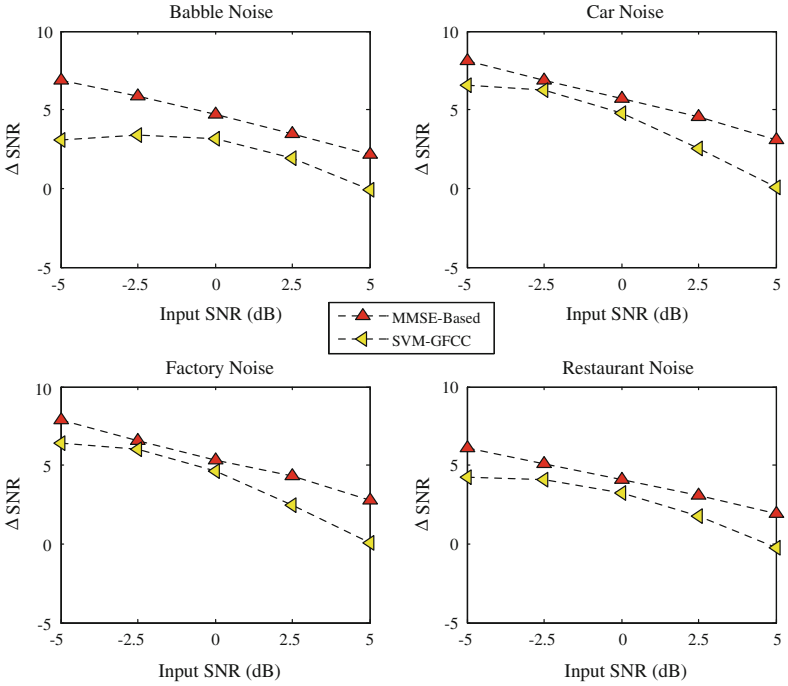


Fig. 13.14 Average SNR-Improvement obtained by processing mixture signals with the proposed MMSE-based mask estimation method and SVM-GFCC approach at different input SNRs and noise types

low SNR values, but investigating the results of other tests shows clearly that, in general, the performance of the proposed MMSE-based mask estimation method is superior in all noisy scenarios and input SNR values. The results demonstrate also that the proposed mask estimation approach has good generalization ability to unseen noises.

13.6 Summary

The ideal binary mask (IBM) has been assigned as a computational goal in computational auditory scene analysis (CASA). However, there are two problems with employing IBM in source separation applications. First, an optimum LC for a certain SNR may not be appropriate for other SNRs. Second, binary weighting may cause some parts or regions of the synthesized speech to be discarded at the output. We interpret these regions as deep artificial gaps. This is specifically the case in high frequency regions where the high energy of noise may cause the loss of unvoiced parts of speech resulting in the degradation of speech intelligibility.

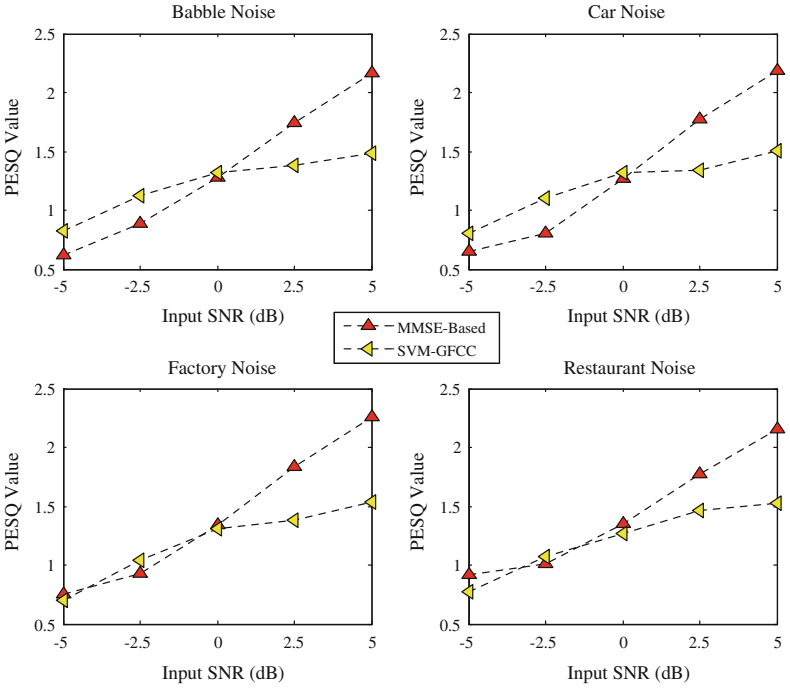


Fig. 13.15 Average PESQ values obtained by processing mixture signals with the proposed MMSE-based mask estimation method and SVM-GFCC approach at different input SNRs and noise types

In this chapter, a novel approach in designing an auditory-based mask is proposed which can be used in the source separation problem. Here, an ideal multi-threshold mask (IMM) is designed and employed in the speech enhancement application to obtain a high quality and intelligible target signal from a mixture.

The performance evaluation of the proposed multi-threshold mask is conducted in two steps. In the first step, the performance of the ideal multi-threshold mask (IMM) is assessed against the ideal binary mask (IBM) by applying these ideal masks directly to the input noisy mixture to generate the enhanced speech signal. In the second step, and to use the potential benefits of IMM in the real-world monaural conditions, a minimum mean-square error (MMSE)-based approach is proposed to estimate IMM. Systematic evaluations and comparisons with an SVM-based classifier for estimating binary mask show that the proposed estimation method of IMM improves substantially the performance of the conventional speech enhancement systems.

It is known that there is a relationship between intelligibility and labeling errors in IBM estimation [6]. Accuracy and HIT-FA are two important criteria to assess the intelligibility of resynthesized signals in such systems which are based on the estimation of binary mask. The HIT rate is defined as the percentage of the target-dominated units in the IBM labeled correctly and the FA rate refers to the percentage

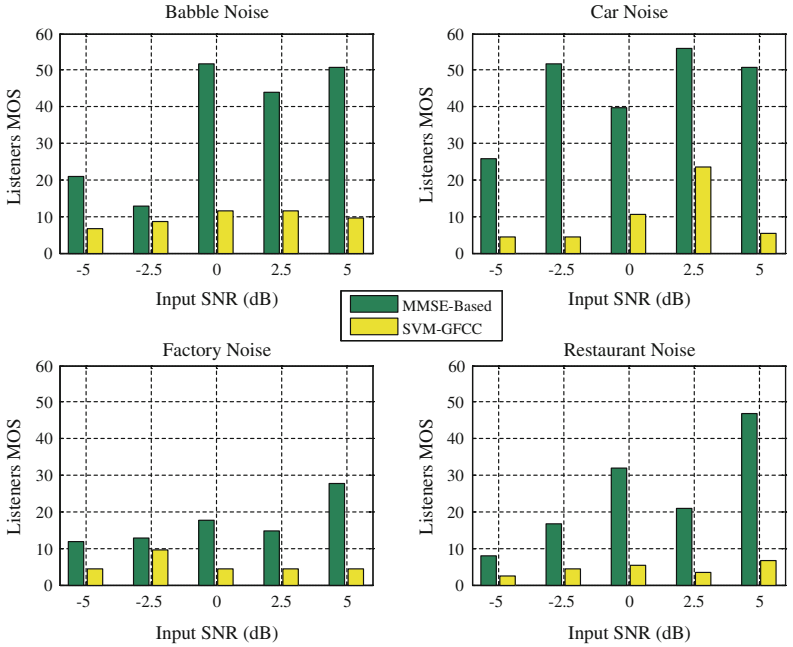


Fig. 13.16 The MUSHRA listening test results obtained by processing mixture signals with the proposed MMSE-based mask estimation method and SVM-GFCC approach at different input SNRs and noise types

of the interference-dominated units in the IBM labeled wrongly. The Accuracy is calculated as the percentage of correctly labeled units with respect to the IBM. As stated before, it is expected that employing the multi-threshold mask improves the intelligibility score of the input noisy signal, especially at high frequency ranges of the spectrum. However, in this work, we have not conducted any formal intelligibility tests to assess our proposed multi-threshold mask in the monaural speech enhancement system. This is because variable weights are employed in IMM, as opposed to the hard limiting weights (i.e., 0 or 1) taken in IBM. Therefore, here conventional criteria such as Accuracy and HIT-FA are not directly applicable. As the future work, we are working on developing appropriate intelligibility measures to evaluate the proposed IMM in source separation systems.

References

1. Bregman, A.S.: Auditory Scene Analysis. MIT, Cambridge (1955)
2. Wang D.L., Brown G.J.: Computational Auditory Scene Analysis: Principles, Algorithms, and Applications. Wiley-IEEE Press, Hoboken (2006)

3. Wang D.L.: On ideal binary mask as the computational goal of auditory scene analysis. In: P. Divenyi (ed.) Speech Separation by Humans and Machines, pp. 181–197. Kluwer Academic Publishers, Norwell (2005)
4. Patterson R.D., Nimmo-Smith I., Holdsworth J., Rice P.: An Efficient Auditory Filterbank Based on the Gammatone Function. Report No. 2341, MRC Applied Psychology Unit, Cambridge (1985)
5. Brungart, D., Chang, P.S., Simpson, B.D., Wang, D.L.: Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J. Acoust. Soc. Am. **120**(6), 4007–4018 (2006)
6. Li, N., Loizou, P.C.: Factors influencing intelligibility of ideal binary-masked speech: implications for noise reduction. J. Acoust. Soc. Am. **123**(3), 1673–1682 (2008)
7. Sawada H., Araki S., Makino S.: A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 139–142 (2007)
8. Mandel, M.I., Weiss, R.J., Ellis, D.P.W.: Model-based expectation-maximization source separation and localization. IEEE Trans. Audio Speech Lang. Process. **18**(2), 382–394 (2010)
9. Alinaghi A., Wang W., Jackson P.J.B.: Spatial and coherence cues based time-frequency masking for binaural reverberant speech separation. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) (2013)
10. Anzalone, M.C., Calandruccio, L., Doherty, K.A., Carney, L.H.: Determination of the potential benefit of time-frequency gain manipulation. Ear Hear. **27**(5), 480–492 (2006)
11. Cao, S., Li, L., Wu, X.: Improvement of intelligibility of ideal binary-masked noisy speech by adding background noise. J. Acoust. Soc. Am. **129**(4), 2227–2236 (2011)
12. Fletcher, H.: Speech and Hearing in Communication. D. Van Nostrand Company, New York (1958)
13. Dewey, G.: Relative Frequency of English Speech Sounds. Harvard University Press, Cambridge (1923)
14. Rothauser, E.H., Chapman, W.D., Guttman, N., Hecker, M.H.L., Nordby, K.S., Silbiger, H.R., Urbanek, G.E., Weinstock, M.: Ieee recommended practice for speech quality measurements. IEEE Trans. Audio Electroacoust. **17**, 225–246 (1969)
15. Noisex-92. http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html,(2014)
16. Ephraim, Y., Cohen, I.: Recent advancements in speech enhancement. In: Dorf, R.C. (ed.) The Electrical Engineering Handbook, 3rd edn. CRC Press, Boca Raton (2006)
17. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. Acoustics Speech Signal Process. **32**(6), 1109–1121 (1984)
18. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoustics Speech Signal Process. **23**(2), 443–445 (1985)
19. Fort, G., Lambert-Lacroix, S.: Classification using partial least squares with penalized logistic regression. Bioinformatics **21**(7), 1104–1111 (2005)
20. Weintraub M.: A Theory and Computational Model of Auditory Monaural Sound Separation, Ph.D. Thesis, Stanford University (1985)
21. Hu, Y., Loizou, P.C.: Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Lang. Process. **16**(1), 229–238 (2008)
22. ITU-T.: Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs, Series P: Telephone Transmission Quality Recommendation P.862, ITU, 1.4 (2001)
23. ITU-R.: Recommendation BS.1534-1: Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems (2001)

24. Vincent E.: MUSHRAM: A MATLAB Interface for MUSHRA Listening Tests. Available on http://c4dm.eecs.qmul.ac.uk/downloads/,(2014)
25. Hu K., Wang D.L.: SVM-based separation of unvoiced-voiced speech in cochannel conditions. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 4545–4548 (2012)
26. Wang, Y., Han, K., Wang, D.L.: Exploring monaural features for classification-based speech segregation. IEEE Trans. Audio Speech Lang. Process. **21**(2), 270–279 (2013)

Chapter 14
REPET for Background/Foreground Separation in Audio

Zafar Rafii, Antoine Liutkus and Bryan Pardo

Abstract Repetition is a fundamental element in generating and perceiving structure. In audio, mixtures are often composed of structures where a repeating background signal is superimposed with a varying foreground signal (e.g., a singer overlaying varying vocals on a repeating accompaniment or a varying speech signal mixed up with a repeating background noise). On this basis, we present the *REpeating Pattern Extraction Technique (REPET)*, a simple approach for separating the repeating background from the non-repeating foreground in an audio mixture. The basic idea is to find the repeating elements in the mixture, derive the underlying repeating models, and extract the repeating background by comparing the models to the mixture. Unlike other separation approaches, REPET does not depend on special parameterizations, does not rely on complex frameworks, and does not require external information. Because it is only based on repetition, it has the advantage of being simple, fast, blind, and therefore completely and easily automatable.

14.1 Introduction

Figure–ground perception is the ability to segregate a scene into a foreground component (figure) and a background component (ground). In vision, the most famous example is probably the *Rubin vase*: depending on one's attention, one would perceive either a vase or two faces [19]. In auditory scene analysis [2], different cues

Z. Rafii (✉) · B. Pardo
Northwestern University, Evanston, IL, USA
e-mail: zafarrafii@u.northwestern.edu

B. Pardo
e-mail: pardo@northwestern.edu

A. Liutkus
Inria, PAROLE, Villiers-lès-Nancy, France
e-mail: antoine.liutkus@inria.fr

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_14,
© Springer-Verlag Berlin Heidelberg 2014

can be used to segregate foreground and background: loudness (e.g., the foreground signal is louder), spatial location (e.g., the foreground signal is in the center of the stereo field), or timbre (e.g., the foreground signal is a woman speaking).

Unlike fixed images (e.g., Rubin vase), audio has also a temporal dimension that can be exploited for segregation. Particularly, auditory scenes are often composed of a background component that is more stable or repeating in time (e.g., air conditioner noise or footsteps), and a foreground component that is more variable in time (e.g., a human talking or a saxophone solo). The most notable examples are probably seen (or rather heard) in music. Indeed, musical works are often organized into structures where a varying melody is overlaid on a repeating background (e.g., rapping over a repeating drum loop or playing a solo over a repeating chord progression). This implies that there should be patterns repeating in time that could be used to discriminate the background from the foreground in an auditory scene.

Repetition also appears as an exploitable cue for source separation in audio. By identifying and extracting the repeating patterns (e.g., drum loop or guitar riff), we show that it is possible to separate the repeating background from the non-repeating foreground in an audio mixture. This idea is supported by recent findings in cognitive psychology which showed that human listeners are able to segregate individual audio sources if they repeat across different mixtures, even in the absence of other cues (e.g., spatial location) and without a prior knowledge of the sources [10].

In this chapter, we present the *REpeating Pattern Extraction Technique (REPET)*, a simple method that uses repetition as a basis for background/foreground separation in audio. The basic idea is to find the repeating elements in the mixture, derive the underlying repeating models, and extract the repeating background by comparing the models to the mixture. The rest of this chapter is organized as follows.

In Sect. 14.2, we present the original REPET. The original REPET aims at identifying and extracting the repeating patterns in an audio mixture, by estimating a period of the underlying repeating structure and modeling a segment of the periodically repeating background [13, 16]. The idea can be loosely related to background subtraction, a technique used in computer vision for separating moving foreground objects from a fixed background scene in a sequence of video frames [12].

In Sect. 14.3, we present the adaptive REPET. The original REPET works well when the repeating background is relatively stable (e.g., a verse or the chorus in a song); however, the repeating background can also vary over time (e.g., a verse followed by the chorus in the song). The adaptive REPET is an extension of the original REPET that can handle varying repeating structures, by estimating the time-varying repeating periods and extracting the repeating background locally, without the need for segmentation or windowing [9].

In Sect. 14.4, we present *REPET-SIM*. The REPET methods work well when the repeating background has periodically repeating patterns (e.g., jackhammer noise); however, the repeating patterns can also happen intermittently or without a global or local periodicity (e.g., frogs by a pond). REPET-SIM is a generalization of REPET that can also handle non-periodically repeating structures, by using a similarity matrix to identify the repeating elements [14, 15].

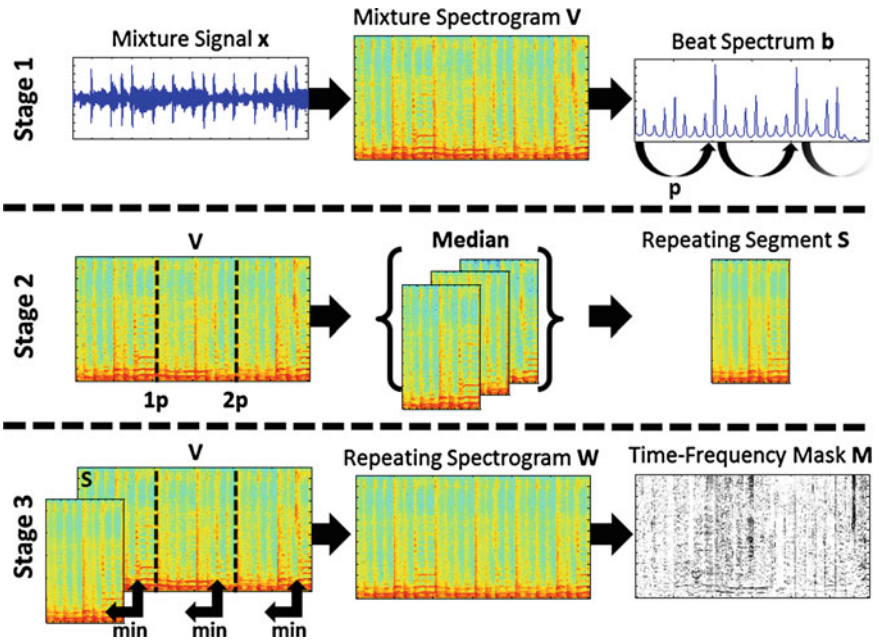


Fig. 14.1 Overview of the original REPET. *Stage 1* calculation of the beat spectrum b and estimation of a repeating period p. *Stage 2* segmentation of the mixture spectrogram V and calculation of the repeating segment model S. *Stage 3* calculation of the repeating spectrogram model W and derivation of the soft time–frequency mask M

14.2 REpeating Pattern Extraction Technique

The original REPET aims at identifying and extracting the repeating patterns in an audio mixture, by estimating a period of the underlying repeating structure and modeling a segment of the periodically repeating background [13, 16].

The original REPET can be summarized in three stages (see Fig. 14.1): (1) identification of a repeating period (see Sect. 14.2.1), (2) modeling of a repeating segment (see Sect. 14.2.2), and (3) extraction of the repeating structure (see Sect. 14.2.3).

14.2.1 Repeating Period Identification

Periodicities in a signal can be found by using the autocorrelation, which is the cross-correlation of a signal with itself. The function basically measures the similarity between a segment and a lagged version of itself over successive time lags.

Given a mixture signal x, we first compute its short-time Fourier transform (STFT) X using windows of N samples. We then derive the magnitude spectrogram V by

taking the absolute value of the elements of X, after discarding the symmetric part (i.e., the frequency channels above half the sampling frequency).

We then compute the autocorrelation over time for each frequency channel of the power spectrogram V^2 (i.e., the element-wise square of V) and obtain the matrix of autocorrelations A. We use V^2 to emphasize peaks of periodicity in A. If x is stereo, V^2 can be averaged over the channels. The overall self-similarity b of x is then obtained by taking the mean over the rows of A. We finally normalize b by dividing it by its first term (i.e., time lag 0). The calculation of b is shown in Eq. 14.1.

$$
\begin{aligned}
A(i,l) &= \frac{1}{m-l+1} \sum_{j=1}^{m-l+1} V(i,j)^2 V(i,j+l-1)^2 \\
b(l) &= \frac{1}{n} \sum_{i=1}^{n} A(i,l) \quad \text{then } b(l) = \frac{b(l)}{b(1)} \\
&\text{for } i = 1 \ldots n \quad \text{where } n = \frac{N}{2} + 1 = \text{number of frequency channels} \\
&\text{for } l = 1 \ldots m \quad \text{where } m = \text{number of time frames.}
\end{aligned}
\tag{14.1}
$$

The idea is very similar to the *beat spectrum* introduced in [7], except that no similarity matrix is explicitly calculated here, and the dot product is used in lieu of the cosine similarity. Pilot experiments showed that this method allows for a clearer visualization of the underlying periodically repeating structure in the mixture. For simplicity, we will refer to b as the beat spectrum for the remainder of this chapter.

Once the beat spectrum b is calculated, the first term which measures the similarity of the whole signal with itself (i.e., time lag 0) is discarded. If periodically repeating patterns are present in x, b would form peaks that are periodically repeating at different period rates, unveiling the underlying periodically repeating structure of the mixture, as exemplified in the top row of Fig. 14.1.

We then use a period finder to estimate the repeating period p from b. One approach can be to identify the period in the beat spectrum that has the highest mean accumulated energy over its integer multiples (see Algorithm 1 in [16]). Another approach can be to find the local maximum in a given lag range of the beat spectrum (see source codes online[1]).

The calculation of the beat spectrum b and the estimation of the repeating period p are illustrated in the top row of Fig. 14.1.

14.2.2 Repeating Segment Modeling

Once the repeating period p is estimated, we use it to segment the mixture spectrogram V into r segments of length p. We then take the element-wise median of the r

[1] http://music.eecs.northwestern.edu/research.php?project=repet

segments and obtain the repeating segment model S, as exemplified in the middle row of Fig. 14.1. The calculation of the repeating segment model S is shown in Eq. 14.2.

$$\begin{aligned} &S(i,j) = \underset{k=1\ldots r}{\text{median}}\left\{V(i, j+(k-1)p)\right\} \\ &\text{for } i = 1\ldots n \quad \text{and} \quad j = 1\ldots p \\ &\text{where } p = \text{period length} \quad \text{and} \quad r = \text{number of segments.} \end{aligned} \tag{14.2}$$

The rationale is that, if we assume that the non-repeating foreground has a sparse and varied time–frequency representation compared with the time–frequency representation of the repeating background, time–frequency bins with small deviations at period rate p would most likely represent repeating elements and would be captured by the median model. On the other hand, time–frequency bins with large deviations at period rate p would most likely be corrupted by non-repeating elements (i.e., outliers) and would be removed by the median model.

The segmentation of the mixture spectrogram V and the calculation of the repeating segment model S are illustrated in the middle row of Fig. 14.1.

14.2.3 Repeating Structure Extraction

Once the repeating segment model S is calculated, we use it to derive a repeating spectrogram model W, by taking the element-wise minimum between S and each of the r segments of the mixture spectrogram V, as exemplified in the bottom row of Fig. 14.1. The calculation of the repeating spectrogram model W is shown in Eq. 14.3.

$$\begin{aligned} &W(i, j+(k-1)p) = \min\left\{S(i,j), V(i, j+(k-1)p)\right\} \\ &\text{for } i = 1\ldots n, \quad j = 1\ldots p, \quad \text{and} \quad k = 1\ldots r \end{aligned} \tag{14.3}$$

The idea is that, if we assume that the non-negative mixture spectrogram V is the sum of a non-negative repeating spectrogram W and a non-negative non-repeating spectrogram $V - W$, then we must have $W \leq V$, element-wise.

Once the repeating spectrogram model W is calculated, we use it to derive a soft time–frequency mask M, by normalizing W by the mixture spectrogram V, element-wise. The calculation of the soft time–frequency mask M is shown in Eq. 14.4.

$$\begin{aligned} &M(i,j) = \frac{W(i,j)}{V(i,j)} \quad \text{with } M(i,j) \in [0,1] \\ &\text{for } i = 1\ldots n \quad \text{and} \quad j = 1\ldots m \end{aligned} \tag{14.4}$$

The rationale is that time–frequency bins that are likely to repeat at period rate p in V would have values near 1 in M and would be weighted toward the repeating

background. On the other hand, time–frequency bins that are not likely to repeat at period rate p in V would have values near 0 in M and would be weighted toward the non-repeating foreground.

We could further derive a binary time–frequency mask by setting time–frequency bins in M with values above a chosen threshold $t \in [0, 1]$ to 1, while the rest is set to 0. Pilot experiments showed that the estimates sound better when using a soft time–frequency mask.

The time–frequency mask M is then symmetrized and multiplied to the STFT X of the mixture x, element-wise. The estimated background signal is obtained by inverting the resulting STFT into the time domain. The estimated foreground signal is obtained by simply subtracting the background signal from the mixture signal.

The calculation of the repeating spectrogram model W and the derivation of the soft time–frequency mask M are illustrated in the bottom row of Fig. 14.1.

Experiments on a data set of song clips showed that the original REPET can be effectively applied for music/voice separation [13, 16], performing as well as two state-of-the-art methods, one based on a pitch-based method [8] and the other based on non-negative matrix factorization (NMF) and a source-filter model [3]. Experiments showed that REPET can also be combined with other methods to improve background/foreground separation; for example, it can be used as a preprocessor to pitch detection algorithms to improve melody extraction [16], or as a postprocessor to a singing voice separation algorithm to improve music/voice separation [17].

The time complexity of the original REPET is $O(m \log m)$, where m is the number of time frames in the spectrogram. The calculation of the beat spectrum takes $O(m \log m)$, since it is based on the autocorrelation which is itself based on the fast Fourier transform (FFT), while the median filtering takes $O(m)$ (Fig. 14.2).

Figure 14.2 shows an example of music/voice separation using the orginal REPET. The mixture is a female singer (foreground) singing over a guitar accompaniment (background). The guitar has a repeating chord progression that is stable along the song. The spectrograms and the mask are shown for 5 s and up to 2.5 kHz. The file is Tamy—Que Pena Tanto Faz from the task of professionally produced music recordings of the Signal Separation Evaluation Campaign (SiSEC).[2]

The original REPET can be easily extended to handle varying repeating structures, by simply applying the method along time, on individual segments or via a sliding window (see also Sect. 14.3). For example, given a window size and an overlap percentage, the local repeating backgrounds can be successively extracted using the original REPET; the whole repeating background can then be reconstructed via overlap-add [16].

Experiments on a data set of full-track real-world songs showed that this method can be effectively applied for music/voice separation [16], performing as well as a state-of-the-art method based on NMF and a source-filter model [3]. Experiments also showed that there is a trade-off for the window size in REPET: if the window is too long, the repetitions will not be sufficiently stable; if the window is too short, there will not be sufficient repetitions [16].

[2] http://sisec.wiki.irisa.fr/tikiindex.php?page=Professionally+produced+music+recordings

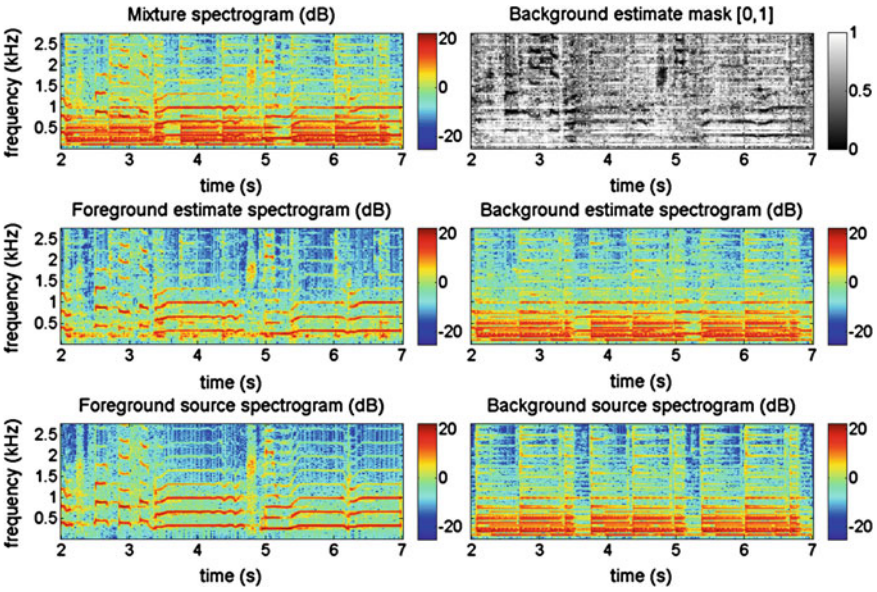


Fig. 14.2 Example of music/voice separation using the orginal REPET

14.3 Adaptive REPET

The original REPET works well when the repeating background is relatively stable (e.g., a verse or the chorus in a song); however, the repeating background can also vary over time (e.g., a verse followed by the chorus in the song). The adaptive REPET is an extension of the original REPET that can handle varying repeating structures, by estimating the time-varying repeating periods and extracting the repeating background locally, without the need for segmentation or windowing [9].

The adaptive REPET can be summarized in three stages (see Fig. 14.3): (1) identification of the repeating periods (see Sect. 14.3.1), (2) modeling of a repeating spectrogram (see Sect. 14.3.2), and (3) extraction of the repeating structure (see Sect. 14.3.3).

14.3.1 Repeating Periods Identification

The beat spectrum helps to find the global periodicity in a signal. Local periodicities can be found by computing beat spectra over successive windows. A *beat spectrogram* thus helps to visualize the variations of periodicity over time.

Given a mixture signal x, we first compute its magnitude spectrogram V (see Sect. 14.2.1). Given a window size $w \leq m$, where m is the number of time frames in

V, we then compute for every time frame j in V, the beat spectrum b_j of the local magnitude spectrogram V_j centered on j (see Sect. 14.2.1). We then concatenate the b_j's into the matrix of beat spectra B. To speed up the calculation of B, we can also use a step size s, and compute the b_j's every s frames only, and derive the rest of the values through interpolation. The calculation of B is shown in Eq. 14.5.

$$
\begin{aligned}
&V_j(i,h) = V(i, h + j - \lceil \frac{w+1}{2} \rceil) \\
&A_j(i,l) = \frac{1}{w-l+1} \sum_{h=1}^{w-l+1} V_j(i,h)^2 V_j(i,h+l-1)^2 \quad \text{and} \\
&b_j(l) = \frac{1}{n} \sum_{i=1}^{n} A_j(i,l) \\
&B(l,j) = b_j(l) \\
&\text{for } i = 1 \ldots n \quad \text{where } n = \frac{N}{2} + 1 = \text{number of frequency channels} \\
&\text{for } h = 1 \ldots w \quad \text{where } w = \text{window size} \\
&\text{for } j = 1 \ldots m \quad \text{and} \quad l = 1 \ldots m \quad \text{where } m = \text{number of time frames.}
\end{aligned}
\tag{14.5}
$$

The idea of the beat spectrogram was also introduced in [7], except that no similarity matrix is explicitly calculated here, and the dot product is used in lieu of the cosine similarity. For simplicity, we will refer to B as the beat spectrogram for the remainder of this chapter.

Once the beat spectrogram B is calculated, the first row (i.e., time lags 0) is discarded. If periodically repeating patterns are present in x, B would form horizontal lines that are periodically repeating vertically, unveiling the underlying periodically repeating structure of the mixture, as exemplified in the top row of Fig. 14.3. If variations of periodicity happen over time in x, the horizontal lines in B would show variations in their vertical periodicity.

We then use a period finder to estimate for every time frame j, the repeating period p_j from the beat spectrum b_j in B (see Sect. 14.2.1). To speed up the estimation of the p_j's, we can also use a step size s, and compute the p_j's every s frames only, and derive the rest of the values through interpolation.

The calculation of the beat spectrogram B and the estimation of the repeating periods p_j's are illustrated in the top row of Fig. 14.3.

There is no one method to compute the beat spectrum/spectrogram or to estimate the repeating period(s). We proposed to compute the beat spectrum/spectrogram using the autocorrelation and estimate the repeating period(s) using a local maximum finder (see source codes online[3]). In [9], the beat spectrogram was derived by computing the power spectrograms of the frequency channels of the power spectro-

[3] http://music.eecs.northwestern.edu/research.php?project=repet

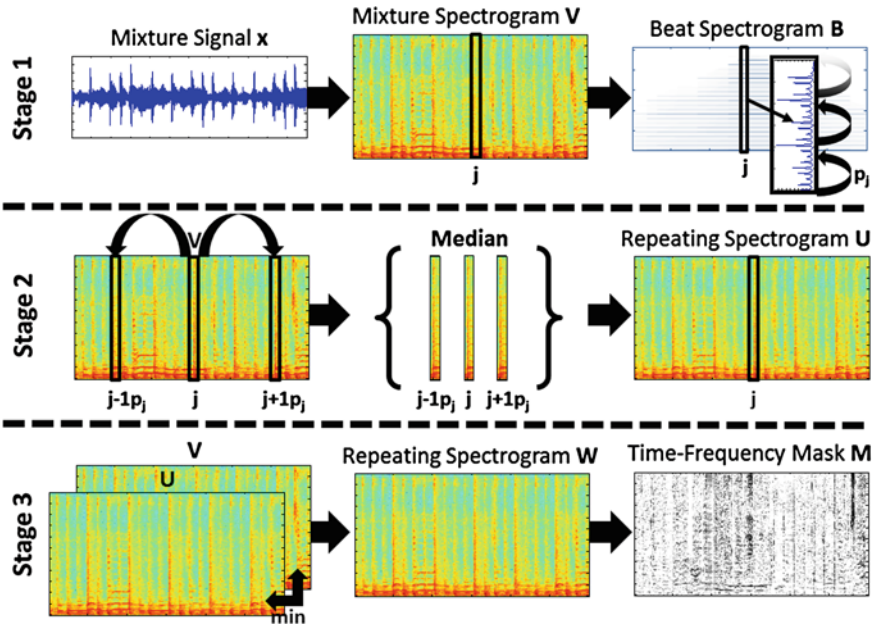


Fig. 14.3 Overview of the adaptive REPET. *Stage 1* calculation of the beat spectrogram B and estimation of the repeating periods p_j's. *Stage 2* filtering of the mixture spectrogram V and calculation of an initial repeating spectrogram model U. *Stage 3* calculation of the refined repeating spectrogram model W and derivation of the soft time–frequency mask M

gram of the mixture, and taking the element-wise mean of those power spectrograms; the repeating periods were estimated by using dynamic programming.

14.3.2 Repeating Spectrogram Modeling

Once the repeating periods p_j's are estimated, we use them to derive an initial repeating spectrogram model U. For every time frame j in the mixture spectrogram V, we derive the corresponding frame j in U by taking for every frequency channel, the median of the k frames repeating at period rate p_j around j, where k is the maximum number of repeating frames, as exemplified in the middle row of Fig. 14.3. The calculation of the initial repeating spectrogram model U is shown in Eq. 14.6.

$$\begin{aligned} &U(i,j) = \underset{l=1\ldots k}{\text{median}}\left\{V\left(i, j + (l - \lceil \frac{k}{2} \rceil) p_j\right)\right\} \\ &\text{for } i = 1\ldots n \quad \text{and} \quad \text{for } j = 1\ldots m \\ &\text{where } k = \text{maximum number of repeating frames} \\ &\text{where } p_j = \text{period length for frame } j. \end{aligned} \tag{14.6}$$

The rationale is that, if we assume that the non-repeating foreground has a sparse and varied time–frequency representation compared with the time–frequency representation of the repeating background, time–frequency bins with small deviations at their period rate p_j would most likely represent repeating elements and would be captured by the median model. On the other hand, time–frequency bins with large deviations at their period rate p_j would most likely be corrupted by non-repeating elements (i.e., outliers) and would be removed by the median model.

The filtering of the mixture spectrogram V and the calculation of the initial repeating spectrogram model U are illustrated in the middle row of Fig. 14.3.

Note that, compared with the original REPET that uses the same repeating period for each time frame of the mixture spectrogram (see Sect. 14.2), the adaptive REPET uses a different repeating period for each time frame, so that it can also handle varying repeating structures where the repeating period can also change over time.

14.3.3 Repeating Structure Extraction

Once the initial repeating spectrogram model U is calculated, we use it to derive a refined repeating spectrogram model W, by taking the element-wise minimum between U and the mixture spectrogram V, as exemplified in the bottom row of Fig. 14.3. The calculation of the refined repeating spectrogram model W is shown in Eq. 14.7.

$$W(i, j) = \min\left\{U(i, j), V(i, j)\right\} \quad \text{for } i = 1 \ldots n \quad \text{and} \quad j = 1 \ldots m \tag{14.7}$$

The idea is that, if we assume that the non-negative mixture spectrogram V is the sum of a non-negative repeating spectrogram W and a non-negative non-repeating spectrogram $V - W$, then we must have $W \leq V$, element-wise (see also Sect. 14.2.3).

Once the refined repeating spectrogram model W is calculated, we use it to derive a soft time–frequency mask M (see Sect. 14.2.3).

The calculation of the refined repeating spectrogram model W and the derivation of the soft time–frequency mask M are illustrated in the bottom row of Fig. 14.3.

Experiments on a data set of full-track real-world songs showed that the adaptive REPET can be effectively applied for music/voice separation [9], performing as well as a state-of-the-art method based on multiple median filtering of the mixture spectrogram at different frequency resolutions [5] (Fig. 14.4).

The time complexity of the adaptive REPET is $O(m \log m)$, where m is the number of time frames in the spectrogram. The calculation of the beat spectrogram takes $O(m \log m)$, since it is based on the beat spectrum (see Sect. 14.2.3), while the median filtering takes $O(m)$.

Figure 14.4 shows an example of music/voice separation using the adaptive REPET. The mixture is a male singer (foreground) singing over a guitar and drums

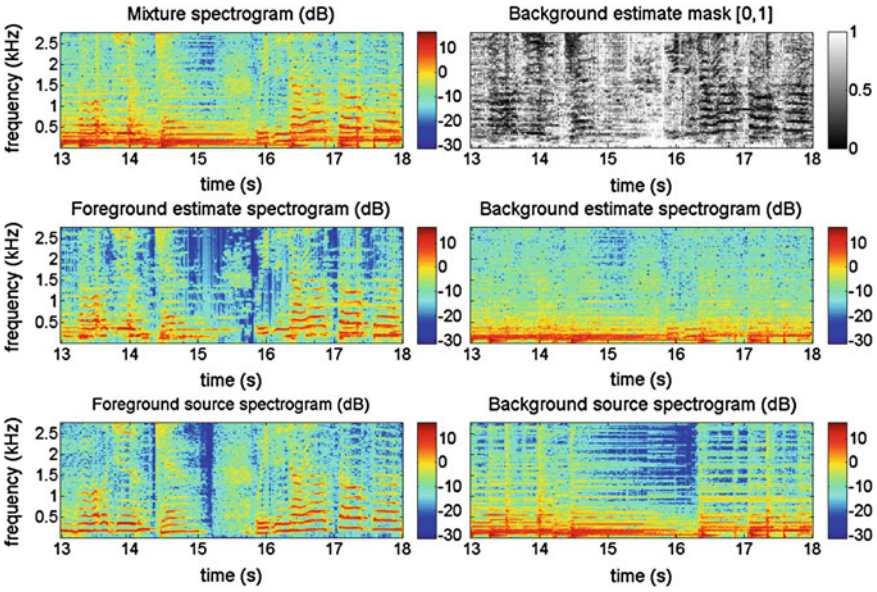


Fig. 14.4 Example of music/voice separation using the adaptive REPET

accompaniment (background). The guitar has a repeating chord progression that changes around 15 s. The spectrograms and the mask are shown for 5 s and up to 2.5 kHz. The file is Another Dreamer—The Ones We Love from the task of professionally produced music recordings of SiSEC.[4]

14.4 REPET-SIM

The REPET methods work well when the repeating background has periodically repeating patterns (e.g., jackhammer noise); however, the repeating patterns can also happen intermittently or without a global or local periodicity (e.g., frogs by a pond). REPET-SIM is a generalization of REPET that can also handle non-periodically repeating structures, by using a similarity matrix to identify the repeating elements [14, 15].

REPET-SIM can be summarized in three stages (see Fig. 14.5): (1) identification of the repeating elements (see Sect. 14.4.1), (2) modeling of a repeating spectrogram (see Sect. 14.4.2), and (3) extraction of the repeating structure (see Sect. 14.4.3).

[4] http://sisec.wiki.irisa.fr/tikiindex.php?page=Professionally+produced+music+recordings

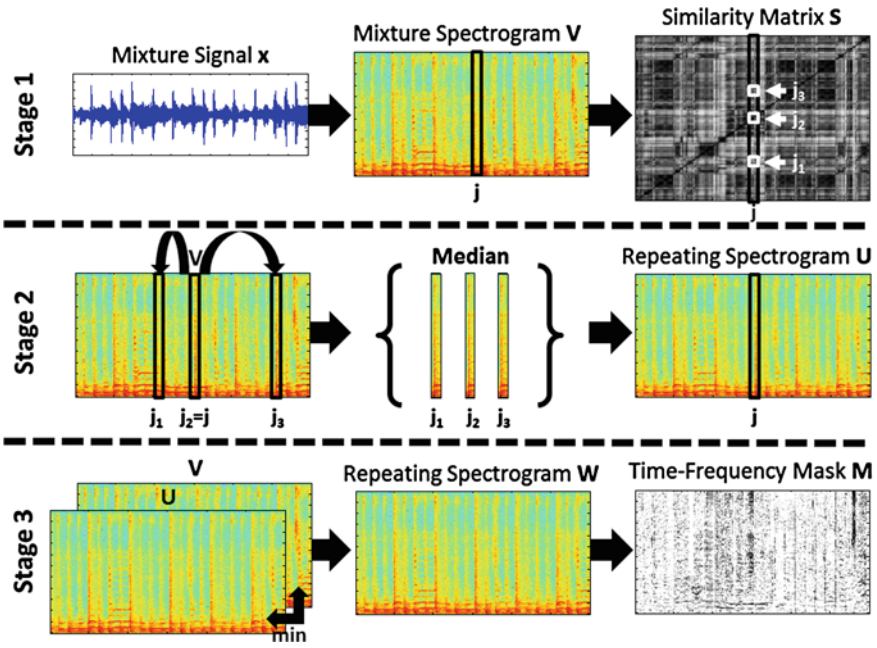


Fig. 14.5 Overview of REPET-SIM. *Stage 1* calculation of the similarity matrix S and estimation of the repeating elements j_k's. *Stage 2* filtering of the mixture spectrogram V and calculation of an initial repeating spectrogram model U. *Stage 3* calculation of the refined repeating spectrogram model W and derivation of the soft time–frequency mask M

14.4.1 Repeating Elements Identification

Repeating/similar elements in a signal can be found by using the *similarity matrix*, which is a two-dimensional representation where each point (a, b) measures the similarity between any two elements a and b of a given sequence.

Given a mixture signal x, we first compute its magnitude spectrogram V (see Sect. 14.2.1). We then compute the similarity matrix S by multiplying transposed V and V, after normalization of the columns of V by their Euclidean norm. In other words, each point (j_a, j_b) in S measures the cosine similarity between the time frames j_a and j_b of V. The calculation of the similarity matrix S is shown in Eq. 14.8.

$$S(j_a, j_b) = \frac{\sum_{i=1}^{n} V(i, j_a) V(i, j_b)}{\sqrt{\sum_{i=1}^{n} V(i, j_a)^2} \sqrt{\sum_{i=1}^{n} V(i, j_b)^2}}$$

$$\text{where } n = \frac{N}{2} + 1 = \text{number of frequency channels} \qquad (14.8)$$

$$\text{for } j_a = 1 \ldots m \quad \text{and} \quad j_b = 1 \ldots m$$

$$\text{where } m = \text{number of time frames.}$$

The idea of the similarity matrix was introduced in [6], except that the magnitude spectrogram and the cosine similarity are used here in lieu of the mel-frequency cepstrum coefficients (MFCC) and the dot product, respectively as the audio parametrization and the similarity measure. Pilot experiments showed that this method allows for a clearer visualization of the repeating structure in x.

Once the similarity matrix S is calculated, we use it to identify the repeating elements in the mixture spectrogram V. If repeating elements are present in x, S would form regions of high and low similarity at different times, unveiling the underlying repeating structure of the mixture, as exemplified in the top row of Fig. 14.5.

We then identify for every time frame j in V, the frames j_k's that are the most similar to frame j and save them in a vector of indices J_j. The rationale is that, if we assume that the non-repeating foreground has a sparse and varied time–frequency representation compared with the time–frequency representation of the repeating background, the repeating elements unveiled by the similarity matrix should be those that basically compose the underlying repeating structure.

We can add the following parameters when identifying the repeating elements in the similarity matrix: t, the minimum similarity between a repeating frame and frame j; d, the minimum distance between two consecutive repeating frames; k, the maximum number of repeating frames for a frame j.

The calculation of similarity matrix S and the estimation of the repeating elements j_k's are illustrated in the top row of Fig. 14.5.

14.4.2 Repeating Spectrogram Modeling

Once the repeating elements j_k's are identified, we use them to derive an initial repeating spectrogram model U. For every time frame j in the mixture spectrogram V, we derive the corresponding time frame j in U by taking for every frequency channel, the median of the repeating frames j_k's given by the vector of indices J_j, as exemplified in the middle row of Fig. 14.5. The calculation of the initial repeating spectrogram model U is shown in Eq. 14.9.

$$\begin{aligned} &U(i,j) = \underset{l=1\ldots k}{\text{median}}\left\{V(i, J_j(l)\right\} \\ &\text{where } J_j = j_1 \ldots j_k = \text{indices of repeating frames} \\ &\text{where } k = \text{maximum number of repeating frames} \\ &\text{for } i = 1\ldots n \quad \text{and} \quad \text{for } j = 1\ldots m. \end{aligned} \tag{14.9}$$

The rationale is that, if we assume that the non-repeating foreground has a sparse and varied time–frequency representation compared with the time–frequency representation of the repeating background, time–frequency bins with small deviations within their repeating frames j_k's would most likely represent repeating elements and would be captured by the median model. On the other hand, time–frequency

bins with large deviations within their repeating frames j_k's would most likely be corrupted by non-repeating elements (i.e., outliers) and would be removed by the median model.

The filtering of the mixture spectrogram V and the calculation of the initial repeating spectrogram model U are illustrated in the middle row of Fig. 14.5.

Note that, compared with the REPET methods that look for periodically repeating elements for each time frame of the mixture spectrogram (see Sects. 14.2 and 14.3), REPET-SIM also looks for non-periodically repeating elements for each time frame, so that it can also handle non-periodically repeating structures where repeating elements can also happen intermittently.

14.4.3 Repeating Structure Extraction

Once the initial repeating spectrogram model U is calculated, we use it to derive a refined repeating spectrogram model W, as exemplified in the bottom row of Fig. 14.5 (see Sect. 14.3.3).

Once the refined repeating spectrogram model W is calculated, we use it to derive a soft time–frequency mask M (see Sect. 14.2.3).

The calculation of the refined repeating spectrogram model W and the derivation of the soft time–frequency mask M are illustrated in the bottom row of Fig. 14.5.

Experiments on a data set of full-track real-world songs showed that REPET-SIM can be effectively applied for music/voice separation [14], performing as well as a state-of-the-art method based on multiple median filtering of the mixture spectrogram at different frequency resolutions [5], and the adaptive REPET [9].

Note that FitzGerald proposed a method very similar to REPET-SIM, except that he computed a distance matrix based on the Euclidean distance and he did not use a minimum distance parameter [4].

The time complexity of the REPET-SIM is $O(m^2)$, where m is the number of time frames in the spectrogram. The calculation of the similarity matrix takes $O(m^2)$, since it is based on matrix multiplication, while the median filtering takes $O(m)$ (Fig. 14.6).

Figure 14.6 shows an example of noise/speech separation using REPET-SIM. The mixture is a female speaker (foreground) speaking in a town square (background). The square has repeating noisy elements (passers-by and cars) that happen intermittently. The spectrograms and the mask are shown for 5 s and up to 2 kHz. The file is dev_Sq1_Co_B from the task of two-channel mixtures of speech and real-world background noise of the SiSEC.[5]

REPET-SIM can be easily implemented online to handle real-time computing, particularly for real-time speech enhancement. The online REPET-SIM simply processes the time frames of the mixture one after the other given a buffer that temporally stores past frames. For every time frame being processed, the similarity

[5] http://sisec.wiki.irisa.fr/tiki-index.php?page=Two-channel+mixtures+of+speech+and+realworld+background+noise

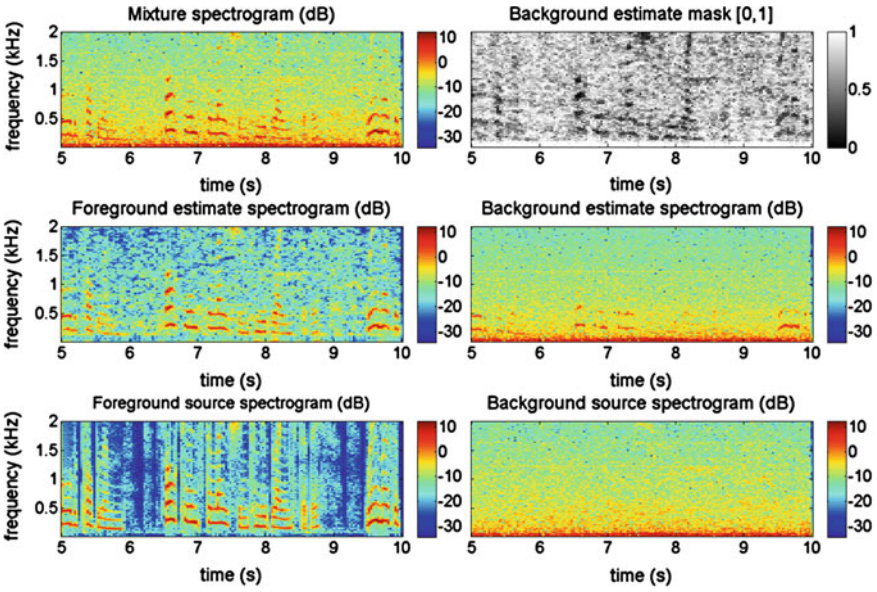


Fig. 14.6 Example of noise/speech separation using REPET-SIM

is calculated with the past frames stored in the buffer. The median is then taken between the frame being processed and its most similar frames for every frequency channel, leading to the corresponding time frame for the repeating background [15].

Experiments on a data set of two-channel mixtures of one speech source and real-world background noise showed that the online REPET-SIM can be effectively applied for real-time speech enhancement [15], performing as well as different state-of-the-art methods, one based on independent component analysis (ICA) [11], one based on the degenerate unmixing estimation technique (DUET) [20] and a minimum-statistics-based adaptive procedure [18], and one based on time differences of arrival (TDOA) and a multichannel Wiener filtering [1].

14.5 Conclusion

In this chapter, we presented *REPET*, a simple method that uses repetition as a basis for background/foreground separation in audio. In Sect. 14.2, we have presented the original REPET that aims at identifying and extracting the repeating patterns in an audio mixture, by estimating a period of the underlying repeating structure and modeling a segment of the periodically repeating background. In Sect. 14.3, we have presented the adaptive REPET, an extension of the original REPET that can directly handle varying repeating structures, by estimating the time-varying repeating periods and extracting the repeating background locally, without the need for segmentation or

windowing. In Sect. 14.4, we have presented REPET-SIM, a generalization of REPET that can also handle non-periodically repeating structures, by using a similarity matrix to identify repeating elements.

Experiments on various data sets showed that REPET can be effectively applied for background/foreground separation, performing as well as different state-of-the-art approaches, while being computationally efficient. Unlike other separation approaches, REPET does not depend on special parameterizations, does not rely on complex frameworks, and does not require external information. Because it is only based on repetition, it has the advantage of being simple, fast, blind, and therefore completely and easily automatable.

More information about REPET, including source codes, audio examples, and related publications, can be found online.[6] This work was in part supported by NSF grant number IIS-0812314.

References

1. Blandin, C., Ozerov, A., Vincent, E.: Multi-source TDOA estimation in reverberant audio using angular spectra and clustering. Signal Process. **92**(8), 1950–1960 (2012)
2. Bregman, A.S.: Auditory Scene Analysis. MIT Press, Cambridge (1990)
3. Durrieu, J.L., David, B., Richard, G.: A musically motivated mid-level representation for pitch estimation and musical audio source separation. IEEE J. Sel. Top. Sig. Process. **5**(6), 1180–1191 (2011)
4. FitzGerald, D.: Vocal separation using nearest neighbours and median filtering. In: 23rd IET Irish Signals and Systems Conference. Maynooth, Ireland (2012)
5. FitzGerald, D., Gainza, M.: Single channel vocal separation using median filtering and factorisation techniques. ISAST Trans. Electron. Signal Process. **4**(1), 62–73 (2010)
6. Foote, J.: Visualizing music and audio using self-similarity. In: 7th ACM International Conference on Multimedia, pp. 77–80. Orlando, FL, USA (1999)
7. Foote, J., Uchihashi, S.: The beat spectrum: a new approach to rhythm analysis. In: IEEE International Conference on Multimedia and Expo, pp. 881–884. Tokyo, Japan (2001)
8. Hsu, C.L., Jang, J.S.R.: On the improvement of singing voice separation for monaural recordings using the MIR-1K dataset. IEEE Trans Audio Speech Lang. Process. **18**(2), 310–319 (2010)
9. Liutkus, A., Rafii, Z., Badeau, R., Pardo, B., Richard, G.: Adaptive filtering for music/voice separation exploiting the repeating musical structure. In: 37th International Conference on Acoustics, Speech and Signal Processing. Kyoto, Japan (2012)
10. McDermott, J.H., Wrobleski, D., Oxenham, A.J.: Recovering sound sources from embedded repetition. Proc Nat. Acad. Sci. U.S.A. **108**(3), 1188–1193 (2011)
11. Nesta, F., Matassoni, M.: Robust automatic speech recognition through on-line semi blind source extraction. In: CHIME 2011 Workshop on Machine Listening in Multisource Environments, pp. 18–23. Florence, Italy (2011)
12. Piccardi, M.: Background subtraction techniques: a review. In: IEEE International Conference on Systems, Man and Cybernetics. The Hague, The Netherlands (2004)
13. Rafii, Z., Pardo, B.: A simple music/voice separation system based on the extraction of the repeating musical structure. In: 36th International Conference on Acoustics, Speech and Signal Processing. Prague, Czech Republic (2011)

[6] http://music.eecs.northwestern.edu/research.php?project=repet

14. Rafii, Z., Pardo, B.: Music/voice separation using the similarity matrix. In: 13th International Society for Music Information Retrieval. Porto, Portugal (2012)
15. Rafii, Z., Pardo, B.: Online REPET-SIM for real-time speech enhancement. In: 38th International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada (2013)
16. Rafii, Z., Pardo, B.: REpeating Pattern Extraction Technique (REPET): A simple method for music/voice separation. IEEE Trans. Audio Speech Lang. Process **21**(1), 71–82 (2013)
17. Rafii, Z., Sun, D.L., Germain, F.G., Mysore, G.J.: Combining modeling of singing voice and background music for automatic separation of musical mixtures. In: 14th International Society for Music Information Retrieval. Curitiba, PR, Brazil (2013).
18. Rangachari, S., Loizou, P.C.: A noise-estimation algorithm for highly non-stationary environments. Speech Commun. **48**(2), 220–231 (2006)
19. Rubin, E.: Synsoplevede Figurer. Gyldendal, Skive (1915)
20. Özgür Yilmaz, Rickard, S.: Blind separation of speech mixtures via time–frequency masking. IEEE Trans. Signal Process. **52**(7), 1830–1847 (2004)

Chapter 15
Nonnegative Matrix Factorization Sparse Coding Strategy for Cochlear Implants

Hongmei Hu, Guoping Li, Mark E. Lutman and Stefan Bleeck

Abstract With the development of new speech processors and algorithms, the majority of cochlear implant (CI) users benefit from their device, however, the average performance of most CI users still falls below normal hearing (NH) listeners, and speech quality and intelligibility generally deteriorate in the presence of background noise. Cochlear implants require efficient speech processing to maximize information transfer to the brain, especially in noise. Our current work is to improve the performance of CIs in noisy environments by developing new speech processing strategies. In this chapter, a nonnegative matrix factorization (NMF)-based speech coding strategy is introduced, where the speech is first transferred to the time–frequency domain via a 22-channel filter bank and the envelope in each frequency channel is extracted; and then the NMF SPARSE strategy is applied on these envelopes. The algorithm was evaluated by objective and subjective experiments, and the results were compared to the standard CI speech processing strategy (Advanced Combination Encoder, ACE).

H. Hu (✉) · G. Li · M. E. Lutman · S. Bleeck
Institute of Sound and Vibration Research, University of Southampton,
Southampton SO17 1BJ, UK
e-mail: huhongmei.hu@gmail.com; hongmei.hu@uni-oldenburg.de

G. Li
e-mail: lgp@soton.ac.uk

M. E. Lutman
e-mail: mel@isvr.soton.ac.uk

S. Bleeck
e-mail: Bleeck@gmail.com

H. Hu
Medical Physics, University of Oldenburg and Cluster of Excellence "Hearing4all",
26129 Oldenburg, Germany

H. Hu
Department of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, China

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_15,
© Springer-Verlag Berlin Heidelberg 2014

A vocoder simulation study with six participants showed that the proposed sparse NMF strategy can outperform ACE, especially at low SNR for both speech intelligibility and quality.

15.1 Introduction

Cochlear implants (CIs) are electrical devices that help to restore hearing to the profoundly deaf. The main principle of CIs is to stimulate the auditory nerve via electrodes surgically inserted into the inner ear. With the development of new speech processors and algorithms, CI users benefit more and more from CIs [1], some of them to a degree that allows them to communicate via telephone without much difficulty. However, the average speech perception performance of CI users decreases dramatically in the presence of background noise. One potential reason is the limitation of the CI electrical hearing system, such as reduced dynamic range and frequency resolution in the impaired auditory system compared to the normal hearing system, limited electrodes numbers and inaccurate channel selection methods in the current CI systems, and so on. Thus there are several bottlenecks in this electrical stimulation system, which only allows limited acoustic information to be transmitted to the auditory neurons [2]. There are currently two main ways that speech processing algorithms improve CI performance: one focuses on noise reduction by trying to enhance speech and suppress noise, such as model-based and non-model-based noise reduction algorithms [3, 4]; the other focuses on developing new cochlear coding strategies [5–7] to make good use of the limited dynamic range in the impaired auditory system. Several speech processing strategies [5, 7–11] were proposed in our group to improve the speech intelligibility in CIs to partly overcome these bottlenecks in the CI system. A nonnegative matrix factorization (NMF)-based speech processing strategy is presented in this chapter.

Nonnegative matrix factorization [12, 13] has recently attracted interest at the intersection of many scientific and engineering disciplines, such as image processing, pattern classification, blind source separation, speech enhancement, and speech separation [3, 4, 6, 14–36]. NMF is useful for transforming high-dimensional data sets into a lower dimensional space [12, 24]. Basically, given a nonnegative matrix $\mathbf{Z}$, NMF is a method to factorize $\mathbf{Z}$ into two nonnegative matrices. Motivated by the nonnegativity of the envelopes in CI channels, which results in firing of auditory neurons, a sparse coding strategy based on NMF is proposed in this chapter to improve the performance of CI users in noisy environments [9, 10]. In this application, $\mathbf{Z}$ is a matrix that consists of the envelopes of CI channels, named *envelopegram* here. Considering the computation complexity of NMF and an envisaged real-time implementation in the future, a basic NMF method with a sparse constraint [37] is applied.

15.2 Nonnegative Matrix Factorization

Given a nonnegative matrix $\mathbf{Z}$ with observations z_{ij}, NMF is a method to factorize $\mathbf{Z}$ into the NMF basis matrix $\mathbf{W}$ and component matrix $\mathbf{H}$ so that $\mathbf{Z} \approx \mathbf{WH}$. To do the factorization, a cost function $D(\mathbf{Z}||\mathbf{WH})$ is usually defined and minimized. There are many possibilities for defining the cost function and various procedures for performing the consequence minimization [13, 27, 28, 32]. Since the basic NMF allows a large degree of freedom, different types of regularizations have been used in the literature to derive meaningful factorizations for a specific application. In a general notation the following minimization is performed: $[\widehat{\mathbf{W}}, \widehat{\mathbf{H}}] = \arg\min_{\mathbf{W},\mathbf{H}}[D(\mathbf{Z}||\mathbf{WH}) + f(\mathbf{W}) + g(\mathbf{H})]$, where $f(\mathbf{W})$ and $g(\mathbf{H})$ are regularity functions for basis matrix $\mathbf{W}$ and NMF component matrix $\mathbf{H}$. The most common regularizations are motivated by the sparseness of the signal [20, 29, 30, 38] and the correlation of the signal over time [20, 33]. A sparseness constraint based on the relation between L_1 and L_2 norm is proposed in [38]. A faster algorithm is introduced in [39] to implement NMF using the same constraint and also a new sparseness constraint is given by direct controlling of the number of nonzero elements (L_0 norm). In this chapter, the Euclidean distance-based NMF (EUC-NMF) will be combined with a L_1—regularized least squares sparseness penalty function through a least absolute shrinkage and selection operator (LASSO) framework, i.e., the sparsity is measured by L_1 norm [37, 38].

15.2.1 Sparse EUC-NMF

In our application, $\mathbf{Z}$ is the envelope of CI-channels in multiple frequency bands. NMF is applied to factorize the envelope matrix into two matrices consisting respectively of NMF basis vectors $\mathbf{W}$ and the NMF components $\mathbf{H}$ that represent the activity of each basis vector over time.

As standard NMF usually provides sparseness of its components to a certain degree, an additional sparseness constraint is applied to explicitly control the sparsity of the NMF component matrix $\mathbf{H}$. In the future it might be preferable to optimize the trade-off between the sparseness and reconstruction of each individual CI user. The L_1 norm of $\mathbf{H}$ is used as the sparsity measure and the optimization algorithm proposed by Hoyer [37, 38] is applied to obtain nonnegative matrices $\mathbf{W}$ and $\mathbf{H}$.

Let $\mathbf{Z}$ denote an $N \times M$ envelope matrix of one analysis block where N and M indicate the number of channels and the number of frames, respectively. Given the nonnegative envelope matrix $\mathbf{Z}$, we aim to obtain the basis matrix $\mathbf{W}$ and component matrix $\mathbf{H}$ such that

$$D(\mathbf{Z}||\mathbf{WH}) = \tfrac{1}{2}\,\|\mathbf{Z} - \mathbf{WH}\|_2^2 + \lambda g(\mathbf{H}) \tag{15.1}$$

is minimized, under the constraints $\forall_{i,j,k} : w_{ik} \geq 0, h_{kj} \geq 0, \lambda \geq 0$, where $\mathbf{W} = \begin{bmatrix} w_{11} & \dots & w_{1K} \\ \vdots & \ddots & \vdots \\ w_{N1} & \dots & w_{NK} \end{bmatrix}_{N \times K}$, $\mathbf{H} = \begin{bmatrix} h_{11} & \dots & h_{1M} \\ \vdots & \ddots & \vdots \\ h_{K1} & \dots & h_{KM} \end{bmatrix}_{K \times M}$, $\mathbf{w}_i$ denotes the ith column of $\mathbf{W}$, $g(\mathbf{H}) = \sum_{k=1}^{K} \sum_{j=1}^{M} h_{kj}$.

The parameter λ in Eq. (15.1) is an important factor that handles the compromise between the NMF approximation and the sparsity. One contribution of this paper is to show how to choose λ heuristically to maximize the performance of the whole algorithm by objective evaluation methods and subjective perceptional psychophysical experiments.

As proposed by Hoyer [37, 38], an iterative algorithm is implemented to minimize the cost function in (15.1), in which the basis matrix $\mathbf{W}$ and the component matrix $\mathbf{H}$ are updated by gradient descent and multiplicative update rules respectively.

15.3 Applicability of NMF in the Envelope Domain

In this section, the effect of sparse NMF is shown in processing the noisy speech in the envelope domain.

15.3.1 Speech Test Materials

Speech materials were single words from 20 vocabulary sets of 80 words that were composed as rhyme tests (e.g., SUN, SUB, SUD, SOME or GET, BET, WET, YET) [40]. This material was used in a variant of the four-alternative auditory feature test (FAAF) [41]. Figure 15.1 shows the waveforms of four clean words in one set (BIN, PIN, DIN, TIN). Figure 15.2 shows the corresponding envelopes of 22 channels, the *envelopegram*, where the x-axis is the time frame bins. Since the patterns of different words are more distinguishable in the envelope domain (Fig. 15.2) than the waveform domain (Fig. 15.1) and the envelopes are nonnegative, NMF should work properly for the envelope matrix.

15.3.2 NMF Analysis on the Envelopegram

For the purpose of demonstration, NMF was applied to the whole *envelopegram* with dimension 22 * T of each word individually, where T is the number of the short-time frames in each word. Assuming sample rate is $fs = 16$ kHz, the length of the word

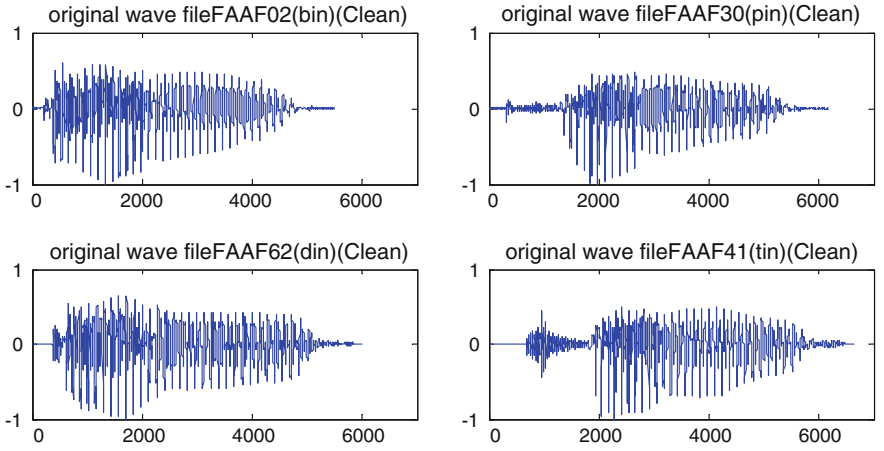


Fig. 15.1 Waveforms of four example sounds (BIN, PIN, DIN, TIN) in the time domain (This figure is reproduced from Hu et al.'s paper [10])

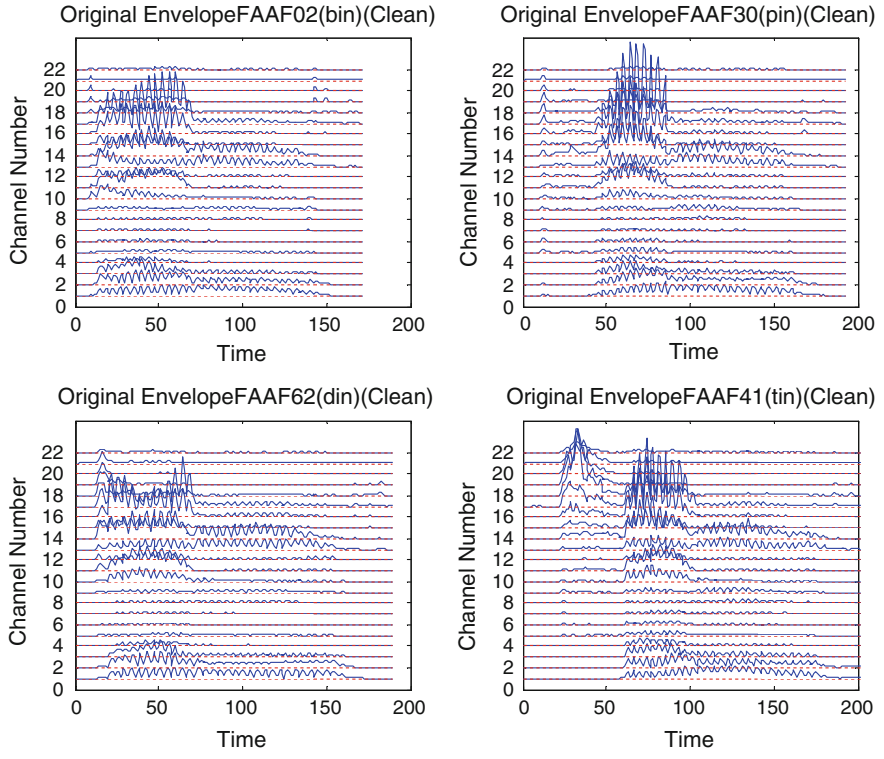


Fig. 15.2 *Envelopegram* of the corresponding CI envelopes from the four sounds shown in Fig. 15.1 (This figure is reproduced from Hu et al.'s paper [10])

is L samples, then $T \approx L/(0.25 * 128)$ with 128 samples' frame length and 75 % overlap between each frame. Five basis vectors were obtained for each *envelopegram*. Figure 15.3a shows the component matrix, which determines the activity of different basis vectors over time. Figure 15.3b shows the basis vectors for different words. Note that although the basis vectors are different for each word, the component matrices reflect similar patterns along time dimension for all the words, but not necessarily in the same order of basis number. In the following section, the effect of the number of the components in the reconstruction of the *envelopegram* is further investigated.

15.3.3 Reconstructed Envelope

Figure 15.4 shows the reconstruction of the envelopes with different components for the word "DIN". This analysis illustrates that: (1) the representation in the NMF domain is sparser than in the time domain, indicating that NMF can reconstruct speech with reduced information by choosing only few components. In this example, components 1 and 4 alone can reconstruct most of the envelope information (see Fig. 15.4 top left panel). This reflects that speech has a high degree of redundancy and only few components are necessary to reconstruct an intelligible speech signal [42, 43]. (2) The inherent correlation in the speech signal is conserved in the component matrix after applying NMF. As illustrated in the top-left panel of Fig. 15.4 and in Fig. 15.3a, the NMF components (the activity of basis vectors) tend to be continuous over time; in other words, if a basis vector is active (meaning that its corresponding coefficient is relatively large in the component matrix) at a specific time frame, it will often remain active for several time frames. This might be used as additional factor for improving iteration speed and speech reconstruction in the future. In this study the *envelopegram* is factorized by the NMF into the basis and component matrices where some components correspond to the speech source and others correspond to the noise source. The application of sparse NMF can be interpreted by assuming that the smaller NMF components correspond either to the noise basis vectors, or they do not contribute significantly to the intelligibility of speech. By normalizing each basis vector to unit norm and by applying sparseness constraint to the factorization, the small NMF components are removed and hence a sparser signal will be obtained while effectively performing noise reduction and reducing redundancy. The proposed algorithm can therefore possibly enhance the speech intelligibility by increasing the sparseness of the reconstructed signal.

15.4 Sparse NMF Strategy for CIs

Speech has a high degree of redundancy [42, 43] and the human auditory system has the ability to understand speech based on partial information or in difficult environments. Several models, like the glimpsing [43] and binary masking [44] theo-

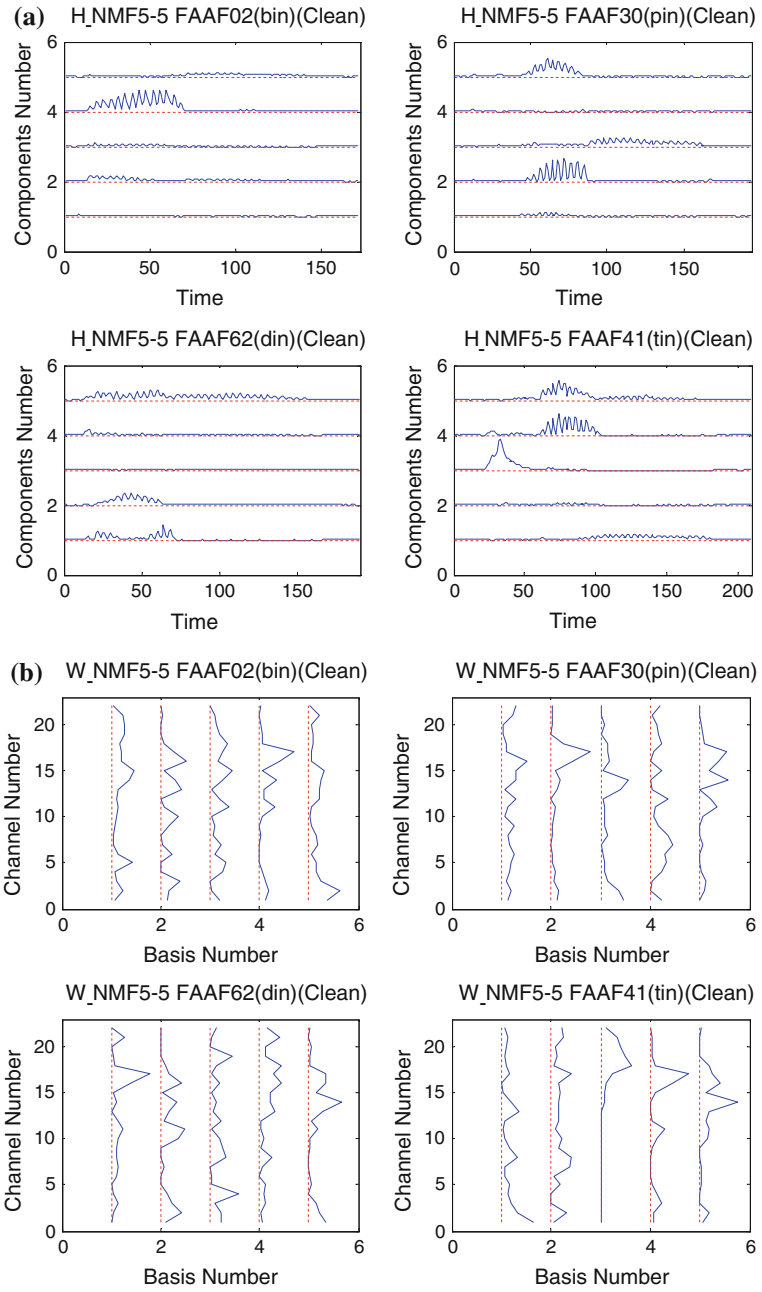


Fig. 15.3 The component matrices (**a**) and the basis matrices (**b**) of the example words "bin", "pin", "din", and "tin" (These figures are reproduced from Hu et al.'s paper [10])

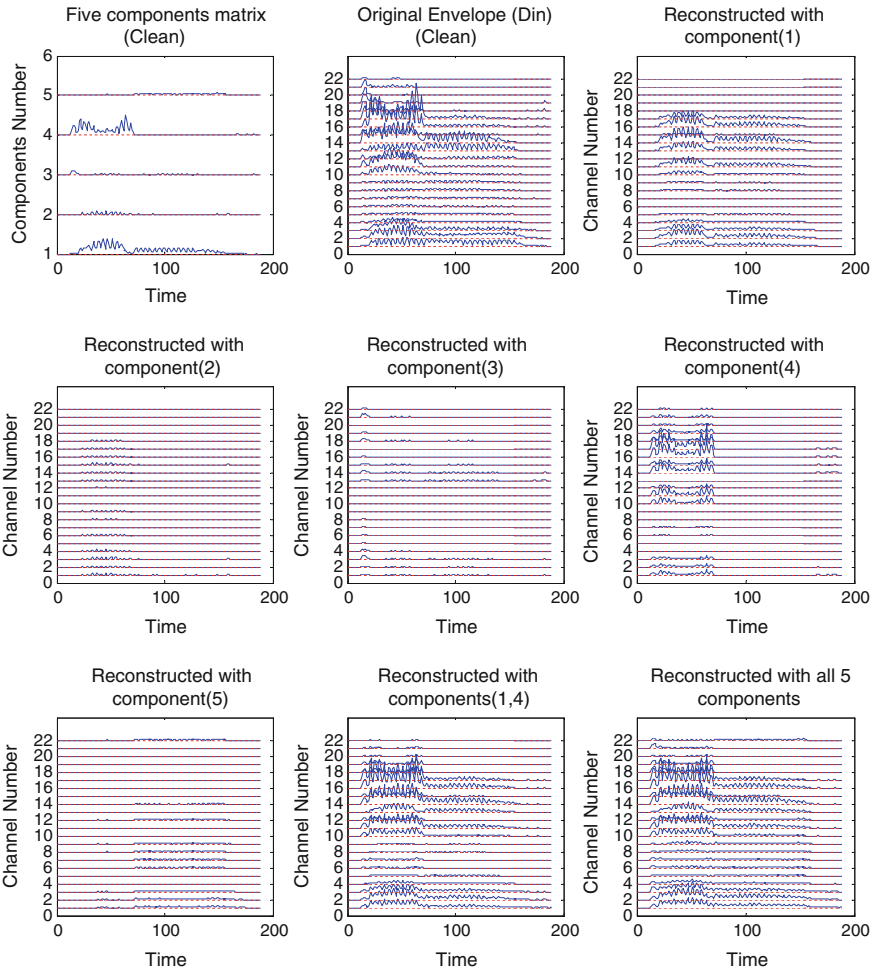


Fig. 15.4 An example of the reconstruction with different components (This figure is reproduced from Hu et al.'s paper [10])

ries have tried to explain and model this phenomenon. Existing CI strategies, such as continuous interleaved sampling (CIS), spectral peak (SPEAK), and advanced combination encoder (ACE), already take advantage of the redundancy properties of speech by selecting only few channels or only using envelope information for stimulation. Li et al. [45] demonstrated that these strategies deliver stimulation in a sparse representation of the speech. Our former work [5, 8, 45] further introduced a SPARSE strategy, in which an independent component analysis (ICA)-based sparse algorithm is applied to the spectral envelope. The redundancy properties of speech were investigated using the SPARSE strategy and tested with objective and subjective measures at various SNRs [5, 8, 45]. It was shown that the SPARSE strategy

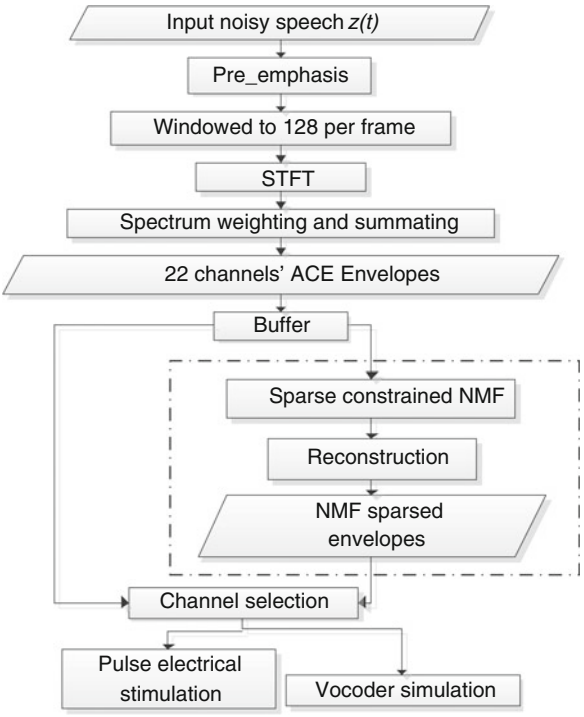


Fig. 15.5 The proposed sparse NMF strategy and ACE strategy

can improve speech intelligibility for CI users even with very limited familiarity [5, 45]. To further investigate the contribution of more efficient speech representation approaches to the performance of CIs, a sparse NMF strategy is introduced and described in detail in this section, which aims to further improve the CI performance in noisy environments.

Suppose $z(t)$ is the measured noisy speech signal, $z_{i,j}$ is the envelope-time bin in the ith channel of the jth frame, which is calculated by weighting and summating the short time Fourier transfer (STFT) spectrum according to the ACE strategy [46]. $\mathbf{Z}$ is an $N \times M$ *envelopegram*, where each column is the $N = 22$ channel envelope bin, $M = 10$ is the number of frames used in each analysis block, which is the same as the one used in [8–10] in order to provide the same input signal in each analysis block and short enough to allow real-time implementation.

Figure 15.5 shows the flowchart of the sparse NMF algorithm. The first steps are identical to the standard ACE strategy. The new blocks in the dashed frame ("sparse constrained NMF", "reconstruction", and "sparse NMF processed envelopes") indicate the changes that are made in addition to ACE. The blocks indicate steps of processing: The pre-emphasis filter attenuates low frequencies and amplifies high frequencies, to compensate for the -6 dB/octave natural slope in the long-term speech spectrum. After transforming the input speech signal into a spectrogram,

the 22-channel *envelopegram* is extracted by summing the power at frequency bins within each band. The sparse NMF algorithm is then applied to the *envelopegram* on a block-by-block basis by buffering a certain number of continuous frames in each channel. The envelopes are reconstructed from the modified sparse NMF components. Finally, appropriate channels are selected for stimulation in a real CI or to obtain a vocoder simulation that can be tested in experiments with NH listeners.

15.5 Two-Step Sparsity Level Selection Procedure

The sparsity constraint parameter λ in Eq. (15.1) controls the level of sparsity as a compromise between the NMF approximation and sparsity, which is an important factor for the sparseness and ultimately for the speech processing performance. Because it is not possible to determine an optimal value from the first principles, we developed a two-step parameter selection procedure and evaluated it in detail in [9]. This procedure works in two stages combining objective measurements with subjective experiments: in the first stage, various objective measurements are used to select a range of possible λ values; then, in the second stage, the final value of λ is determined in subjective experiments.

Vocoder simulations have been widely used as a valuable tool in CI research to simulate the perception of a CI user in experiments using NH participants [19, 47, 48]. In vocoder studies, the signal of a CI is simulated by reconstructing an acoustical signal based on the spectrum envelope [47]. Although the simulations cannot model individual CI users' performance perfectly, it has been shown that these simulations are a good model for real CI perception, specifically for speech perception, predicting the pattern and trends in performance observed in CI users [19].

Here, in order to evaluate the performance of the sparse NMF algorithms, the test data are vocoder simulated signals which are either produced by the ACE strategy (control condition) or the sparse NMF strategy. Bamford-Kowal-Bench (BKB) sentences [49] were used in both the objective and subjective experiments. BKB sentence lists are standard British speech materials with 21 lists. Each list contains 50 keywords in 16 sentences.

15.5.1 Step 1: Objective Measures and Results

Babble noise was added to the speech material at three different long-term signal-to-noise ratios (SNR) (0, 5 and 10 dB); five objective evaluation methods were selected and results were calculated for a wide range of $\lambda = [0.01 : 0.01 : 0.2]$. The results of the objective measures were used to set a smaller range of λ for a further subjective experiment.

15.5.1.1 Kurtosis

One of the most important goals of the algorithm is to sparsify the stimuli. A simple measure of sparseness is the kurtosis [5] based on Eq. (15.2):

$$K = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^4 - 3 \tag{15.2}$$

where x is the amplitude, μ is the mean, and σ is the standard deviation of the signal. For a normalized Gaussian (non-sparse) distribution with $\mu = 0$ and $\sigma = 1$, the kurtosis is by definition $K = 0$; for other signals the kurtosis may be larger than zero for a super-Gaussian or smaller than 0 for a sub-Gaussian process. If the kurtosis becomes larger, the sparseness of the stimulus increases.

Figure 15.6 shows the kurtosis values of the vocoded speech reconstructed from the ACE and sparse NMF strategies, for clean and noisy conditions at three SNR levels (0, 5, and 10 dB) respectively. The value of sparseness takes the vocoded output waveforms as a whole and calculates the kurtosis of the entire time series. The maximum kurtosis values in each noise condition are marked with red ellipses in Fig. 15.6, the corresponding λ values are defined as the optimal λ according to kurtosis. These results confirm that the outputs of sparse NMF algorithms are sparser than that of the ACE algorithm.

15.5.1.2 Normalized Covariance Metric

The normalized covariance metric (NCM) measure is similar to the speech transmission index (STI) and is a widely used measure of speech intelligibility [50]. NCM is based on the covariance between the input and output envelope signals. For computing the NCM measure, the stimulus was first bandpass filtered into k bands spanning the signal bandwidth. The envelope of each band was computed using the Hilbert transform, anti-aliased using low-pass filtering, and then down-sampled to f_d Hz, thereby limiting the envelope modulation frequencies to $0 - f_d$ Hz. The NCM measure is expected to correlate highly with the intelligibility of vocoded speech due to the similarities in the NCM calculation and CI processing strategies; both use information extracted from the envelopes in a number of frequency bands while discarding fine-structure information [51, 52].

15.5.1.3 Short-Time Objective Intelligibility

The short-time objective intelligibility (STOI) is a recent improvement of traditional objective measures [53]. The STOI calculation is based on a correlation coefficient between the temporal envelopes of the clean and degraded speech in short-time overlapping segments. The input of STOI is the clean and the processed signal in

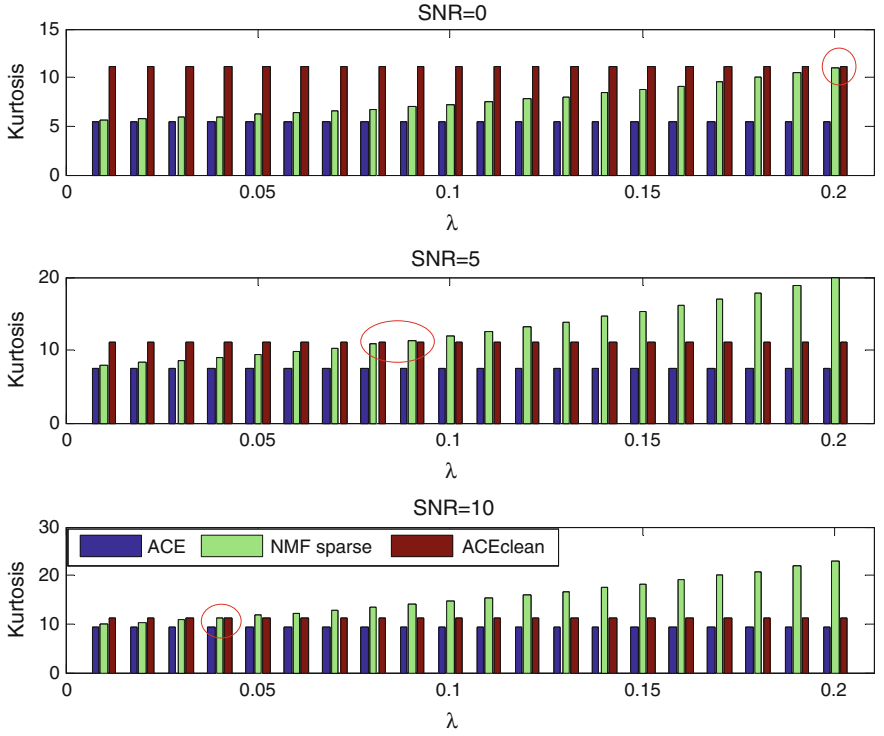


Fig. 15.6 Kurtosis of speech processed by three strategies at three SNR levels of 0, 5, and 10 dB

the time domain, and the output is a scalar value that has a monotonic relation with the average intelligibility of the processed signal [53]. In our case, the first input is the vocoded sparsified signal and the second is the corresponding vocoded signal of the clean speech.

15.5.1.4 Segmental SNR and SNR

Both SNR and frame-based segmental SNR are used as objective measures of speech quality as follows [54–56]:

$$\text{Segsnr} = \frac{10}{M}\sum_{m=0}^{M-1}\log\frac{\sum_{n=Nm}^{Nm+N-1} s_c^2(n)}{\sum_{n=Nm}^{Nm+N-1}[s_n(n)-s_c(n)]^2} \tag{15.3}$$

where s_n and s_c denote processed noisy and clean speech, respectively, M is the frame number, N is the frame length in points per frame.

Figure 15.7 shows the NCM, STOI, segmental SNR (Segsnr), and SNR of speech at two SNR levels (5 and 10 dB) processed by ACE, sparse NMF. The maximum objective measure values in each noise condition were marked with red ellipses in Fig. 15.7, the corresponding λ values were defined as the optimal λ according to different objective measures.

Results shown in Figs. 15.6 and 15.7 demonstrate that for different objective evaluations, different values of λ should be set to achieve optimal results. The question is which objective evaluation method is best suited to predict an optimal λ value. To answer this question, first a subjective experiment was performed with NH listeners to determine the relation between objective evaluation results and subjective intelligibility experiment results.

15.5.2 Step 2: Subjective Speech Reception Threshold Test

The speech reception threshold (SRT) is the established method for measuring speech perception and has been shown to faithfully represent speech perception ability [57]. A two-up one-down adaptive procedure was used to find the SNR required for 70.7 % correct recognition in each condition. A sentence was classified to be correctly identified when at least two keywords were correctly repeated. The noise was babble noise, and the SNR level varied adaptively with a 1 dB step size. To enable comparison with subjective results, speech recognition results were assessed using a method to provide a speech-in-noise threshold in dB [58]. Two psychophysical experiments were performed: first a "pilot" SRT experiment to determine an optimal λ range, and finally a "formal" SRT experiment. All participants were normal hearing (within -10 to 15 dB HL) native English speakers with no previous experience of BKB sentence lists. All experiments were performed in a sound-isolated room with the sounds presented binaurally through a Sennheiser HDA 200 headphone with a Creek OBH-21SE headphone amplifier. The vocoded BKB sentence lists were presented by a female speaker. The sample rate was 16 kHz. The participants were trained with vocoded clean BKB sentences to familiarize with the test procedure. All experiments were approved by the Human Experimentation Safety and Ethics Committee, Institute of Sound, and Vibration Research, University of Southampton, UK.

15.5.2.1 Pilot SRT Experiment

In the pilot SRT experiment, six NH participants (four males, two females, and aged 18–26) participated. Four sparse NMF strategies with different λ were tested. Table 15.1 shows different test conditions. In conditions 1, 2, and 3, the vocoded sounds were reconstructed from the NMF envelope with $\lambda = 0.08$ (called "NMF008"), 0.13 ("NMF013"), and 0.18 ("NMF018"), respectively, for all SNRs (from -1 to 10 dB in the SRT adaptive procedure, with 1 dB step size, i.e., $[-1 : 1 : 10]$). In condition 4 ("NMFcomb"), different λ were applied within the SNR range, e.g.,

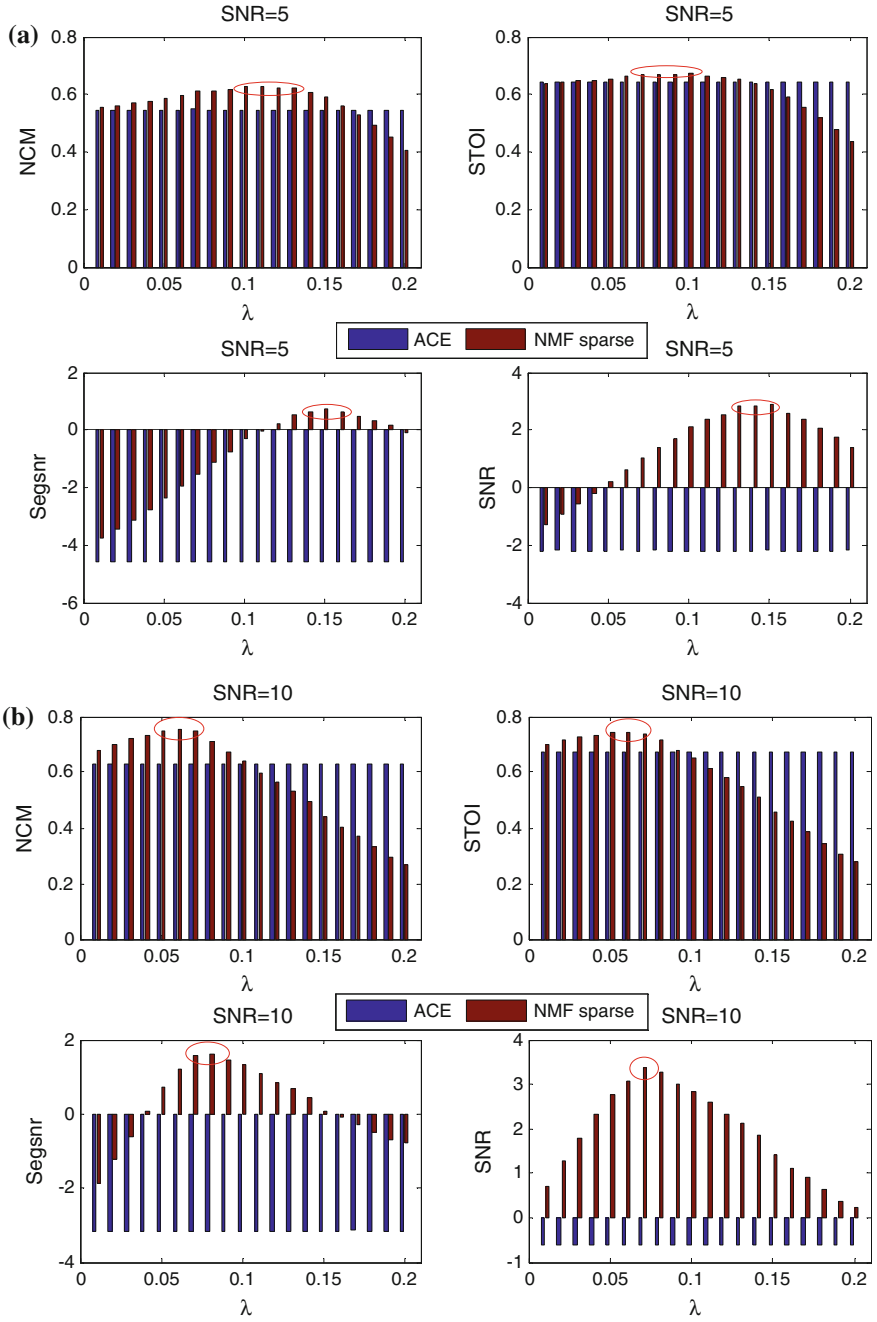


Fig. 15.7 NCM, STOI, Segsnr, and SNR of speech processed by different strategies at two SNR levels of 5 and 10 dB

Table 15.1 The pilot subjective experiment conditions

Condition	Strategy	λ	SNR(dB)
1	NMF008	0.08	−1 : 1 : 10
2	NMF013	0.13	−1 : 1 : 10
3	NMF018	0.18	−1 : 1 : 10
4	NMFcomb	0.08	7, 8, 9, 10
		0.13	3, 4, 5, 6
		0.18	−1, 0, 1, 2

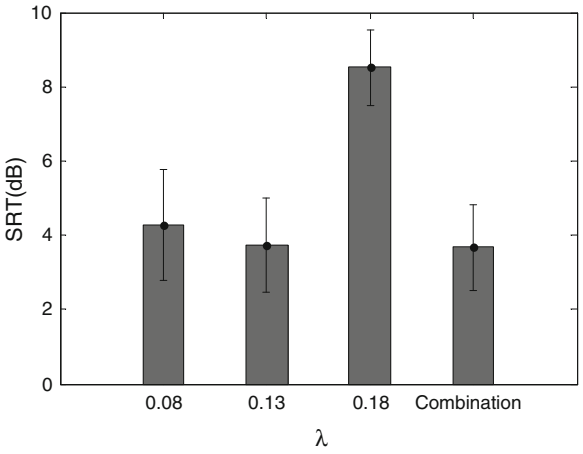


Fig. 15.8 The pilot subjective experiment results for all four conditions. The vertical axis is the SRT in dB, the lower SRT means the better performance. The error bars indicate 1 standard deviation

$\lambda = 0.08$ when SNR was 7–10 dB, $\lambda = 0.13$ when SNR was 3–6 dB, and $\lambda = 0.18$ when SNR was between −1 and 2 dB. These values were obtained according to the SNR-dependent optimization value of λ shown in Figs. 15.6 and 15.7.

Figure 15.8 shows the SRT results of all four conditions from six participants in the pilot SRT experiments (the higher SRT value means the worse intelligibility performance). A one-way repeated-measures analysis of variance (ANOVA) with Fisher's least significant difference (LSD) post hoc test shows that the effects of different λ are significant [$F(3, 15) = 19.033$, $p < 0.001$]. NH listeners perform significantly worse in condition 3 compared to all other conditions. Although performance in condition 2 and condition 4 is higher than condition 1, this difference is not significant.

The optimized λ for better SRT therefore probably lies between 0.08 and 0.13, which means both NCM and STOI, especially NCM can in some instance predict the performance of intelligibility for noise vocoded speech in such cases.

Table 15.2 The "formal" subjective experiment conditions

Condition	Strategy	λ
1	ACE	
2	NMF008	0.08
3	NMF010	0.10
4	NMF013	0.13

Table 15.3 The paired-compared win/loss number

Strategy	ACE	NMF008	NMF010
NMF008	1:9		
NMF010	8:2	9:1	
NMF013	9:1	8:1, 1:1	5:5

15.5.2.2 "Formal" SRT Experiment

After narrowing down the range of optimal λ in the first experiment to λ between 0.08 and 0.13, we further evaluated the sparse NMF strategy within this range in a speech intelligibility experiment. This SRT experiment test was also designed to compare the sparse NMF strategies with the ACE strategy.

Ten new NH (six males, four females, and aged 18–26) were recruited. All participants were native English speakers with no previous experience of BKB sentence lists. The ACE strategy and three NMF strategies with different sparsity conditions were tested. Table 15.2 shows the description of different conditions. In condition 1, the ACE strategy was used in conditions 2–4, the vocoded sound was reconstructed from NMF envelopes with $\lambda = 0.08$ ("NMF008"), 0.10 ("NMF010"), and 0.13 ("NMF013") for all SNR (from -1 to 10 dB in the SRT adaptive procedure). The procedure was identical to the pilot experiment.

The result reveals a large individual performance difference. To understand averaged results, a paired "win/loss" numbers analysis is shown in Table 15.3: A "win" is marked when one strategy produces better results compared to another. Table 15.3 demonstrates that the ACE strategy outperforms the NMF008 strategy for 9 out of 10 subjects (9:1), while both NMF010 and NMF013 strategies outperform the ACE strategy (8:2) and there is no difference between NMF010 and NMF013. On average, there was a 0.74 dB improvement for NMF010 and a 0.92 dB improvement for NMF013 compared to the ACE strategy. A one-way repeated-measures ANOVA with LSD post hoc test shows that the differences between the strategies are significant [$F(3, 27) = 7.13$, $p < 0.05$]. The following comparisons are significantly different: NMF010 $<$ ACE($p = 0.037$), NMF013 $<$ ACE($p = 0.012$), NMF010 $<$ NMF008($p = 0.003$) and NMF013 $<$ NMF008($p = 0.006$).

15.6 Subjective Quality Experiments with the Two-Step Selected NMF Sparsity Level

To further evaluate the sparse NMF strategies with selected sparsity constraint parameters $\lambda = 0.1$ and 0.13, subjective quality experiments in different SNRs were performed to compare the performance of the NMF010 and NMF013 sparse strategies with the ACE strategy.

15.6.1 Material and Methods

Five NH subjects were recruited (all male, aged between 20 and 26 years) in this experiment. Three conditions were tested at three different SNRs (0, 5, and 10 dB). Four speech conditions were also compared, they were (a) ACE processed vocoded clean speech ("ACE clean"), (b) ACE processed vocoded noisy speech ("ACE noisy"), (c) and (d) sparse NMF processed noisy speech with $\lambda = 0.1$ and 0.13 respectively ("NMF010", "NMF013"). Each speech group consisted of the same seven individual BKB sentences with the corresponding SNR and the named processing strategies, which were vocoded and concatenated into one long presentation as testing speech.

A multi-comparison preference rating test was introduced to evaluate the quality of the speech, in which the global speech quality is evaluated for each session, i.e., each SNR (0, 5, and 10 dB). Participants were asked to rate the presentations by giving a score between 0 and 100 according to their perceived general quality (higher = better). The participants were allowed to repeat the speech stimuli as often as they wanted and they could give identical scores when unable to rate differently. The aim of this experiment was to give an indication of whether the sparse NMF strategy can improve the quality of the noisy speech and which sparsity level was preferred.

15.6.2 Subjective Quality Experiment Results

The results from the quality experiment are shown in Fig. 15.9 where the vertical axis is the quality score from 0 to 100. Results from all participants are shown. The different filling patterns correspond to the different coding strategies.

Figure 15.9a shows the overall speech quality test results for each individual participant in different SNR sessions. It shows that all subjects rate the ACE clean speech highest quality. More interestingly, all subjects prefer NMF processed speech to the corresponding ACE noisy speech in conditions 0 and 5 dB, and three out of five prefer at 10 dB SNR. Figure 15.9b shows the range of scores in all conditions. A one-way repeated-measures ANOVA with Fisher's LSD posthoc test shows that the effects of the different strategies on the quality performance are significant $[F(3, 12) = 38.3, p < 0.001]$. Both sparse NMF010 and NMF013 significantly

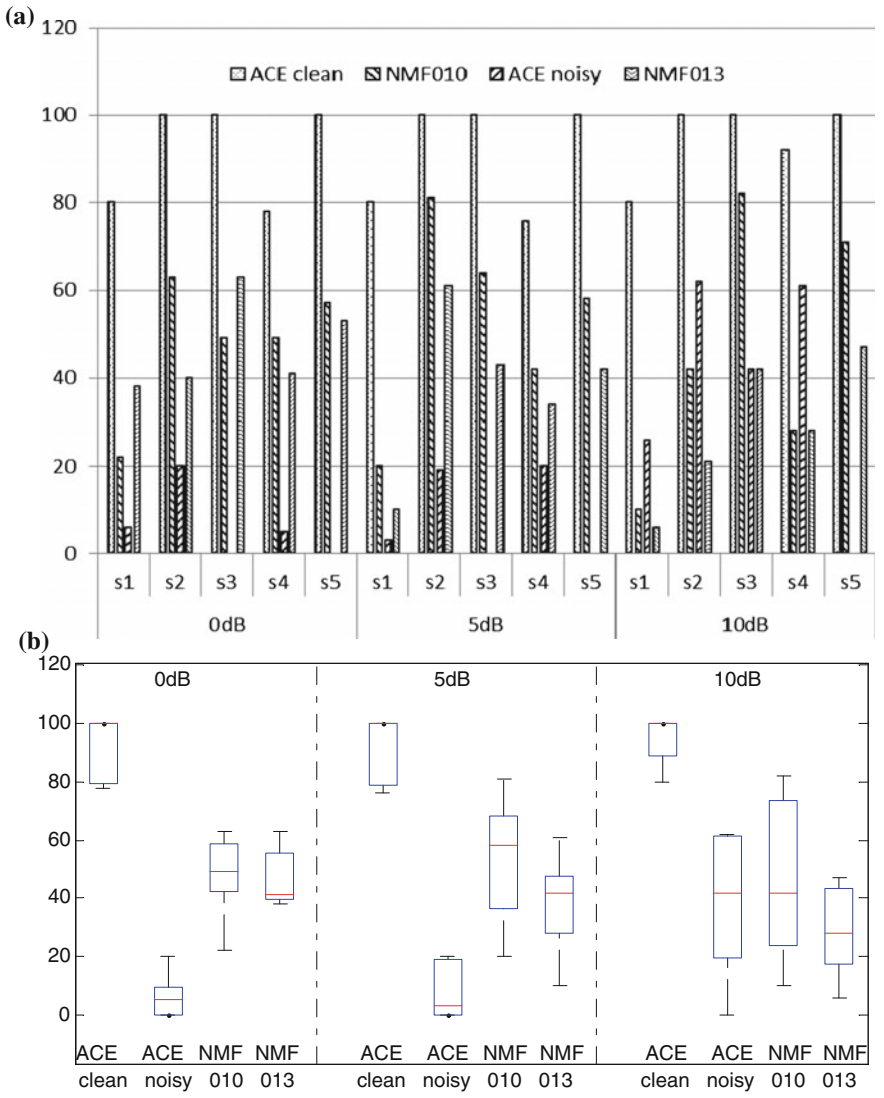


Fig. 15.9 Quality experiment scores of individual subjects (a) and boxplots (b)

improved the quality of vocoded speech compared to the noisy ACE strategy for 0 and 5 dB ($p < 0.05$), but there is no statistically significant improvement for 10 dB. There is no significant difference between NMF010 and NMF013 for 0 and 10 dB, while NMF010 significantly ($p < 0.05$) outperforms NMF013 at 5 dB. These results indicate that the sparse NMF speech processing strategy is able to improve both speech intelligibility and quality for CIs and further evaluation in CI users is necessary.

15.7 Conclusions

A novel CI coding strategy is proposed in which sparse NMF is applied to the envelopes of CI-channels in order to improve the performance of CIs in noisy environment. As demonstrated, the signal *envelopegram* is sparsified in the NMF domain, and only a few basis vectors are active for each time frame. A two-step parameter selection procedure was developed to choose the sparsity constraint parameter by combining objective measures and SRT. Subjective listening experiments demonstrated that the proposed sparse NMF strategy can outperform the existing ACE strategy when using appropriate sparsity, especially at low SNR. This is evident for both speech intelligibility and quality, at least as far as can be gauged from NH listeners and noise vocoder CI simulation. Speech intelligibility in the sparse NMF strategy benefits from noise reduction more than ACE, because only the key parts of the signal are chosen for reconstruction. However, at high SNRs, speech quality becomes more important and distortion caused by over sparsification may increase listening effort. The sparse NMF strategy shows promise for achieving better speech perception for CI users. To further improve the performance of the proposed sparse NMF, it is suggested to combine the sparse NMF algorithm with an SNR-dependent sparsity constraint.

Acknowledgments This work was funded by the European Commission within the Marie Curie ITN AUDIS (grant PITNGA-2008-214699) and Cochlear Europe when Hongmei Hu worked in the Institute of Sound and Vibration Research, University of Southampton, UK. It is currently partly supported by EU FP7 under ABCIT grant agreement (No. 304912). The authors thank Professor Arne Leijon, NasserMohammadiha, and Jalil Taghia for their collaboration on part of the work during her visit in Sound and Image Processing Lab, KTH, Stockholm, Sweden. The authors thank Cochlear Europe for providing software for the ACE algorithm and all the participants for their hard work.

References

1. Wilson, B., Dorman, M.: The surprising performance of present-day cochlear implants. IEEE Trans. Biomed. Eng. **54**(6), 969–972 (2007)
2. Greenberg, S., Ainsworth, W., Popper, A., Fay, R.: Speech Processing in the Auditory System: An Overview. Springer Handbook of Auditory Research, vol. 18, pp. 1–62. Springer, New York (2004)
3. Hussain, A., Chetouani, M., Squartini, S., Bastari, A., Piazza, F.: Nonlinear Speech Enhancement: An Overview. Progress in Nonlinear Speech Processing, vol. 4391, pp. 217–248. Springer, New York (2007)
4. Roberts, W., Ephraim, Y., Lev-Ari, H.: A Brief Survey of Speech Enhancement, chap. 20, pp. 1–11. CRC Press, Boca Raton (2006)
5. Li, G.: Speech perception in a sparse domain. Ph.D. Dissertation, University of Southampton (2008)
6. Nie, K., Drennan, W., Rubinstein, J.: Cochlear Implant Coding Strategies and Device Programming, chap. 33, pp. 389–394. People's Medical Publishing House, Shelton (2009)
7. Li, G., Lutman, M.: Sparse stimuli for cochlear implants. In: 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, 25–29 Aug 2008

8. Hu, H., Li, G., Chen, L., Sang, J., Lutman, M., Bleeck, S.: Enhanced sparse speech processing strategy for cochlear implants. In: 19th European Signal Processing Conference (EUSIPCO 2011), Barcelona, Spain, pp. 491–495, Aug 29–Sept 2 2003
9. Hu, H., Mohammadiha, N., Taghia, J., Leijon, A., Lutman, M., Wang, S.: Spasity level in a non-negative matrix factorization based speech strategy in cochlear implants. In: 19th European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania, pp. 2432–2436, 27–31 Aug 2012
10. Hu, H., Sang, J., Lutman, M., Bleeck, S.: Non-negative matrix factorization on the evelope matrix in cochlear implant. In: 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, Canada, pp. 7790–7794, 26–31 May 2012
11. Hu, H., Krasoulis, A., Lutman, M., Bleeck, S.: Development of a real time sparse non-negative matrix factorization module for cochlear implants by using xPC target. Sensors **13**, 13861–13878 (2013)
12. Lee, D., Seung, H.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788–791 (1999)
13. Lee, D., Seung, H.: Algorithms for non-negative matrix factorization. In: 25th Annual Conference on Neural Information Processing Systems, NIPS 2011. MIT Press, pp. 556–562 (2001)
14. Berouti, M., Schwartz, R., Makhoul, J.: Enhancement of speech corrupted by acoustic noise. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1979), Washington, DC, USA, pp. 208–211, 2–4 Apr 1979
15. Ephraim, Y., Malah, D.: Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. **32**(6), 1109–1121 (1984)
16. Lockwood, P., Boudy, J., Blanchet, M.: Non-linear spectral subtraction (nss) and hidden markov models for robust speech recognition in car noise environments. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1992), San Francisco, CA, USA, vol. 1, pp. 265–268, 23–26 Mar 1992
17. Gannot, S., Burshtein, D., Weinstein, E.: Iterative and sequential kalman filter-based speech enhancement algorithms. IEEE Trans. Speech Audio Process. **6**(4), 373–385 (1998)
18. Martin, R.: Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Speech Audio Process. **9**(5), 504–512 (2001)
19. Loizou, P.C.: Speech processing in vocoder-centric cochlear implants, 2006th edn, vol. 26, pp. 109–143. Karger, Basel (2006)
20. Virtanen, T.: Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. IEEE Trans. Audio Speech Langn. Process. **15**(3), 1066–1074 (2007)
21. Hendriks, R., Gerkmann, T.: Noise correlation matrix estimation for multi-microphone speech enhancement. IEEE Trans. Audio Speech Lang. Process. **20**(1), 223–233 (2012)
22. Smaragdis, P., Brown, J.: Non-negative matrix factorization for polyphonic music transcription. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 177–180 (2003)
23. Spratling, M.: Learning image components for object recognition. J. Mach. Learn. Res. **7**, 793–815 (2006)
24. Potluru, V., Calhoun, V.: Group learning using contrast nmf : application to functional and structural mri of schizophrenia. In: IEEE International Symposium on Circuits and Systems (ISCAS 2008), Washington, DC, USA, pp. 1328–1331, 18–21 May 2008
25. Shashanka, M., Raj, B., Smaragdis, P.: Probabilistic latent variable models as nonnegative factorizations. Comput. Intell. Neurosci. **2008**, 9 (2008)
26. Mohammadiha, N., Gerkmann, T., Leijon, A.: A new linear mmse filter for single channel speech enhancement based on nonnegative matrix factorization. In: 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, vol. 16–19, pp. 45–48, Oct 2011
27. Cichocki, A., Zdunek, R., Amari, S.: New algorithms for non-negative matrix factorization in applications to blind source separation. In: 2006 IEEE International Conference on Acoustics,

Speech and Signal Processing, ICASSP 2006 Proceedings, vol. 5, p. V. Toulouse, France, 14–19 May 2006
28. Zdunek, R., Cichocki, A.: Fast nonnegative matrix factorization algorithms using projected gradient approaches for large-scale problems. Comput. Intell. Neurosci. **2008**, 13 (2008)
29. Rennie, S., Hershey, J., Olsen, P.: Efficient model-based speech separation and denoising using non-negative subspace analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, CA, USA, 30 Mar–4 Apr 2008, pp. 1833–1836 (2008)
30. Schmidt, M.: Single-channel source separation using non-negative matrix factorization. Ph.D. Dissertation, Technical University of Denmark, Denmark (2008)
31. Cichocki, A., Zdunek, R., Phan, A., Amari, S.: Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. Wiley, Chichester (2009)
32. Fevotte, C., Bertin, N., Durrieu, J.: Nonnegative matrix factorization with the itakura-saito divergence: with application to music analysis. Neural Comput. **21**(3), 793–830 (2009)
33. Mysore, G., Smaragdis, P., Raj, B.: Non-negative hidden markov modeling of audio with application to source separation. In: Proceedings of the 9th International Conference on Latent Variable Analysis and Signal Separation, ser. LVA/ICA'10, pp. 140–148. Springer, Berlin, Heidelberg (2010)
34. Wang, J., Lai, S., Li, M.: Improved image fusion method based on nsct and accelerated nmf. Sensors **12**(5), 5872–5887 (2012). http://www.mdpi.com/1424-8220/12/5/5872
35. Wang, W.: Squared euclidean distance based convolutive non-negative matrix factorization with multiplicative learning rules for audio pattern separation. In: Proceedings of the 7th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2007), Cairo, Egypt, 15–18 Dec 2007, pp. 347–352 (2007)
36. Wang, W., Cichocki, A., Chambers, J.: A multiplicative algorithm for convolutive non-negative matrix factorization based on squared euclidean distance. IEEE Trans. Signal Process. **57**(7), 2858–2864 (2009)
37. Hoyer, P.: Non-negative sparse coding. In: Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing, Valais, Switzerland, 4–6 Sept 2002, pp. 557–565 (2002)
38. Hoyer, P.: Non-negative matrix factorization with sparseness constraints. J. Mach. Learn. Res. **5**, 1457–1469 (2004)
39. Morup, M., Madsen, K., Hansen, L.: Approximate l0 constrained non-negative matrix and tensor factorization. In: IEEE International Symposium on Circuits and Systems, ISCAS 2008, Washington, DC, USA, 18–21 May 2008, pp. 1328–1331 (2008)
40. Lutman, M., Clark, J.: Speech identification under simulated hearing-aid frequency response characteristics in relation to sensitivity, frequency resolution, and temporal resolution. J. Acoust. Soc. Am. **80**(4), 1030–1040 (1986)
41. Foster, J., Haggard, M.: Faaf—an efficient analytical test of speech perception. In: Proceedings of the Institute of Acoustics, pp. 1A3: 9–12
42. Kasturi, K., Loizou, P., Dorman, M., Spahr, T.: The intelligibility of speech with "holes" in the spectrum. J. Acoust. Soc. Am. **112**(3), 1102–1111 (2002)
43. Cooke, M.: A glimpsing model of speech perception in noise. J. Acoust. Soc. Am. **119**, 1562–1573 (2006)
44. Wang, D., Kjems, U., Pedersen, M., Boldt, J., Lunner, T.: Speech intelligibility in background noise with ideal binary time–frequency masking. J. Acoust. Soc. Am. **125**(4), 2336–2347 (2009)
45. Li, G., Lutman, M., Wang, S., Bleeck, S.: Relationship between speech recognition in noise and sparseness. Int. J. Audiol. **51**(2), 75–82 (2012)
46. Patrick, J., Busby, P., Gibson, P.: The development of the nucleus freedom cochlear implant system. Trends Amplif **10**(4), 175–200 (2006)
47. Shannon, R., Zeng, F., Kamath, V., Wygonski, J., Ekelid, M.: Speech recognition with primarily temporal cues. Science **270**(5234), 303–304 (1995)

48. Stone, M., Fullgrabe, C., Moore, B.: Benefit of high-rate envelope cues in vocoder processing: effect of number of channels and spectral region. J. Acoust. Soc. Am. **124**(4), 2272–2282 (2008)
49. Bench, J., Kowal, A., Bamford, J.: The bkb (bamford-kowal-bench) sentence lists for partially-hearing children. Br J Audiol **13**(3), 108–12 (1979)
50. Steeneken, H.: A physical method for measuring speech transmission quality. J. Acoust. Soc. Am. **67**(1), 318 (1980)
51. Chen, F., Loizou, P.: Analysis of a simplified normalized covariance measure based on binary weighting functions for predicting the intelligibility of noise-suppressed speech. J. Acoust. Soc. Am. **128**(6), 3715–3723 (2010)
52. Goldsworthy, R., Greenberg, J.: Analysis of speech-based speech transmission index methods with implications for nonlinear operations. J. Acoust. Soc. Am. **116**(6), 3679–3689 (2004)
53. Taal, C., Hendriks, R., Heusdens, R., Jensen, J.: An algorithm for intelligibility prediction of time and frequency weighted noisy speech. IEEE Trans. Audio Speech Lang. Process. **19**(7), 2125–2136 (2011)
54. Hansen, J., Pellom, B.: An effective quality evaluation protocol for speech enhancement algorithms. In: Proceedings of the International Conference on Speech and Language Processing, vol. 7, pp. 2819–2822, Nov 30–Dec 4 (1998)
55. Loizou, P.C.: Speech Enhancement: Theory and Practice. CRC Press, Boca Raton (2007)
56. Hu, Y., Loizou, P.: Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Lang. Process. **16**(1), 229–238 (2008)
57. Plomp, R., Mimpen, A.: Improving the reliability of testing the speech reception threshold for sentences. Int. J. Audiol. **18**(1), 43–52 (1979)
58. Dahlquist, M., Lutman, M., Wood, S., Leijon, A.: Methodology for quantifying perceptual effects from noise suppression systems. Int. J. Audiol. **44**(12), 721–732 (2005)

Chapter 16
Exploratory Analysis of Brain with ICA

Rubén Martín-Clemente

Abstract This chapter introduces the use of independent component analysis (ICA) in the study of electroencephalographic (EEG) data. Though the main application of ICA is in the context of denoising, we prefer to focus our attention to the independent components of artifacts-free EEG data. The interpretation of these independent components is still controversial, and we outline the more accepted alternatives. An introduction to the results obtained when applying ICA to evoked potentials (EPs) and event-related potentials (ERPs) is presented, as well as an explanation of the ICA of natural images and its relationship with models of visual cortex is also presented. This chapter is written as a general introduction to the subject for those who want to get started in the main topics.

16.1 Introduction

Independent Component Analysis (ICA) is a multivariate technique that enables us to linearly transform a given random vector into a vector of (maximally) independent components. In the last decade, ICA has been widely used in biomedical applications: e.g., for the detection of the fetal electrocardiogram [20, 45, 64–66], in the analysis and classification of heartbeats [10, 12, 57, 71], in functional magnetic resonance imaging (fMRI) [22, 36, 52], for the development of brain computer interfaces [27, 80], in photoplethysmography [54, 56], in electromyography [14, 44], for the diagnostic of scoliosis [1], in the modeling of metabolic processes [59], *et cetera*. ICA is also closely related to the blind source separation problem.

This chapter reviews the use of ICA in the study of brain and, specifically, electroencephalogram (EEG), which records the brain's electrical activity. Our aim is to provide an introduction for those who want to get started in the main points. The

R. Martín-Clemente (✉)
Department of Signal Processing and Communications, University of Seville, Sevilla, Spain
e-mail: ruben@us.es

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_16,
© Springer-Verlag Berlin Heidelberg 2014

chapter is organized as follows: first of all, we provide basic background information on the structure and function of the brain. The application of ICA to EEG data is reviewed in Sect. 16.4, with special emphasis in the interpretation of the independent components, in the use of ICA for denoising the data, in the search for the sources of the electromagnetic fields in the brain, and in the study of the so-called evoked and event-related potentials. We focus in these specific analyses because ICA has demonstrated well its effectiveness for all of them. The ICA of natural images has attracted great attention in recent years, due to its ability to explain certain characteristics of the simple cells in the visual cortex, and is explained in Sect. 16.5. In Sect. 16.6, we present some algorithms specifically devised for the analysis of the EEG and, finally, the last Section is devoted to present some conclusions.

16.2 Background of Brain Structure and Function

The brain is the part of the central nervous system that gives rise to thought and consciousness, interprets the *stimuli* from the environment, and controls and coordinates other organs of the body. It is made up of 15–33 billion neurons and more than 100 billion nerves. There are two kinds of tissue in the central nervous system: *grey matter* and *white matter*. *Grey matter* consists of closely packed neural cell bodies, and can be regarded as the information processing part of the central nervous system. *Grey matter* is found at the *cerebral cortex* and also at the surfaces of the *cerebellum*, the *brainstem*, the *basal ganglia* and the *limbic system* (these terms are explained below). *White matter* is a vast system of neural connections that contains the nerve fibers (*axons*) that communicate the regions of the brain to each other.

Our brain is composed of three specialized parts that collaborate together: the *cerebrum* (see Sect. 16.2.1), the *cerebellum*, and the *brain stem* (see Fig. 16.1):

- The *brain stem* is the link between the spinal cord and the rest of the brain. It performs many basic reflex functions, contributing to the control of the cardiac and respiratory functions and maintaining the consciousness.
- The *cerebellum* is at the back of the brain and regulates the muscular activity. It is responsible for accurate movement coordination, motor learning, equilibrium, posture, balance, and muscle tone. The *cerebellum* does not decide to make the movements, but executes the motor commands from the *cerebrum*, calibrating the actions and position according to the information received from the muscles and the inner ear.

The brain is bathed in cerebrospinal fluid, surrounded and protected by a layer of tissues called *meninges*, the blood–brain barrier, and the bones of the skull (*cranium*).

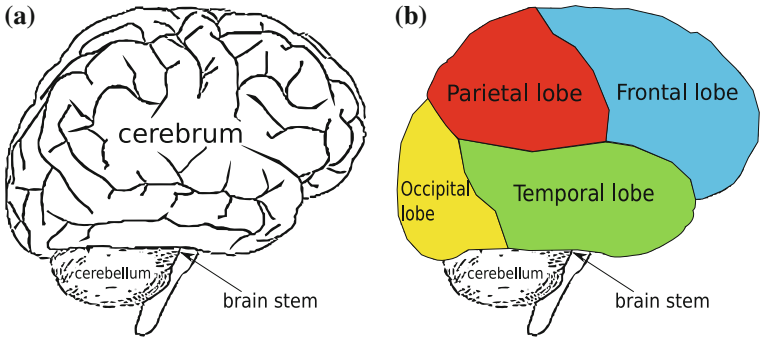


Fig. 16.1 Parts of the brain. **a** Structure of the brain. **b** Brain cortex

16.2.1 The Cerebrum

The *cerebrum* is the dominant part of the brain and comprises two (more or less symmetric) left and right hemispheres, connected by a large *white matter* structure called *corpus callosum*. The *cerebrum* may itself be divided into three subregions (see Fig. 16.1):

1. The *cerebral cortex*.
2. The *basal ganglia*.
3. The *limbic system*.

The outermost layer of brain cells is called *cerebral cortex* and is made up of *grey matter*. Thinking and voluntary movements begin in the *cortex*. The cortex is only 1.5–4.5 mm deep and, due to its special interest, it will be described in some detail in Sect. 16.2.2. Under the cortex we find a large mass of *white matter*, within which a number of clusters of neurons (*grey matter*) called *basal ganglia* are found. The *basal ganglia* are involved in perception, attention, *motivation* and motor functions. Basal ganglia also have an important role in controlling eye movements. Finally, the *limbic system* (also called the "emotional brain") consists of several nerve pathways incorporating subcortical structures located on top of the *brain stem*, including the *hippocampus*,[1] the *hypothalamus*,[2] the *amygdala*[3] and the *thalamus*.[4] The *limbic system* controls our emotions and plays an important role in learning, memory, control of appetite, and in the regulation of hormones.

---

[1] The *hippocampus* plays an important role in the formation of new memories, and in spatial orientation. It also seems to be related to behavioral inhibition.

[2] The *hypothalamus* is involved in emotion and endocrine function control, hunger, and sleep–wake cycle regulation, among other tasks. It also controls the pituitary gland.

[3] The *amygdala* is involved in memory, emotion, and fear.

[4] The *thalamus* regulates auditory, somatosensory, and visual sensory information. *All sensory stimuli*, with the exception of smell, is received in the cortex after passing through the thalamus.

Interestingly, it has been suggested that the *cerebral cortex* performs unsupervised learning, the *basal ganglia* are devices for reinforcement learning, and the *cerebellum* performs supervised learning.

16.2.2 The Cerebral Cortex

The *cortex* is the outermost layer of brain cells, and deserves special attention. It is a very thick layer of neural tissue, composed of a narrow convoluted margin of grey substance.

The cortex is a continuous sheet of grey matter. Note, however, that it is conventionally divided in each hemisphere into four *lobes*, named after the bones under which they are located (see Fig. 16.1b):

1. The *frontal lobe*. Under the forehead.
2. The *parietal lobe*. Under the top of the head, above the ears.
3. The *temporal lobe*. Above ears and immediately behind and below the frontal lobe.
4. The *occipital lobe*. At the back of the head.

Different lobes of the cortex have different functions. Basically, these functions can be grouped into three major categories: cognitive (language, thinking, and interpretation of the world), motor (functions related to the control of voluntary movements), and sensory functions (the ability to process the information from our senses):

- The *frontal lobe* is associated with higher cognitive functions (personality, reasoning, and judgement) and, in collaboration with the *basal ganglia* and the *parietal lobe*, is also responsible for motor functions (e.g., the primary *motor cortex* is located at the posterior part of the frontal lobe). Broca's area, whose functions are linked to speech production, is also in the *frontal lobe*.
- The *parietal lobe* integrates the main somatosensory receptive areas, i.e., those related to the sense of touch, and its functions also include spatial orientation or the ability to read and write. Left part of the parietal lobe has also the ability to understand numbers and solve mathematical problems.
- The part of the cortex responsible for processing sound is mainly at the *temporal lobe* (the Wernicke's area, which is usually above the left ear, plays a key role in the comprehension of language). *Temporal lobes* also control visual and verbal memories.
- The part of the cortex that processes visual information (i.e., the *primary visual cortex*) is located at the *occipital lobe*.

Let us finish with a true curiosity: each cerebral hemisphere controls mainly the opposite side of the body and, interestingly, left part of the *cerebrum* seems to be responsible for numerical and scientific thinking, and written and spoken language; by contrast, the right part of the *cerebrum* seems to be linked to artistic capabilities and imagination.

16.2.3 The Electroencephalogram

In a sense, trying to understand the inner working of the brain through the EEG is comparable to trying to understand the mechanisms of a motor through the motor noise. The EEG mainly arises from the postsynaptic currents in the pyramidal neurons of the cortex. Pyramidal neurons are the most abundant type of neuron in the cortex, and receive their name from the similarity between the cell body (*soma*) and a pyramid. Every neuron receives inputs from many others. In each communication, the "transmitter" neuron is called *presynaptic*, and the "receiver" neuron is called *postsynaptic* (the *synapse* is the point of connection between the neurons). When two neurons communicate, a flow of positively charged ions, the postsynaptic current, is generated from the presynaptic cell to the postsynaptic cell (that current also produces a voltage, called *postsynaptic potential*, across the membrane of the postsynaptic neuron). In practice, hundreds, if not thousands, of postsynaptic currents combine in the neuron and, if their sum pass a threshold, an *action potential* occurs. The action potential is a short spike (1 ms) that propagates through the axon to other neurons, generating new *postsynaptic currents*. The summation of the electric fields associated with the synchronous *postsynaptic currents* of millions of neurons can be measured at the scalp, giving the EEG. More precisely, the EEG is a record over time of the differences of potential between different locations on the surface of the head.

Figure 16.2 shows the standard location of the electrodes for EEG recording. As an example, Fig. 16.3 shows typical voltage waveforms as can be measured at these locations: in this figure, note that the EEG is not "clean", but rather is contaminated by a number of artifacts, e.g., a "bump" artifact appears at $t = 2\,\mathrm{s}$ in the frontal electrodes most probably due to the fact that the subject has blinked or moved the eyes (see Sect. 16.4.4).

16.3 Overview of EEG Signal Processing

EEG signal processing (see [72] for a book of reference) usually comprises three steps:

1. Noise reduction.
2. Feature extraction.
3. Feature classification.

Some comments are in order.

16.3.1 Noise Reduction

The EEG signal measurements are usually contaminated by several types of noise and artifacts, for example, electrocardiogram artifacts and eye-induced artifacts. Eye blinks, for example, elicit a large potential difference between the cornea and the retina that can be one order of magnitude larger than the EEG (see Fig. 16.3).

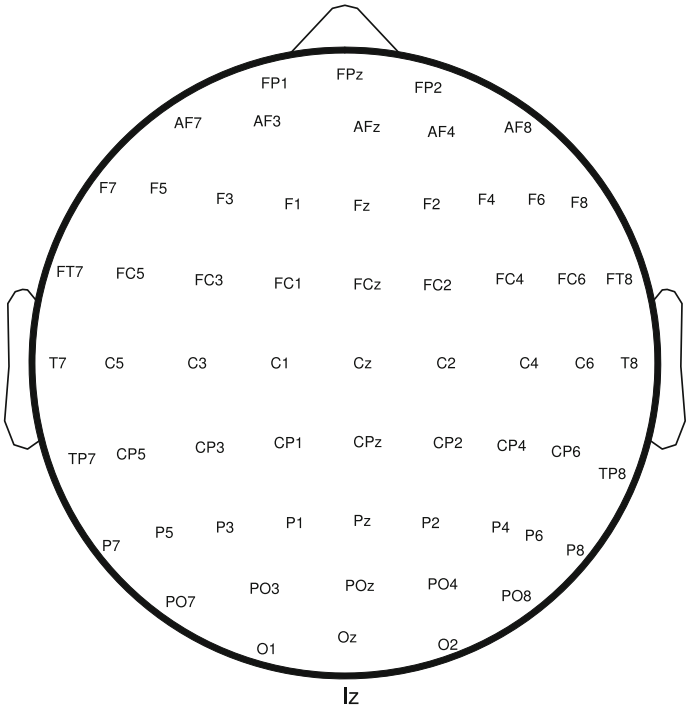


Fig. 16.2 Standard placement of electrodes for EEG recording. Letters "*F*", "*T*", "*P*" and "*O*", respectively, mean *frontal*, *temporal*, *parietal*, and *occipital* lobe (see Fig. 16.1b). The '*C*' letter stands for *central*, and letter "*z*" (zero) refers to an electrode placed on the center line. Electrodes on the *right hemisphere* are numbered with even numbers, and odd-numbers are used on the *left hemisphere*. "*Fp*" refers to the frontal polar sites

The bandwidth of the EEG is from about 1 to 100 Hz, although we rarely go beyond 50 Hz in clinical practice. Most of the noise can be suppressed by applying low-pass filters. DC and baseline drifts can be eliminated using high-pass filters (1 Hz cutoff frequency), and powerline harmonics can be removed with a comb filter. If the subjects under test do not maintain their eyes closed during the recording of the EEG, additional processing is required to eliminate eye-blink artifacts. Adaptive filtering has been used for this task, where the necessary reference signals are taken from electrodes located in the vicinity of the eyes. Adaptive filtering can be also used to eliminate electrocardiogram (ECG) artifacts.

Of course, as the reader well knows, ICA is a valuable tool for denoising and removing artifacts. In fact, denoising and removing artifacts seem to be the primary use for ICA in EEG signal processing. More information will be given in Sect. 16.4.4.

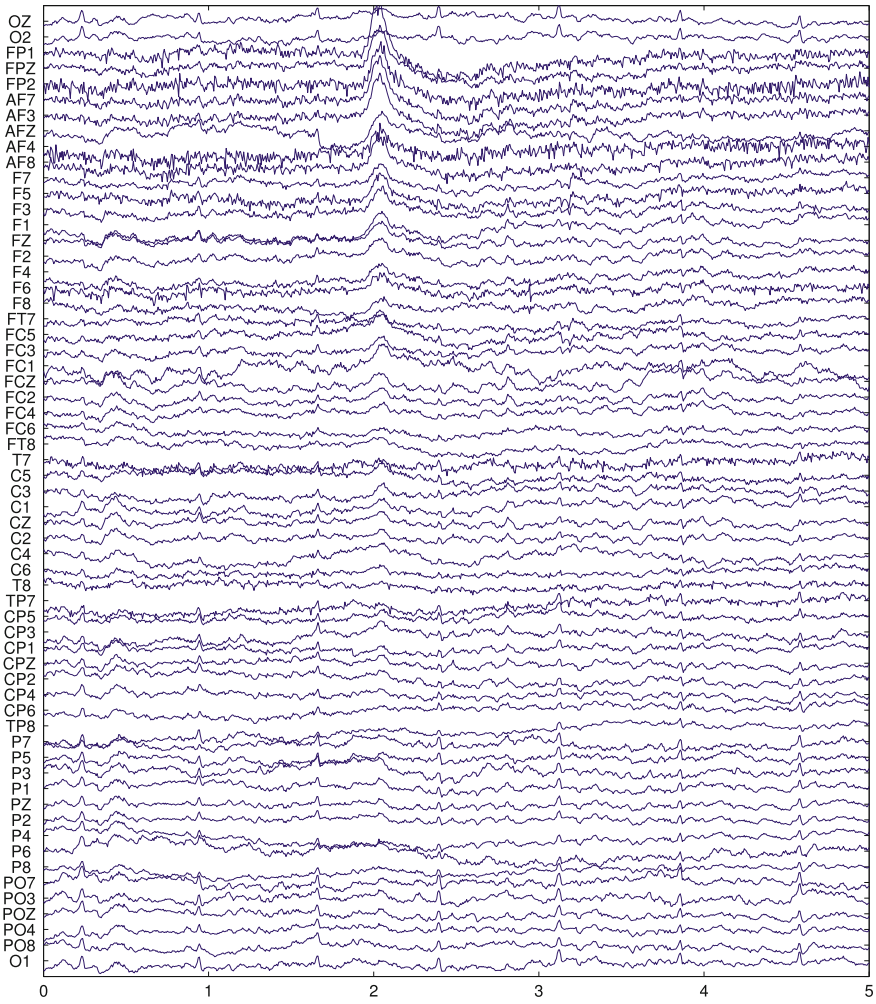


Fig. 16.3 EEG data. The figure represents 5 s of 61 raw EEG channels, obtained from a healthy subject . Data was obtained from the Physionet database (http://www.physionet.org/pn4/eegmmidb/). The placement of the electrodes, as well as an explanation of the nomenclature used for the channels, can be seen in Fig. 16.2. The *horizontal axis* represents time in seconds. The ICA of these data is presented in Fig. 16.4

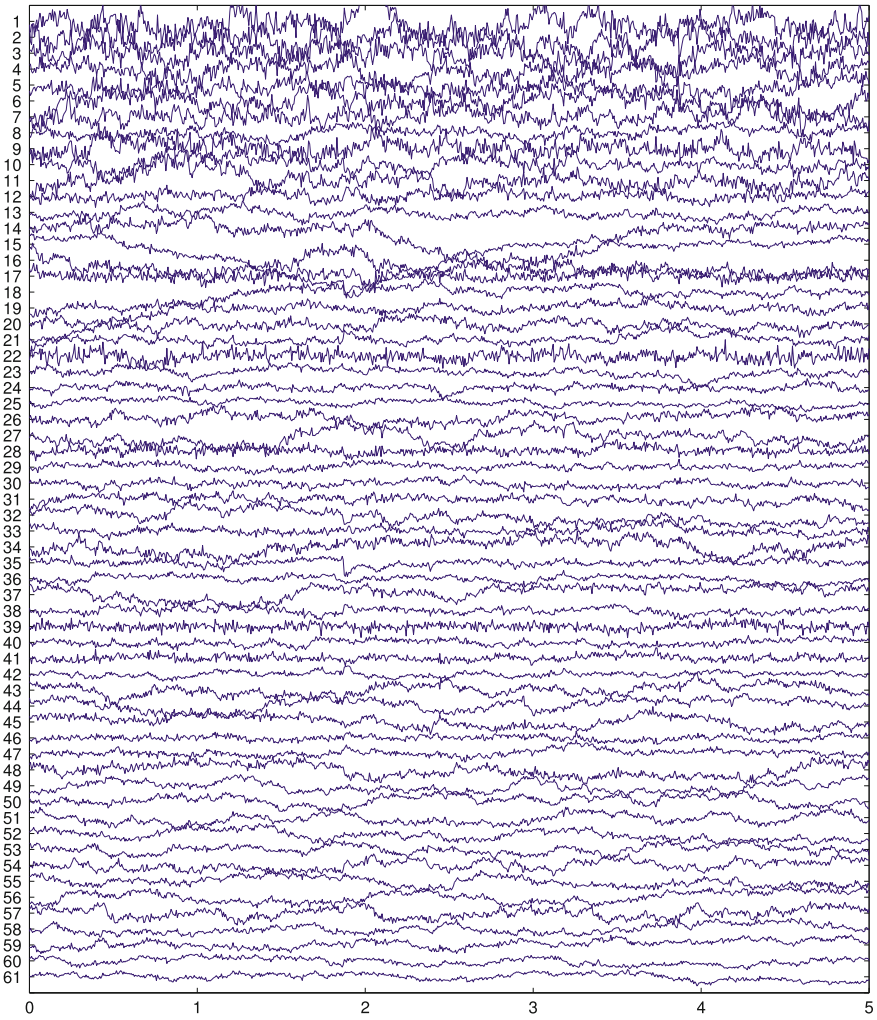


Fig. 16.4 Independent components of the data shown in Fig. 16.3. ICA was performed by using the *Infomax* algorithm [7]. Scalp maps and equivalent current dipoles (ECDs) of these independent components are shown in Fig. 16.5

16.3.2 Feature Extraction

After removing noise and artifacts, the second step in EEG signal processing usually consists in extracting relevant features out of the EEG signals.

Since the EEG is highly nonstationary in nature, feature extraction can be performed only after prior segmentation of the signals into short segments, usually not longer than a few seconds. Features are then extracted from each one of them. Within

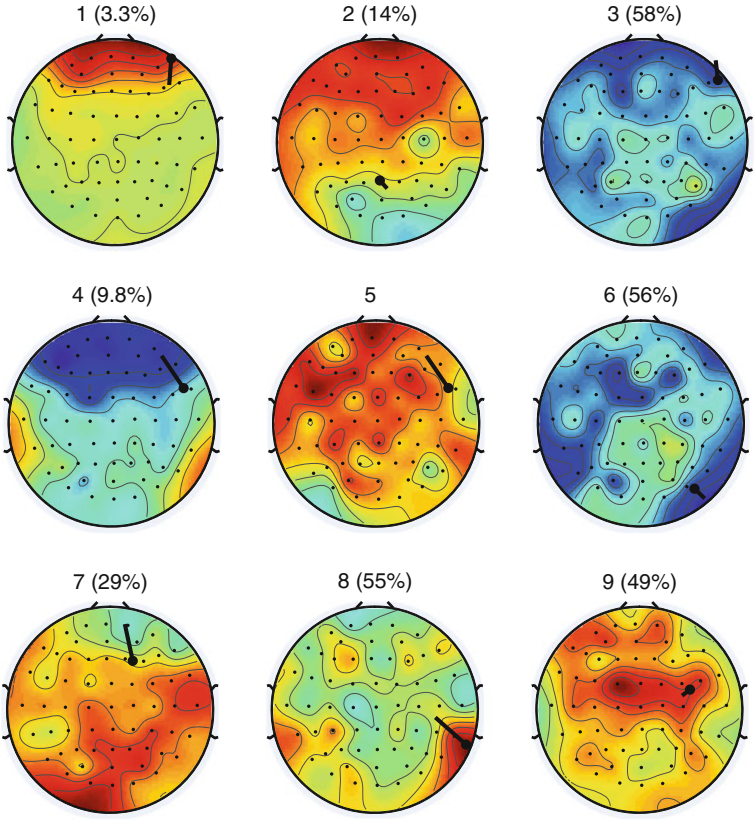


Fig. 16.5 This figure shows the scalp topographies and the current equivalent dipoles (ECDs) of some of the independent components in Fig. 16.4 (the number between parentheses is an indicator of the residual error in the estimation of the ECD). The physiological origin of the independent components may be determined from this information. The "*dots*" indicate the location of the electrodes

each segment, the signals are considered to be stationary and can then be described by suitable probability distributions. The major problem is, of course, to determine the initial and final time instants of each segment. Usually, the data is first divided into short-time frames and statistics, such as the kurtosis, are computed for each frame. Denoting $s(n)$ the value of the test statistic in the nth frame, if $|s(n) - s(n-1)|$ is greater than a predefined threshold, we assume that the "border" that separates two consecutive segments is located in between the nth frame and the preceding $(n-1)$th frame.

Features can be selected in several ways. There exists time-dependent features (mean and peak values of the EEG signals, energy, higher order statistics, entropy, autoregressive (AR) parameters, Lyapunov exponents, ...), frequency-dependent features (power spectral density (PSD) values, band powers, ...),

time–frequency-dependent features (matching pursuit, coefficients of the wavelet decomposition, ...), and so on. Spatial-based features are particularly interesting. The most important spatial-based feature is the localization *at a given time* of the regions inside the brain in which the *postsynaptic currents*[5] are more active. This feature provides valuable information on the functioning of brain and, also, on several diseases and abnormalities. ICA has revealed itself as an useful preprocessing tool for this task (see Sect. 16.4.2).

Notice finally that, to take into account the time-course variation of the EEG's characteristics, it is usual in EEG signal processing to concatenate the features from several different time segments into a single feature vector.

16.3.3 Feature Classification

Finally, we try to classify the features into different classes that, in turn, correspond to different brain activities. For example, epileptic seizures produce a series of sharp spikes in the EEG. Their second- and higher order statistics may be classified to determine automatically the type and severity of the epileptic attack or, even, to distinguish between a true epileptic seizure and a nonepileptic attack.

Linear classifiers, such as Fisher's linear discriminants and support vector machines (SVM), are probably the most popular classification methods in EEG signal processing. Linear classifiers use hyperplanes to separate the data into classes. Fisher's linear discriminant assumes that the data is gaussian distributed, and (roughly speaking) obtains the separating hyperplanes by maximizing the distance between certain projections of representative members of the classes. As an alternative, SVMs select the hyperplanes by maximizing the distance to the classes. Interestingly, SVMs also enable us to define nonlinear decision boundaries by previously mapping the data to another space of higher dimensionality.

Other classifiers used in EEG signal processing include multilayer perceptrons, Bayes classifiers or Hidden Markov Model (HMM) classifiers. Nearest neighbor classifiers are also popular when unsupervised learning is required. Finally, note that several classifiers can be combined to obtain a better performance using, for example, voting algorithms such as bagging or boosting.

16.4 The ICA of EEG data

The use of ICA for studying brain dynamics greatly follows from the seminal work [60] by Makeig and co-workers. A good survey of these and other authors' contributions can be found in [50, 77, 78]. For simplicity, we shall focus mainly on the analysis of the electroencephalogram (EEG), but essentially the same applies to the ICA of magnetoencephalogram (MEG) data. Also note that there exists an excellent

[5] See Sect. 16.2.3.

and freely available Matlab toolbox, called EEGLAB, that can be used to process EEG data in many ways (www.sccn.ucsd.edu/eeglab/). This software has been used to generate nearly all of the figures in this chapter.

16.4.1 Interpretation of ICA

In EEG signal processing, unfortunately, ICA raises more questions than we can answer. Let us list some open problems below:

- What does ICA do? This is at least controversial: since no part of the brain functions completely independent from the others, how can ICA generate physiologically plausible component waveforms [61]? All we can actually expect is that ICA will perform a decomposition of the EEG recordings into temporally independent components. "Temporally independent components" is often interpreted by neurobiologists as signals having "maximally distinct" waveforms. The effective number of independent components contributing to the EEG is *a priori* unknown, and may vary from one subject to another even under the same conditions.
- Have the "independent components" got a definite physical origin? Actually, their origin may be distributed across many brain regions and, moreover, is *a priori* unknown. Each independent component can come from the linear combination of postsynaptic currents spread around all the brain. Having said that, it is very interesting that, in many cases, the independent components seem to be linked to physically compact areas of the brain (see Sect. 16.4.2).
- What does ICA actually do? Makeig et al. consider that ICA actually reveals a system of synchronous but independent electromagnetic activity within relatively large independent EEG domains [63]. In other words, ICA defines transient brain networks (that may be distributed, linked, and even interpenetrated) whose electromagnetic activity is concurrent and independent, and all together make up the EEG data. This is a different but complementary perspective of the brain to that adopted by traditional neuroscience. Note that ICA is not actually concerned with the spatial location of those brain networks, if this has sense, but with the information they provide. What to do with this information, and how to integrate it with other approaches, is an interesting line of open research.
- What is ICA currently useful for? In any case, ICA has demonstrated its effectiveness as a preprocessing tool: definitively, ICA is able to remove a wide range of artifacts (see Sect. 16.4.4) and is of great assistance in modeling the electromagnetic fields in the brain (see Sect. 16.4.2). Moreover, the ICA decomposition facilitates the analysis and classification of the so-called evoked and event-related potentials (EPs and ERPs) (see Sect. 16.4.3). Finally, although not directly connected with the study of the EEG, we would like to mention that there are strong similarities between the processing of images in the human visual system and ICA (see Sect. 16.5).

Table 16.1 EEG frequency bands

Name	Frequency (Hz)	Characteristics
α	8–13	Present in sleep relaxation and usually when eyes are closed. They mainly originate in the occipital lobes
β	14–30	Associated with consciousness and reasoning. Sensitive to medications
θ	4–8	Present during light sleep. Theta waves arise in the cortex or in the hyppocampus
δ	<4	Present when in deep sleep. They can originate from the cortex or in the thalamus
γ	>30	May be associated to high-level information processing
μ	8–13	Present when the body is at rest. Unlike the α wave, which is related to the visual cortex, μ waves are associated to the motor cortex

16.4.1.1 Characteristics of the Independent Components

Having identified the ICA model,

$$\mathbf{x} = \mathbf{A}\,\mathbf{s},$$

where $\mathbf{x}$ contains the signals recorded by the electrodes and $\mathbf{s}$ is the vector of independent components, the columns of the mixing matrix $\mathbf{A}$ give the relative strength of each component at each electrode. A graphical representation of these strengths, depicted at the location of the corresponding electrodes on a cartoon head model, is called *scalp map* or *scalp topography* of the independent component (see Fig. 16.5). It should be noted that as important as the waveform of the independent component is its associated scalp map: the physical origin of the components can be often identified by these maps (e.g., eye activity is located mainly at frontal sites [50]).

Moving on to other issues, it is well known that the normal EEG waveforms can be classified into six patterns: alpha, beta, delta, gamma, mu, and theta (see Table 16.1). The frequency analysis of the independent components shows that gamma band and near DC dynamics appear to be less well represented than activity in intermediate frequency bands [2]. Recent papers include a study of the reliability of the independent components when ICA is trained on insufficient data, that can be found in [26].

16.4.2 Identifying the Electromagnetic Brain Sources

We have already mentioned (see Sect. 16.2.3) that the EEG is a record of the electrical activity of the brain that arises from the postsynaptic currents in the pyramidal neurons of the cortex. A postsynaptic current appears to an external observer as if it were generated by a current dipole. When many neurons are active, dipoles with the same

orientation sum to form a single large current dipole, which is usually referred to as an "equivalent current dipole" (ECD). Interestingly, areas with a diameter up to 3 cm can be accurately modeled by a single ECD. The potential due to a current dipole of moment $\mathbf{p}(t)$ at a point specified by a radius vector $\mathbf{r}$ originated at the position of the dipole is

$$v(t) = \mathbf{p}(t) \cdot \frac{\mathbf{r}}{4\,\pi\,\sigma\,|\mathbf{r}|^3}$$

where σ is the permittivity of the medium. Denoting $\mathbf{e}_i$, $i = 1, 2, 3$, the orthonormal basis vectors in the three-dimensional space and letting $\{s_1(t), s_2(t), s_3(t)\}$ be the coordinates of $\mathbf{p}(t)$ in this basis, i.e., $\mathbf{p}(t) = \sum_{i=1}^{3} s_i(t)\mathbf{e}_i$, it follows that

$$v(t) = \sum_{i=1}^{3} a_i\, s_i(t)$$

where $a_i = \mathbf{e}_i \cdot \mathbf{r}/(4\,\pi\,\sigma\,|\mathbf{r}|^3)$. The signals recorded at the electrodes $v_1(t), \ldots, v_N(t)$ are modeled as the superposition of the potentials due to a large number of dipoles:

$$\begin{aligned} v_1(t) &= a_{11}\, s_1(t) + \cdots + a_{1M}\, s_M(t) + n_1(t) \\ &\vdots \\ v_N(t) &= a_{N1}\, s_1(t) + \cdots + a_{NM}\, s_M(t) + n_M(t) \end{aligned}$$

where $s_i(t)$, $i = 1, \ldots, M$ denote the dipoles' coordinates ($M >> N$) and $n_i(t)$ considers the contribution of noise. Inferring the number, spatial localization, and orientation of the ECDs on the cortical surface helps to identify the areas responsible for those brain activities which are of interest, but it is a very difficult inverse problem (one of the main difficulties arising from the fact that the electrodes actually record a mixture of the contributions of all dipoles).

ICA has not been designed to solve the above-mentioned inverse problem (among other things because we have no guarantee that the s_i are independent). Nevertheless, since ICA is able to remove a wide range of artifacts (see Sect. 16.4.4), it has proven to be an efficient preprocessing step that makes easier the localization of the ECDs [15, 34, 70, 75]. Most importantly—and here we refer back to the previous sections—many independent components have scalp maps that are perfectly compatible with an origin in a single equivalent current dipole or in a pair of dipoles [21]. It follows that determining the ECDs that generate those scalp maps may be much better conditioned than solving directly the original inverse problem. As an example, Fig. 16.5 shows the scalp topographies and the current equivalent dipoles (ECDs) of some of the independent components shown in Fig. 16.4. Most importantly, we can assume that the independent components originate at the locations of these ECDs. In this way, we can link the independent components to physically compact regions of the brain.

16.4.3 Evoked and Even-Related Brain Potentials

External stimuli cause the brain to produce electrical potentials known as *evoked potentials* and *even-related potentials* (EPs and ERPs in the future). Measurement of EPs/ERPs involves recording the EEG while stimuli (e.g., sound burst or light flashes) is presented. Usually, EPs/ERPs are signals of very low amplitude (μV) that cannot be discerned by the naked eye from the background EEG activity. For this reason, the stimulus is repeated many times and the segments (or *epochs*) of EEG preceding and immediately following each stimulus presentation are collected and summed together, causing random noise to be canceled. The difference between EPs and ERPs is conceptual: while EPs directly reflect the basic processing of the stimulus and occur early in time, ERPs involve later and more complex processes in higher brain structures. Furthermore, EPs usually require to average more epochs than ERPs.

Multiple studies of EPs/ERPs have benefited from the use of ICA, and we will review a few for illustration [9, 11, 16, 17, 28, 47, 55, 62, 79, 80]. Makeig et al. [62] decomposed ERPs, which were recorded in response to visual stimuli, into three meaningful independent components with physically plausible scalp maps. The time–frequency characteristics of the independent components were related to those of an ERP called *P300*.[6] Jentzsch [47] conducted an experiment in which subjects were instructed to press buttons in response to some property of a visual stimulus, and ICA was applied to auditory grand average ERPs.[7] The independent component amplitudes appeared to be sensitive to the hand used in the response, and the components themselves turned out to be quite similar to P300 and N1 waves.[8] Xu et al. [80] also proposed an algorithm for the P300 ERP detection. Basically, ICA was applied to raw EEG data and those independent components more consistent with the P300 wave were first identified, and then projected back to the scalp. By doing so, the signal-to-noise ratio of P300 was increased, and the wave was then easily detected.

Bishop et al. [9] were interested in the process of maturing of the auditory system. They analyzed auditory grand average ERPs elicited by tones in children between 7 and 11 years. For all age groups, two major independent components were found in the data, which mapped on to the projections of single equivalent dipoles located on the temporal lobe. Interestingly, one of the generators was tangentially oriented and showed substantial changes between 7 and 11 years, whereas the other generator was radially oriented and did not show age changes.

---

[6] P300 (also called P3 or *late positive component* or LPC) is a reliable positive ERP that peaks at approximately 300 ms after the presentation of relevant or infrequent stimuli. It has two subcomponents, P3a and P3b, which respectively originate from frontal and parietal lobes. P3a is associated with the response to a change in the environment, while the amplitude of P3b is inversely proportional to the probability of the stimulus. P300 also seems to be correlated with decision-making processes.

[7] The term "grand average" means that the author averaged together epochs from many subjects.

[8] N1 is a large EP that appears in visual discrimination tasks.

Müller et al. [67] studied event-related MEG recordings, where a single patient was subject to combined auditory and vibrotactile stimulation, generated with a loudspeaker that was also coupled to a balloon that was held by the subject with both hands. ICA was able to separate the somatosensory and the auditory brain responses, and the scalp maps of the independent components were in good agreement with the field patterns of conventional ECDs. Furthermore, these ECDs were located precisely in the brain regions expected to be activated by the respective stimuli. The most interesting part of the paper, however, is that in which the authors discuss the effects of overlearning: while averaging the event-related responses is required to remove the background EEG activity and increase the signal-to-noise ratio, the number of data points available for the ICA algorithms decreases to the same extent, so that the independent components are prone to suffer from overlearning or overfitting. Overlearning produces independent components that are zero almost everywhere except for a single spike or "bump" when HOS-based algorithms are used [73], or independent components with sinusoidal spurious components when SOS-based ICA methods are employed.[9] As a solution, the authors propose to reduce the dimensionality of the data and an additional resampling-based method to evaluate the reliability of the results. Wang et al. [79] used ICA to select the optimal electrode pair, in the sense of enhancing the signal-to-noise ratio, and detect visual EPs.

16.4.3.1 Analyzing Single-Trial EPs/ERPs

However, averaging EPs/ERPs has several disadvantages. The most important one is that it eliminates the trial-to-trial temporal variability between EPs/ERPs, even though this variability may reflect changes in subject state and reveal information about brain dynamics [61]. When applied to single-trial EPs/ERPs, ICA gives distinctive results that cannot be obtained by conventional approaches: Jung et al. [51], e.g., describe the ICA decomposition of single-trial 31-channel ERP epochs[10] from 28 normal, 10 autistic, and 12 brain lesion subjects, all of whom were asked to participate in visual attention tasks and to press a button each time they saw a circle appear on the screen. ICA separated out:

1. Blink-related artifacts and eye movement components.
2. Independent components whose activation was time-locked to the visual stimuli. When projected back to the scalp[11] and then summed to estimate their contri-

[9] HOS = Higher Order Statistics. SOS = Second-Order Statistics.

[10] The number of epochs ranged from 300 to 700.

[11] Given the ICA model $\mathbf{x} = \mathbf{A}\,\mathbf{s}$, where $\mathbf{x}$ stands for the observations and $\mathbf{s}$ represents the independent components, we project back to the scalp these independent components simply by setting the other independent components to zero. In other words, the observations are reconstructed considering only the contribution of the independent components time-locked to the visual stimuli.

butions to the average response, they accounted for nearly all of the P1 and N1 peaks.[12]
3. Independent components clearly time-locked to the button press. After being realigned to the median response time and projected back to the scalp, the sum of these independent components was closely related to P300 ERPs.
4. Independent components whose behavior is similar to that of μ brain waves (see Table 16.1). These independent components decrease following the button press.
5. Spatially overlapping independent components accounting for α band activity (see again Table 16.1), and that show a variety of relationships to the stimuli and the subject responses.
6. Nonevent-related background EEG activity.

In conclusion, ICA enhances the amount and quality of the information that can be extracted from ERP data. The authors report that ICA facilitates the analysis and classification (successful clustering experiments are reported) of the different types of response, allowing the study of the interactions between the ERPs and the ongoing EEG activity, as well as a better understanding of the brain dynamics.

16.4.4 Denoising

It should not be surprising that ICA is primarily used as a blind source separation technique for the removal of artifacts such as those caused by blinking, eye muscle movement (electrooculogram or EOG), facial muscle movements, cardiac activity, etc [6, 18, 19, 23, 30, 35, 43, 46, 48, 49, 53, 70, 74, 76]. The idea is simply to reconstruct the EEG data as follows:

$$\mathbf{x}_d = \mathbf{A}\,\mathbf{s}_0$$

where $\mathbf{x}_d$ is the denoised EEG vector and $\mathbf{s}_0$ is the vector of independent components, in which we have set the artifactual components to zero.

Let us present a simple example. Figure 16.4 shows real EEG data (data were collected for 1 min though only 5 s are shown for clarity). The EEG is contaminated by several artifacts. Specifically, there is an strong eye activity in the frontal electrodes (FP1 and so on): for example, an ocular artifact is clearly visible at $t = 2$ s—observe, for example, that the short duration of the deflections is compatible with blinking. There is another interfering signal, more visible at the occipital and parietal electrodes (O1 and so on), that is (more or less) periodic with a period slightly lower than 1 s. It is a "peaky" signal that seems to be an electrocardiogram (ECG) artifact.

Figure 16.6 shows the distribution of the voltage at the head surface at $t = 2$ s and, for comparison, at $t = 3$ s (when there are no visible artifacts). The plots confirm that the voltage concentrates over the frontal scalp when an ocular artifact

[12] P1 (or P100) is an EP sensitive to visual discrimination tasks that peaks at 100–130 ms after stimulus presentation and is modulated by attention.

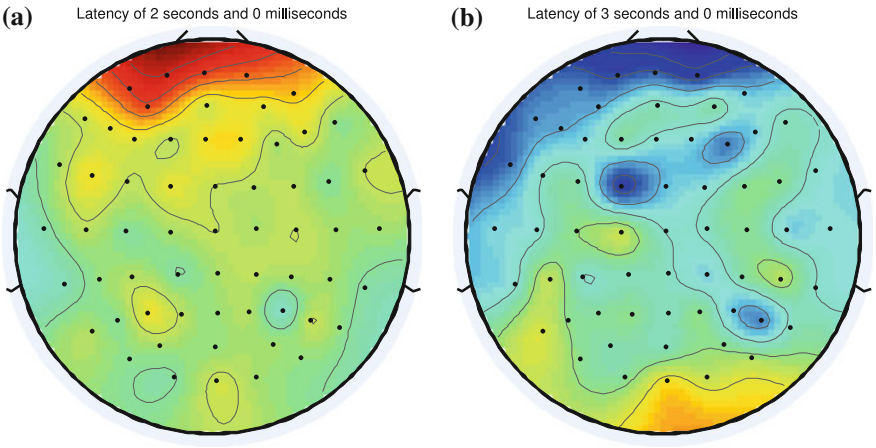


Fig. 16.6 Voltage distribution at the head surface. **a** Voltage at $t = 2\,\text{s}$. **b** Voltage at $t = 3\,\text{s}$

is present. First of all, we rejected the independent components whose scalp maps are similar to Fig. 16.6 (such as, e.g., the independent component 1, see Fig. 16.5). These components are assumed to be responsible for the ocular artifacts. By so doing, we obtained the denoised EEG data shown in Fig. 16.7. Figure 16.8 plots the power spectra of the independent components, showing a large peak around 60 Hz. This is not a typical EEG frequency, and we consider it to be the "signature" of an artifact (probably, it corresponds to the aforementioned ECG artifact or, perhaps, to noise line). The figure also shows that the components 1, 2, 4, 6, and 9 are the components which contribute the most at 60 Hz. After rejecting them, we finally obtain the "cleaned" EEG data depicted in Fig. 16.9.

In the previous example, we identified the artifactual components by visual inspection. The automatic identification of the artifacts seems to be a more powerful approach, and we will briefly review here three representative ideas:

Escudero et al. [23] obtained satisfactory results in denoising MEG data from 11 healthy elderly subjects. They propose a few criteria for the identification of the artifactual components. Cardiac signals, for example, have highly asymmetric density functions and also tend to be leptokurtic (*supergaussian*), so that they can be discriminated by their skewness and kurtosis coefficients (which are expected to take large values). On the other hand, power line noise and ocular artifacts can be easily detected by examining their frequency characteristics and scalp maps.

Shao et al. [74] also extract several features from the independent components and use a support vector machine (SVM) to classify them as inherent brain activities or artifacts. For each independent component s_i, six extracted features are defined as follows:

1. The ratio between the maximum peak amplitude and the variance of the independent component: $f_1 = \max(|s_i|)/\sigma_{s_i}^2$ (ocular artifacts, e.g., have a large amplitude).

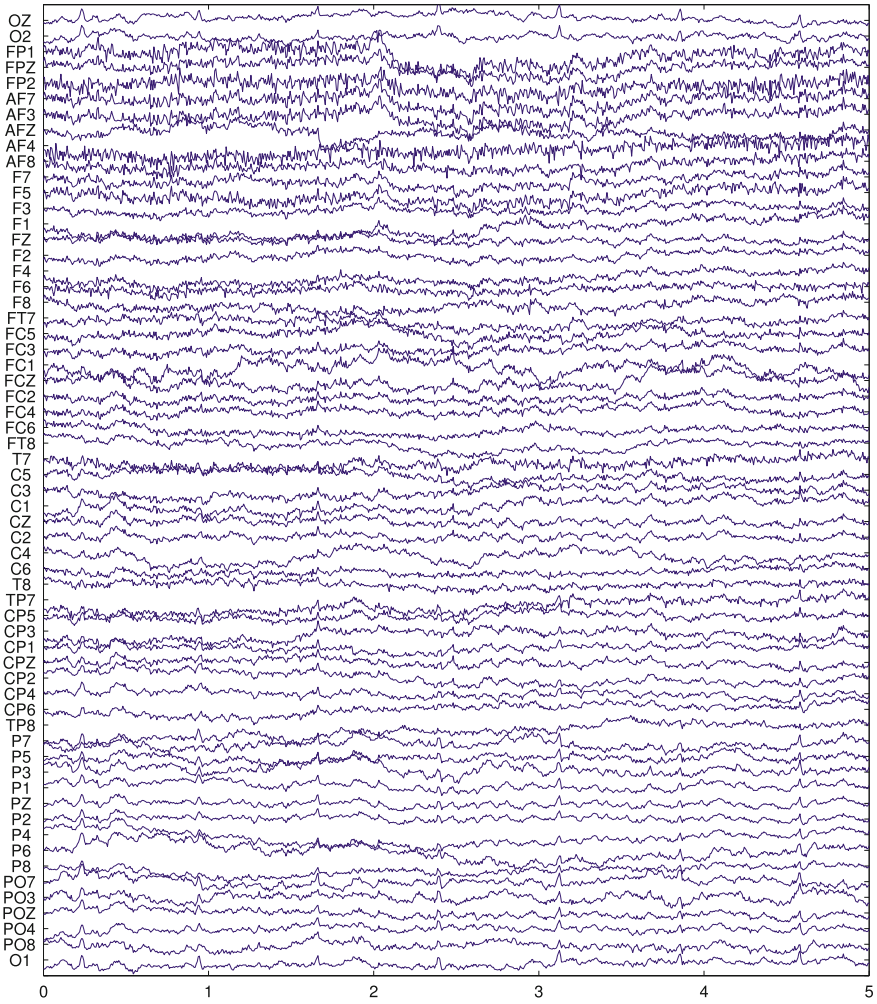


Fig. 16.7 EEG data after rejecting the independent components associated with ocular artifacts

2. The normalized skewness: $f_2 = |E[s_i^3]|/\sigma_{s_i}^3$ (as explained above, the distribution of cardiac artifacts is highly asymmetric).
3. The variance of the scalp map of s_i: $f_3 = \text{var}(\mathbf{a}_i/\|\mathbf{a}_i\|)$, where $\mathbf{a}_i$ is the ith column of the mixing matrix (it seems that the scalp map of the cardiac artifacts has a low variance).
4. A measure (i.e., the Kullback-Leibler divergence) of the difference between the probability density function of the independent component and that of a representative EOG artifact.
5. The Kullback-Leibler divergence of the probability of the independent component from that of a reference cardiac artifact.

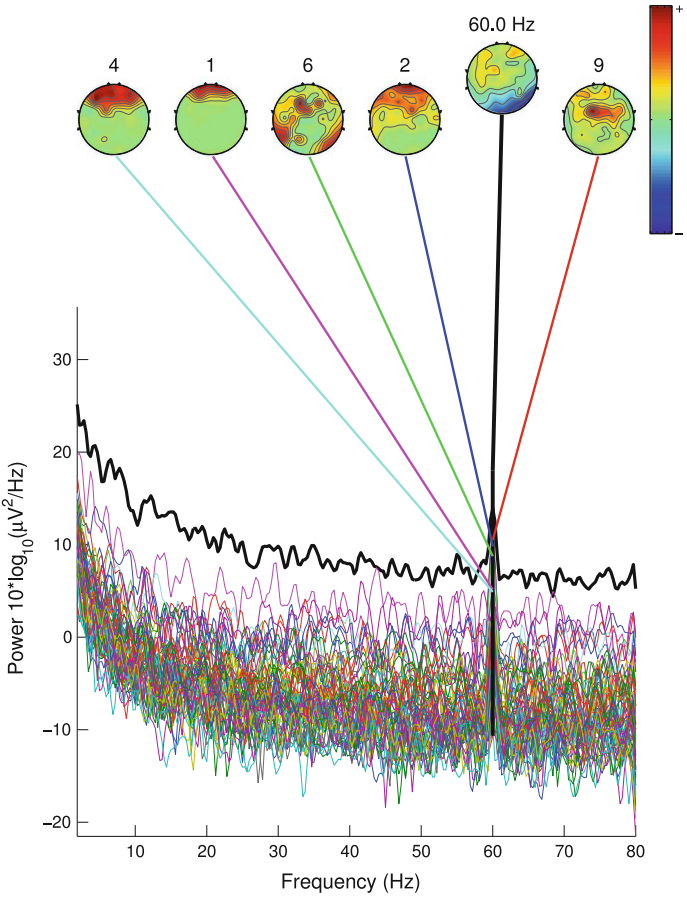


Fig. 16.8 Power spectra of the independent components and distribution of the voltages over the surface of the head at 60 Hz. The figure also shows that the independent components 1, 2, 4, 6, and 9 contribute the most at 60 Hz

6. The cross correlation between the independent component and a set of eye-blinking dominated EEG channels (namely, Fp1, Fp2, F3, F4, O1, and O2, see Fig. 16.2).

Along the same lines, Dammers et al. [18] propose another criteria for the automated classification of the independent components as either valid data or noise. For example, the detection of cardiac artifacts is performed in [18] as follows: after a bandpass filtering of the independent component under test (using different frequency bands that cover the spectrum of the ECG, namely, 2–4, 4–8, 8–16, and 10–20 Hz), its normalized phase is calculated by the formula

$$\Phi(t) = \psi(t)/(2\pi) \bmod 1$$

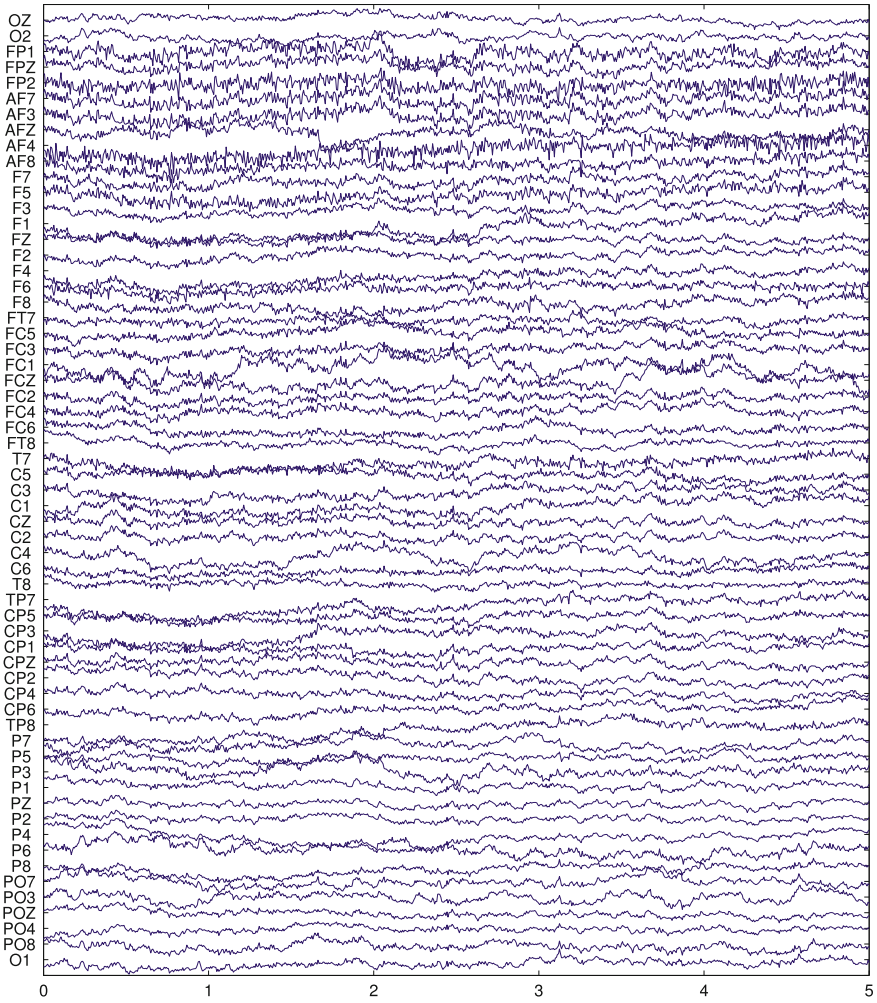


Fig. 16.9 EEG data after removing the independent components associated with ocular artifacts and those that contribute the most at 60 Hz

where $\psi(t)$ is the instantaneous phase of the independent component, obtained by the Hilbert transform. The normalized phase is then divided into segments of 1 s around the R-peaks of the ECG signal. Cardiac artifacts are synchronous with the ECG, and hence different segments are expected to have nearly identical normalized phases. In other words: all segments have the same values at the same time or, in other words, samples at the same time point are identical. The distribution of the samples is then degenerate, i.e., a Dirac delta. On the contrary, when the independent component is not a cardiac artifact, according to the principle of maximum entropy, we can assume that the samples are uniformly distributed (the uniform distribution is

the maximum entropy distribution among all distributions supported in the interval $[0, 2\pi]$). A statistical test is then used to quantify the deviation of the distribution from the uniform distribution. The authors claim that the proposed criterion is highly sensitive for identification of weak components caused by cardiac activity.

16.5 ICA of Natural Images

Hubel and Wiesel received the Nobel Prize after showing that certain neurons of the primary visual cortex (the so-called *simple cells*) give their maximum response in the presence of visual stimuli consisting of *localized and oriented structures* [37, 38], i.e., the neurons respond only if a line in a particular direction (an "edge") enters their receptive fields.[13] As one moves through the visual cortex in the occipital lobe, one finds columns of neurons that have approximately the same receptive field location, but with different orientation selectivities. Its an important problem for neuroscience to understand the reasons for this organization in the visual sensory system (why are cells directionally dependent?).

Natural images are highly redundant (i.e., nearby pixels are strongly correlated). Barlow suggested that all sensory systems, including the visual one, aim to remove the redundancy in the input data, trying to minimize the amount of information to be processed, and hypothesized that the activation of each neuron in the sensory system should be as statistically independent from the others as possible [3–5]. Furthermore, Field [24, 25] argued that the responses of the neurons of the primary visual cortex should be *sparsely distributed*.

How do we perform ICA in image processing? The observed data vectors $\mathbf{x}_i$ are obtained after the vectorization of a large number of $M \times N$ pixel patches selected randomly from the images.[14] The ICA decomposition of the data can be written as:

$$\begin{aligned} \mathbf{x}_i &= \mathbf{A}\,\mathbf{s}_i \\ &= \sum_k \mathbf{a}_k\, s_{ik} \end{aligned}$$

where $\mathbf{a}_k$ denotes the kth column of the mixing matrix $\mathbf{A}$, and s_{ik} is the ith sample of the kth independent component. Vectors $\mathbf{a}_k$ are often called *basis vectors*, since they provide a *generative model* of the data. These *basis vectors* can be also plotted as $M \times N$ images by an inverse-vectorization operation. When we do that, we get an interesting surprise—and here we refer back to the previous paragraphs: the images

[13] Each cell in the visual cortex responds only to the presence of light in a well-defined part of the retina, called the *receptive field* of the cell. The part of the visual scene projected on that area of the retina is also called "receptive field". Roughly speaking, we may think that the job of the cell is to report to the rest of the brain what is happening in that little area.

[14] The pixels of each pack are stacked one under the other to form the associated $MN \times 1$ observation vector.

of the basis vectors resemble "edges" with different orientations, lengths, and widths (Fig. 16.10). Furthermore, the distribution of the independent components is sparse, as expected, in the sense that most of the values are close to zero and only a few of them are significantly large. In other words, and very roughly speaking, each patch of the image seems to be formed with only a few simple lines.[15] Confirming what Barlow and Field had predicted, only a few neurons are therefore activated at a time.

These results are not sensitive to the choice of algorithm used. They were first described by Bell and Sejnowski [8], which employed *Infomax* [7]. Similar results have been obtained using *FastICA* [39]. Well before the emergence of ICA, Hancock et al. [31] proposed a redundancy reduction approach based on *Principal Component Analysis* (PCA) only. However, they failed in modeling the receptive fields of the simple cells: according to their results, only a few basis vectors matched oriented and localized patterns. Olshausen and Field [68, 69] proposed an unsupervised learning algorithm that attempted to find a factorial code of independent visual features, generating a set of bases that presented similar properties to the receptive fields of simple cells, i.e., most of them also showed localized and oriented "edges".

Recent works include [41, 42], where it is proposed a model of spatial organization of the ICA bases that attempts to imitate the retinotopic organization [29] of the visual cortex, and the papers [13, 40], where the authors analyze the similarities between the processing of color images in the human visual system processing and ICA.

16.6 Semi-Blind ICA of Brain Data

Most researchers use traditional ICA *blind* algorithms for the analysis of brain signals. Nevertheless, we wish to draw attention to three representative approaches [19, 33, 46] that exploit the available *a priori* knowledge about the data. As a matter of fact, there exists in many cases *a priori* information about the artifacts that contaminate the data: power line interferences, for example, are at 50/60 Hz and its harmonics, cardiac artifacts are synchronized with heart activity, eye activity is located mainly at frontal sites, etc. The use of this information seems to be a promising possibility.

16.6.1 Exploiting the Temporal Structure of the Brain Signals

De Clercq et al. [19] use canonical correlation analysis (CCA) for muscle artifact removal in EEG, as follows: given the zero-mean observation vector $\mathbf{x}(t)$, the idea is to force the source estimates to be maximally correlated with $\mathbf{x}_1(t) = \mathbf{x}(t-1)$. Thus they pretend to enforce the generation of maximally autocorrelated sources, since it is known that brain sources have a high autocorrelation whereas muscle activity is

[15] Actually, this statement has to be taken with care: all basis functions (more or less) equally contribute to many image patches.

Fig. 16.10 Typical ICA image-basis obtained from 12 × 12 patches

similar to white noise, due to its broader frequency spectrum. The idea is to search for the vectors $\mathbf{w}$ and $\mathbf{w}_1$ that maximize the objective function:

$$\rho(x(t), x_1(t)) = \frac{E[x(t)\, x_1(t)]}{\sqrt{E[x^2(t)]\, E[x_1^2(t)]}}$$

where $x(t) = \mathbf{w}^T\, \mathbf{x}(t)$ and $x_1(t) = \mathbf{w}_1^T\, \mathbf{x}_1(t)$. After some algebra, it is found that $\mathbf{w}$ is an eigenvector of the matrix:

$$\mathbf{C}_{xx}^{-1}\,\mathbf{C}_{xx_1}\,\mathbf{C}_{x_1x_1}^{-1}\mathbf{C}_{xx_1},$$

where $\mathbf{C}_{xx}$ and $\mathbf{C}_{x_1x_1}$ are the auto-covariance matrices of $\mathbf{x}(t)$ and $\mathbf{x}_1(t)$, respectively, and $\mathbf{C}_{xx_1}$ is the cross-covariance matrix of $\mathbf{x}(t)$ and $\mathbf{x}_1(t)$. The source estimates are then simply given by

$$\mathbf{w}^T\,\mathbf{x}(t).$$

Each eigenvector of the matrix gives a different source estimate, and the eigenvectors corresponding to the lowest eigenvalues are expected to generate the muscle artifacts. Experiments show that the algorithm is superior to traditional approaches and other ICA techniques based on higher order statistics.

16.6.2 Using a Temporal Reference

James et al. [46] used a reference signal $r(t)$ which incorporates the *a priori* information to guide the search for the independent components. Given the observation vector $\mathbf{x}$, the following criterion is used in [46]:

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{w}) \\ \text{subject to} \quad & g(\mathbf{w}) \leq 0 \\ \text{and} \quad & E[y^2] = 1 \\ \text{and} \quad & E[r^2] = 1 \end{aligned}$$

where $f(\mathbf{w})$ is the following approximation to the negentropy of the estimated independent component $y = \mathbf{w}^T\,\mathbf{x}$ [39]:

$$f(\mathbf{w}) = \{E[G(y)] - E[G(v)]\}^2$$

where v is a zero-mean unit-variance Gaussian random variable, $G(\cdot)$ can be any nonquadratic function, and

$$g(\mathbf{w}) = \varepsilon - E[r(t)\,y(t)]$$

measures the similarity between $r(t)$ and $y(t)$, with ϵ being a threshold.[16] This is a constrained optimization problem that can be solved through a Newton-like algorithm [58]. Interestingly, experiments show that the exact waveform of the reference signals is not very important, provided that the temporal features of interest are captured. For example, a good reference signal for the ECG artifact can be simply obtained by passing the contaminated data through a peak detector that highlights the

[16] Interestingly, in the original formulation of the algorithm [58], $g(\mathbf{w})$ is defined as $g(\mathbf{w}) = E[r(t)\,y(t)] - \epsilon$.

R waves. As $g(\mathbf{w})$ is a correlation-based measure, the reference signal $r(t)$ and the independent component must be aligned in time. The authors address this problem by repeatedly applying the method with the reference shifted one sample from one experiment to the next, until the correlation between $r(t)$ and the estimated source signal $y(t)$ attains its maximum value.

16.6.3 Using Spatial Constraints

Hesse et al. [33] noted that the scalp maps of some expected source signals may be approximately calculated *a priori* from previous data or using, for example, dipole models. This information may be used as a constraint on the mixing matrix $\mathbf{A}$, assuming that

$$\mathbf{A} = [\mathbf{A}_c, \mathbf{A}_u]$$

where $\mathbf{A}_c$ are columns subject to those constraints, and $\mathbf{A}_u$ contains unconstrained columns. Roughly speaking, the algorithm may be as follows:

1. Execute one step of some iterative ICA algorithm to find an estimate $\hat{\mathbf{A}}$ of the mixing matrix $\mathbf{A}$.
2. Enforce the constraints on the estimate $\hat{\mathbf{A}}$ of $\mathbf{A}$, ensuring that $\hat{\mathbf{A}}$ is of full column rank.
3. Return to 1 until convergence.

The second step can be performed in several ways: for example, the columns of $\mathbf{A}_c$ may directly overwrite the corresponding columns of $\hat{\mathbf{A}}$. Given a column $\mathbf{a}_c$ of $\mathbf{A}_c$ and the corresponding column $\hat{\mathbf{a}}_c$ of $\hat{\mathbf{A}}$, a "softer" and alternative procedure may be to overwrite $\hat{\mathbf{a}}_c$ with

$$p\,\mathbf{a}_c + (1 - p)\,\hat{\mathbf{a}}_c$$

whereas p is chosen so that angle between $\mathbf{a}_c$ and the new $\hat{\mathbf{a}}_c$ is below some threshold [32]. Note that the final constrained source signals may not be statistically independent among themselves. Having said that, when applied to EEG recorded during an epileptic seizure (called *ictal EEG*), the algorithm obtains a coherent and physiologycally plausible decomposition of the data. The authors also report good results in removing ocular artifacts.

16.7 Concluding Remarks

ICA has undoubtedly proven to be a useful tool for removing artifacts from the EEG data. The interpretation of the true "brain components", however, is still controversial and seems to be an exciting open field for research. The ICA of natural images has also revealed interesting connections with the early models of the visual cortex

and the characterization of the so-called simple cells. Finally, the use of *a priori* information about the brain sources to help the ICA algorithms is a third promising line of research.

This chapter has introduced the use of ICA in the study of electroencephalographic (EEG) data. We hope to achieved our goal of writing a general and accessible introduction to the problem for those who want to get started in the main topics. We refer the reader to the references for a second and more profound insight into this exciting subject.

Acknowledgments This work was supported by the Andalusian Regional Government (under the program entitled Programa de Proyectos de Excelencia) under project P07-TIC-02865. We are also grateful to S. Makeig and A. Delorme for their permission to include graphs and figures generated with EEGLAB in the manuscript.

References

1. Adankon, M., Chihab, N., Dansereau, J., Labelle, H., Cheriet, F.: Scoliosis follow-up using non-invasive trunk surface acquisition. IEEE Trans. Biomed. Eng. **60**(8), 2262–2270 (2013)
2. Anemüller, J., Sejnowski, T., Makeig, S.: Complex independent component analysis of frequency-domain electroencephalographic data. Neural Netw. **16**(9), 1311–1323 (2003)
3. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. In: Rosenblith, W.A. (ed.) Sensory Communication, pp. 217–234. MIT Press, Cambridge (1961)
4. Barlow, H.B.: Unsupervised learning. Neural Comput. **1**, 295–311 (1989)
5. Barlow, H.B.: Redundancy reduction revisited. Netw. Comput. Neural Syst. **12**, 241–253 (2001)
6. Barros, A., Vigario, R., Jousmaki, V., Ohnishi, N.: Extraction of event-related signals from multichannel bioelectrical measurements. IEEE Trans. Biomed. Eng. **47**(5), 583–588 (2000)
7. Bell, A.J., Sejnowski, T.J.: An information maximisation approach to blind separation and blind deconvolution. Neural Comput. **7**, 1129–1159 (1995)
8. Bell, A.J., Sejnowski, T.J.: Edges are the 'independent components' of natural scenes. Adv. Neural Inf. Process. Syst. **9**, 831–837 (1997)
9. Bishop, D., Anderson, M., Reid, C., Fox, A.: Auditory development between 7 and 11 years: an event-related potential study. PLoS ONE **6**, e18, 993 (2011)
10. Can, Y., Kumar, B., Coimbra, M.: Heartbeat classification using morphological and dynamic features of ECG signals. IEEE Trans. Biomed. Eng. **59**(10), 2930–2941 (2012)
11. Cao, J., Zhao, L., Cichocki, A.: Visualization of dynamic brain activities based on single-trial MEG and EEG data analysis. In: Lecture Notes in Computer Science, vol. 3973, pp. 531–540. Springer, Berlin (2006)
12. Castells, F., Rieta, J., Millet, J., Zarzoso, V.: Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias. IEEE Trans. Biomed. Eng. **52**(2), 258–267 (2005)
13. Caywood, M., Willmore, B., Tolhurst, D.: Independent components of color natural scenes resemble v1 neurons in their spatial and color tuning. J. Neurophysiol. **91**, 2854–2873 (2004)
14. Chen, X., He, C., Wang, Z., McKeown, M.: An IC-PLS framework for group corticomuscular coupling analysis. IEEE Trans. Biomed. Eng. **60**(7), 2022–2033 (2013)
15. Chen, Y., Akutagawa, M., Katayama, M., Zhang, Q., Kinouchi, Y.: Ica based multiple brain sources localization. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pp. 1879–1882 (2008)
16. Chen, Z., Ohara, S., Cao, J., Vialatte, F., Lenz, F., Cichocki, A.: Statistical modeling and analysis of laser-evoked potentials of electrocorticogram recordings from awake humans. Comput. Intell. Neurosci. **2007**, 1–24 (2007)

17. Chin-Teng, L., I-Fang, C., Li-Wei, K., Yu-Chieh, C., Sheng-Fu, L., Jeng-Ren, D.: EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment. IEEE Trans. Biomed. Eng. **54**(7), 1349–1352 (2007)
18. Dammers, J., Schiek, M., Boers, F., Silex, C., Zvyagintsev, M., Pietrzyk, U., Mathiak, K.: Integration of amplitude and phase statistics for complete artifact removal in independent components of neuromagnetic recordings. IEEE Trans. Biomed. Eng. **55**(10), 2353–2362 (2008)
19. De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., Van Huffel, S.: Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. IEEE Trans. Biomed. Eng. **53**(12), 2583–2587 (2006)
20. De Lathauwer, L., De Moor, B., Vandewalle, J.: Fetal electrocardiogram extraction by blind source subspace separation. IEEE Trans. Biomed. Eng. **47**(5), 567–572 (2000)
21. Delorme, A., Palmer, J., Onton, J., Oostenveld, R., Makeig, S.: Independent EEG sources are dipolar. PloS one **7**(2), e30135 (2012)
22. Dyrholm, M., Goldman, R., Sajda, P., Brown, T.: Removal of BCG artifacts using a non-kirchhoffian overcomplete representation. IEEE Trans. Biomed. Eng. **56**(2), 200–204 (2009)
23. Escudero, J., Hornero, R., Abasolo, D., Fernández, A., Lopez-Coronado, M.: Artifact removal in magnetoencephalogram background activity with independent component analysis. IEEE Trans. Biomed. Eng. **54**(11), 1965–1973 (2007)
24. Field, D.J.: Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A **4**, 2379–2394 (1987)
25. Field, D.J.: What is the goal of sensory coding? Neural Comput. **6**, 559–601 (1994)
26. Groppe, D., Makeig, S., Kutas, M.: Identifying reliable independent components via split-half comparisons. Neuroimage **45**(4), 1199–2011 (2009)
27. Grosse-Wentrup, M., Buss, M.: Multiclass common spatial patterns and information theoretic feature extraction. IEEE Trans. Biomed. Eng. **55**(8), 1991–2000 (2008)
28. Guimaraes, M., Wong, D.K., Uy, E., Grosenick, L., Suppes, P.: Single-trial classification of MEG recordings. IEEE Trans. Biomed. Eng. **54**(3), 436–443 (2007)
29. Guyton, A.C., Hall, J.E.: Textbook of medical physiology. Saunders, Philadelphia (2000)
30. Hae-Jeong, P., Do-Un, J., Kwang-Suk, P.: Automated detection and elimination of periodic ECG artifacts in EEG using the energy interval histogram method. IEEE Trans. Biomed. Eng. **49**(12), 1526–1533 (2002)
31. Hancock, P.J., Baddeley, R., Smith, L.: The principle components of natural images. Neural Comput. **3**, 61–72 (1992)
32. Hesse, C., James, C.: The fastica algorithm with spatial constraints. IEEE Signal Process. Lett. **12**(11), 792–795 (2005)
33. Hesse, C., James, C.: On semi-blind source separation using spatial constraints with applications in EEG analysis. IEEE Trans. Biomed. Eng. **53**(12), 2525–2534 (2006)
34. Hild, K., Nagarajan, S.: Source localization of EEG/MEG data by correlating columns of ICA and lead field matrices. IEEE Trans. Biomed. Eng. **56**(11), 2619–2626 (2009)
35. Hu, S., Stead, M., Worrell, G.: Automatic identification and removal of scalp reference signal for intracranial EEGs based on independent component analysis. IEEE Trans. Biomed. Eng. **54**(9), 1560–1572 (2007)
36. Hualiang, L., Correa, N., Rodriguez, P., Calhoun, V., Adali, T.: Application of independent component analysis with adaptive density model to complex-valued fMRI data. IEEE Trans. Biomed. Eng. **58**(10), 2794–2803 (2011)
37. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in cat's visual cortex. J. Physiol. **160**, 106–154 (1962)
38. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of the monkey striate cortex. J. Physiol. **195**, 215–243 (1968)
39. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. **10**(3), 626–634 (1999)
40. Hyvärinen, A., Gutmann, M., Hoyer, P.: Statistical model of natural stimuli predicts edge-like pooling of spatial frequency channels in v2. BMC Neurosci. (2005). doi:10.1186/1471-2202-6-12

41. Hyvärinen, A., Hoyer, P., Hurri, J.: Extensions of ICA as models of natural images and visual processing. In: Proceedings of 4th International Symposium on Independent Component Analysis and Blind, Signal Separation (ICA2003) (2003)
42. Hyvärinen, A., Hoyer, P., Inki, M.: Topographic ICA as a model of v1 receptive fields. Neural Netw. **4**, 83–88 (2000)
43. Iriarte, J., Urrestarazu, E., Valencia, M., Alegre, M., Malanda, A.: Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. J. Clin. Neurophysiol. **20**, 249–257 (2003)
44. Irimia, A., Richards, W., Bradshaw, L.: Comparison of conventional filtering and independent component analysis for artifact reduction in simultaneous gastric EMG and magnetogastrography from porcines. IEEE Trans. Biomed. Eng. **56**(11), 2611–2618 (2009)
45. Jafari, M., Chambers, J.: Fetal electrocardiogram extraction by sequential source separation in the wavelet domain. IEEE Trans. Biomed. Eng. **52**(3), 390–400 (2005). doi:10.1109/TBME.2004.842958
46. James, C., Gibson, O.: Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis. IEEE Trans. Biomed. Eng. **50**(9), 1108–1116 (2003)
47. Jentzsch, I.: Independent component analysis separates sequence-sensitive ERP components. Int. J. Bifurcat. Chaos **14**(2), 667–678 (2004)
48. Joyce, C., Gorodnitsky, I., Kutas, M.: Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. Psychophysiology **41**, 313–325 (2004)
49. Jung T-P. adn Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J.: Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin. Neurophysiol. **11**, 1754–1758 (2000)
50. Jung, T.P., Makeig, S., McKeown, M., Bell, A.J., Lee, T.W., Sejnowski, T.: Imaging brain dynamics using independent component analysis. Proc. IEEE **89**(7), 1107–1122 (2001)
51. Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J.: Analysis and visualization of sigle-trial event-related potentials. Hum. Brain Mapp. **14**(3), 166–185 (2001)
52. Kadah, Y.: Adaptive denoising of event-related functional magnetic resonance imaging data using spectral subtraction. IEEE Trans. Biomed. Eng. **51**(11), 1944–1953 (2004)
53. Kelly, J., Siewiorek, D., Smailagic, A., Collinger, J., Weber, D., Wang, W.: Fully automated reduction of ocular artifacts in high-dimensional neural data. IEEE Trans. Biomed. Eng. **58**(3), 598–606 (2011)
54. Kim, B., Yoo, S.: Motion artifact reduction in photoplethysmography using independent component analysis. IEEE Trans. Biomed. Eng. **53**(3), 566–568 (2006)
55. Koutras, A., Kostopoulos, G., Ioannides, A.: Exploring the variability of single trials in somatosensory evoked responses using constrained source extraction and RMT. IEEE Trans. Biomed. Eng. **55**(3), 957–969 (2008)
56. Krishnan, R., Natarajan, B., Warren, S.: Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data. IEEE Trans. Biomed. Eng. **57**(8), 1867–1876 (2010)
57. Langley, P., Rieta, J., Stridh, M., Millet, J., Sornmo, L., Murray, A.: Comparison of atrial signal extraction algorithms in 12-lead ECGs with atrial fibrillation. IEEE Trans. Biomed. Eng. **53**(2), 343–346 (2006)
58. Lu, W., Rajapakse, J.C.: ICA with reference. In: Proceedings of 3rd International Conference on Independent Component Analysis and Blind, Signal Separation, pp. 120–125 (2001)
59. Mabrouk, R., Dubeau, F., Bentabet, L.: Dynamic cardiac PET imaging: extraction of time-activity curves using ICA and a generalized gaussian distribution model. IEEE Trans. Biomed. Eng. **60**(1), 63–71 (2013)
60. Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.: Independent component analysis of electroencephalographic data. In: D. Touretzky, M. Mozer (eds.) Advances in Neural Information Processing Systems, vol. 8, pp. 145–151. MIT Press, Cambridge (1996)
61. Makeig, S., Onton, J.: Oxford Handbook of Event-Related Potential Components, chap. ERP features and EEG dynamics: and ICA perspective, pp. 51–88. Oxford University Press, New York (2012)

62. Makeig, S., Westerfield, M., Jung, T.P., Covington, J., Townsend, J., Sejnowski, T.J., Courhesne, E.: Functionally independent components of the late positive event-related potential during visual spatial attention. J. Neurosci. **19**(7), 2665–2680 (1999)
63. Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.: Dynamic brain sours of visual evoked responses. Science **295**(5555), 690–694 (2002)
64. Martín-Clemente, R., Camargo-Olivares, J.: Independent Component Analysis for Audio and Biosignal Applications, Chap. ICA-Based Fetal Monitoring, pp. 247–268. Intech, Vienna (2012)
65. Martín-Clemente, R., Camargo-Olivares, J., Hornillo-Mellado, S., Elena, M., Román, I.: Fast technique for noninvasive fetal ECG extraction. IEEE Trans. Biomed. Eng. **58**(2), 227–230 (2011)
66. McCubbin, J., Robinson, S., Cropp, R., Moiseev, A., Vrba, J., Murphy, P., Preissl, H., Eswaran, H.: Optimal reduction of MCG in fetal MEG recordings. IEEE Trans. Biomed. Eng. **53**(8), 1720–1724 (2006)
67. Müller, K.R., Vigário, R., Meinecke, F., Ziehe, A.: Blind source separation techniques for decomposing event-related brain signals. Int. J. Bifurcat. Chaos **14**(2), 773–791 (2004)
68. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381**, 607–609 (1996)
69. Olshausen, B.A., Field, D.J.: Natural image statistics and efficient coding. Network **7**, 333–339 (1996)
70. Ossadtchi, A., Baillet, S., Mosher, J., Thyerleid, D., Sutherling, W., Leahy, R.: Automated interictal spike detection and source localization in magnetoencephalography using independent component analysis and spatio-temporal clustering. Clin. Neurophysiol. **115**, 508–522 (2004)
71. Phlypo, R., Zarzoso, V., Lemahieu, I.: Atrial activity estimation from atrial fibrillation ECGs by blind source extraction based on a conditional maximum likelihood approach. Med. Biol. Eng. Comput. **48**(5), 483–488 (2010)
72. Sanei, S., Chambers, J.A.: EEG Signal Processing. Wiley, New York (2007)
73. Särelä, J., Vigário, R.: Overlearning in marginal distribution-based ICA: analysis and solutions. J. Mach. Learn. Res. **4**, 1447–1469 (2003)
74. Shi-Yun, S., Kai-Quan, S., Chong-Jin, O., Wilder-Smith, E., Xiao-Ping, L.: Automatic EEG artifact removal: a weighted support vector machine approach with error correction. IEEE Trans. Biomed. Eng. **56**(2), 336–344 (2009)
75. Tang, A., Pearlmutter, B., Malaszenko, N., Phung, D.: Independent components of magnetoencephalography: localization. Neural Comput. **14**, 1827–1858 (2002)
76. Vigario, R.: Extraction of ocular artefacts from EEG using independent component analysis. Clin. Neurophysiol. **103**, 395–404 (1997)
77. Vigário, R., Oja, E.: BSS and ICA in neuroinformatics: from current practices to open challenges. IEEE Rev. Biomed. Eng. **1**, 50–61 (2008). doi:10.1109/RBME.2008.2008244
78. Vigário, R., Sarela, J., Jousmiki, V., Hainen, M., Oja, E.: Independent component approach to the analysis of eeg and meg recordings. IEEE Trans. Biomed. Eng. **47**, 589–593 (2000)
79. Wang, W., Zhang, Z., Gao, X., Gao, S.: Lead selection for SSVEP-based brain computer interface. In: Proceedings of the 26th International Conference on Engineering in Medicine and Biology Society (EBS 04), pp. 4507–4510 (2004)
80. Xu, N., Xiaorong, G., Hong, B., Xiaobo, M., Shangkai, G., Fusheng, Y.: BCI competition 2003 data set IIB: Enhancing p 300 wave detection using ICA-based subspace projections for BCI applications. IEEE Trans. Biomed. Eng. **51**(6), 1067–1072 (2004). doi:10.1109/TBME.2004.826699

Chapter 17
Supervised Normalization of Large-Scale Omic Datasets Using Blind Source Separation

Andrew E. Teschendorff, Emilie Renard and Pierre A. Absil

Abstract Biotechnological advances in genomics have heralded in a new era of quantitative molecular biology whereby it is now possible to routinely measure over tens of thousands of molecular features (e.g., gene expression levels) in hundreds if not thousands of patient samples. A key statistical challenge in the analysis of such large omic datasets is the presence of confounding sources of variation, which are often either unknown or only known with error. In this chapter, we present a supervised normalization method in which Blind Source Separation (BSS) is applied to identify the sources of variation, and demonstrate that this leads to improved statistical inference in subsequent supervised analyses. The statistical framework presented here will be of interest to biologists, bioinformaticians and signal processing experts alike.

17.1 Introduction

Omic and sequencing technologies have revolutionized the biomedical field [40]. With these technologies, it is now possible, at a reasonable economic cost, to measure the levels of molecular entities, for instance, gene expression, genome-wide,

A. E. Teschendorff (✉)
Statistical Cancer Genomics, UCL Cancer Institute, 72 Huntley Street, London WC1E 6BT, UK
e-mail: a.teschendorff@ucl.ac.uk

A. E. Teschendorff
CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences, Shanghai Institute for Biological Sciences, 320 Yue Yang Road, Shanghai 200031, China

E. Renard · P. A. Abseil
Department of Mathematical Engineering, ICTEAM Institute, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
e-mail: emilie.renard@uclouvain.be

P. A. Abseil
e-mail: absil@inma.ucl.ac.be

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_17,
© Springer-Verlag Berlin Heidelberg 2014

in cellular specimens from large numbers of patients [8]. Analysis of these large genomic, more generally refered to as "omic", datasets promises to provide the advances and biomarkers, which are urgently needed in the biomedical field, heralding in the new age of personalized medicine [34]. However, a serious obstacle in translating these mammoth amounts of data into biomedical advances is the presence of confounding factors, both technical and biological [21]. Recent studies [21, 43] have shown that technical confounding factors, generally refered to as batch effects, for instance the date in which a sample was processed, are omnipresent in omic datasets, affecting even some of the highest-profile studies such as The Cancer Genome Atlas [46], or the 1,000 Genomes Project [7]. Some estimates indicate that in any given study up to 80 % of measured molecular features can correlate with unwanted technical factors [21]. Furthermore, not adjusting for confounding factors can adversely impact statistical inference, compromising sensitivity and specificity [20, 45].

There are many reasons why these batch effects arise. Specially, in the case of large-scale studies profiling hundreds to thousands of samples, samples will inevitably have been processed on either different dates, by different laboratories or personnel, or on different plates or chips. Laboratory conditions can vary between dates affecting the biological measurements, or the quality of the profiling technology may also vary significantly from batch to batch. Moreover, profiled samples may come from patients treated at different medical centers, and therefore the way samples were handled (e.g., time from sampling to storage) may introduce further variation (see e.g., [25]). All of these factors have been shown to introduce unwanted variation in the data, and since *"the more you measure the more can go wrong"*, it is clear that large scale studies are particularly vulnerable to such confounding factors. On the other hand, it is worth pointing out that large-scale studies are also much better placed than small sample-size studies at adjusting for confounding factors. For instance, it is easier to detect and subsequently correct for a single chip/plate effect if there are many other chips/plates in the study that have performed well since the latter can then serve as controls.

The statistical design of a study is of critical importance in trying to prevent the potentially adverse effects of confounding factors on downstream statistical inference. Clearly, the statistical design of a study must be such so as to ensure that a number of specific research questions can be properly addressed. This typically requires that samples be distributed randomly across batches, ensuring balanced numbers of specific phenotypes across them. Thus, in comparing phenotypes A and B, one would randomize these across batches ensuring balanced numbers of A and B in each batch. However, it is not unusual for unbalanced designs to arise as a result of samples dropping out, in turn caused by logistical or quality control issues. This is particularly true for large-scale studies where logistical or quality control issues almost inevitably arise. These unbalanced designs can then have a dramatic negative impact on statistical inference if adjustment for the technical sources of variation is not performed. Thus, (large-scale) studies with an initial perfect study design may still be hampered by confounding factors.

There are a number of other key issues to mention in connection with confounding factors. First, it is clear that the potential impact of confounding factors will depend on the signal-to-noise ratio. This in turn depends on numerous study-specific factors, including the phenotype of interest, the nature of the confounding variation and the tissue type being profiled. For instance, if one is measuring DNA methylation, a covalent modification of DNA that can affect the activity of nearby genes [9], and if the comparison is between normal and cancer tissue, then it is likely that batch effects can be ignored, since DNA methylation changes associated with cancer are generally of a large magnitude (high signal-to-noise ratio limit) [46]. On the other hand, if the Epigenome-wide Association Study (EWAS) [31] measuring DNA methylation is being conducted in whole blood tissue [24], then this is likely to involve small effect sizes in relation to the technical sources of variation like chip effects, or biological factors such as age. For instance, in Rakyan et al. [31], the authors report a genomic site with a DNA methylation pattern in whole blood that correlates with smoking status, involving small 5–10 % shifts in average methylation between cases and controls. Such 5–10 % shifts could in principle be also caused by batch/chip effects. Similarly, such small shifts in average DNA methylation levels could be due to relatively small changes in blood cell type composition, which in turn could be caused by differences in the age of the sampled individuals [43]. Thus, techniques like Singular Value Decomposition (SVD) are specially useful for omic data since they easily allow approximate relative quantification of the variance associated with different sources of variation [43].

A second important issue is that the way in which statistical inference is affected strongly depends on how the confounders are correlated to the phenotype of interest (POI) [19]. Clearly, a confounding factor which is anti-correlated to a POI will dampen the statistical significance, while positive correlations will lead to overoptimistic results. An orthogonal confounder of large variability in relation to the POI signal will similarly compromise the statistical significance and lead to a large false negative rate (FNR). Thus, when analyzing omic data it is important to be aware of these different potential scenarios and generation of P-value histograms is strongly recommended as a means of detecting the strength and type of confounding [19].

Last but not least, confounding sources of variation can be of a very different nature, directly influencing the type of statistical adjustment procedure to be used. For instance, some confounders like plate or date, are examples of known confounders in the sense that we know exactly on which date and on which plate a given sample was processed, as these are factors that are normally recorded in an experiment. In this case, adjustment with (Bayesian) regression models, which use the confounders as explicit covariates, is possible and indeed fairly popular [16]. However, surprisingly often confounders are only known with uncertainty or error. For instance, in DNA methylation studies conducted with the Illumina Infinium beadchips, samples need to be preprocessed using a bisulfite conversion step, which translates epigenetic changes into genetic ones allowing these to be measured on the beadchip [4]. This conversion step is variable between samples and although the conversion efficiency can be measured using control probes on the beadchip, this measurement is subject to error. As another example, we have observed components of variation in DNA

methylation data associated with the season in which samples were collected. Season can be viewed as a surrogate for temperature, which is the more likely causal factor, yet the exact temperature to which the samples were exposed to during transportation from medical centers to the central processing lab was not recorded. At the other extreme, we may have confounders which are completely unknown, or there is no correlated known factor that could be used as surrogate. All these considerations are important in the context of this chapter, because clearly in the latter two scenarios, explicit adjustment for confounders is neither advisable or possible. Hence, BSS techniques are needed to infer these confounders from the data itself. On the other hand, as we shall see, known confounders also become useful in the BSS context, since they can be used to objectively evaluate the quality of blind source separation.

It is paramount to stress again the importance of adjusting for confounding factors, as not doing so can seriously reduce the effective power of the studies, or lead to unacceptably large false discovery rates [21, 45]. Thus, there is an urgent need for powerful statistical methods to be applied in the biomedical field to help address these significant challenges. To further motivate a BSS-based approach to statistical inference, we emphasize that it is only natural to view any biological omic dataset as an interference pattern, with some sources of variation reflecting the biological phenotype of interest, and others reflecting the effects of technical factors. Therefore, BSS methods are optimally placed to infer such sources of variation.

Indeed, BSS methods have already been extensively applied to omic data, but only as a means of performing dimensional reduction to identify *biological* sources of variation [12, 18, 22, 23, 28, 42, 49], and, secondly, as a means of performing feature selection and classification [14]. Specific popular BSS algorithms include Independent Component Analysis (ICA) [15] and non-negative matrix factorisation (NMF) [13], which have been applied to diverse data types, from gene expression [42] to DNA methylation data [51], including even mutational data [1] and multidimensional cancer genomic profiles [50]. The earliest studies already demonstrated that BSS methods like ICA and NMF lead to substantial improvements in modeling biological sources of variation and that these improvements are mainly due to the sparse (supergaussian) nature of the underlying biological sources [18, 42].

In contrast, relatively few BSS applications have focused on the problem of artifact removal in biomedical data, which is surprising given that technical sources of variation are omnipresent in such data and that they can so negatively affect statistical inference. We would also argue that the application of BSS methods to identify and remove technical artifacts in real omic data provides a substantially better framework in which to objectively evaluate BSS algorithms. There are several reasons for this. First, biological sources of variation such as activity of a molecular signaling pathway are "fuzzy" objects and only rarely can be used as defining a ground truth. On the other hand, technical artifacts are sometimes well known to the experimentalist performing the study and hence, as explained above, these can be exploited to assess the quality of BSS separation. Indeed, we recently demonstrated the feasibility of this conceptual framework for assessing BSS methods in a proof-of-principle study, analyzing both DNA methylation and gene expression data [45]. In that work, we proposed an algorithm called Independent Surrogate Variable Analysis

(ISVA), based on ICA, for performing supervised normalization in the presence of confounding factors [45], demonstrating its superiority over non-BSS based alternatives. The main purpose of this chapter is therefore to demonstrate that BSS methods can lead to substantial improvements in statistical inference in large omic datasets, thanks to a more efficient deconvolution of the confounding sources of variation. Our secondary aim is to increase the awareness among the BSS community of the importance of this fairly novel BSS application to artifact removal in biomedical omic data, and thus provide a fertile ground for interdisciplinary cross-pollination.

This chapter is organized as follows. First, because most of the examples considered in this chapter are drawn from studies in DNA methylation, we provide the reader with a brief introduction to DNA methylation and the Illumina Infinium Beadarray technology, a technology that allows genome-wide measurements of this epigenetic mark. In the subsequent section, we provide a number of examples of confounding variation in omic data and describe their negative impact on downstream statistical inference, including examples where methods based on explicit adjustment of confounders cannot be applied. In Sect. 17.3, we describe the problem of performing supervised analysis in the background of confounding factors, introducing and reviewing the SVA framework of Leek et al. [19, 20]. We argue theoretically why SVA may break down and why a BSS method is needed to avoid the pitfalls associated with SVA. This motivates the ISVA algorithm [45], which we review in the next subsection. In Sect. 17.4, we validate ISVA on simulated data and demonstrate the need for adjustment of confounding factors. In Sect. 17.5, we compare ISVA to SVA in modeling beadchip effects in real omic data. Section 17.6 provides a rigorous evaluation of ISVA on eight real omic datasets, using the non-BSS SVA method as well as another method based on explicit adjustment as benchmarks. In the final section, we briefly explore the performance of a generalized BSS algorithm in modeling beadchip effects. We end with conclusions and suggestions for further research.

17.2 DNA Methylation and the Illumina Infinium Beadarray Technology

DNA methylation refers to the covalent attachment of a methyl CH_3 group to DNA cytosines, normally, but not exclusively, in the context of a CG dinucleotide, refered to as a CpG [9]. There are about 30 million of such CpG sites in the human genome, most of which are methylated. These 30 million CpG sites represent in fact an underenrichment of CpGs in the human genome. In some genomic regions however, the density of CpGs is much higher than normal, and these are refered to as CpG islands. Roughly, about 60 % of gene promoters fall within CpG islands and most of these are normally unmethylated. Thus, whereas most of the genome is methylated, many of the promoter CpG islands are unmethylated in the normal state.

DNA methylation is important for a number of reasons. It is not only essential for embryonic development, but is also key in developmental processes [9]. Very recently, it has been demonstrated that differentially methylated regions between diverse normal cell types are enriched for transcription factor binding sites, supporting the view that DNA methylation is associated with how accessible the DNA is to transcription factors. Thus, hypomethylation, i.e., loss of DNA methylation, allows transcription factor proteins to more easily bind to DNA in order to initiate developmental differentiation programs. The DNA methylation state at the gene promoter is also a key determinant of the gene's activity, i.e., its gene expression level, with promoter hypermethylation normally associated with gene silencing [9]. DNA methylation is particularly important in diseases like cancer, where it is significantly altered [11, 17]. Indeed, a key cancer hallmark is the hypermethylation of CpG island promoters, whilst most of the cancer genome undergoes widespread hypomethylation. These deregulations in DNA methylation may lead respectively, to underexpression/silencing of key tumor suppressor genes, or overexpression of oncogenes (tumor promoting genes).

DNA methylation can be measured fairly accurately using a number of different technologies. In this chapter, we will be considering DNA methylation data generated using the Infinium beadarray technology from Illumina [4]. In particular, we will be considering a version of this technology, called Infinium 27k, that allows measurement of DNA methylation at over 27,000 CpG sites, mostly located within gene promoters of approximately 14,000 genes. The beadarray consists of a set of probes that interrogate the methylation state at each of these 27,000 sites. For each CpG site, there are two sets of probes, one designed to match the methylated version of the allele, while the other matches the unmethylated version. This is made possible by treating the DNA with bisulfite, prior to hybridisation to the beadarray. During bisulfite conversion, unmethylated cytosines are converted into uracil and then thymine upon DNA amplification (i.e., $uC \rightarrow T$), whereas methylated cytosines are protected and remain cytosines (i.e., $mC \rightarrow C$). Thus, an epigenetic difference can be translated into a genetic one, which is then easily measured using probes on the beadarray as described. While the methylation state of a given CpG site in a given diploid cell can take only three values ($0 =$ both alleles unmethylated, $1 =$ only one of the alleles is methylated, $2 =$ both alleles are methylated), in practice, measurement is taken over many thousands of cells, with the methylation state also being potentially variable between cells. Hence, methylation at a single CpG site in a given sample taken from an individual is quantified in terms of a β-distributed quantity, $\beta = M/(U + M)$, where M and U denote the intensities of the methylated and unmethylated versions of the allele, as estimated from the respective probes on the array. By construction, this β-value lies between 0 (unmethylated) and 1 (fully methylated).

A number of important features of the Illumina methylation beadarrays are worth mentioning. First, a maximum of 12 samples can be measured on any given beadchip. As with any technology, the quality of beadchips can vary from batch to batch. Also, the DNA quality of a sample can vary significantly, which would subsequently affect β-value estimates. For these reasons, the beadchips are equipped with a number

of control probes, each designed to measure the quality of a particular aspect of the assay. For instance, bisulfite conversion efficiency (BSC) could vary between samples, causing biases in the β-values, and this can be assessed using built-in control probes which measure the efficiency of bisulfite conversion.

17.3 Confounding Factors in Large-Scale Omic Studies

In order to illustrate the nature and impact of the problem posed by confounding factors, we consider two examples. These examples are taken from two separate DNA methylation studies generated with the Infinium 27k technology. Let us consider our first example. This is a DNA methylation dataset of whole blood samples from 187 individuals with type-1 diabetes, including both sexes, and with individuals drawn from two underlying cohorts. This particular dataset was used to test if DNA methylation changes correlate with the age of the individual at sample draw, thus age is here the POI [44]. The 187 samples were distributed over 17 different beadchips with at most 12 samples per beadchip. A SVD of the 27,578 $\times$ 187 row-centered (rows label CpGs) data matrix was performed to assess the nature of the largest sources of variation. As can be seen in Fig. 17.1, it is only the fifth component of variation that correlates with the POI (i.e., age), with the top components correlating with other factors such as sex, BSC and (bead)chip. Furthermore, it can be seen that the fifth component also correlates with chip indicating that this could be a potential confounder. This example further illustrates that technical or other biological variation can be of orders of magnitude larger than the effect size of interest.

As a second example, we consider a DNA methylation dataset of 48 samples, consisting of 30 normal samples from the cervix and 18 representing an intraepithelial cervical neoplasia of grade 2 or higher (CIN2+) (a preinvasive cancer condition). Here too, a SVD on the row-centered data matrix, reveals that it is only the third, fourth, and fifth components that correlate with biological factors such as age or CIN2+ status (Fig. 17.2a–b). Furthermore, unsupervised clustering of the samples does not lead to segregation of the samples according to CIN2+ status, as one would have expected on biological grounds (Fig. 17.2c). This example also illustrates that the top component of variation is correlating with an unknown factor, possibly spatial artifacts on the chips but which are also largely independent of chip. The key point to appreciate here is that there is no surrogate known factor that we can use to model this confounding source of variation, and hence explicit adjustment for this confounder using a multivariate regression model in which the confounder is included as a covariate is not possible [16].

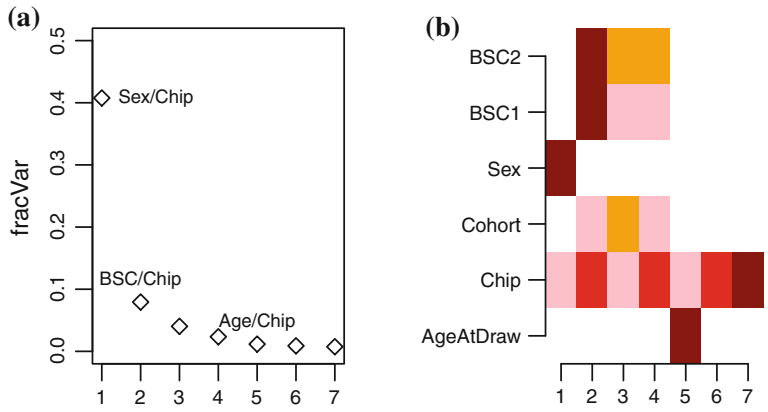


Fig. 17.1 **a** Relative fraction of variation carried by each of the seven significant singular vectors of a SVD, as measured relative to the total variation in the data. Number of significant singular vectors was estimated using Random Matrix Theory (RMT) [45]. Some of the singular values are labeled according to which confounders the corresponding singular vectors are correlated to, as shown in panel **b**. **b** Heatmap of P-values of association between the seven significant singular vectors and the phenotype of interest (here age at sample draw) and confounding factors (Chip, cohort, sex, and bisulphite conversion (BSC) efficiency controls 1 and 2). P-values were estimated using linear ANOVA models in the case of chip, cohort and sex, while linear regressions were used for age and BSC efficiency. Color codes: $P < 1e-10$ (*brown*), $P < 1e-5$ (*red*), $P < 0.001$ (*orange*), $P < 0.05$ (*pink*), $P > 0.05$ (*white*)

17.4 Supervised Normalization by SVA and ISVA

The previous examples illustrate some of the difficulties that confounding factors can pose in statistical analyses. One of the common tasks in omic data analysis is to perform a supervised analysis in which we seek to identify features associated with a phenotype of interest. Clearly, such task may be compromised by the presence of confounding factors, specially if the confounder is unknown or if it is only known subject to error, since in these cases we can't adjust for them explicitly. Thus, one desires a statistical framework in which to perform supervised analysis (i.e., feature selection) in the presence of uncertain or unknown confounding factors. We refer to this supervised analysis problem as "supervised normalization" in the sense that the normalization of the data is performed as part of the supervised analysis and is therefore dependent on the phenotype of interest. So far, only two algorithms, SVA [19, 20] and ISVA [45] have been proposed to address this problem in the context of omic data, where by definition the number of features is relatively large.

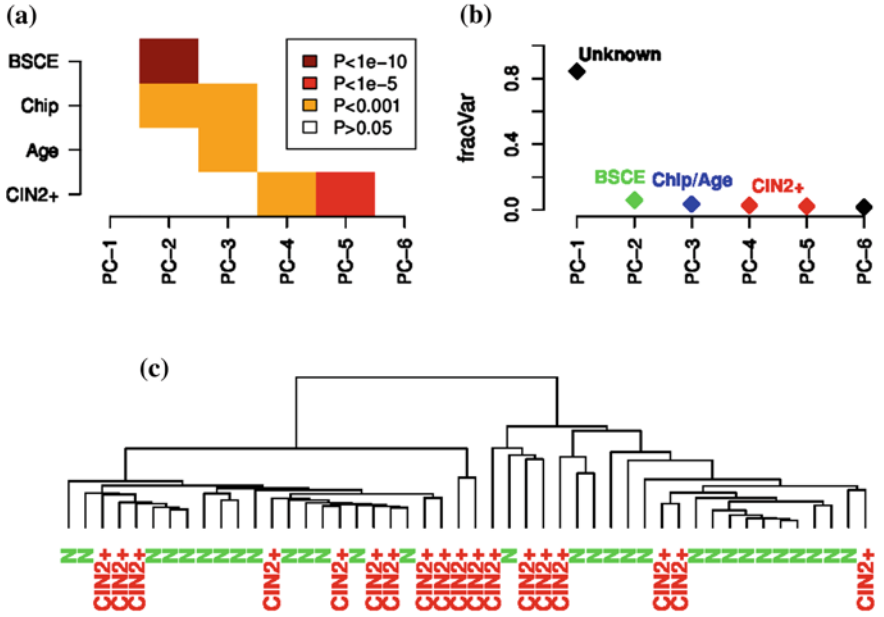


Fig. 17.2 Confounding variation in a DNA methylation dataset of 30 normal cervical samples and 18 cervical intraepithelial neoplasias of grade 2 or higher (CIN2+). **a** Relative fraction of variation carried by each of the six significant singular vectors of a SVD, as measured relative to the total variation in the data. Number of significant singular vectors was estimated using Random Matrix Theory (RMT) [45]. Some of the singular values are labeled according to which confounders the corresponding singular vectors are correlated to, as shown. **b** Heatmap of P-values of association between the six significant singular vectors and the phenotypes of interest (here CIN2+ status and age at sample draw) and confounding factors (Chip and bisulphite conversion efficiency (BSCE)). P-values were estimated using linear ANOVA models in the case of chip and CIN2+ status, while linear regressions were used for age and BSC efficiency. Color codes: $P < 1e-10$ (*brown*), $P < 1e-5$ (*red*), $P < 0.001$ (*orange*), $P > 0.05$ (*white*). **c** Hierarchical clustering of the 48 samples over the 5,000 most variable probes

17.4.1 Surrogate Variable Analysis

Leek and Storey proposed an ingenious solution to the problem posed above, known as SVA [19, 20], which we now describe. Let us assume that we have a data matrix, X_{ij}, with i ($i = 1, \ldots, p$) labeling the features (genes, CpGs,...) and j ($j = 1, \ldots, n$) labeling the samples, with $p \gg n$. Furthermore, we assume that each row of X has been mean centered, and that we have a POI encoded by a vector $\mathbf{y} = \{y_1, \ldots, y_n\}$. As in [20] we may allow for a general function of the phenotype vector, so that the starting model for SVA takes the form

$$X_{ij} = f_i(y_j) + \varepsilon_{ij}. \tag{17.1}$$

Typically, $f_i(y)$ would be a function of the form $f_i = b_i F(y)$ with b_i a feature specific regression parameter (to be estimated) and F representing a general link function. Thus, SVA starts by performing univariate regressions, leading to estimates $\hat{b}_i$ as well as an estimate of the error matrix ϵ, which we shall call the residual variation matrix, $R \equiv \hat{\epsilon}$. Componentwise, $R_{ij} \equiv X_{ij} - \hat{f}_i(y_j)$. SVA then proceeds by performing a SVD of the residual variation matrix

$$R = UDV^T. \tag{17.2}$$

Thus, the singular vectors of the SVD capture variation which is orthogonal to the variation associated with the POI. This residual variation is therefore likely to be associated with other biological factors, not of direct interest, or with experimental factors, all of which constitute potential confounders. SVA provides a prescription for the construction of surrogate variables, v_k ($k = 1, \ldots, K$ with $K < n$), in terms of the singular vectors (i.e., the column vectors of V) of this SVD [20]. In the final step, feature selection is performed using the modified regression model

$$X_{ij} = f_i(y_j) + \sum_{k=1}^{K} \lambda_{ki} v_{kj} + \epsilon'_{ij}. \tag{17.3}$$

with the rows of ϵ' now uncorrelated [19].

In the above framework, it is key to realize that SVA hinges on a big assumption, which is that we have a perfect, or at least a sufficiently accurate model $F(y)$ describing the data, such that the residual variation encapsulated by the matrix R does not contain any biological variation of interest (see left part of Fig. 17.3). In this case, the only requirement on the surrogate variables describing the confounding variation is that they span the residual variation space. We note that there is in fact no requirement for the surrogate variables (SVs) to align with (i.e., precisely model) the confounding factors.

However, now consider an alternative, and, as we shall see later, a more realistic scenario, where model $F(y)$ is imperfect. For instance, we may be using a linear function F when the relation between data and POI is highly nonlinear. In this case, residual biological variation of interest may be present in R (see right part of Fig. 17.3). In such a scenario, we would want our SVs to align with the confounding factors and not with the residual biological variation, since otherwise inclusion of this in the subsequent adjusted supervised analysis (Eq. 17.3) would lead to a reduced biological signal. Later we shall see examples of this happening. Hence, in this more realistic scenario, we need to choose SVs that span a *subspace* of R, i.e., one that is also orthogonal to the residual biological variation. This in turn means that we need an algorithm that can more accurately deconvolve the confounding sources from the residual biological variation. As one might expect (and we shall see examples of this later), the SVD used in SVA can not accurately deconvolve these different sources of variation. This motivates the introduction of BSS methods in the context of supervised normalization.

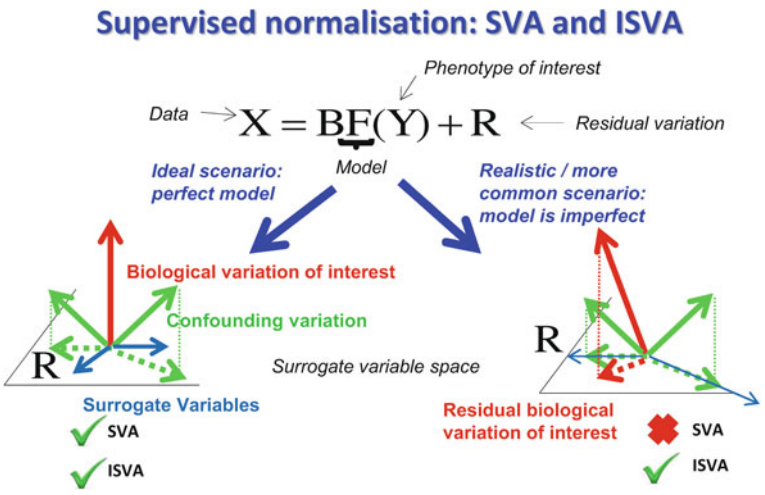


Fig. 17.3 Surrogate Variable Analysis (SVA) begins by performing a regression of the data matrix, X, against the phenotype of interest, Y, specified through a possibly nonlinear function $F(Y)$. In the equation above, B denotes regression parameters, whereas R denotes the residual variation, i.e., the variation in the data not explained by the phenotype of interest under the specified model F. Under such a model, there are two possible scenarios. In the ideal scenario (*left pointing arrow*), $F(Y)$ models the data perfectly in the sense that the residual variation space, depicted by the plane R, contains no residual biological variation of interest. In this case, the surrogate variables, which are estimated from a SVD of R, and are indicated by blue arrows, don't need to align with the confounding factors (*green arrows*), as they are only required to span the same plane R. However, in the more realistic scenario, there could be imperfections in the model $F(Y)$ (e.g., using a linear model when the relationship between X and Y is nonlinear), which in turn could lead to residual biological variation (*red arrow*) in the residual variation space R. In this case, we need to choose surrogate variables that align with the confounders and "avoid" the residual biological variation of interest, since otherwise using the whole space R in the subsequent adjustments will lead to loss of biological signal. Thus, in this scenario, we need to select an appropriate subspace of R and only use this subspace for the subsequent adjustments and supervised analysis. ISVA uses ICA instead of PCA/SVD in the decomposition of R, thus allowing to infer surrogate variables that better model the confounding sources of variation. Geometrically, this means that the independent surrogate variables align significantly better with the confounders and the residual biological variation, thus allowing an appropriate subspace of R to be selected. This subspace should not contain any residual biological variation and ICA is key to achieving this

17.4.2 Independent Surrogate Variable Analysis

Motivated by the discussion above, we seek a BSS method that can more accurately infer the sources of variation in the estimated residual matrix R. The generalization of SVA in which a BSS method is used to decompose R is called ISVA [45]. Although many BSS methods exist, in [45] we considered one of the simplest versions of ICA, the "fastICA" algorithm [15]. Thus, as with SVA, there are three parts to the ISVA algorithm: (i) detection of confounding/unmodeled factors (steps 1–4),

(ii) construction of surrogate variables (SVs) (steps 5–10), and (iii) final feature selection using the SVs as covariates.

In detail, the steps in ISVA are:

1. Construction of the residual variation matrix by removing the variation associated with the phenotype of interest: $R_{ij} \equiv X_{ij} - \hat{f}_i(y_j)$.
2. We estimate the intrinsic dimensionality, K, of the residual variation matrix using RMT [29]. This gives the number of components as input to the ICA algorithm.
3. Perform ICA on R: $R = SA + \epsilon$, with S a $p \times K$ source matrix and A a $K \times n$ mixing matrix. We point out that in this formulation of ICA, the statistical independence requirement is imposed on the columns of S. We denote the columns of S and rows of A by S_k and A_k, respectively.
4. We regress A_k to each X_i ($i = 1, \ldots, p$) and calculate P-values of association p_i.
5. From this P-value distribution, we estimate the FDR using the q-value method [38] and select the features with $q < 0.05$. If the number of selected features is less than 500, we select the top 500 features (based on P-values). Let r_k denote the number of selected features.
6. We construct the reduced $r_k \times n$ data matrix X_r obtained by selecting the features in previous step.
7. Perform ICA on X_r using K independent components: $X_r = S_r A_r + \epsilon_r$. Find the column k^* of A_r that best correlates (absolute correlation) with A_k.
8. Set the SV $v_k = (A_r)_{k^*}$. The purpose of steps 4–8 is to regularize the estimates and thus avoid overfitting as explained in [20].
9. Repeat steps 4–8 for each significant independent component, A_k, obtained in step-3.
10. Perform SV subspace selection using a SV selection criterion. Let K^* denote the set of selected SVs.
11. Finally, we run the model

$$X_{ij} = f_i(y_j) + \sum_{k \in K^*} \lambda_{ki} v_{kj} + \epsilon'_{ij}. \tag{17.4}$$

and perform feature selection using a FDR (q-value) estimation procedure [38] and a nominal q-value threshold of say 0.05.

As formulated above, there are three differences between ISVA [45] and SVA [19]. First, ISVA uses RMT to estimate the dimensionality, in contrast to SVA which uses an explicit randomization procedure [20]. This difference is, however, not of major consequence [45]. Second, ISVA uses ICA in step-3 instead of SVD. Third, ISVA incorporates a SV subspace selection step (step-10) using a SV selection criterion that we shall discuss in detail in Sect. 17.7.4. This step is absolutely key to the improved inference that ISVA offers, and we point out here that the use of a BSS method in step-3 is also key to facilitating the choice of SV subspace in step-10. Finally, we remark that any BSS technique could be used to model the sources of variation in

R (step-3), and thus the ISVA framework can be easily generalized to incorporate more sophisticated BSS algorithms.

17.5 Validation of SVA and ISVA on Simulated Data

Before exploring the SVA and ISVA algorithms in the context of real data, it is illuminating to first compare their performance on simulated data. The simulation model is exactly the one considered in [45], and for completeness we provide full details here again in the appendix. Briefly, we generated synthetic data matrices with 2,000 features and 50 samples and considered the case of two confounding factors (CFs) in addition to the primary POI. The primary phenotype is a binary variable y with 25 samples in one class ($y = 0$) and the other half with $y = 1$. Similarly, each confounding factor is assumed to be a binary variable affecting one half of the samples (randomly selected). We further assume 10 % of features (200 features) to be true positives (TPs) discriminating the two phenotypic classes. We model the confounding factors as follows: each confounding factor is assumed to affect 10 % of features with a 25 % overlap with the TPs (i.e., 50 of the 200 TPs are confounded by each factor). Without loss of generality, noise is modeled by a Gaussian of mean zero and unit variance $N(0, 1)$. We further assume that the POI is associated with an effect size $e_y (= \Delta\mu/\sigma)$ of 1, i.e., the difference in the means between the phenotypes, $\Delta\mu$, equals the standard deviation, σ, within each group. Effect sizes of the two confounders are assumed to be equal to e_{CF} and we define the relative effect size as $e_R \equiv e_{CF}/e_y = e_{CF}$. We here consider the case $e_R = 2$ corresponding to a situation where the confounding factors are associated with a larger variance than the POI. The simulation model is run a total of 100 times and for each run we record the following measures (using an estimated FDR threshold of 0.05): the sensitivity (SE), the positive predictive value (PPV), the sensitivity of TPs specifically affected by the confounding factors (SE-A), and the overall correlation (R^2-values) to the CFs. For the first three measures, we also compare SVA and ISVA to a simple linear regression method that does not do any adjustment for the confounding factors (LR). Results are shown in Fig. 17.4.

From this figure, we can make the following observations. First, the PPV is high for all methods, and is in line with the estimated FDR (=1-PPV) of 0.05 used in performing feature selection. Second, we can see that the power of the study is reduced if no adjustment is made for the confounding factors. Indeed, we can see that, focusing on those true positive features which are corrupted by confounding variation, the sensitivity to retrieve these features is improved approximately twofold by using SVA or ISVA. Third, ISVA and SVA perform similarly on simulated data, despite the fact that ISVA reconstructs the confounding factors at substantially higher R^2 values. Thus, the simulated data nicely illustrates the "perfect model" scenario depicted in the left side of Fig. 17.3. Since the data are simulated with the same model that is subsequently used to run the univariate regression, the residual variation matrix R contains no residual biological variation, hence it does not matter if the SVs align

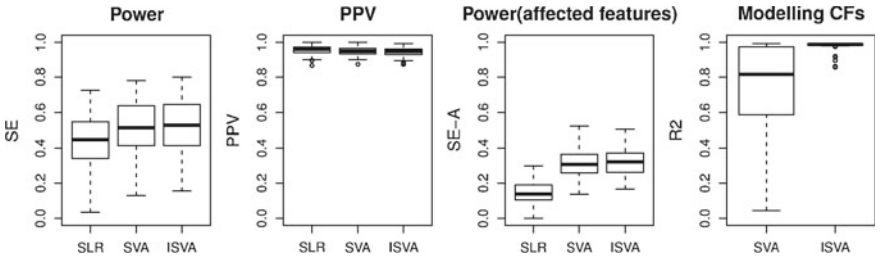


Fig. 17.4 Feature selection performance metrics of different algorithms over 100 runs of the synthetic data ran with $e_R = 2$. The algorithms for feature selection are SVA, ISVA, and a simple linear regression without adjustment for confounders (SLR). For a given estimated FDR threshold of 0.05, we compare the sensitivity/power (SE), the positive predictive value (PPV), the sensitivity to detect true positives which are affected by confounders (SE-A), and the average R^2-value between confounders and the best correlated surrogate variable. See Appendix for further details of simulation model

with the confounders. The main requirement is for the SVs to span the space R, and hence similar results are obtained using the SVs from SVA or ISVA, since in both cases, the SVs span the same space.

17.6 Improved Modeling of Confounding Factors in Omic Data by BSS Methods

In the previous section, we have seen how ISVA models the confounding factors much better than SVA. The aim of this section is to demonstrate that ISVA also leads to improved modeling of the confounding sources of variation in real data. Later, in the subsequent section, we shall see how this translates into improved feature selection. Once again, we consider DNA methylation data and as confounding factor we consider the beadchip. Illumina Infinium beadchips can accommodate up to 12 samples per chip, hence there are enough samples for beadchip effects to be assessed. Importantly, it is always known which samples were profiled on which beadchip, hence this is an example of a known confounder and thus it can be used to objectively assess the quality of blind source separation. As a benchmark we consider SVA which uses SVD/PCA to decompose the residual variation matrix. As shown in Fig. 17.5, the surrogate variables inferred using ISVA model the beadchip effects substantially better than those inferred using SVA, as indicated by the significantly higher R^2 values. For further examples, we refer the reader to [45].

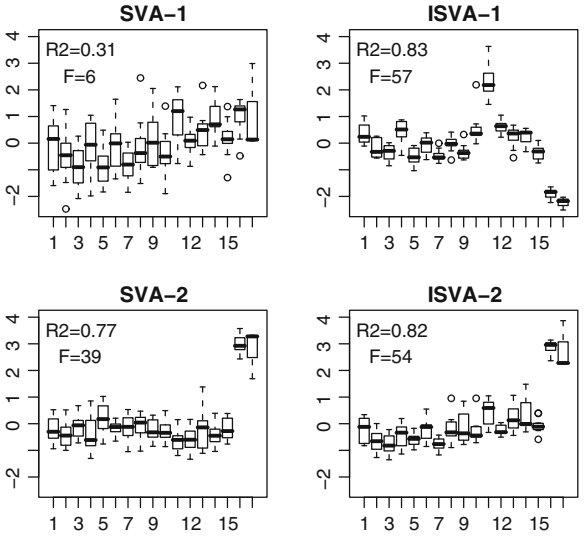


Fig. 17.5 Comparison of ISVA to SVA in identifying beadchip effects in the DNA methylation dataset from [3]. The weights (y-axis) of the two surrogate variables that most significantly associated with beadchip effects are plotted against beadchip number (x-axis), for SVA and ISVA separately. To compare the identifiability of beadchip effects, we provide the R^2 and F-statistics of a linear ANOVA model with beadchip number as the independent variable

17.7 Improved Feature Selection Using ISVA

We have seen that ISVA can model confounding sources of variation substantially better than SVA. This in turn should lead to improved statistical inference, e.g., feature selection, at least in those scenarios where it is necessary to select a surrogate variable subspace, as explained in Sect. 17.3. To demonstrate this, we first provide a number of real data examples where SVA breaks down. Subsequently, we show how ISVA circumvents the problem, leading to substantially improved statistical inference.

17.7.1 SVA Breakdown in mRNA Expression Data

In order to demonstrate that SVA can break down, we consider a real dataset with a known biological signature: it is well known that many genes implicated in cell proliferation and the cell-cycle are differentially expressed between high and low grade cancers [26, 32, 36, 41]. The grade of a cancer refers to the level of differentiation of the cancer cells, with high-grade cancers exhibiting a less differentiated state, whilst low-grade cancers are more differentiated in the sense that they are

more similar to normal (healthy) tissue, which is a highly differentiated state compared to the undifferentiated stem cells that they are derived from. Thus, high-grade cancers are generally more aggressive and correspondingly are also characterized by a higher expression of cell proliferation and cell-cycle genes. This cell proliferation gene expression signature is a universal signature, able to distinguish high grade from low-grade cancers, irrespective of tissue type [26, 32, 36, 41]. Thus, given a gene expression dataset of high and low grade cancers, selecting features (genes) that best discriminate low and high grade cancers should lead to significant enrichment of genes implicated in the cell-cycle and cell proliferation. The enrichment of a top ranked list of discriminatory genes for any gene ontology can be assessed using a Fisher's exact test, as done for instance in [43], a procedure known generally as Gene Set Enrichment Analysis (GSEA) [39]. If a feature selection method were to not yield significant enrichment for cell-cycle or cell proliferation genes, one would conclude that the feature selection procedure has failed to retrieve the known biological signature. Thus, in what follows we consider "grade" as the POI and we aim to show that SVA breaks down, not being able to retrieve the cell proliferation/cell-cycle enrichment due to the presence of confounding factors.

Specifically, we consider the case of breast cancer. There are two main subtypes of breast cancer: estrogen receptor positive (ER+) and estrogen receptor negative (ER−) breast cancer [48]. This stratification of breast cancers reflects the levels of expression of the estrogen receptor gene, *ESR1*, with ER− breast cancers showing absent expression of *ESR1*. Thus, in ER+ breast cancer, *ESR1* expression and activity is high, which results in the overexpression of genes within the *ESR1* signaling pathway. We note that these *ESR1* signaling genes are different from the cell-cyle/cell-proliferation ones. Now, it is well known that most ER− breast cancers are of high grade, whilst ER+ breast cancers can be either high or low grade [41]. Thus, if the aim is to identify genes whose expression correlates with grade, ER−status may be seen as a biological confounder, since the distribution of ER+ and ER− tumors will differ between low and high-grade cancers. Furthermore, it is also well known that low and high grade ER+ breast cancers do not differ in terms of the level of *ESR1* expression and ER−signaling [26, 36, 41]. Hence, this means that in the task of identifying genes that are associated with grade, any gene set enrichment must be specific to cell-cycle and should not include terms involved in ER−signaling. In other words, if feature selection for grade associated genes also leads to enrichment of ER−signaling genes, then this indicates confounding by ER−status. Although here the confounder is biological, this does not matter for the sake of comparing algorithms, and indeed the biological framework considered here provides a nice testing ground for the SVA and ISVA algorithms.

As expression data, we consider the data from four independent breast cancer studies [5, 26, 35, 36], as used in [45]. In these datasets, besides ER−status, we also consider tumor size as a potential biological confounder. We note that in these datasets potential technical confounders such as batch effects are unknown. The P-values of the GSEA of the top ranked grade-associated genes against cell-cycle and ER−signaling terms are given in Table 17.1 for genes selected using SVA and a

Table 17.1 Grade associated expression differences: in each mRNA expression dataset and for each method (LR + CFs, SVA, ISVA) we give the number of confounding factors (CFs) or SVs used as covariates in the regression analysis, the number of genes differentially expressed with histological grade (nDEGs) at a false discovery rate threshold of 0.05 (FDR < 0.05), and the *P*-value of enrichment (Hypergeometric/Fisher test) of cell-cycle and estrogen upregulated gene (ESR1-UP) categories among these differentially expressed genes

	LR + CF	SVA	ISVA
Dataset(Sotiriou)			
nCF/SV	2	4	4
nDEGs	491	0	607
P-value(Cell-cycle)	**6e-18**	1	**5e-16**
P-value(ESR1-UP)	0.03	1	0.14
Dataset(Loi)			
nCF/SV	2	19	5
nDEGs	829	0	146
P-value(Cell-cycle)	**5e-37**	1	**7e-24**
P-value(ESR1-UP)	0.90	1	0.61
Dataset(Schmidt)			
nCF/SV	2	27	15
nDEGs	2364	0	451
P-value(Cell-cycle)	**3e-25**	1	**5e-19**
P-value(ESR1-UP)	**7e-4**	1	0.14
Dataset(Blenkiron)			
nCF/SV	2	20	8
nDEGs	1292	1	829
P-value(Cell-cycle)	**2e-25**	1	**7e-27**
P-value(ESR1-UP)	**7e-4**	1	0.31

Confounding factors here are ER status and tumor size. In bold face we indicate those *P*-values that are significant after adjustment for multiple-testing

feature selection method that uses ER−status and tumor size as explicit covariates in the linear regression model (LR + CF).

Based on this table, we can make two important observations. First, in three datasets, SVA predicts no differentially expressed genes between low and high grade breast cancer, a result which is in complete disagreement with extensive biological knowledge [26, 32, 41]. As a result of this, none of the biological terms cell-cycle or ER−signaling are enriched. Second, performing feature selection using a multivariate linear regression model with ER−status and size as explicit covariates (LR + CF) leads to many differentially expressed genes (DEGs) in every dataset. Correspondingly, we observe strong enrichment of the cell-cyle term among these genes, consistent with biological knowledge. However, we also observe that ER−signaling is significantly enriched in 2 out of 4 studies, hence the enrichment for cell-cycle genes is nonspecific. This means that explicit adjustment for the confounders has not fully eliminated the effect of one confounder (ER−status) and hence we can conclude that the list of DEGs contains many false positives associated with ER−signaling. This contamination of ER−signaling genes is likely to be due to the

Table 17.2 Age-associated CpGs: in each dataset and for each method (LR + CFs, SVA, ISVA) we give the number of CFs or SVs used as covariates in the regression analysis, the number of CpGs differentially methylated with age (nDMCs) (FDR < 0.05 for Datasets T1D and UKOPS1, FDR < 0.3 for Datasets UKOPS2 and WBBC), the number of these that are hypermethylated with age and that map to polycomb group targets (nPCGTs), and the *P*-value of PCGT enrichment among age-hypermethylated CpGs (Hypergeometric test)

Dataset(T1D)	LR + CF	SVA	ISVA
nCF/SV	4	4	6
nDMCs	440	688	902
nPCGTs	96	110	148
P-value	**4e-32**	**7e-26**	**2e-34**
Dataset(UKOPS1)			
nCF/SV	3	18	6
nDMCs	267	4	232
nPCGTs	75	1	59
P-value	**4e-24**	0.27	**2e-19**
Dataset(UKOPS2)			
nCF/SV	3	21	8
nDMCs	20	201	225
nPCGTs	4	15	29
P-value	0.001	0.01	**3e-7**
Dataset(WBBC)			
nCF/SV	3	15	6
nDMCs	564	185	469
nPCGTs	84	19	64
P-value	**7e-22**	0.01	**3e-11**

The CFs in each dataset are described in Appendix. In bold-face we indicate those *P*-values that are significant after adjustment for multiple-testing

fact that the immunohistochemically determined ER−status of the samples is only approximate, i.e., the confounder is subject to error. Thus, neither method, SVA or LR + CF, succeeds in yielding specific enrichment of cell-cycle genes among the genes associated with grade.

17.7.2 SVA Breakdown in DNA Methylation Data

As a second example, we consider DNA methylation data. A large number of studies have now unequivocally demonstrated that promoter DNA methylation of a specific class of genes, known generally as PolyComb Group Targets (PCGTs), increases with the age of the tissue (see e.g., [27, 30, 44]). Hence, feature selection for CpGs in gene promoters undergoing age-associated increases in DNA methylation should be enriched of PCGTs. Table 17.2 shows the results of applying SVA and a linear regression method that uses confounders as explicit covariates (LR + CF).

We can see that in only one of the four datasets (T1D set), does SVA convincingly retrieve the age-PCGT DNA methylation signature. In the other three datasets, the P-value of enrichment is either not significant or would fail significance after correction for multiple testing. In contrast, linear regression with explicit adjustment for confounders (see Appendix for the nature of the explicit confounders) convincingly captures the biological signature in 3 out of 4 datasets.

17.7.3 Residual Biological Variation

The results presented above clearly demonstrate a pitfall of the SVA algorithm: it can fail to retrieve a well-known and extensively validated association between a molecular signature and a phenotype of interest. The most plausible explanation for why this happens is that residual biological variation is being interpreted as confounding variation leading to a "dampening" of the biological signal (see Fig. 17.3). To show that this is indeed what is happening we can study the correlations between the surrogate variables and the biological as well as confounding factors. The statistical significance of these correlations is best shown as a heatmap. This is shown for the four DNA methylation datasets considered in Table 17.2 in Fig. 17.6. From this figure and Table 17.2 we can see that in all three datasets where SVA fails to clearly capture the age-PCGT DNA methylation signature, that in all three of them there is residual variation correlating with age. Conversely, in the one dataset where there is no residual variation correlating with age (i.e., T1D set), SVA retrieves the biological signature. Thus, this example clearly illustrates that the scenario of residual biological variation arising due to imperfections in the modeling, as depicted in Fig. 17.3, is indeed fairly common.

17.7.4 The Need for Surrogate Variable Subspace Selection

The above two examples in gene expression and DNA methylation data demonstrate the need to perform adjustment on a surrogate variable subspace, since otherwise one risks "peeling" away biological variation of interest. In the case where there is no residual biological variation it should be clear that it does not matter what basis (i.e., surrogate variables) we use to span the surrogate variable subspace. In other words, it should not matter whether we use SVs constructed from principal components (SVA) or from the independent components (ISVA). However, in the scenario where biological variation of interest is present in the residual variation matrix R, we need to select surrogate variables that "align" with the true confounders and which avoid as much as possible the directions defined by the residual biological variation. This then requires a BSS method to better deconvolute the effects of the confounders and this residual biological variability. However, application of a BSS method to R only yields a decomposition of R into a number of independent "sources" and does not,

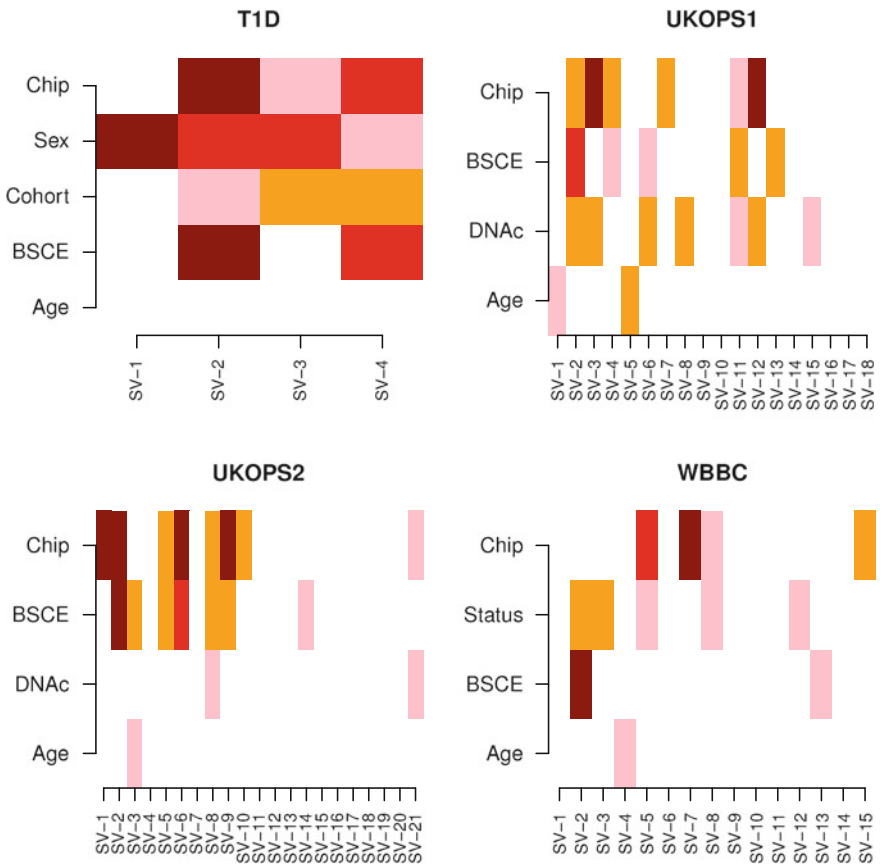


Fig. 17.6 Heatmap of P-values of association between the surrogate variables (SVs) inferred using SVA and the confounders and phenotype of interest (age). P-values were estimated from linear ANOVA in the case of categorical confounders (e.g Chip, Sex, Cohort) and from linear regressions in the case of continuous variables (age, BSC efficiency-BSCE and DNA concentration-DNAc). Color codes: $P < 1e-10$ (*darkred*), $P < 1e-5$ (*red*), $P < 0.001$ (*orange*), $P < 0.05$ (*pink*), $P > 0.05$ (*white*)

on its own, provide a prescription for subspace selection. Hence, how do we select this subspace?

The previous example discussed in Table 17.2 and Fig. 17.6 provides a possible prescription for how to perform the subspace selection, namely, only those SVs should be included that do not correlate significantly with the phenotype of interest. But what if SVs correlate significantly with both the POI and a confounder? In this scenario, it is unclear whether to include these SVs in the final feature selection procedure (i.e., step-11). The surrogate variable selection step therefore remains an outstanding problem.

Table 17.3 Surrogate Variable Selection: there are four possible case scenarios to consider depending on the R^2_{vf} values between surrogate variable v and factor f, as shown

Scenarios	POI($f = b$)	CF($f = t$)	ISVA
Case-1	$P_{vb} < 0.001$	$P_{vt} > 0.001$	Exclude
Case-2	$P_{vb} > 0.001$	$P_{vt} < 0.001$	Include
Case-3	$P_{vb} < 0.001$	$P_{vt} < 0.001$	Include if $R^2_{vb} < R^2_{vt}$
Case-4	$P_{vb} > 0.001$	$P_{vt} > 0.001$	Normally include

POI phenotype of interest ($f = b$), *CF* technical confounder ($f = t$). P_{vf} denotes the P-value of the association between SV v and factor f. Final column indicates whether the SV v should be included in the final adjustment step of ISVA or not. A conservative Bonferroni threshold of 0.001 is used to call statistical significance since the number of SVs is typically on the order of $\sim$10

Here we propose a simple heuristic to the subspace selection problem, which we can only justify a posteriori, by showing that it leads to successful retrieval of the known biological signatures. For each of the SVs and for each factor (biological or technical) we first compute a model fit R^2 value, using an appropriate linear or nonlinear model framework. Let R^2_{vf} denote the R^2 value between surrogate variable v and factor f. Further, let b denote the POI factor, and t denote a generic technical factor. Then, there are four possible cases to consider, as indicated in Table 17.3. In case-1, the surrogate variable correlates significantly only with the POI, and hence it ought to be excluded as remarked earlier. Conversely, if the surrogate variable correlates significantly with a technical factor but not with the POI, then the corresponding SV should be included. In the third case, where the SV correlates significantly with both the POI and a technical CF, we use the model selection criterion

$$R^2_{vb} < R^2_{vt} \tag{17.5}$$

to include only those where the correlation with the technical factor is stronger. The rationale for this criterion is that if the variation described by v correlates more strongly with the POI, then it is more likely that this variation is genuinely associated with the POI, and hence this component should be excluded. The final case corresponds to a scenario where the SV does not correlate with any known factor, in which case it is also unclear whether to include the SV or not. In principle, one must allow for the possibility of complete unknown (i.e., hidden) factors, in which case the SV should be included. On the other hand, exclusion could be argued on grounds of small variability and inaccuracies in dimensionality estimation.

Before demonstrating that the simple procedure presented in Table 17.3 works, we need to discuss further what may seem as a serious drawback with the above heuristic, as it requires some knowledge of the technical confounding factors. Given that BSS methods are ideally suited to the scenario where sources of variation are unknown, does this then pose an intrinsic limitation to the ISVA method? The answer is no. To understand this, we first note that BSS methods are useful also in circumstances where confounders are only known with error, since in such cases it would be better to model the effects of the confounders from the data itself. In this case, the simple

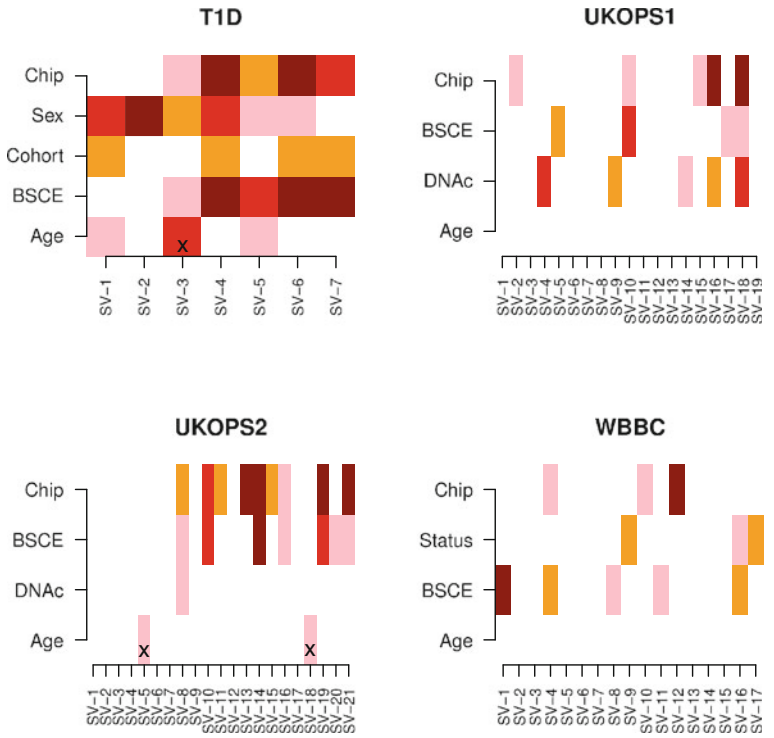


Fig. 17.7 Heatmap of P-values of association between the surrogate variables (SVs) inferred using ISVA and the confounders and phenotype of interest (age). P-values were estimated from linear ANOVA in the case of categorical confounders (e.g., Chip, Sex, Cohort) and from linear regressions in the case of continuous variables (age, BSC efficiency-BSCE and DNA concentration-DNAc). Color codes: $P < 1e-10$ (*darkred*), $P < 1e-5$ (*red*), $P < 0.001$ (*orange*), $P < 0.05$ (*pink*), $P > 0.05$ (*white*)

SV subspace selection step described above can be applied. Second, the scenario where confounders are known, or only known subject to error, constitutes the most common scenario. Last but not least, SVs not correlating with any factor (case-4) may still be included in the adjustment, as the main requirement is to avoid including SVs that correlate strongly with the POI.

17.7.5 The ISVA Solution

Let us now see how ISVA resolves the problematic issues that we encountered earlier with SVA. We first consider the four DNA methylation datasets considered in Table 17.2 and Fig. 17.6. In Fig. 17.7 we show the heatmap of associations between SVs constructed from ISVA with the same confounders.

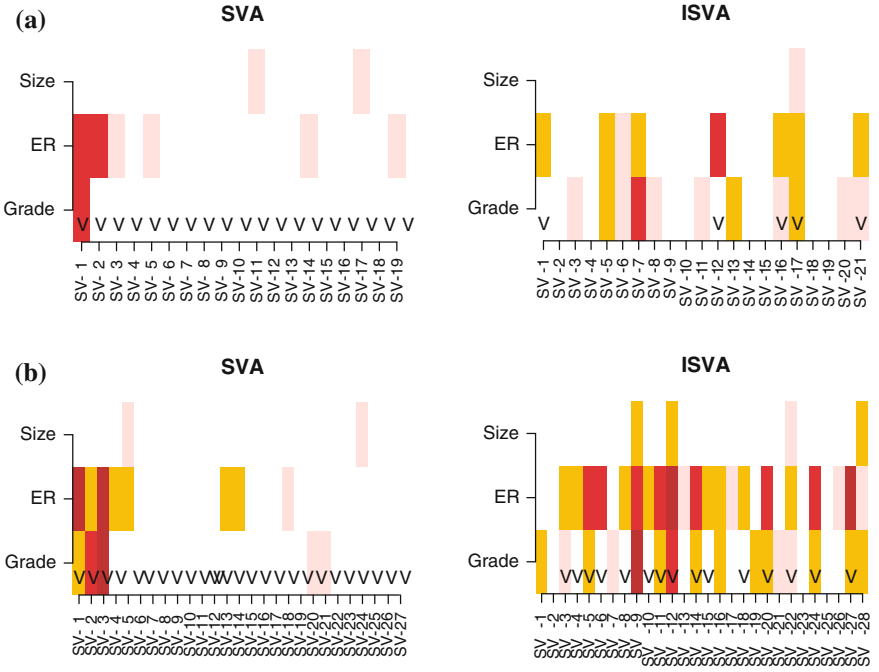


Fig. 17.8 Heatmap of P-values of association between the surrogate variables (SVs) inferred using SVA and ISVA and the confounders (ER−status and tumor size) and the phenotype of interest (Grade). **a** Dataset Loi, **b** Dataset Schmidt. P-values were estimated from linear regressions. Color codes: $P < 1e-10$ (*darkred*), $P < 1e-5$ (*red*), $P < 0.001$ (*orange*), $P < 0.05$ (*pink*), $P > 0.05$ (*white*). "V" indicates SVs selected for adjustment in SVA or ISVA

Note how in two datasets (UKOPS1 and WBBC) there is no residual biological variability associated with age (the POI). In the UKOPS2 set, there are two SVs that correlate marginally with age, and importantly they do not correlate with any other factor, hence these are not included in step-11 of ISVA. In the T1D set, there are three SVs that correlate with age, but only one of these (SV-3) is excluded, because the other two (SV-1 and SV-5) correlate more strongly with potential confounders such as Sex, Cohort, BSCE, and Chip. As seen in Table 17.2, ISVA with the above prescription for SV subspace selection, leads to significant enrichment of PCGTs in all four DNA methylation datasets. Thus, using ISVA the known biological signature is successfully retrieved in all sets.

It could be argued that the key step is the SV subspace selection, and not the BSS algorithm *per se*. To show how the use of ICA facilitates the SV subspace selection, we return to the example of mRNA expression data with grade as the POI and ER−status playing the role of confounder. Table 17.1 shows the results obtained by ISVA. In comparison to SVA, we can see that ISVA leads to specific enrichment of cell-cycle genes (i.e., ER−signaling genes are not enriched), clearly indicating that confounding by ER−status has been successfully removed. As we can see from Fig. 17.8, this improved feature selection can be attributed to a more accurate

deconvolution of residual variation associated with grade from that associated with ER−status. As illustrated in Fig. 17.8a, SV-1 in SVA is equally strongly correlated with grade and ER−status, indicating inaccurate deconvolution. In contrast, with ISVA, the SVs correlating most strongly with ER (SV-12) and grade (SV-7) are distinct, thus facilitating SV subspace selection and subsequently allowing improved feature selection. Similarly, in Fig. 17.8b, SV-3 in SVA is selected for adjustment yet it correlates very strongly with grade. In contrast, in ISVA the SV correlating most strongly with grade (SV-9) does so much more strongly than with ER−status, and hence this SV is not included in the subsequent adjustment. The effect of ER in the residual variation space is captured by other SVs (SV-12, 20, 24, 27) which do not correlate as strongly with grade, and these are therefore included in the adjustment. Thus, in these two examples, the BSS method is key since it allows more accurate deconvolution of the different sources of variation in the residual variation space. Even if a SV subspace selection step is incorporated into SVA (using the same heuristic criterion as for ISVA), we would still select problematic SVs since PCA does not allow accurate deconvolution of the different sources of variation (see [45] for results of this modified SVA).

17.8 Modeling of Confounding Factors with Generalized BSS Algorithms

In the previous sections, we have seen how a simple BSS method (fastICA) can lead to substantial improvements in modeling confounding factors as well as to an improved deconvolution of the biological and confounding factors, both of which are important, and which subsequently lead to improved feature selection in supervised analysis problems. We have also provided an objective evaluation framework in which to assess and compare the different algorithms.

It is therefore of interest to consider more sophisticated BSS methods, since these might offer further improvements in statistical inference. In doing so, the first question to address is whether modeling of confounders is improved using these more advanced BSS methods. One particular generalization of ICA which is of interest to study concerns the statistical independence assumption, which so far has been applied to the columns of the source matrix S. In other words, given the residual matrix R of dimension $p \times n$, we applied ICA in the context

$$R = S_1 A + \epsilon \tag{17.6}$$

with the inference required to minimize a residual sum of squares subject to the constraint that the K $p-$dimensional column vectors of S_1 be as statistically independent as possible. However, as shown in previous studies [37, 47], a dual interpretation/implementation is possible, whereby statistical independence is imposed on the rows of the mixing matrix A. This dual problem can be expressed as:

$$\begin{aligned} R^T &= A^T S_1^T + \epsilon \\ &= \tilde{S}_2 \tilde{A} + \epsilon \end{aligned} \tag{17.7}$$

where statistical independence is now imposed on the columns of $\tilde{S}_2$ which is a matrix of dimensionality $n \times K$. As shown in [2, 33, 37, 47], it is possible to formulate a "spatio-temporal" or bi-dimensional ICA,

$$R = S_1 S_2^T + \epsilon \tag{17.8}$$

in which statistical independence is favored across both features ("time") and samples ("space"), by means of an overall cost function, C_f, defined as a weighted linear combination of the cost functions used to solve Eqs. 17.6 and 17.7, i.e.,

$$C_f = (1-a)C_{f_1} + aC_{f_2} \tag{17.9}$$

More formally, the specific bi-dimensional ICA algorithm we consider here [2, 33, 47] starts with a SVD of the row and column centered (residual) data matrix R, so $R = UDV^T$, with corresponding estimation of the dimensionality K (using as before RMT). One then constructs the reduced matrix $R_K = U_K D_K V_K^T$ where the first K columns of U and V have been selected corresponding to the top K singular values of D. This reduced matrix can then be rewritten as

$$R_K = \underbrace{U_K D_K W^{-1}}_{S_1} \underbrace{W V_K^T}_{S_2^T} \tag{17.10}$$

with W an invertible matrix of size $K \times K$. Finally, we seek to optimize the matrix W such that the fourth-order cumulants of S_1 and S_2 are as diagonal as possible, i.e., minimizing

$$C_f(W) = \left(a \sum_i \text{Off}\left(C_i(S_2^T)\right) + (1-a) \sum_i \text{Off}\left(C_i(S_1^T)\right) \right) \tag{17.11}$$

where $\text{Off}(Y)$ returns the sum of squares of the off-diagonal elements of Y, and the C_i are fourth-order cumulants. Imposing that W is orthogonal leads to a formulation which can be solved by means of the JADE algorithm [6]. We note however that this formulation of bi-dimensional ICA differs slightly from that of [33, 47], as the second term in the contrast function involves $\left(C_i(S_1^T)\right)$ instead of $\left(C_i(S_1^T)\right)^{-1}$. Minimizing one or the other pursues the same goal, namely statistical independence for columns of S_1. This novel formulation however allows us to treat both extreme cases on an equal footing: $a = 1$ corresponds to JADE applied on $R_K^T = S_2 S_1^T$ whereas $a = 0$ corresponds to JADE applied on $R_K = S_1 S_2^T$. Thus, the cost function can be interpreted as a weighted linear combination of two 'jade-like' cost functions.

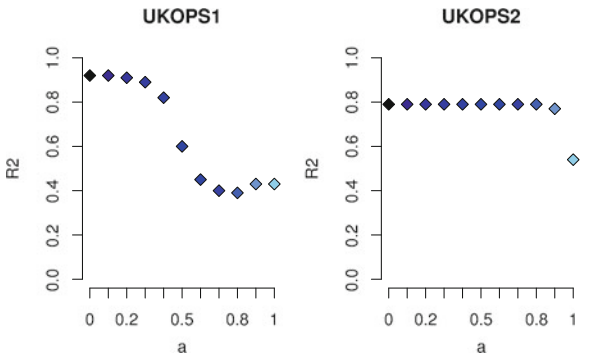


Fig. 17.9 Modeling of beadchip effects by bi-dimensional ICA in two DNA methylation datasets. y-axis labels the R^2 value of the component correlating best with the beadchip as assessed using a linear ANOVA model. x-axis labels the parameter a in Eq. 17.11

Given the above formulation of bi-dimensional ICA, it is of interest to study the effect of the parameter a on the quality of BSS. Since beadchip effects provide an objective framework in which to assess the quality of the BSS, we focus on how well these effects are modeled by the family of bi-dimensional ICA algorithms above. For simplicity, we consider the unsupervised problem in which the ICA decomposition is done on the data matrix X itself.[1] Figure 17.9 shows the results, indicating that in terms of modeling beadchip effects, ICA is best run with values of a close to zero. This corresponds to imposing statistical independence of the sources across features, as implemented in the fastICA version of the ISVA algorithm.

17.9 Conclusions

In this chapter, we have presented and discussed the problem that confounding factors pose in large omic datasets. Since feature selection is a common task in the analysis of such large datasets, it is paramount to have statistical methods in place that can perform supervised analysis and feature selection in the background of such confounding factors, specially when these are uncertain or unknown. We have seen how BSS methods are necessary in this context, since there is a requirement to accurately model confounding factors and to deconvolve these from variation associated with the phenotype of interest. We have presented an algorithm, ISVA, which uses a BSS technique (ICA) to perform a supervised normalization of the data and have shown that it offers a more sound statistical framework in which to perform feature selection than a competing non-BSS tool based on PCA.

As mentioned earlier, it is possible to consider any BSS algorithm within the ISVA framework. One of the most straightforward generalizations of the fastICA algorithm used in our ISVA implementation is to relax the statistical independence

[1] Instead of the residual variation matrix R which requires specification of the POI and is thus supervised.

assumption, but to simultaneously impose partial statistical independence along the dual "sample"-space, resulting in a bi-dimensional ICA. However, we have seen that, at least in terms of modeling beadchip effects, that the original implementation (i.e., imposing statistical independence across features) is optimal. This could be due to the sources across features being well described by sparse distributions or by the fact that statistical independence is best assessed using the larger feature space.

Although the bi-dimensional ICA did not lead to improved modeling of beadchip effects, it is nevertheless of interest to investigate this and other BSS algorithms in the ISVA context. For instance, it could well be that other types of confounding factors are best modeled using bi-dimensional ICA or ICA algorithms that also allow for skewed sources of variation [37, 47]. Exact known confounders (like beadchip effects) allow for objective assessment of BSS in real data, yet unfortunately, not many such factors exist. On the other hand, the number of beadchips in studies can vary substantially, thus allowing assessment of the BSS methods at least in relation to statistical properties such as kurtosis, which would vary for beadchip effects depending on the overall sample size of the study. Thus, a beadchip effect affecting 12 samples out of 120 samples (10 beadchips) will exhibit different statistical properties to one in a study of only 36 samples.

Besides the detailed modeling of the sources, another key challenge faced in ISVA is the SV subspace selection step. Although we have presented a simple heuristic selection criterion, which, as we have seen, successfully retrieves the known biological signatures in diverse real datasets, the criterion itself is not applicable to the case where confounders are complete unknowns (i.e., hidden). In fact, this remains an outstanding statistical challenge since (1) the presence of biological variation of interest in the matrix of residuals is almost always inevitable and (2) it is entirely plausible that some of this variation is driven by hidden confounding factors and hence that the associated SVs should be included in the final regression model.

The results on eight real datasets presented here however, conclusively demonstrate that a SV selection step is absolutely necessary to arrive at the correct biological conclusion, yet in other datasets where the biological truth is unknown, the SV selection criterion used here could falter due to hidden confounding factors. In other words, in the eight real datasets considered here we can be fairly certain that the data is not subject to substantial hidden (i.e., completely unknown) confounding variation, since otherwise our SV selection criterion would not have led to the retrieval of the known biological signatures.

With this chapter we hope to engage biologists, bioinformaticians, and signal processing experts alike. The problem that confounding factors pose in the statistical analysis of omic data is both challenging and critical to the ultimate success of large-scale genomic and epigenomic studies aiming to identify the much needed disease biomarkers. Further research in this area is therefore urgently needed.

Acknowledgments AET was supported by a Heller Research Fellowship. This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office.

Appendix

Simulated Data

We simulated data matrices with 2,000 features and 50 samples and considered the case of two confounding factors (CFs) in addition to the primary phenotype of interest. The primary phenotype is a binary variable I_1 with 25 samples in one class ($I_1 = 0$) and the other half with $I_1 = 1$. Similarly, each confounding factor is assumed to be a binary variable affecting one half of the samples (randomly selected). For a given sample s we thus have a 3-tuple of indicator variables $I_s = (I_{1s}, I_{2s}, I_{3s})$ where I_2 and I_3 are the indicators for the two confounding factors. Thus, samples fall into 8 classes. For instance, if $I_s = (0, 0, 0)$ then this sample belongs to phenotype class 1 and is not affected by the two confounding factors. Similarly, $I_s = (0, 1, 0)$ means that the sample belongs to class 1 and is affected by the first confounding factor but not the second.

We assume 10% of features (200 features) to be TPs discriminating between the two phenotypic classes. We model the confounding factors as follows: each confounding factor is assumed to affect 10% of features with a 25% overlap with the TPs (i.e 50 of the 200 TPs are confounded by each factor). Let J_g denote the indicator variable of feature g, so J_g is a 3-tuple (J_{1g}, J_{2g}, J_{3g}) with J_{1g} an indicator for the feature to be a true positive, and J_{2g} (J_{3g}) an indicator for the feature to be affected by the first (second) confounding factor. Thus, the space of features is also divided into eight groups. Furthermore, let (e_1, e_2, e_3) denote the effect sizes of the primary variable and the two confounding factors respectively, where we assume for simplicity that $e_2 = e_3$. Without loss of generality, we further assume that noise is modeled by a Gaussian of mean zero and unit variance $N(0, 1)$. Thus, for a given sample s we draw data values for the various feature groups as follows:

1. $J_g = (0, 0, 0)$: null unaffected features

$$p(x|I_s) \sim \delta_{J_g,000} N(0, 1)$$

2. $J_g = (0, 1, 0)$ or $(0, 0, 1)$: null features affected by only one CF

$$\begin{aligned} p(x|I_s) \sim \delta_{J_g,010} \big\{ & \delta_{I_s,x1z} N(e_2, 1) \\ & + \delta_{I_s,x0z} N(0, 1) \big\} \\ & + \delta_{J_g,001} \big\{ \delta_{I_s,xy1} N(e_3, 1) \\ & + \delta_{I_s,xy0} N(0, 1) \big\} \end{aligned}$$

3. $J_g = (0, 1, 1)$: null features affected by the two CFs

$$
\begin{aligned}
p(x|I_s) \sim \delta_{J_g,011}\big\{&\delta_{I_s,x11}N(e_2+e_3,1)\\
&+\delta_{I_s,x01}N(e_3,1)\\
&+\delta_{I_s,x10}N(e_2,1)\\
&+\delta_{I_s,x00}N(0,1)\big\}
\end{aligned}
$$

4. $J_g = (1, 0, 0)$: true positives not affected by CFs

$$
\begin{aligned}
p(x|I_s) \sim \delta_{J_g,100}\big\{&\delta_{I_s,0yz}N(0,1)\\
&+\delta_{I_s,1yz}(\pi_{-1}N(-e_1,1)+\pi_1 N(e_1,1))\big\}
\end{aligned}
$$

5. $J_g = (1, 0, 1)$ or $(1, 1, 0)$: true positives affected by one CF

$$
\begin{aligned}
p(x|I_s) \sim \delta_{J_g,101}\big\{&\delta_{I_s,0y0}N(0,1)+\delta_{I_s,0y1}N(e_3,1)\\
&+\delta_{I_s,1y0}(\pi_{-1}N(-e_1,1)+\pi_1 N(e_1,1))\\
&+\delta_{I_s,1y1}(\pi_{-1}N(-e_1+e_3,1)\\
&+\pi_1 N(e_1+e_3,1))\big\}\\
\sim \delta_{J_g,110}\big\{&\delta_{I_s,00z}N(0,1)+\delta_{I_s,01z}N(e_2,1)\\
&+\delta_{I_s,10z}(\pi_{-1}N(-e_1,1)+\pi_1 N(e_1,1))\\
&+\delta_{I_s,11z}(\pi_{-1}N(-e_1+e_2,1)\\
&+\pi_1 N(e_1+e_2,1))\big\}
\end{aligned}
$$

6. $J_g = (1, 1, 1)$: true positives affected by all CFs

$$
\begin{aligned}
p(x|I_s) \sim \delta_{J_g,111}\big\{&\delta_{I_s,000}N(0,1)\\
&+\delta_{I_s,010}N(e_2,1)+\delta_{I_s,001}N(e_3,1)\\
&+\delta_{I_s,011}N(e_2+e_3,1)\\
&+\delta_{I_s,101}(\pi_{-1}N(-e_1+e_3,1)\\
&+\pi_1 N(e_1+e_3,1))\\
&+\delta_{I_s,110}(\pi_{-1}N(-e_1+e_2,1)\\
&+\pi_1 N(e_1+e_2,1))\\
&+\delta_{I_s,111}(\pi_{-1}N(-e_1+e_2+e_3,1)\\
&+\pi_1 N(e_1+e_2+e_3,1))\big\}
\end{aligned}
$$

where in the above $\delta_{x'y'z',xyz}$ denotes the triple Kronecker delta: $\delta_{x'y'z',xyz} = 1$ if and only if $x' = x$, $y' = y$ and $z' = z$, otherwise $\delta_{x'y'z',xyz} = 0$, and (π_{-1}, π_1) are weights satisfying $\pi_{-1} + \pi_1 = 1$. In our case, we used $\pi_1 = \pi_{-1} = 0.5$.

DNA Methylation Data (Whole Blood Tissue)

In all datasets, age is the phenotype of interest. (i) T1D: this DNAm dataset consists of 187 blood samples from patients (94 women and 93 men) with type-1 diabetes. This set served as validation for a DNAm signature for aging [44]. We take BSCE, beadchip, cohort, and sex as potential confounding factors. Samples were distributed over 17 beadchips; (ii) UKOPS1: this DNAm set consists of 108 blood samples from healthy postmenopausal women which served as controls for the UKOPS study [43]. Confounding factors in this study include BSCE, beadchip and DNA concentration (DNAc). Samples were distributed over 10 beadchips; (iii) UKOPS2: This is similar to Dataset2 but consists of 145 blood samples from healthy postmenopausal women distributed over 36 beadchips (i.e., approximately four healthy samples per chip, the other eight blood samples per chip were from cancer cases) [43]; (iv) WBBC: This dataset consists of whole blood samples from a total of 84 women (49 healthy and 35 women with breast cancer). Samples were distributed over seven beadchips, and confounders are BSCE, status (cancer/healthy), and beadchip.

Breast Cancer mRNA Expression Data

The mRNA expression profiles are all from primary breast cancers and three of the datasets were profiled on Affymetrix platforms, while another was profiled on an Illumina Beadchip. Normalized data were downloaded from GEO (http://ncbi.nlm.nih.gov/), and probes mapping to the same Entrez ID identifier were averaged. Sotiriou: 14,223 genes and 101 samples [36]; Loi: 15,736 genes and 137 samples [26]; Schmidt: 13,292 genes and 200 samples [35]; Blenkiron: 17,941 genes and 128 samples [5]. In these datasets, we take histological grade as the phenotype of interest and consider estrogen receptor status and tumor size as potential confounders. Cell-cycle-related genes are known to discriminate low and high grade breast cancers irrespective of estrogen receptor status [26, 36]. Therefore, we compare the algorithms in their ability to detect specifically cell-cycle-related genes and not estrogen-regulated genes. To this end, we focused attention on two gene sets, one representing cell-cycle-related genes from the Reactome http://www.reactome.org, and another representing estrogen receptor (*ESR1*) upregulated genes [10]. The cell-cycle set showed negligible overlap with the *ESR1* gene set, however, we removed the few overlapping genes to ensure mutual exclusivity of the cell-cycle and *ESR1* sets.

References

1. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., Stratton, M.R.: Deciphering signatures of mutational processes operative in human cancer. Cell Rep. **3**(1), 246–259 (2013)
2. Baufays, H.: Unification de techniques de sparation aveugle de sources avec application l'analyse de l'expression des gnes. Ecole Polytechnique de Louvain, Master thesis with Prof. P.-A. Absil (2011)
3. Bell, C.G., Teschendorff, A.E., Rakyan, V.K., Maxwell, A.P., Beck, S., Savage, D.A.: Genome-wide dna methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. BMC Med. Genomics **3**, 33 (2010)
4. Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., Gunderson, K.L.: Genome-wide DNA methylation profiling using the infinium assay. Epigenomics **1**(1), 177–200 (2009)
5. Blenkiron, C., Goldstein, L.D., Thorne, N.P., Spiteri, I., Chin, S.F., Dunning, M.J., Barbosa-Morais, N.L., Teschendorff, A.E., Green, A.R., Ellis, I.O., Tavar, S., Caldas, C., Miska, E.A.: Microrna expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biol. **8**(10), R214 (2007)
6. Cardoso, J.F.: High-order contrasts for independent component analysis. Neural Comput. **11**(1), 157–192 (1999)
7. Consortium 1000 Genomes Project, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A.: An integrated map of genetic variation from 1,092 human genomes. Nature **491**(7422), 56–65 (2012)
8. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Grf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Brresen-Dale, A.L., Brenton, J.D., Tavar, S., Caldas, C., Aparicio, S.: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature **486**(7403), 346–352 (2012)
9. Deaton, A.M., Bird, A.: Cpg islands and the regulation of transcription. Genes Dev. **25**, 1010–1022 (2011)
10. Doane, A.S., Danso, M., Lal, P., Donaton, M., Zhang, L., Hudis, C., Gerald, W.L.: An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. Oncogene **25**(28), 3994–4008 (2006)
11. Feinberg, A.P., Vogelstein, B.: Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature **301**(5895), 89–92 (1983)
12. Frigyesi, A., Veerla, S., Lindgren, D., Hoglund, M.: Independent component analysis reveals new and biologically significant structures in micro array data. BMC Bioinformatics **7**, 290 (2006)
13. Gao, Y., Church, G.: Improving molecular cancer class discovery through sparse non-negative matrix factorization. Bioinformatics **21**(21), 3970–3975 (2005)
14. Huang, D.S., Zheng, C.H.: Independent component analysis-based penalized discriminant method for tumor classification using gene expression data. Bioinformatics **22**(15), 1855–1862 (2006)
15. Hyvaerinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
16. Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics **8**(1), 118–127 (2007)
17. Jones, P.A., Baylin, S.B.: The epigenomics of cancer. Cell **128**(4), 683–692 (2007)
18. Lee, S.I., Batzoglou, S.: Application of independent component analysis to microarrays. Genome Biol. **4**(11), R76 (2003)
19. Leek, J.T., Storey, J.D.: A general framework for multiple testing dependence. Proc. Natl. Acad. Sci. USA **105**(48), 18, 718–18, 723 (2008)
20. Leek, J.T., Storey, J.D.: Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. **3**(9), 1724–1735 (2007)

21. Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., Irizarry, R.A.: Tackling the widespread and critical impact of batch effects in high-throughput data. Nat. Rev. Genet. **11**(10), 733–739 (2010)
22. Liao, J.C., Boscolo, R., Yang, Y.L., Tran, L.M., Sabatti, C., Roychowdhury, V.P.: Network component analysis: reconstruction of regulatory signals in biological systems. Proc. Natl. Acad. Sci. USA **100**(26), 15,522–15,527 (2003)
23. Liebermeister, W.: Linear modes of gene expression determined by independent component analysis. Bioinformatics **18**(1), 51–60 (2002)
24. Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., Shchetynsky, K., Scheynius, A., Kere, J., Alfredsson, L., Klareskog, L., Ekstrm, T.J., Feinberg, A.P.: Epigenome-wide association data implicate dna methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. **31**(2), 142–147 (2013)
25. Liu, N.W., Sanford, T., Srinivasan, R., Liu, J.L., Khurana, K., Aprelikova, O., Valero, V., Bechert, C., Worrell, R., Pinto, P.A., Yang, Y., Merino, M., Linehan, W.M., Bratslavsky, G.: Impact of ischemia and procurement conditions on gene expression in renal cell carcinoma. Clin. Cancer Res. **19**(1), 42–49 (2013)
26. Loi, S., Haibe-Kains, B., Desmedt, C., Lallemand, F., Tutt, A.M., Gillet, C., Ellis, P., Harris, A., Bergh, J., Foekens, J.A., Klijn, J.G., Larsimont, D., Buyse, M., Bontempi, G., Delorenzi, M., Piccart, M.J., Sotiriou, C.: Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J. Clin. Oncol. **25**(10), 1239–1246 (2007)
27. Maegawa, S., Hinkal, G., Kim, H.S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L.A., Issa, J.P.: Widespread and tissue specific age-related dna methylation changes in mice. Genome Res. **20**(3), 332–340 (2010)
28. Martoglio, A.M., Miskin, J.W., Smith, S.K., MacKay, D.J.: A decomposition model to track gene expression signatures: preview on observer-independent classification of ovarian cancer. Bioinformatics **18**(12), 1617–1624 (2002)
29. Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A., Guhr, T., Stanley, H.E.: Random matrix approach to cross correlations in financial data. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. **65**(6), 066,126 (2002)
30. Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M., Leslie, R.D., Deloukas, P., Spector, T.D.: Human aging-associated dna hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res. **20**(4), 434–439 (2010)
31. Rakyan, V.K., Down, T.A., Balding, D.J., Beck, S.: Epigenome-wide association studies for common human diseases. Nat. Rev. Genet. **12**(8), 529–541 (2011)
32. Rhodes, D.R., Chinnaiyan, A.M.: Integrative analysis of the cancer transcriptome. Nat. Genet. **37**, S31–S37 (2005)
33. Sainlez, M., Absil, P.-A., Teschendorff, A. Gene expression data analysis using spatiotemporal blind, source separation. In: Proceedings of ESANN'2009, pp. 159–164. (2009)
34. Sawyers, C.L.: The cancer biomarker problem. Nature **452**(7187), 548–552 (2008)
35. Schmidt, M., Bhm, D., von Trne, C., Steiner, E., Puhl, A., Pilch, H., Lehr, H.A., Hengstler, J.G., Klbl, H., Gehrmann, M.: The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res. **68**(13), 5405–5413 (2008)
36. Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., Desmedt, C., Larsimont, D., Cardoso, F., Peterse, H., Nuyten, D., Buyse, M., Van de Vijver, M.J., Bergh, J., Piccart, M., Delorenzi, M.: Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J. Natl. Cancer Inst. **98**(4), 262–272 (2006)
37. Stone, J.V., Porrill, J., Porter, N.R., Wilkinson, I.D.: Spatiotemporal independent component analysis of event-related fmri data using skewed probability density functions. Neuroimage 15 (2002)

38. Storey, J.D., Tibshirani, R.: Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA **100**(16), 9440–9445 (2003)
39. Subramanian, A,. Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA **102**(43), 15, 545–15, 550 (2005)
40. Swanton, C., Caldas, C.: From genomic landscapes to personalized cancer management-is there a roadmap? Ann. N. Y. Acad. Sci. **1210**, 34–44 (2010)
41. Teschendorff, A.E., Naderi, A., Barbosa-Morais, N.L., Caldas, C.: Pack: profile analysis using clustering and kurtosis to find molecular classifiers in cancer. Bioinformatics **22**(18), 2269–2275 (2006)
42. Teschendorff, A.E., Journe, M., Absil, P.A., Sepulchre, R., Caldas, C.: Elucidating the altered transcriptional programs in breast cancer using independent component analysis. PLoS Comput. Biol. **3**(8), e161 (2007)
43. Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Gayther, S.A., Apostolidou, S., Jones, A., Lechner, M., Beck, S., Jacobs, I.J., Widschwendter, M.: An epigenetic signature in peripheral blood predicts active ovarian cancer. PLoS ONE **4**(12), e8274 (2009)
44. Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Weisenberger, D.J., Shen, H., Campan, M., Noushmehr, H., Bell, C.G., Maxwell, A.P., Savage, D.A., Mueller-Holzner, E., Marth, C., Kocjan, G., Gayther, S.A., Jones, A., Beck, S., Wagner, W., Laird, P.W., Jacobs, I.J., Widschwendter, M.: Age-dependent dna methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome Res. **20**(4), 440–446 (2010)
45. Teschendorff, A.E., Zhuang, J., Widschwendter, M.: Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. Bioinformatics **27**(11), 1496–1505 (2011)
46. The Cancer Genome Atlas Research Network: Integrated genomic analyses of ovarian carcinoma. Nature **474**(7353), 609–615 (2011)
47. Theis, F., Gruber, P., Keck, I., Meyer-Bäse, A., Lang, E.: Spatiotemporal blind source separation using double-sided approximate joint diagonalization. In: Proceedings of EUSIPCO 2005, Antalya, Turkey (2005)
48. Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Yu, J., Jatkoe, T., Berns, E.M., Atkins, D., Foekens, J.A.: Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet **365**(9460), 671–679 (2005)
49. Zhang, X.W., Yap, Y.L., Wei, D., Chen, F., Danchin, A.: Molecular diagnosis of human cancer type by gene expression profiles and independent component analysis. Eur. J. Hum. Genet. **13**(12), 1303–1311 (2005)
50. Zhang, S., Liu, C.C., Li, W., Shen, H., Laird, P.W., Zhou, X.J.: Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. Nucleic Acids Res. **40**(19), 9379–9391 (2012)
51. Zhuang, J., Widschwendter, M., Teschendorff, A.E.: A comparison of feature selection and classification methods in dna methylation studies using the illumina infinium platform. BMC Bioinformatics **13**, 59 (2012)

Chapter 18
*Feb*ICA: Feedback Independent Component Analysis for Complex Domain Source Separation of Communication Signals

A. K. Kattepur and F. Sattar

Abstract In this chapter, an effective blind source separation (BSS) algorithm is applied to solve the co-channel interference problem in wireless communication systems. Algorithms developed for this purpose must not only have the capability of working in the complex domain and improving output signal to interference plus noise ratio (SINR), but also have relatively low computational complexity. We propose a fast Fourier transform (FFT)-based algorithm called feedback independent component analysis (*Feb*ICA) that is able to blindly separate complex modulated digital signals. By applying this algorithm to communication signals, it is observed that it has the advantages of SINR gain improvement as well as low computational complexity. The performance of the *Feb*ICA algorithm is shown to be better than the joint approximate diagonalization of eigen-matrices (JADE) algorithm in terms of the output SINR and requires lower computational complexity than the analytical constant modulus algorithm (ACMA). The algorithm is also shown to be more robust with increasing number of sources compared to other algorithms. The separation performance by using the collected field data has also been demonstrated.

18.1 Introduction

Blind source separation (BSS) algorithms are used for separating individual sources from their mixtures with minimal *a priori* information about the source signals or

A. K. Kattepur (✉)
Inria Paris - Rocquencourt, Domaine de Voluceau, Le Chesnay, France
e-mail: ajay.kattepur@inria.fr

F. Sattar
Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada
e-mail: fsattar@uwaterloo.ca

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_18,
© Springer-Verlag Berlin Heidelberg 2014

their mixing process. This technique has been used with significant success in various fields such as speech and music processing, sonar, biomedical and financial data [1].

Blind signal processing is statistically based on independent component analysis (ICA) techniques which are based on the assumptions that the original signals are independent and non-Gaussian in nature [2]. The basic instantaneous source separation problem in the time domain can be described as:

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \tag{18.1}$$

Here, $\mathbf{X}$ is the observed mixed signal, $\mathbf{A}$ is the mixing matrix, $\mathbf{S}$ is the source signal, and $\mathbf{N}$ is the additive noise. The objective of any blind source separation algorithm is to generate an unmixing matrix $\mathbf{W}$ such that the resulting signal $\mathbf{Y}$ will be a close estimate of the original source signal $\mathbf{S}$.

$$\mathbf{Y} = \mathbf{WX} \tag{18.2}$$

A number of BSS algorithms widely used include FastICA [3], Infomax [4] and joint approximate diagonalization of eigen-matrices (JADE) [5]. They make use of the second-order or higher order statistics to estimate the unmixing matrix $\mathbf{W}$ in order to recover the original sources.

In this chapter, we address the BSS problem applicable to wireless communications. Besides multipath fading, the ability of many practical wireless communications to reliably detect and extract information from the received signals is largely affected by the band noise and co-channel transmission interference. Due to the multipath channel effects, they arrive at the receiver as angularly spread interference sources. The received signals are also corrupted by noise from receiver electronics as well as man-made equipments (e.g., microwave ovens, electrical lamps, motors, overhanging power lines, etc). Typically, these noise sources have non-Gaussian statistics and are directional in nature. The use of BSS could then help to solve the existing multipath fading problem caused by the external noise and interferences, and thereby improve the performance of the wireless system by increasing the signal to interference plus noise ratio (SINR) as well as capacity of the system. This is a seemingly important yet relatively unexplored problem and needs more work to be done in this area.

As the BSS algorithms are applied on the modulated signals (Gaussian minimum shift keying for global system of mobile, for instance), they must be able to separate signals in the complex domain. Moreover, this must be done under real-time processing constraints with added computational complexity heavily deteriorating system performance. In this context, JADE [5] and analytical constant modulus algorithms (ACMA) [6] are the two commonly used algorithms to perform BSS in complex domain. JADE algorithm makes use of Jacobi optimization to extract the source signals, while ACMA performs singular value decomposition (SVD) to extract the independent components. However, the limitation of these methods lies in their large computational complexity which makes them not suitable for real-time processing.

To tackle this problem, an iterative fast Fourier transform (FFT)-based feedback independent component analysis (*Feb*ICA) algorithm is proposed here based on feedback architecture which not only has relatively low computational complexity (specially for a large number of interfering sources), but able to separate also the complex digitally modulated signals. We also demonstrate that this system can work with nearly singular mixing matrices, which are typically observed in multiple-input and multiple-output (MIMO) communication systems. Another motivation of this chapter lies on the developing and comparing the performance of different classes of such algorithms, that could be applied in future to other fields such as biomedical applications (e.g., electroencephalographic recording using wireless).

The proposed *Feb*ICA algorithm belongs to the class of algorithms that make use of information maximization. By using a nonlinear function with updated weight vectors, this type of algorithms relies on the mutual information providing a simple learning rule. The performance of the *Feb*ICA algorithm has been demonstrated in terms of computational complexity as well as SINR improvements. Even after reducing complexity of source separation, the *Feb*ICA algorithm does not suffer from deteriorating output SINR performance as seen in other BSS algorithms [7]. Hence, it can be efficiently used for real-time separation of communication signals corrupted by co-channel interference. Also, the proposed scheme is found to be more robust with an increasing number of sources as compared to other methods. The separation performance for the collected field data from GSM signals has also been demonstrated.

The organization of the chapter is as follows. In Sect. 18.2, the proposed *Feb*ICA algorithm as well as the theoretical framework of the method are presented. The convergence analysis concerning the stability of the proposed algorithm and the computational complexities of the method are included in Sect. 18.3. In Sect. 18.4, the detailed results and the performances of the proposed *Feb*ICA algorithm and the comparison with other methods are presented in terms of output SINR as well as computational complexities. Related work on source separation in the complex domain, including applications for communication systems, is presented in Sect. 18.5. Finally, Sect. 18.6 presents the conclusion and future work.

18.2 *Feb*ICA Algorithm

The proposed *Feb*ICA algorithm is an extension of the adaptive neural network approach proposed in [8], which has been used to separate a mixture of odor sources. The approach in [8] which has been originally used to estimate the olfactory perception of odors in animals is modified here to update the complex domain unmixing matrices for source separation. It has been done by adopting the weight updates and a gradient ascent learning rule for the application of source separation as used previously in [9]. Thus, the key contribution lies in proposing an algorithm for source separation in complex domain which has not been done yet within adaptive neural networks framework applied to source separation in communication systems.

Note that among the well-known measures of BSS performance including kurtosis, negentropy, and mutual independence of source signals [1], in this chapter, we exploit the mutual independence of sources as used in [4, 10]. Thus, making use of the criterion for mutual independence, the general expression for sequential updating of weight vectors based on the global gradient update rule, can be shown by (18.3):

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \boldsymbol{\eta}(t)[\mathbf{I} - f(y(t))g^T(y(t))]\mathbf{W}(t) \tag{18.3}$$

where $\mathbf{W}$ is the unmixing matrix, $\boldsymbol{\eta}$ is the learning rate, $f(y)$ is a function of the observed mixture, and g is a nonlinearity (For detailed derivation, please see the appendix). Thus, the crux of our source separation problem lies in the use of a cost function for measuring independence and an appropriate optimization criterion for updating the weights.

The proposed *Feb*ICA algorithm is presented by the steps described as follows:

1. Initialize complex weights $\mathbf{W}$ of the *Feb*ICA algorithm. This can either be done randomly or by using the unmixing matrix estimate from the JADE or ACM algorithms. The feedback weights $\mathbf{W}_{fb}$ are then initialized by the off-diagonal elements of the weights $\mathbf{W}$ given by:

$$\mathbf{W}_{fb}(n,m) = \mathbf{W}(n,m) \qquad \forall n \neq m \tag{18.4}$$

   where n and m represent the row and columns of the matrices, respectively. The complex weights will eventually converge to the desired unmixing matrix, while the feedback weights are used to control the convergence rates.
2. The FFT γ and the row-wise mean μ of the observed signal matrix $\mathbf{X}$ (with M columns) are given by:

$$\gamma_{k,m} = \sum_{n=0}^{N-1} \mathbf{x}_{n,m} e^{-j\frac{2\pi}{N}nk} \qquad k = 0, \ldots, N-1 \tag{18.5}$$

$$\mu_m = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}_{n,m} \tag{18.6}$$

   where N is the length of the complex sequence in each row of the observed matrix $\mathbf{X}$ given by $\mathbf{x}_1$,…,$\mathbf{x}_{M-1}$. This Fourier domain application to BSS has been used in [11] and [12]. As shown in [13], by using a Fourier basis, advantages such as compact representations, higher convergence rate, and lower mean squared error may be achieved. Consider the following Fourier basis function:

$$q_n[y(k)] = e^{jn\,\omega\,y(k)} \tag{18.7}$$

   where ω is the system frequency. The gradient ascent rule used in (18.16) can be written as:

$$w_n(k+1) = w_n(k) + \eta \Delta w_n^q(k) \tag{18.8}$$

$$\Delta w^q(k) = \frac{1}{N} \sum_{\iota=0}^{N-1} e(\iota) q_n[y(k)] \tag{18.9}$$

$$e(\iota) = y(\iota+1) - \hat{y}(\iota+1) \tag{18.10}$$

where $\hat{y}(\iota+1)$ is the estimated unmixed signal with every ιth iteration. As shown in [13], the mean square error Ψ with each iteration would drop as:

$$\Psi = \frac{1}{N} \sum_{\iota=0}^{N-1} \left[y(\iota+1) - \hat{y}(\iota+1) \right]^2 \tag{18.11}$$

$$E_r = \frac{\left\| \sqrt{\Psi(k)} - \sqrt{\Psi(k-1)} \right\|}{\left\| \sqrt{\Psi(k)} \right\|} \tag{18.12}$$

Thus, by controlling the relative error E_r, a better accuracy measure for the source separation problem can be achieved. The Fourier basis is also computationally less intensive which is useful in the case of online source separation.

3. Based on (18.3) we introduce a nonlinear function $g(\cdot)$ which is a suitably chosen odd nonlinearity, providing stability in the process of separation. The nonlinear function should be judiciously selected to deal with the super-Gaussian, sub-Gaussian, stationary, and nonstationary signals. A popular choice is a sigmoidal-shaped functions shown in [4, 9]. This nonlinearity in the function also creates a narrow boundary condition that is responsible for distinguishing various independent components.

$$\xi(\gamma_{k,m}, \beta_k, \delta_m) = e^{-\beta_k/\delta_m} . \gamma_{k,m} \tag{18.13}$$

where β and δ are constants and ξ is the enhanced output of the nonlinear function. By creating the narrow nonlinear boundary ξ, the probability of finding a single independent vector within that reduced space increases. Furthermore, the feedback architecture reduces the mutual information between the mixed components which is updated based on a nonlinear gradient. This is a stochastic gradient ascent algorithm which tries to maximize the sum of fourth-order cumulants. The performance criterion is expressed as

$$\mathrm{J}(\mathbf{W}) = \sum_{\iota=1}^{n} E\left\{ f\left(y(\iota) \right) \right\} \tag{18.14}$$

where the function $f(\cdot)$ represents the objective function used in the algorithm. The objective function is chosen to be of the form $f(\cdot) = \ln[\cosh(\cdot)]$ which on

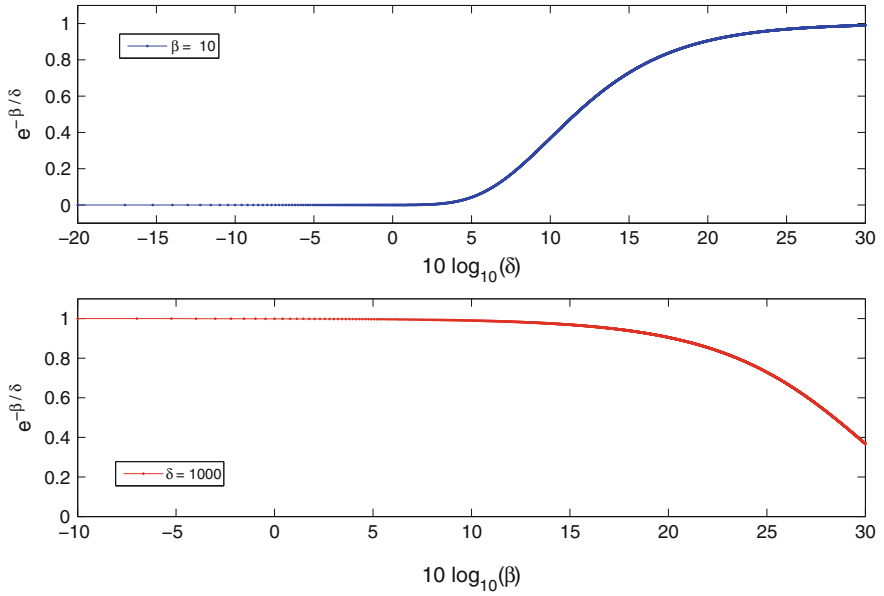


Fig. 18.1 The behavior of the nonlinear operator for various settings of parameters β and δ

differentiation provides the nonlinearity of the form $g(\cdot) = f'(\cdot) = \tanh(\cdot)$. In the learning rule given in (18.3), the nonlinearity $g(\cdot)$ will dominate and the learning rule will converge to a separating matrix $\mathbf{W}$ (Fig. 18.1).

4. Based on the weights $\mathbf{W}$ and feedback weights $\mathbf{W}_{fb}$, we further define an operator ψ which is updated iteratively. This is based on minimizing the mutual information between the original signals. As the fundamental assumption of ICA is independence of sources, the mutual information must tend to zero as the separation progresses.

$$\psi_\iota = \mathbf{W}_\iota \gamma - (\mathbf{W}_{fb})_\iota \xi \tag{18.15}$$

Here, ι represents the iteration count for the updating process.

5. The iterative process of updating the weights is described in terms of the following equations where η is the learning factor. As described in [14], the learning rate function η is of an exponentially decreasing form $e^{(-\omega_k/5)}/250$, where $\omega_k = 2\pi k/N, k = 0, 1, \ldots, N-1$. This is a gradient ascent method of updating the weights by minimizing the mutual information between the signals based on a nonlinear gradient.

$$\Delta \mathbf{W}_\iota = \eta \left(1 + \frac{\xi''}{\xi'} \psi_\iota\right) \mathbf{W}_\iota \tag{18.16}$$

$$\mathbf{W}_\iota = \mathbf{W}_{\iota-1} + \Delta \mathbf{W}_\iota \tag{18.17}$$

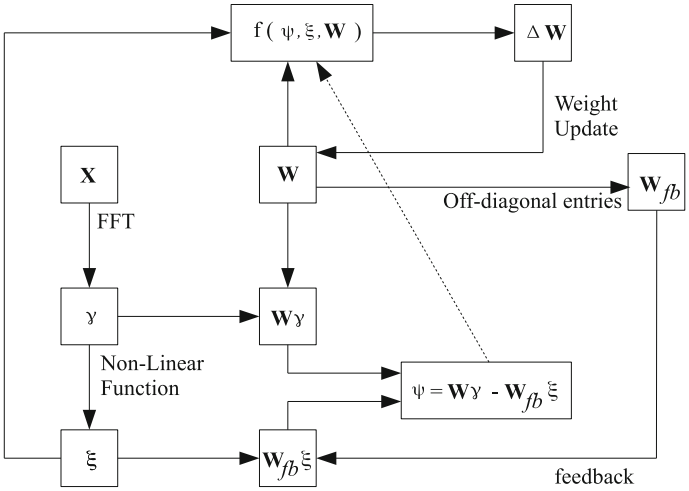


Fig. 18.2 Parameters and operations involved in the *Feb*ICA algorithm

where $\Delta \mathbf{W}_\iota$ refers to the step size of the weight updates and ξ' refers to the derivative of the nonlinear function defined in (18.13). Here, we also update the feedback weights $\mathbf{W}_{fb\iota}$ with the new value of $\mathbf{W}_\iota$ as shown in (18.4).

6. Steps 4 and 5 are repeated till convergence. After the weights converge with the values of ψ, the independent vectors are obtained as the rows of $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} + \mu \tag{18.18}$$

where $\mathbf{Y}$ represents the estimated original signals by the *Feb*ICA algorithm.

The described *Feb*ICA algorithm is consequently represented in Fig. 18.2 to show the interaction of various parameters for weight update.

18.3 Convergence Analysis and Computational Complexities

For fast convergence of the *Feb*ICA algorithm, we need to set the learning rate η such that cost function Ω in (18.19) is minimized with every iteration.

$$\Omega = \eta \left(1 + \frac{\xi''}{\xi'} (\mathbf{W}_\iota \gamma - (\mathbf{W}_{fb})_\iota \xi) \right) \tag{18.19}$$

For the learning rate η, let us refer to (18.16) and take the ratio $\Delta \mathbf{W}_\iota / \mathbf{W}_\iota$ representing Ω as

$$\Omega = \eta \, (1 + f(\xi) \psi_\iota) \tag{18.20}$$

where (18.20) is a simplified function of learning rate η, the nonlinear function $f(\xi)$, and iterative operator ψ_ι. Then we make use of the Newton's method [15] to provide a local minima for the convergence criterion:

$$\Omega' = \eta \left(f'(\xi)\psi_\iota + f(\xi)\psi_\iota' \right) \tag{18.21}$$

$$\Omega'' = \eta \left(f''(\xi)\psi_\iota + 2f'(\xi)\psi_\iota' + f(\xi)\psi_\iota'' \right) > 0 \tag{18.22}$$

where (18.22) represents a *local minima* for the weights to converge. As we make use of a hyperbolic function in (18.13), the boundary values can be used as limiting conditions for the scale η. In order to show an example of the convergence criterion, we make use of a differentiable nonlinear function $\xi = e^{-\beta_k/\delta_m} \cdot \gamma \cong \tanh(\mathbf{z})$ with a sigmoidal shape. The functions in (18.15) and (18.19) then become:

$$\psi_\iota(\mathbf{z}) = \mathbf{W}_\iota \gamma - (\mathbf{W}_{fb})_\iota \tanh(\mathbf{z}) \tag{18.23}$$

$$\begin{aligned} \Omega(\mathbf{z}) &= \eta \left(1 + \frac{\xi''}{\xi'} (\mathbf{W}_\iota \gamma - (\mathbf{W}_{fb})_\iota \tanh(\mathbf{z})) \right) \\ &= \eta \left(1 + \frac{-2\text{sech}^2(\mathbf{z}) \tanh(\mathbf{z})}{\text{sech}^2(\mathbf{z})} (\mathbf{W}_\iota \gamma - (\mathbf{W}_{fb})_\iota \tanh(\mathbf{z})) \right) \end{aligned} \tag{18.24}$$

Producing the derivatives,

$$\Omega'(\mathbf{z}) = \eta \left(-2\mathbf{W}_\iota \gamma \text{sech}^2(\mathbf{z}) + 4(\mathbf{W}_{fb})_\iota \tanh(\mathbf{z}) \text{sech}^2(\mathbf{z}) \right) \tag{18.25}$$

$$\begin{aligned} \Omega''(\mathbf{z}) =& \eta \left(4\gamma \mathbf{W}_\iota \text{sech}^2(\mathbf{z}) \tanh(\mathbf{z}) + 4(\mathbf{W}_{fb})_\iota \text{sech}^4(\mathbf{z}) - 8(\mathbf{W}_{fb})_\iota \text{sech}^2(\mathbf{z}) \right. \\ & \left. \tanh^2(\mathbf{z}) \right) > 0 \end{aligned} \tag{18.26}$$

This leads to the criterion for convergence when $\xi = \tanh(\mathbf{z})$:

$$4\eta \text{sech}^2(\mathbf{z}) \left(\gamma \mathbf{W}_\iota \tanh(\mathbf{z}) + (\mathbf{W}_{fb})_\iota \text{sech}^2(\mathbf{z}) - 2(\mathbf{W}_{fb})_\iota \tanh^2(\mathbf{z}) \right) > 0 \tag{18.27}$$

$$\frac{\gamma \sinh(\mathbf{z}) \cosh(\mathbf{z})}{1 - 2\sinh^2(\mathbf{z})} > \left(\frac{\mathbf{W}_{fb}}{\mathbf{W}} \right)_\iota ; \quad \text{sech}(\mathbf{z}) \neq 0 \tag{18.28}$$

with (18.28) representing a general condition for convergence of the *Feb*ICA algorithm, that relates the ratio $(\mathbf{W}_{fb}/\mathbf{W})$ for iteration ι. If the ratio is less than the specified bounds, the weights converge to the derived unmixing matrix.

One can view this as being analogous to the anti-Hebbian terms used for information maximization weight updates in [4]. In [10], the convergence properties of the tanh function with respect to complex and split-complex infomax

Table 18.1 Floating point operations involved with *Feb*ICA

Equation	Flop count
$\gamma_k = \sum_{n=0}^{N-1} \mathbf{x}_n e^{-j\frac{2\pi}{N}nk}$	$m(d\ log(d))$ for all $d \geq n$
$\mu = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{x}_n$	mn
$\xi = e^{-\beta_k/\delta_m} \cdot \gamma$	mn
$\psi_\iota = \mathbf{W}_\iota \gamma - (\mathbf{W}_{fb})_\iota \xi$	$\mathbf{I}dm^2n$
$\Delta\mathbf{W}_\iota = \eta\left(1 + \frac{\xi''}{\xi'}\psi_\iota\right)\mathbf{W}_\iota$	$\mathbf{I}dm^2n$
*Feb*ICA	$2m[n(1 + \mathbf{I}dm + d\ \log(d))]$

Table 18.2 Computational complexity of BSS techniques

Algorithm	Flop count
*Feb*ICA	$2m[n(1 + \mathbf{I}dm + d\ \log(d))]$
ACMA [6]	$9d^4n + 36m^2n$
JADE [16]	$8n^5 + 2d(m^2 + n^2) + \mathbf{I}(2n^4 + 10n^3 + 30n^2)$
SOBI [16]	$4n^3 + 2m^3 + 4n^2 + 0.5dm^2$
FastICA [16]	$2dm^2 + \mathbf{I}(2n(n + d) + 2.5dn^2)$
INFOMAX [16]	$2dm^2 + \mathbf{I}(n^3 + n^2 + n(5d + 4))$

algorithms are further studied. According to [4], the stochastic gradient leads to $(\mathbf{W}^{-1})_\iota = 2 \ll \tanh(\mathbf{z})\gamma^T \gg_\iota$ for which $\left(\frac{\mathbf{W}_{fb}}{\mathbf{W}}\right)_\iota$ in (18.28) can be expressed as $2(\mathbf{W}_{fb})_\iota \ll \tanh(\mathbf{z})\gamma^T \gg_\iota$, where $\ll\gg$ denotes the projection (scalar dot product) and $\tanh(\mathbf{z}) = \sum_i \mathrm{t}_i \mathbf{z}^{2p+1}$ for $p = 0, 1, 2\ldots$ with t_i being the coefficients coming from the Taylor series expansion of the tanh function. This reduces $\left(\frac{\mathbf{W}_{fb}}{\mathbf{W}}\right)_\iota$ to $2(\mathbf{W}_{fb})_\iota \left(\sum_i \mathrm{t}_i \ll \mathbf{z}_i^{2p+1}\gamma_i \gg\right)_\iota$. Similarly expanding the sinh and cosh terms with Taylor series coefficients s_i and c_i, we obtain:

$$\frac{\sum_i \mathrm{s}_i \ll \mathbf{z}^{2p+1}\gamma_i \gg \sum_i \mathrm{c}_i \mathbf{z}^{2p}}{2\left(1 - 2\left(\sum_i \mathrm{s}_i \mathbf{z}^{2p+1}\right)^2\right)\left(\sum_i \mathrm{t}_i \ll \mathbf{z}_i^{2p+1}\gamma_i \gg\right)_\iota} > \left(\mathbf{W}_{fb}\right)_\iota \tag{18.29}$$

which represents the criterion needed for convergence of weights $\mathbf{W}_{fb}$.

In Table 18.1, the floating point operations involved in various algorithmic steps of our *Feb*ICA method is listed. The computation is determined based on the sizes of the matrices $\mathbf{X}$ $(m \times n)$, $\mathbf{U}$ $(d \times m)$, and the number of iterations for convergence $\mathbf{I}$. The computational complexity of ACM, JADE, FastICA, INFOMAX, and SOBI algorithms has been investigated by [6] and [16], respectively. They are compared with *Feb*ICA for a $m \times n$ unmixing matrix with data length d and maximal iterations $\mathbf{I}$ in Table 18.2.

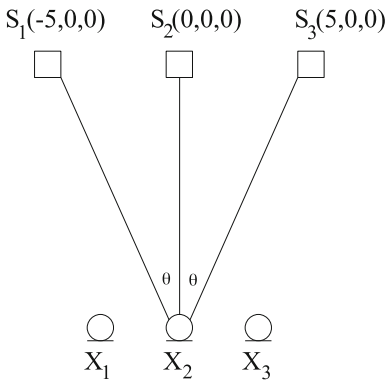


Fig. 18.3 Configuration of the sources S_1, S_2 and S_3 with respect to the sensor array used to generate the mixing process with $\theta = 15°$

18.4 Results and Performance

The results and performances of the proposed method are presented here in terms of output SINR as well as computational complexities (i.e., the number of Flop-count). The proposed method has been tested for the Gaussian minimum shift keying (GMSK) modulated signals which are generated by a simulator developed by Ekstrom and Mikkelsen [17] and distorted with additional white Gaussian noise. In order to develop a realistic model of the mixing process, the mixing matrix consists of the source and sensor array geometry along with the source direction vector to develop the mixing matrix. In this model, the mixing process embeds the source directions in the mixing matrix. The setup consists of 3-sensor ULA (uniform linear array) X_1, X_2, X_3 with half-wavelength spacing and 3 sources positioned as shown in Fig. 18.3. The sources S_1 and S_3 are situated 5 half-wavelengths away from S_2 making an angle of 15° from the normal to the array axis. The observed mixed signals are then separated using the source separation algorithms. An illustrative result of the proposed *Feb*ICA algorithm is shown in Fig. 18.4 and compared with the widely used second-order blind identification (SOBI) algorithm. As shown in Fig. 18.4, the proposed method for a mixture of three GMSK modulated sources with input SNR of 20 dB performs better than the SOBI algorithm. Note that the results in Fig. 18.4 illustrate that algorithms like SOBI that are developed for real-valued signals, fail to perform accurate source separation for complex-valued mixing matrices and signals.

We next compare the SINR improvement provided by various source separation algorithms. Both the nearly singular (when the sources are closed) and nonsingular (when the sources are apart) mixing matrices with varying number of sources and signal length are considered for our evaluation. For example, we have considered the following nonsingular mixing matrix $\mathbf{A}_1$ and nearly singular mixing matrix $\mathbf{A}_2$ with three complex sources. The mixing procedure can be modeled with either: (a) complex-valued instantaneous mixing matrices; (b) complex-discrete Fourier transform of a convolutive mixture (implies separation in the frequency domain).

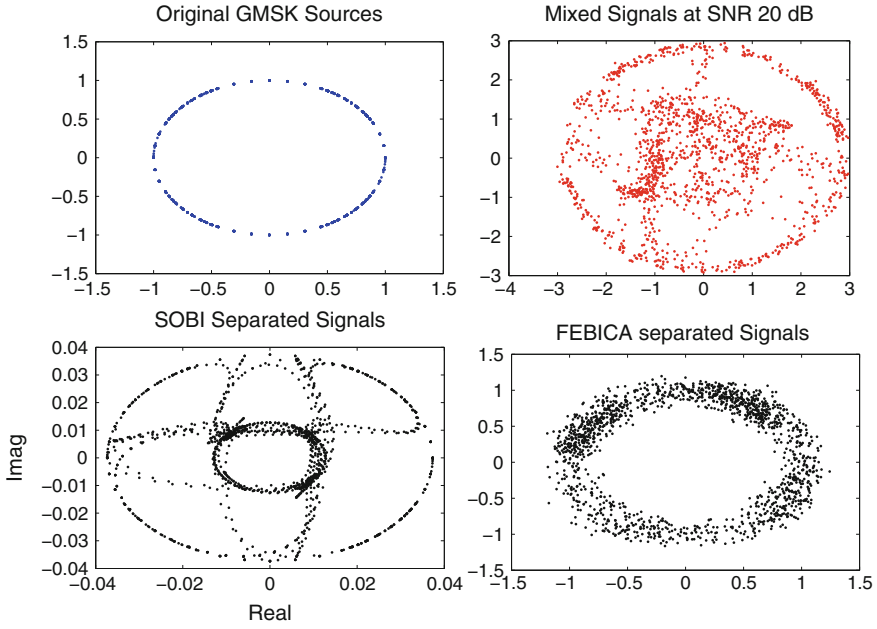


Fig. 18.4 Comparative results of the proposed *Feb*ICA algorithm and the SOBI algorithm [18] for Gaussian minimum shift keying (GMSK) modulated signals

$$\mathbf{A}_1 = \begin{bmatrix} -0.7458 + j0.6661 & -0.9794 + j0.2019 & -1.0000 + j0.0000 \\ 0.0000 + j1.0000 & -0.2291 + j0.9734 & -1.0000 + j0.0000 \\ 0.7458 + j0.6661 & 0.7865 + j0.6176 & -1.0000 - j0.0000 \end{bmatrix} \tag{18.30}$$

$$\mathbf{A}_2 = \begin{bmatrix} 0.1483 - j0.9889 & 0.5158 - j0.8567 & -1.0000 + j0.0000 \\ 0.1213 - j0.9926 & 0.4936 - j0.8697 & -1.0000 + j0.0000 \\ 0.0943 - j0.9955 & 0.4711 - j0.8821 & -1.0000 + j0.0000 \end{bmatrix} \tag{18.31}$$

The eigenvalues of $\mathbf{A}_1$ and $\mathbf{A}_2$ are represented by the diagonal values of the mixing matrices (see (18.32) and (18.33)). Note the lack of distinct eigenvalues for $\mathbf{A}_2$ in (18.33) which represents the nearly singular mixing condition.

$$eig(\mathbf{A}_1) = \begin{bmatrix} -1.4723 + j2.2598 & 0 & 0 \\ 0 & -0.0255 + j0.1070 & 0 \\ 0 & 0 & -0.4771 - j0.7272 \end{bmatrix} \tag{18.32}$$

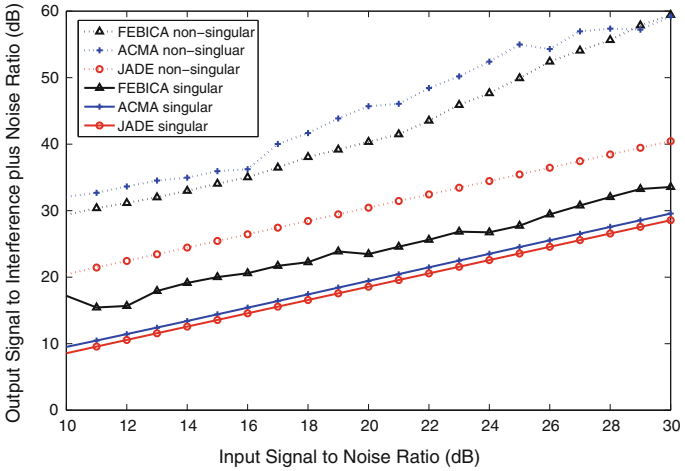


Fig. 18.5 The SINR improvements of various algorithms for different SNR settings with 10 source signals

$$eig(\mathbf{A}_2) = \begin{bmatrix} -0.3770 - j1.8210 & 0 & 0 \\ 0 & 0.0189 - j0.0376 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{18.33}$$

As illustrated in Fig. 18.5, for the case of nonsingular sources, the *Feb*ICA algorithm improves with increase in SINR and outperforms the JADE algorithm. The performance of all the three BSS algorithms drops when the mixing matrix is nearly singular. The reason for this lies in the approximation of the inverse of the nearly singular mixing matrix, that would require additional processing not provided by traditional BSS algorithms. However, as shown in Fig. 18.5, the proposed *Feb*ICA algorithm performs relatively well in the case of singular mixing matrices, attuned to the feedback approach to generating the unmixing matrices. Generally speaking, without BSS algorithm the output SINR becomes significantly lower than the input SNR. However, the incorporation of BSS algorithm in communication systems certainly improves the signal quality making the output SINR well above the input SNR. Further, in Fig. 18.5, it can be noticed that the performance curve of the *Feb*ICA singular seems to be linear, but becomes less consistent for *Feb*ICA nonsingular; similar trends are observed for ACMA singular and ACMA nonsingular. Conditions for statistical independence are valid for nonsingular case when compared to singular (ill-conditioned) case. Thus, higher the input SNR, larger the improvement of output SINR for the nonsingular case. Note that the results for SOBI [18] and FastICA [3] algorithms are not presented, as they are unable to separate complex-valued signals.

We further compare the results of output SINR with increasing number of sources. As noticed in Fig. 18.6, the improvements achieved by the proposed *Feb*ICA algorithm for the complex modulated signals with data size 1480 samples and input SNR of 25 dB are found to be consistent with increasing number of sources. Although

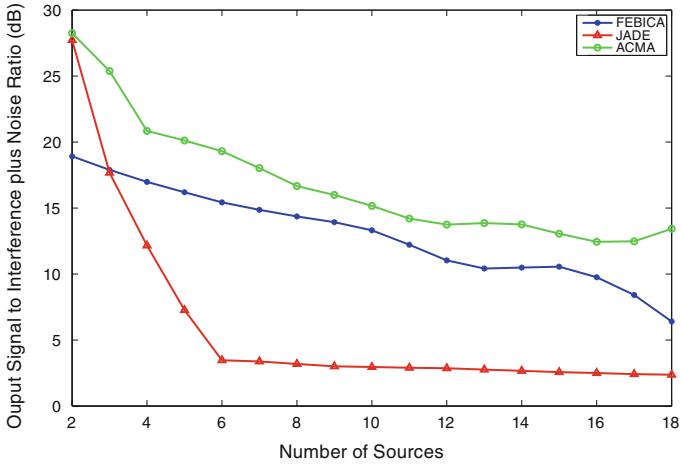


Fig. 18.6 The SINR improvements with increasing number of sources for an input SNR of 20 dB

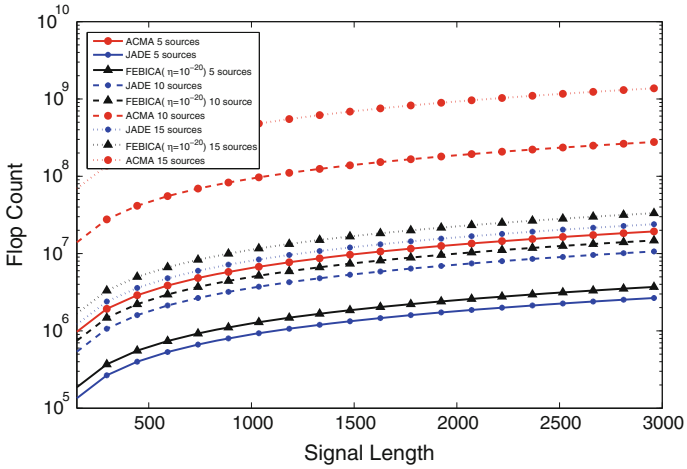


Fig. 18.7 The floating point values of various BSS algorithms for increasing signal length and number of sources

JADE algorithm provides nearly optimal performance when number of sources less is than 4, the performance deteriorates drastically as the number of sources increases. On the other hand, by making use of gradient ascent techniques for optimizing the unmixing weights, the proposed algorithm seems to be more stable for large number of sources without deteriorating the performance. The ACMA is also found quite stable as the number of sources increases. However, it is computationally more intensive as shown in Fig. 18.7. The results by SOBI [18] and FastICA [3] algorithms are not included here due to deteriorating separation performance for complex-valued signals.

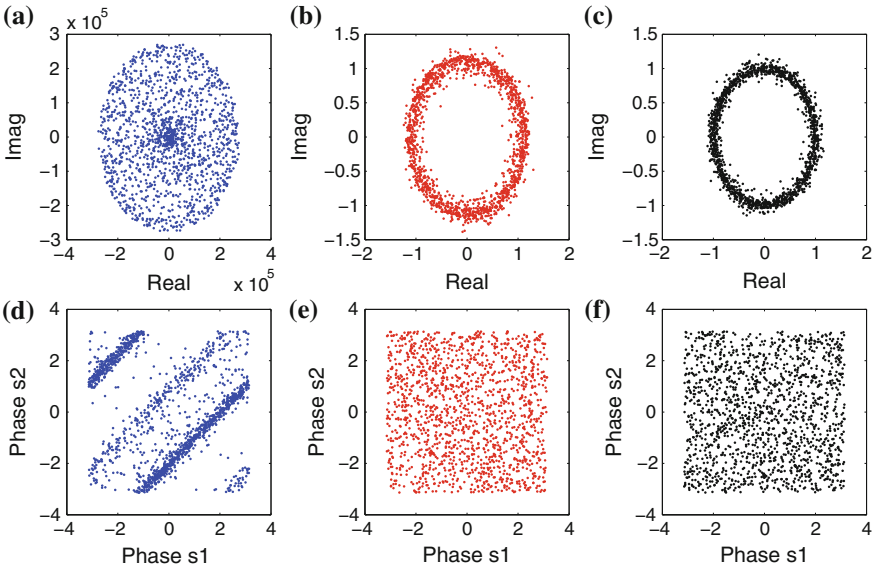


Fig. 18.8 The separation performance of JADE and *Feb*ICA when applied to real field data: **a** observed mixed signal constellation, **b** signal constellation after JADE, **c** signal constellation after *Feb*ICA , **d** relative phase distribution of sources s1 and s2, **e** relative phase distribution after JADE, **f** Relative phase distribution after *Feb*ICA

Figure 18.7 represents the floating point operations associated with various BSS algorithms. We observe that the *Feb*ICA algorithm substantially outperforms the ACMA when the number of sources is more than 6. It performs comparably well with the JADE algorithm especially for large number of sources. Comparing Figs. 18.5 and 18.7, we can see that *Feb*ICA algorithm provides higher SINR gain for GMSK mixtures while maintaining low computational complexity. It can be noted that the computational complexity of *Feb*ICA depends on the number of corresponding iterations, which in turn relates to the threshold specified in (18.28).

The results of the proposed *Feb*ICA algorithm on real field data is presented in Fig. 18.8. The I-Q data of CFSK modulated signals produced by two sources have been sampled at 25 kHz and received by two receivers. Then the data have been sent out in bursts with 80 % of the time frame occupied by data and 20 % left for synchronization. The SNR value has been recorded to be within a range of 30–40 dB. As shown in Fig. 18.8, *Feb*ICA and JADE algorithms are able to restore the signal constellations of the received data to that of a typical GMSK signal. The relative phase between the two separated signals is also randomized, indicating that the output signals become more independent than the observed input data which follows the fundamental assumptions of BSS. Further illustration of the good separation performance of *Feb*ICA is seen in the time-frequency plots of the mixed and separated outputs, as shown in Fig. 18.9.

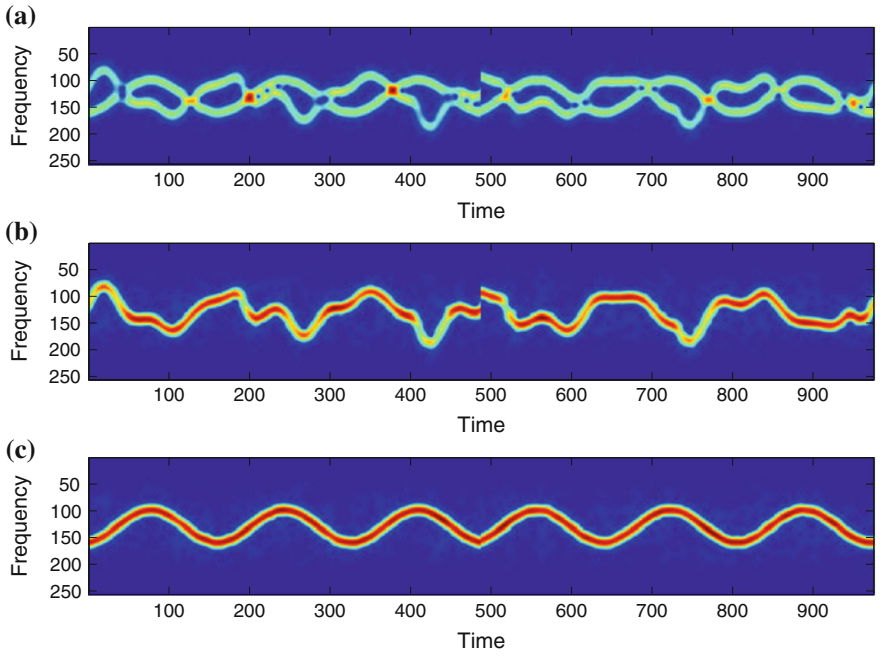


Fig. 18.9 The separation performance of *Feb*ICA when applied to real field data: **a** Observed mixed signal in TF domain **b**, **c** separated signal outputs

18.5 Related Work

In wireless communication systems, co-channel interference may be prevented by applying medium access control (MAC) or random access protocols [19]. By applying cross layer designs, signal processing techniques like blind source separation can be employed along with MAC protocols to improve the performance. As the source separation is applied on digitally modulated signals, it is imperative that the processing is algorithms for separating complex signal mixtures have been described in the literature and the algorithms designed for real-valued data are unable to effectively separate these mixtures. While most of the conventional algorithms like FastICA [3] and Infomax [4] may be applied to real signals, algorithms like JADE [5] and (ACMA) [7] can handle source separation in complex domain. Modified versions of Infomax [10] have been proposed to extend these algorithms to handle complex-valued signals.

While traditional blind source separation algorithms have been applied to speech and music signals, interest in applying these techniques to telecommunication systems has generated interest. It has been applied to multiple aspects of telecommunication systems including blind deconvolution, MIMO channel separation and equalization. In [20], source separation techniques have been applied to estimation

of many parameters belonging to spectrally overlapping sources. The two approaches considered including convolutive unmixing of the channel response function/observed data and exploiting the cyclostationarity of the underlying signals. Dubroca et al. [21] deal with the problem of blind source extraction from a multiple-input/multiple-output (MIMO) convolutive mixtures. In [22], the maximum likelihood method for parameter estimation is shown to provide a unifying framework for deriving blind deconvolution and blind source separation algorithms. The number of communication signals in a sensor array is studied in [23], by exploiting Kolomogorov–Smirnov (K-S) tests along with ICA. In [24], ICA-based equalization approaches are shown to be superior to subspace methods for MIMO systems.

The use of blind source separation to complex domain signals produces some challenges that traditional ICA algorithms cannot handle. In [25], an entropy estimator for complex random variables is applied to separation of complex sources. In a chapter relevant to our work [26], an efficient algorithm to deal with convolutive blind source separation of communication signals in frequency domain is presented by making use of JADE in every frequency bin. In [27], a complex optimum block adaptive ICA (Complex OBA-ICA) is applied to orthogonal frequency division multiplexing (OFDM) to recover user signals in the presence of ICI and channel induced mixing. In [28], a new nonlinear function is proposed for natural gradient separation algorithms (NGSA) to improve separation performance. In [29], an Infomax algorithm has been used for BSS of complex-valued signals in frequency domain where particle swamp optimization (PSO) technique is integrated to find suitable step size for the learning process. Another frequency-domain ICA method using short-time Fourier transform (STFT) is presented in [30] for wireless communications in order to reduce the cross-talk in slow, frequency-selective channels.

In order to evaluate the performance of ICA algorithms for communication signals, [31] sub-Gaussian, Gaussian, and mix users (sub-Gaussian, super-Gaussian, and Gaussian) are generated and then mixed linearly. Separation performance of multiple ICA algorithms such as JADE, Infomax, and Fixed point algorithms are measured by a performance index. In [32], the performances of different types of ICA algorithms (e.g., Infomax, JADE, Pearson-ICA [33], SOBI [18]) are presented for mitigating interferences for the systems based on direct sequence code division multiple access (DS-CDMA) used in commercial cellular networks. In [34], fourth-order cumulant-based separation is applied to multi-user symbol estimation problem in direct sequence code division multiple access (DS-CDMA) systems. Bit error rate (BER) simulations of this algorithm are shown for different number of users, signal to noise ratio (SNR), and different number of symbols per user in comparison with the FastICA algorithm and robust ICA. However, unlike our case, the simulated mixing and separation has been applied to the real domain.

18.6 Conclusion and Future Work

An effective blind source separation algorithm is proposed to reduce the co-channel interference for complex communication signals. The proposed *Feb*ICA scheme performs well in complex domain with low computational complexity and good SINR improvements. When applied to a mixture of co-channel interfering digitally modulated signals, the *Feb*ICA algorithm is shown to outperform the JADE algorithm in terms of SINR improvement. This improvement is seen over a range of input SNR settings as well as in the number of interfering sources. Furthermore, it is shown to have lower computational complexity compared to ACMA, which is less dependent on increasing the signal length as well as the number of sources. It is also stable with increase in the number of sources providing consistent SINR improvements. Thus, it combines the dual advantages of successful interference mitigation and lower computational complexity. This makes the proposed algorithm suitable for practical applications in wireless receivers.

The future work includes to apply our proposed method in vehicular communications [35, 36] where large number of sources (vehicles sensors) are involved and the number of sources are consequently varied. For example, our FFT-based BSS algorithm could be useful to improve real-time vehicle localization for effective traffic monitoring. While the scope of this chapter is centered on the development of a novel algorithm and corresponding performance analysis on communication signals, a detailed comparison with other classes of such complex/frequency domain algorithms (such as [26] and [34]) would also be evaluated in future.

Appendix

Mutual Information Weight Update Criterion

The mutual independence of functions based on joint probability density functions is given by:

$$f_s(s) = \prod_{i=1}^{n} f_i(s_i) \tag{18.34}$$

where $f_i(s_i)$ is the pdf of signal s_i. Based on signal entropy $\mathbf{H}$, the mutual information $\mathbf{I}$ is given by:

$$\mathbf{I}(s_1, s_2, \ldots, s_n) = \sum_{i=1}^{n} \mathbf{H}(s_i) - \mathbf{H}(s) \tag{18.35}$$

$$\mathbf{H}(s_i) = -\int f_i(s_i) \ln f_i(s_i) ds_i \tag{18.36}$$

The objective of BSS is to find an unmixing matrix $\mathbf{W}$, so that $\mathbf{I}(s_i, s_j) = 0$. Traditionally, higher order statistics (HOS) and associated nonlinearities can be used to produce mutual independence [2]. For observed signals $\mathbf{Y} = \mathbf{WX}$, let the transformed function (due to nonlinearities) be given by:

$$z_i = g_i(y_i) \tag{18.37}$$

For a single variable z, the $\mathbf{H}(z)$ is maximized when nonlinearity $g(\cdot)$ is a cumulative density function. In other words, $\mathbf{H}(z)$ is maximized when z has uniform distribution:

$$f_z(z) = \frac{f_y(y)}{\mathrm{d}z/\mathrm{d}y} \tag{18.38}$$

This is of uniform distribution when

$$\frac{\mathrm{d}z}{\mathrm{d}y} = f_y(y) \tag{18.39}$$

For n variables, this may be extended using the Jacobian form $\mathbf{J}$:

$$f_z(z) = \frac{f_y(y)}{|\mathbf{J}|} \tag{18.40}$$

$$\mathbf{J} = \det \begin{bmatrix} \frac{\mathrm{d}z_1}{\mathrm{d}y_1} & \cdots\cdots & \frac{\mathrm{d}z_1}{\mathrm{d}y_n} \\ \vdots & \ddots & \\ \frac{\mathrm{d}z_n}{\mathrm{d}y_1} & \cdots\cdots & \frac{\mathrm{d}z_n}{\mathrm{d}y_n} \end{bmatrix} \tag{18.41}$$

The output joint entropy is then given by:

$$\mathbf{H}(z) = -\mathbf{E}\{\ln f_z(z)\} = -\mathbf{E}\{\ln f_x(x)\} + \mathbf{E}\{\ln |\mathbf{J}|\} = \mathbf{H}(x) + \mathbf{E}\{\ln |\mathbf{J}|\} \tag{18.42}$$

For maximizing $\mathbf{H}(z)$, the updating weights for the unmixing matrix are given by:

$$\Delta\mathbf{W} \propto \frac{\mathrm{d}\mathbf{H}(z)}{\mathrm{d}\mathbf{W}} = \mathbf{E}\left\{\frac{\mathrm{d}\ln|\mathbf{J}|}{\mathrm{d}\mathbf{W}}\right\} \cong \frac{\mathrm{d}}{\mathrm{d}\mathbf{W}} \ln |\mathbf{J}| \tag{18.43}$$

Using the definition of the Jacobian:

$$\mathbf{J} = \det(\mathbf{W}) \prod_{i=1}^{n} \left|\frac{\mathrm{d}z_i}{\mathrm{d}y_i}\right| \tag{18.44}$$

$$\Delta\mathbf{W} \propto \frac{\mathrm{d}}{\mathrm{d}\mathbf{W}} \ln |\det(\mathbf{W})| + \frac{\mathrm{d}}{\mathrm{d}\mathbf{W}} \ln \prod_{i=1}^{n} \left|\frac{\mathrm{d}z_i}{\mathrm{d}y_i}\right| \tag{18.45}$$

It can be proved that:

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{W}} \ln |\det(\mathbf{W})| = \left|\mathbf{W}^T\right|^{-1} \tag{18.46}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{W}} \ln \prod_{i=1}^{n} \left|\frac{\mathrm{d}z_i}{\mathrm{d}y_i}\right| = \frac{\mathrm{d}}{\mathrm{d}\mathbf{W}} \prod_{i=1}^{n} \ln \left|\frac{\mathrm{d}z_i}{\mathrm{d}y_i}\right| = f(y)x^T \tag{18.47}$$

Substituting this into (18.45), we obtain

$$\Delta\mathbf{W} \propto \left|\mathbf{W}^T\right|^{-1} + f(y)x^T \tag{18.48}$$

$$\mathbf{W}_{(k+1)} = \mathbf{W}_{(k)} + \eta(k)[(\mathbf{W}^T)^{-1} + f(y)x^T] \tag{18.49}$$

Multiplying this learning rule by $\mathbf{W}^T\mathbf{W}$, we get the simplified expression:

$$\mathbf{W}_{(k+1)} = \mathbf{W}_{(k)} + \eta(k)[\mathbf{I} + f(y)y^T]\mathbf{W} \tag{18.50}$$

References

1. Hyvarinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Netw. **13**(4–5), 411–430 (2000)
2. Pedersen, M.S., Larsen, J., Kjems, U., Parra, L.C.: A survey of convolutive blind source separation methods. Springer handbook on speech processing and speech communication. Springer, New York (2007)
3. Hyvarinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Comput. **9**, 483–492 (1997)
4. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Comput. **7**, 129–159 (1995)
5. Cardoso, J.F., Souloumiac, A.: Jacobi angles for simultaneous diagonalization. SIAM J. Matrix Anal. Appl. **17**(1), 161–165 (1996)
6. Van der Veen, A.J., Paulraj, A.: An analytical constant modulus algorithm. IEEE Trans. Signal Process **44**(5), 1136–1155 (1996)
7. Van der Veen, A.J., Paulraj, A., Buchner, A., Aichner, R., Kellermann, W.: A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. IEEE Trans. Speech Audio Process **13**(1), 120–134 (2005)
8. Hopfield, J.J.: Olfactory computation and object perception. Proc. Natl. Acad. Sci. USA **88**, 6462–6466 (1991)
9. Jutten, C., Herault, J.: Blind separation of sources, part 1: an adaptive algorithm based on neuromimetic architecture. Sig. Process. **24**, 1–10 (1991)
10. Calhoun, V., Adali, T.: Complex infomax: convergence and approximation of infomax with complex nonlinearities. In: Proceedings of 12th IEEE Workshop on Neural Networks for Signal Processing, pp. 307–316 (2002)
11. Lee, T.W., Bell, A.J., Orglmeister, R.: Blind source separation of real world signals. In: Proceedings of International Conference on Neural Networks, vol. 4, pp. 2129–2134 (1997)
12. Lambert, R.H., Nikias, C.L.: Fast converging methods for multichannel blind equalization or separation of multipath mixtures. In: Proceedings of Military Communications Conference vol. 3, no. 21–24, pp. 854–858 (1996)

13. Karam, M., Fadali, M.S., White, K.: A Fourier/Hopfield neural network for identification of nonlinear periodic systems. In: Proceedings of the 35th Southeastern Symposium on System Theory, pp. 53–57 (2003)
14. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. In: Advances in Neural Information Processing Systems, pp. 757–763. MIT Press, Cambridge (1996)
15. Fletcher, R.: Practical methods of optimization, vol. 1, 2nd edn. Wiley, New York (1987)
16. Kachenoura, A., Albera, L., Senhadji, L., Comon, P.: ICA—a potential tool for brain computer interface systems. IEEE Sig. Process. Mag. **8**, 57–68 (2007)
17. Ekstrom, A.N., Mikkelsen, J.H.: GSMsim—A MATLAB implementation of a GSM simulation platform (Division of Telecommunications). Aalborg University, Denmark (1997)
18. Belouchrani, A., Abed-Meraim, K., Cardoso, J.F., Moulines, E.: A blind source separation technique using second order statistics. IEEE Trans. Sig. Process. **45**(2), 434–444 (1997)
19. Gummalla, A.C., Limb, J.: Wireless medium access control protocols. IEEE Commun. Surv. Tutorials. **3**(2), 2–15 (Second Quarter 2000)
20. Chevalier, P., Chevreuil, A.: Chapter 17—Application to telecommunications. In: Comon, P., Jutten, C. (eds.) Handbook of Blind Source Separation, pp. 683–735. Academic Press, Oxford (2010)
21. Dubroca, R., De Luigi, C., Castella, M., Moreau, E.: A general algebraic algorithm for blind extraction of one source in a MIMO convolutive mixture. IEEE Trans. Signal Process. **58**(5), 2484–2493 (2010)
22. Douglas, S.C., Haykin, S.: On the relationship between blind deconvolution and blind source separation. In: 31st Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 1591–1595 (1997)
23. Choqueuse, V., Yao, K., Collin, L., Burel, G.: Blind detection of the number of communication signals under spatially correlated noise by ICA and K-S Tests. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2397–2400 (2008)
24. Nandi, A.K., Gao, J., Zhu, X.: Independent component analysis—an innovative technique for wireless MIMO OFDM systems. In: 4th International Conference on Computers and Devices for Communication (2009)
25. Li, X., Adali, T.: Complex independent component analysis by entropy bound minimization. IEEE Trans. Circuits Syst. I Regul. Pap. **57**(7), 1417–1430 (2010)
26. Duan, T., Zhang, X.: A solution to blind separation of convolutive communication mixtures in frequency domain. In: 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 2330–2333, April (2012)
27. Ranganathan, R., Yang, T., Mikhael, W.: Intercarrier interference mitigation and multi-user detection employing adaptive ICA for MIMO-OFDM systems in time variant channels. In: IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS) (2011)
28. Liao, H., Li, W., Wei, P.: Blind signal separation based on new nolinear function. In: International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) (2010)
29. Geravanchizadeh, M., Hesam, M.: PSO-based infomax algorithm for frequency-domain blind source separation. In: Iranian Conference on Electrical Engineering (2011)
30. Liu, Y., Mikhael, W.B.: A novel frequency-domain independent component analysis approach for wireless communications. In: International Conference on Electronics, Control and Signal Processing (WSEAS), pp. 187–192, (2005)
31. Parmar, S.D., Unhelkar, B.: Separation performance of ICA algorithms in communication systems. In; International Conference on Multimedia, Signal Processing and Communication Technologies, pp. 142–145, (2009)
32. Parmar, S., Unhelkar, B.: Independent component analysis algorithms in wireless communication systems. In: Mobile Business—Technical, Methodological and Social Perspectives, 2nd edn, pp. 456–463. IGI Publisher (2009)
33. Solvang, H.K., Nagahara, Y., Araki, S., Sawada, H., Makino, S.: Frequency-domain Pearson distribution approach for independent component analysis (FD-Pearson-ICA) in blind source separation. IEEE Trans. Audio Speech Lang. Process. 17(4), 639–649 (2009)

34. Albataineh, Z., Salem, F.: New blind multiuser detection DS-CDMA algorithm using simplified fourth order cumulant matrices. In: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1946–1949 (2013)
35. Andrews, S.: Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communications and cooperative driving. In: Eskandarian, A. (ed.) Handbook of Intelligent Vehicles. Springer, London (2012)
36. Zeletin, R.P., et al.: Applications of vehicular communications. Vehicular-2-X Communication. Springer, Heidelberg (2010)

Chapter 19
Semi-blind Functional Source Separation Algorithm from Non-invasive Electrophysiology to Neuroimaging

Camillo Porcaro and Franca Tecchio

Abstract Neuroimaging, investigating how specific brain sources play a particular role in a definite cognitive or sensorimotor process, can be achieved from non-invasive electrophysiological (EEG, EMG, MEG) and multimodal (concurrent EEG-fMRI) recordings. However, especially for the non-invasive electrophysiological techniques, the signals measured at the scalp are a mixture of the contributions from multiple generators or sources added to background activity and system noise, meaning that it is often difficult to identify the dynamic activity of generators of interest starting from the electrode/sensor recordings. Although the most common method of overcoming this limitation is time-domain averaging with or without source localization, blind source separation (BSS) algorithms are becoming increasingly widely accepted as a way of extracting the different neuronal sources that contribute to the measured scalp signals without trial exclusion. The advantage of BSS or semi-blind source separation (semi-BSS) techniques compared to methods such as time-domain averaging lies in their ability to extract sources exploring the whole time evolving data. Taking into account the whole signal without averaging, it also provides a means suitable to investigate non-phase locked oscillatory processes and single-trial behaviour. This characteristic becomes a crucial issue when investigating combined EEG-fMRI data, particularly when the focus is on neurovascular coupling definitely dependent on single trial variability of the two datasets. In this context, this chapter describes a semi-BSS technique, Functional Source Separation (FSS), which is a tool to identify cerebral sources by exploiting a priori knowledge, such as spectral or evoked activity, which cannot be expressed by sources other than the one to be

C. Porcaro (✉) · F. Tecchio
LET'S-ISTC-CNR, Fatebenefratelli Hospital—Isola Tiberna, Rome, Italy
e-mail: camillo.porcaro@istc.cnr.it

C. Porcaro
Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK

F. Tecchio
Department of Neuroimaging, IRCCS San Raffaele Pisana, Rome, Italy
e-mail: franca.tecchio@istc.cnr.it

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_19,
© Springer-Verlag Berlin Heidelberg 2014

identified (functional *fingerprint*). In other words, FSS allows the identification of specific neuronal pools on the bases of their functional roles, independent of their spatial position.

19.1 Introduction: Relevance and Challenges of Electrophysiology for Neuroimaging

Despite non-invasive electrophysiological techniques such as Electroencephalography (EEG) and Magnetoencephalography (MEG) providing the opportunity to directly measure the electrical activity of large-scale neuronal populations, different challenges exist in characterising this activity, especially at the *single-trial* level [52]. **Single-trial variability and signal-to-noise ratio:** EEG or MEG signals not only reflect activity of the neuronal population of interest (*signal*) but also unrelated activity (*artefact/noise*). Consequently, changes in brain activity over a *single-trial* can be dominated by changes in these artefacts. **Electric/magnetic field propagation:** Electrical potentials and magnetic fields generated by neuronal activity are not only detected close to neuronal sources but also at distant sites. Therefore, each channel derives its signal from more than one source. In this respect, choosing or averaging channels may generate misleading results. Blind Source Separation (BSS) methods such as Independent Component Analysis (ICA), [21, 27, 47] and semi-BSS such as Functional Source Separation (FSS), [47, 54] have been successfully used to separate brain sources from noise and are therefore strong candidates to reduce the above problems. FSS extracts on the basis of a typical behavioural property, which cannot be expressed by other sources than the one to be identified (functional *fingerprint*), independent of the spatial position, are extremely helpful especially in those cases where cerebral plastic changes have altered the location of brain functions with respect to standard anatomical landmarks typical of healthy people [16]. The independence from spatial position also allows the separation of sources close to each other (i.e. primary somatosensory and primary motor representation of the same body district, two hand fingers somatosensory representation).

In particular, FSS uses the simulated annealing algorithm for constraint optimization, allowing non-differentiable contrast function and performing global optimization, while gradient-based algorithms usually employed in ICA only guarantee local optimization. Since brain areas could not always be reasonably assumed independent or uncorrelated, in the FSS procedure the orthogonalization step could also be omitted, producing a non-orthogonal extraction scheme [5]. In this condition, the order of extraction is not significant, because the procedure is applied each time to the original data. A typical functional constraint is applied each time to produce each source. The proposed chapter is the culmination of our research line devoted to source identification grounded on functional behaviour of the cerebral source of interest.

The structure is as follows. In Sect. 19.2, general principles of BSS and semi-BSS with a particular focus on FSS are introduced. In particular, part of this section is devoted to the simulated annealing, the optimization method used by FSS to minimize the cost function is under investigation. The functional constraints, the core of FSS methodology, are described in Sect. 19.3, where we exemplify cases of neuronal pools already identified through FSS (primary somatosensory and motor areas and primary visual area), together with the procedures to assess the 'goodness' of the extraction. Section 19.4 is dedicated to the FSS applied on simultaneous EEG-functional Magnetic Resonance Imaging (fMRI) recordings. How FSS enables the single-trial response measurement is shown in Sect. 19.5 and provides examples from EEG data as well as from technically challenging situations of concurrent EEG-fMRI recordings. Strengths and limitations of FSS procedures, as well as final considerations, are reported in Sect. 19.6.

19.2 Semi-BSS and FSS

19.2.1 Brief Overview

BSS is a multivariate class of computational data analysis techniques for revealing hidden factors that underlie sets of measurements or signals. Thus BSS can be seen as a generative model of *latent variables*, also called *sources* or *factors*, which describes how the observed data are generated by these unknown sources, under the hypothesis that their contributions are linearly mixed (the mixing coefficients are also unknown). A very famous subclass of BSS is ICA, in which the *latent variables* are non-Gaussian distributed and mutually independent. In other words, the aim of ICA is to extract in a *blind* fashion (i.e. without making specific assumptions) meaningful signals that have been linearly mixed, without knowing the original signals or the mixing coefficients.

Based on the observation that when we deal with real-world signals we are never completely *blind*, in that we know (in a more or less detailed and quantitative way) some of their characteristic features, such as the form of their probability densities, their spectral or temporal contents, spatial position, etc. Then the term *blind* is replaced by the term *semi-blind*, and we are in the case of semi-BSS techniques.

19.2.2 ICA a Special Case of BSS

In the simplest noise-free version of the ICA model, a set of signals $\mathbf{x}$ (in the specific case of this chapter simply the recorded MEG/EEG channels) is assumed to be obtained as a linear combination (through an unknown instantaneous mixing matrix $\mathbf{A}$) of statistically independent non-Gaussian sources $\mathbf{s}$. Since the observed

mixed signals will tend to have more Gaussian amplitude distributions, ICA strives to find a separation matrix that maximises the non-Gaussian features of the data, thus optimally separating the signals. For this purpose, we assumed the set of observed EEG/MEG signals to be generated by the mixing model:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{19.1}$$

where $t = 0, 1, 2, \ldots$ is the discrete sampling time; $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]$ is the m-dimensional vector of the observed signal recorded by m electrodes (EEG) or sensors (MEG); $\mathbf{A}$ is an $m \times n$ (with $n \leq m$) unknown full-rank mixing matrix; $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ is the n-dimensional unknown vector of the sources. The model is approached by processing electrode signals by an ICA demixing system described in the form:

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{19.2}$$

where $\mathbf{y}(t) = [y_1(t), \ldots, y_n(t)]^T$ n-dimensional vector of the estimated Independent Components (ICs) and $\mathbf{W}$ is the separation matrix, i.e., the estimate of the inverse of the unknown mixing matrix $\mathbf{A}$, up to permutation and scaling.

$$\mathbf{W} = \hat{\mathbf{A}}^{-1} \tag{19.3}$$

$^{-1}$ should be intended as pseudo-inverse of the matrix in case $n < m$. However, Eq. 19.2 has fewer unknowns than equations, making the estimation rely on additional information, namely the statistical independence of sources. ICA can therefore be cast as an optimization process that maximises independence as described indirectly by a suitable contrast function.

In particular, the ICA assumption is that a set of statistically independent sources $\mathbf{s}$ have been mixed linearly in the recorded data $\mathbf{x}$ by means of a mixing matrix $\mathbf{A}$. The aim is to recover both $\mathbf{s}$ and $\mathbf{A}$ starting from the observation of the linear mixture $\mathbf{x} = \mathbf{A}\mathbf{s}$ without making any particular assumption other than statistical independence of the sources.

19.2.3 FSS a Special Case of Semi-BSS

FSS technique is part of semi-BSS methods [5, 43, 54]. The aim of FSS is to enhance the separation of relevant signals by exploiting some of a priori knowledge without renouncing the advantages of using only information contained in original signal waveforms. FSS, analogous to ICA, models the set of EEG/MEG recorded signals $\mathbf{x}$ as a linear combination of an equal number of sources $\mathbf{s}$ via a mixing matrix $\mathbf{A}$. Differing from other constrained ICA models (for details about these parts of the semi-BSS technique see [23, 26, 60], FSS identifies a single source at a time, building a contrast function for that source that exploits *fingerprint* information associated with the neuronal pool to be identified [43–45, 47, 54]. In general, FSS starts from the original

EEG/MEG data matrix **x** for each source, and returns one functional source (FS) with the required functional property. This scheme gives us the ability to extract the FS that maximises the functional behaviour in agreement with the functional constraint [5, 54]. A modified cost function (with respect to standard ICA) is defined as:

$$F = J + \lambda R \tag{19.4}$$

where J is the statistical constraint normally used in ICA, while R accounts for the a priori information known about the sources. The relative weight of these two parameters can be adjusted via λ [43, Appendix A]. λ has been chosen to both minimize computational time and maximise the functional constraint R (see also Sect. 19.2.3.1). Moreover, the FSS contrast function F is optimized by means of simulated annealing [24], thus allowing prior information about the FS to be described by a non-differentiable function. Furthermore, FSS performs global optimization instead of local optimization (gradient based algorithms) usually employed in ICA. To separate contributions representing different sources, the proposed procedure could be applied in two different modalities: by using an *orthogonal* extraction scheme (as in the basic ICA model); by estimating the first source, and then searching for the second source in the orthogonal space with respect to the first, and so on until the last source is estimated with a stop rule that can be defined according to the data in hand. Since relevant components cannot always be reasonably assumed to be independent or uncorrelated, the FSS procedure using the orthogonalization step could be also completely skipped, producing a *non-orthogonal* extraction scheme. In this condition, the order of extraction is not significant because the procedure is applied each time to the original data. Different constraints are applied each time to produce different sources.

The provided sources are suitable to describe time evolution of on-going activity, which allow for sample single-trial analysis, instead of averaging all sensors channels in specific instants, as is usually done in the standard procedures. Moreover, even if a source is extracted by exploiting a functional constraint related to a specific time portion of the experiment, the corresponding estimated signal could be studied all along the length of the whole session.

In the following subsections, we describe how the parameter λ has been chosen (Sect. 19.2.3.1) and how the simulated annealing works (Sect. 19.2.3.2). Finally, details on the functional constraint R will be given in Sect. 19.3.

19.2.3.1 FSS Contrast Function Settings

In this section, we specify how the values of the parameter λ in Eq. 19.4 was determined. The parameter λ was selected by an initial grid fixed with nine different λ-*values* ($\lambda = 0; 0.01; 0.1; 0.5; 1; 10; 100; 500; 1000$) plus a last condition of $\lambda = 1$ but only activating the functional constraint in Eqs. 19.4, i.e. removing the *J-part* of the contrast function (case named as 'Only Constraint', OC). For each participant (in this specific case 14 volunteers, 9 males and 5 females, mean age 41 $\pm$

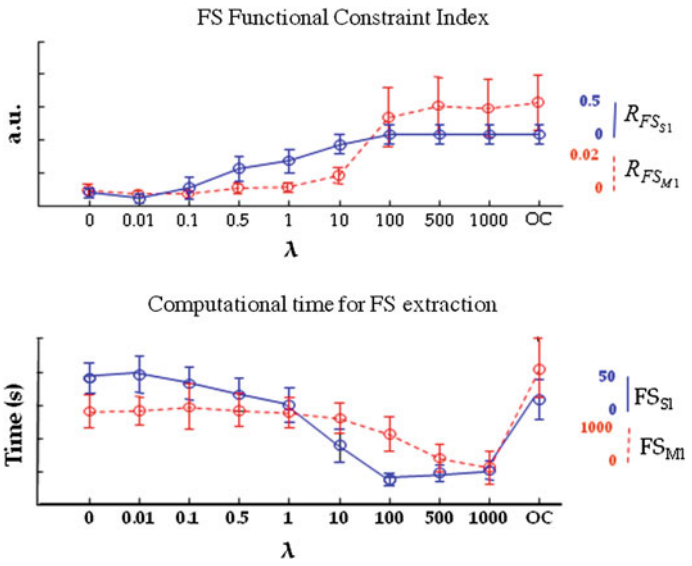


Fig. 19.1 Mean and standard errors across subjects for the tested λ-*values* of two representative functional constraints $R_{FS_{S1}}$ (*solid blue line*) and $R_{FS_{M1}}$ (*dashed red line*) indices (*top*) and of the computational times for the two sources FS_{S1} (*solid blue line*) and FS_{M1} (*dashed red line*) extractions (*bottom*)

15, age range 24–66 years), the two sources (S1–primary sensory area[1]; M1–primary motor area[2]) were extracted, keeping track of the computational times for each case of the grid. The indices $R_{FS_{S1}}$ and $R_{FS_{M1}}$ (see Eqs. 19.8 and 19.9 for formulation) were evaluated for each corresponding source. The parameter λ, consequently, was chosen to both minimize computational times and maximise the *R* indices. Starting from the case $\lambda = 100$, the maximum value of the two indices was reliably reached for all the subjects and the two sources (Fig. 19.1, top). Moreover, looking at computational time distribution (Fig. 19.1, bottom), the λ value minimising the computational effort for the two sources was $\lambda = 1000$. The computational time was estimated for a computer with 3.2 GHz CPU and 1.0 GB RAM, on the data matrix of 28 rows × 240,000 time points. In all cases highlighted in this chapter, we have found that $\lambda = 1000$ is a good compromise to minimise computational time and maximise each functional constraint. Noteworthy, while best solutions corresponded to $\lambda \geq 100$ (i.e. R weighted above two orders of magnitude greater than kurtosis J), we concurrently documented that it is useful not to omit J from functional constraint (OC Fig. 19.1). In fact, we empirically observed that $J = 0$ (OC) required more than three times computational cost than using $\lambda = 1000$.

[1] See Sect. 19.3.1 for more details.

[2] See Sect. 19.3.2 for more details.

19.2.3.2 Optimization by Simulated Annealing

Simulated Annealing (SA) is a well-known global optimization technique [24] used mainly in mechanical statistics. The optimization process is based on the perturbation of a given solution, according to the concepts of temperature, statistical equilibrium and probabilistic acceptance.

The approach has been inspired by the annealing process in metallurgy, a process used to reach the state of minimum energy of a solid (metal). The technique consists of raising the temperature of a metal up to the maximum degree at which the metal melts. At this high temperature, the internal energy reaches a value that permits an atomic rearrangement and moves away from the rigid position to assume a highly disordered configuration typical of the liquid state. To return atoms to the highly ordered crystalline configuration, the temperature of the system must be slowly cooled to allow the atoms to reposition into the crystalline order reaching the corresponding minimum energy state.

Simulating this process is very similar to a combinatory optimization task. For this physical system, the goal is to find some arrangement of atomic particles that minimizes the energy (cost) of the system. The main requirement is the ability to simulate how the system reaches thermodynamic equilibrium at each fixed temperature (obtained by a chosen cooling scheme) used to anneal it.

Based on the laws of static thermodynamics for each value of the temperature (T), the system evolves towards a state of minimum energy (E) and maximum probability (likelihood). However, there is a non-zero probability at which the state is at a higher energy. This non-zero probability is described by the Boltzmann distribution:

$$P(E_i) = \frac{e^{-\frac{E_i}{\mathrm{K_B}T}}}{\sum_m e^{-\frac{Em}{\mathrm{K_B}T}}} \tag{19.5}$$

$\mathrm{K_B}$: is the Boltzmann constant.

The physical process can be successfully simulated by the Metropolis algorithm [33]. The idea, as in iterative improvement, is to propose some random perturbation, such as moving a particle i to a new location j, then to evaluate the resulting energy change $\Delta E = E_j - E_i$ (in the optimization problem the cost function is to be minimized). If the energy is reduced, $\Delta E < 0$, the new configuration has lower energy and is accepted as the starting point for the next move. However, if the energy is increased, $\Delta E > 0$, the move may still happen: the new higher energy configuration may be acceptable with the following probability function:

$$e^{-\frac{\Delta E}{K_B T}} \tag{19.6}$$

The probability used by Metropolis represents the ratio between the probabilities that the system is in state j or state i, based on the Boltzmann distribution (in the optimization problem the K_B parameters is normally fixed to one):

$$\frac{P(E_j)}{P(E_i)} = \frac{e^{-\frac{Ej}{K_B T}}}{\sum_m e^{-\frac{Em}{K_B T}}} \cdot \frac{\sum_m e^{-\frac{Em}{K_B T}}}{e^{-\frac{Ei}{K_B T}}} = e^{-\frac{(Ej - Ei)}{K_B T}} = e^{-\frac{\Delta E}{K_B T}} \tag{19.7}$$

To be noted that at 'high enough' temperatures $P(E_j \backslash E_i, T \to \infty) = e^0 = 1$ all energy states are equally probable. Whereas at 'low enough' temperatures the system is certainly in a state of minimum energy, $P(E_j \backslash E_i, T \to 0) = e^{-\infty} = 0$.

From the above it is clear that the success of the optimization procedure strictly depends on the chosen cooling scheme. If the cooling scheme is slow enough (logarithmic), the algorithm is statistically guaranteed to reach a global optimum (with probability 1). However, such a theoretically correct cooling schedule is too slow to be applied in practice, so a geometric schedule is applied instead (i.e. $T_{t+1} = \alpha T_t$ with α chosen between 0 and 1).

SA optimization has two advantages over traditional techniques used in the ICA model (such as gradient-based): it does not require the use of derivatives and, if properly set, it reaches the global maximum. Although it is considerably slower if compared with traditional techniques, in the FSS procedure this is not a relevant drawback, since usually only a very limited number of sources has to be extracted (in the majority of the cases just one).

19.2.3.3 Simulated Annealing Approach Used in FSS

In the FSS algorithm, the data are whitened using the standard principal component analysis (PCA) approach. For each functional constraint, an initial random $\boldsymbol{w}$ unmixing coefficient vector (in Eq. 19.2) is initialized and the contrast function (in Eq. 19.4) is maximised by perturbing w: an optimal w_{opt} is found at the end of the optimisation process, and the corresponding source is recovered from it. A decrease rate for the temperature is implemented, such that $T_{t+1} = \alpha T_t$, with $\alpha = 0.8$. The algorithm terminates when, comparing the solutions at two consecutive temperatures, the norm of the unmixing coefficients w is under a fixed threshold ($\varepsilon = 10^{-4}$).

We adopted a procedure to automatically set the initial temperature T_0 depending on the data in hand. Starting from a random initial temperature T_R, we keep track of the number of accepted (Acc) and rejected (Rej) state transitions. The ratio ρ = Acc/(Acc + Rej) is computed after the system has reached equilibrium, and the following criterion to set up T_0 is used: if $\rho < 0.8$, the system is not warm enough and the optimisation is not reliable, so we set $T_R = 1.5T_R$; if $\rho > 0.9$, the system is considered too warm and the optimisation may take more time than needed, so we set $T_R = 0.9T_R$; if $0.8 \geq \rho \geq 0.9$, then $T_0 = T_R$.

19.3 FSS: Functional Constraints Expressing a Functional Fingerprint of the Source of Interest

By the described method, FSS optimizes a wide variety of functional constraints, each one typical of the source to be extracted (R quantity in Eq. 19.4).

19.3.1 Primary Sensory Hand Area

In the case of the primary sensory areas, we exploited the knowledge that around 20 ms the primary somatosensory cortex (S1) is recruited by a galvanic stimuli delivered to a hand nerve at the wrist [3, 20]. Thus, FSS identifies the S1 region devoted to the districts innervated by the median nerve maximising this 20 ms response (named S1, see Fig. 19.2, top left).

The functional constraint $R_{\mathrm{FS_{S1}}}$ is defined as [43, 44, 48]:

$$R_{\mathrm{FS_{S1}}} = \sum_{t_k-\Delta_1 t_k}^{t_k+\Delta_2 t_k} |\mathrm{EA}(t)| - \sum_{6}^{11} |\mathrm{EA}(t)| \tag{19.8}$$

where EA stands for evoked activity, which is computed by averaging the signal epochs of the source FS_{S1} triggered on the stimulus ($t = 0$); t_k is the time point with the maximum field potential on the maximal original MEG/EEG channels, individually selected about k ms after the stimulus onset; $\Delta_1 t_k$ ($\Delta_2 t_k$) = time point corresponding to a signal amplitude of 50 % of the maximal value before (after) t_k (grey area in Fig. 19.2, top left); the baseline (no response) is computed in the time interval from 6 to 11 ms.

19.3.2 Primary Motor Hand Area

To identify the source in the primary motor area devoted to the control of hand movements (named M1, Fig. 19.2, bottom left), the coupling of cortical and muscular rhythmic oscillations in the β band was exploited (grey area in Fig. 19.2, bottom left). In fact, it has been demonstrated that the component of the synchronized cortical activity, coupled with synchronous rhythmic motor-unit firing (assessed by surface EMG) is within this band. Cortico-muscular coherence relates to the patterns of motor output and sensory input, both in healthy subjects [18] and in patients with motor disorders [9, 10]. The corresponding functional constraint to obtain the M1 source was [43]:

$$R_{\mathrm{FS_{M1}}} = \sum_{\omega_{\max}-\Delta_1\omega_{\max}}^{\omega_{\max}+\Delta_2\omega_{\max}} Coh(\omega) \tag{19.9}$$

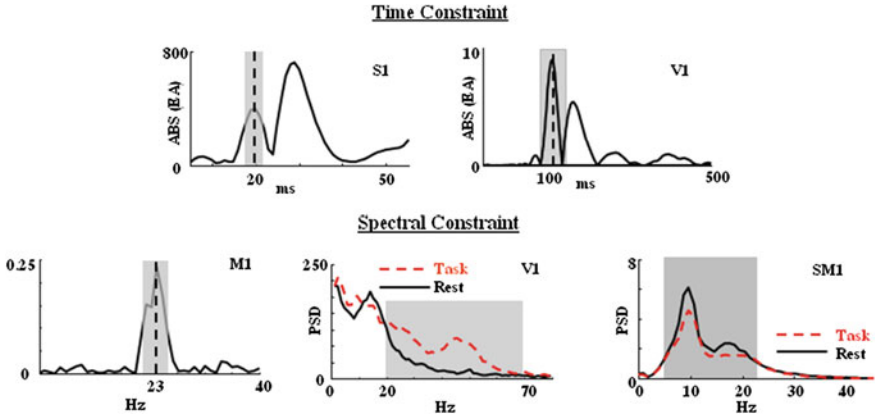


Fig. 19.2 *Time Constraints* (*Top*) Representation of the quantities maximised by the functional constraint to obtain the Functional Source (FS) in different primary brain areas. (*Top Left*) Primary sensory area (S1). The *grey* areas indicate the time interval where the responsiveness to stimulation (Evoked Activity, EA) is maximised (around 20 ms after contralateral median nerve stimuli, known to be generated by S1); (*Top Right*) Primary visual area (V1) is maximised around 100 ms after the visual stimuli (known to be generated by V1). Spectral constraints (*Bottom*). The *grey* area indicates the frequency interval around 23 Hz where the cortico-muscular coherence is maximised, known to be mainly generated in M1 (*bottom left*). (*Bottom centre*) Power Spectral Density (PSD) of the FS in the Task\Rest condition. The *grey* area indicates the frequency interval from 20 to 70 Hz where the spectral difference between Task (visual stimulation) and Rest (fixation point) is maximised (obtaining V1). (*Bottom right*) The functional constraint exploited mu-rhythm (8–25 Hz) reactivity that occurs in contralateral sensorimotor areas during uni-manual motor tasks (SM1)

where *Coh* (coherence function) is a function of frequency ω, obtained for each ω as the amplitude of the cross-spectrum between the M1 source signal and the rectified EMG, normalized by the root mean square of the power spectral densities of these two signals; $\Delta_1\omega_{\max}$ ($\Delta_2\omega_{\max}$) is the frequency point corresponding to a coherence amplitude of 50 % of the maximal value between [13.5–33] Hz (called $\omega_{\max}$) before (after) $\omega_{\max}$.

19.3.3 FS Evaluation for Primary Somatosensory and Motor Areas

To evaluate the *goodness* of functionally separated sources, three criteria were used: functional source behaviour, functional source position and discrepancy.

19.3.3.1 Functional Source Behaviour

The identified sources displayed reactivity properties during movement and during galvanic median nerve stimulation (see Fig. 19.3). In particular, FS_{S1} showed the

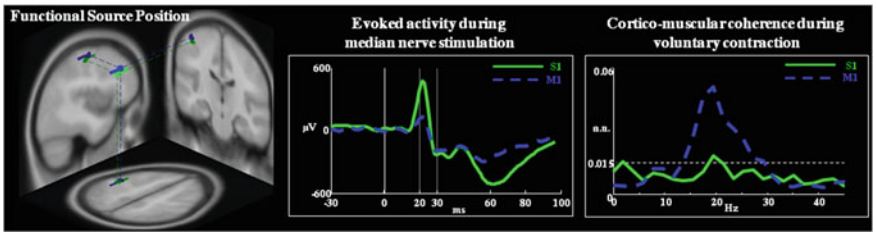


Fig. 19.3 (*Left*) Position, direction and orientation of the ECD corresponding to S1 and M1, in the axial, coronal and sagittal views of the MNI brain template. (*Centre*) Time course of the stimulus-averaged FS_{S1} (*solid green line*, S1) and FS_{M1} (*dashed blue line*, M1) in the −30 to 100 ms time period following the galvanic stimulation of the right hand. (*Right*) Cortico-muscular coherence between each of the two sources and the rectified EMG in the frequency interval [0, 45] Hz during isometric contraction of the right opponent thumb muscle. The *dashed horizontal line* indicates the confidence limit

maximum of responsiveness to median nerve stimulation at around 20 ms (Fig. 19.3, centre), while the source FS_{M1} displayed higher response at around 30 ms than at 20 ms (Fig. 19.3, centre). FS_{M1} showed definitely higher cortico-muscular coherence than FS_{S1} during handgrip (Fig. 19.3, right).

19.3.3.2 Functional Source Position

To investigate the spatial position of the FS_{S1} and FS_{M1} extracted sources, they were separately retro-projected to obtain their field distributions, by:

$$MEG\_\mathrm{rec}_{\mathrm{FS}_y} = \mathbf{a}_{\mathrm{FS}_y}\mathrm{FS}_y \tag{19.10}$$

with $FSy = FS_{S1}$ or FS_{M1} and $\mathbf{a}_{FS_y}$ is the column vector of the matrix $\mathbf{A}$ Eq. 19.1.

Source localization was performed using an equivalent current dipole (ECD) model, with a forward model consisting of four concentric conductive spheres (routine DIPFIT2 [36] of EEGLAB, available at http://www.sccn.ucsd.edu/eeglab [15]). EEGLAB expresses ECD position in Talairach coordinates and projects them onto the Montreal Neurological Institute (MNI) brain template (Fig. 19.3, left).

19.3.3.3 Discrepancy

To check for the level of residual response to the nerve stimulation after source extraction we have introduced the *discrepancy* estimate, defined as the evoked activity of the original data minus the data retro-projected as expressed by Eq. 19.10 (see Fig. 19.4).

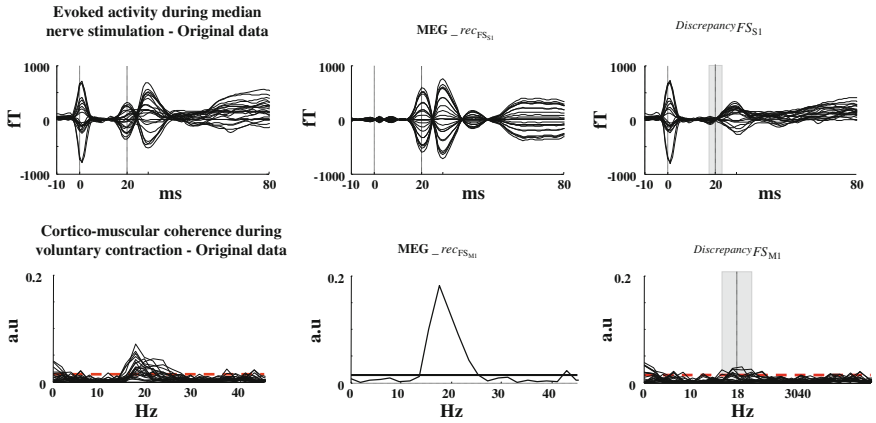


Fig. 19.4 In one representative subject. (*Top*) Superimposition of the channels averaged on median nerve stimuli, in the time window $[-10, 80]$ ms], $t = 0$ the stimulus arrival being at wrist (*vertical solid line*). The time points corresponding to M20 component is indicated (*vertical dashed lines*). (*Top left*) Original data. (*Top centre*) Retro-projected data with only the FS_{S1} source. (*Top right*) Original data minus reconstructed data with only FS_{S1} source. The *grey* area indicates the time interval where the functional constraints are calculated. Note that FS_{S1} well explain the generated field at their respective latencies. (*Bottom*) Superimposition of all channels' coherences with the rectified EMG in the frequency window [0, 45] Hz. The confidence limit is indicated (0.015, *horizontal dashed line*). (*Bottom left*) Original data. (*Bottom centre*) Retro-projected channels with only FS_{M1} source. All channels display the same coherence with the EMG signal; this is because all the channels obtained by retro-projecting only one FS display the same time evolution, unless a multiplicative factor and the coherence are independent from the signals amplitude. (*Bottom right*) Original MEG data minus reconstructed data with only FS_{M1} source. The *grey* area indicates the frequency interval where the functional constraint is calculated

19.3.4 Primary Visual Areas

This section examines the ability of FSS to extract sources specifically related to the Visual Evoked Potential (VEP/P100) and Gamma Band Activity (GBA) elicited by a full reversing checkerboard. The relationship between the low frequency VEP and the high frequency (γ: 30–90 Hz) GBA, both of which can be generated by simple visual stimuli such as checkerboards and gratings, remains unclear. In particular, two different functional constraints were examined in this section to extract the visual activity using FSS: temporal (see Fig. 19.2, top right) and spectral constraints (see Fig. 19.2, bottom centre). The FSS temporal constraint maximised the activity around the P100 of the VEP, and the spectral constraint maximised the difference in the GBA between the rest and the task periods (see Fig. 19.2, bottom centre).

Temporal functional constraint The functional constraint R_{FS} was defined as:

$$R_{FS_{P100}} = \sum_{t_k-\Delta_1 t_k}^{t_k-\Delta_2 t_k} |\mathrm{EA}(t)| - \sum_{t=-100}^{0} |\mathrm{EA}(t)| \tag{19.11}$$

with the evoked activity, EA, computed by averaging signal epochs of the source FS_{P100}, triggered on the visual stimulation ($t = 0$); t_k is the time point with the maximum electric potential around 100 ms after the stimulus onset on the maximal original EEG channel; $\Delta_1 t_k (\Delta_2 t_k)$ is the time point corresponding to a signal amplitude of 50 % of the maximal value before (after) t_k. The baseline was computed in the time interval from −100 to 0 ms. The precise value of each latency t_k was chosen for each subject, corresponding to the maximum electric potential in the time interval of interest (80–120 ms).

Spectral functional constraint To investigate the GBA, the following ad-hoc functional constraint R_{FS} was used:

$$R_{FS_\gamma} = \frac{\sum_\gamma \mathrm{PSD}^{\mathrm{Task}}_{FS(\gamma)} - \sum_\gamma \mathrm{PSD}^{\mathrm{Rest}}_{FS(\gamma)}}{\sum_\gamma \mathrm{PSD}^{\mathrm{Rest}}_{FS(\gamma)}} \tag{19.12}$$

This constraint computes the difference in the PSD between *Task* (from 0 to 5 s of each trial, $t = 0$ corresponding to the stimulus onset) and *Rest* (from −5 to 0 s of each trial) periods in the γ frequency band (30–90 Hz). This difference is then normalised with respect to the GBA in the *Rest* period [4, 47].

19.3.5 FS Evaluation for Primary Visual Areas

The purpose of applying different functional constraints was threefold. Firstly, to determine which technique was able to provide the best characterization of the GBA. Secondly, to provide a degree of validation to the comparison of the GBA and the VEP. Each of the techniques extracts a different part of the raw signal that is dependent on the assumptions underlying the decomposition. In particular, the $\mathrm{EEG\_rec}_{FS\gamma}$ sources were explicitly intended to identify activity in the gamma band, whilst the $\mathrm{EEG\_Rec}_{\mathrm{FSP100}}$ source employed a temporal constraint designed to maximise the VEP. If similar conclusions regarding the relationship between the VEP and GBA are drawn from examination of these different sources then some confidence can be gained that they represent a realistic interpretation of the underlying data. Finally, the use of FSS with multiple constraints allows the relationship between the constraints to be studied directly. In this case, the temporal constraint centred on the P100 of the VEP selectively identified the most probable generator of that peak. The question could then be asked as to whether there is any evidence that this source also generates GBA. Conversely, a completely orthogonal spectral constraint centred on the gamma band was used to select the generator of the GBA, and the low frequency behaviour of that source examined. Only if there is a genuine relationship between the VEP and the GBA will the activity of the two sources be similar. These results provide clear evidence that the neuronal pools generating the VEP and GBA have close spatial relationships.

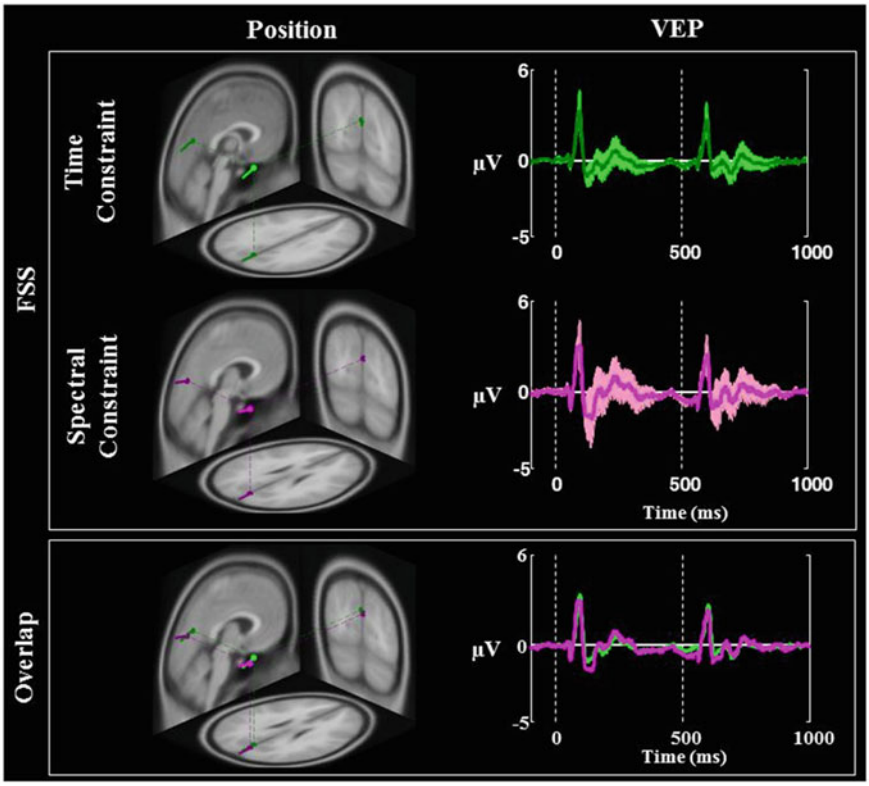


Fig. 19.5 For the grand average, dipole localization and VEP are shown for each method (EEG_rec$_{FSP100}$—first row and EEG_rec$_{FS\gamma}$—second row). The *envelope* indicates plus and minus one standard deviation around the VEP mean. For the dipole fit, position and orientation of the ECD are shown superimposed on the MNI brain template in axial, coronal and sagittal views. The last row shows the overlap across the methods for the VEP and dipole source localization

To evaluate the goodness of the source extractions in the primary visual area, source localization, evoked activity and time frequency analysis have been used. In order to facilitate comparison, the analysed data were taken from a single occipital electrode (POO1, the electrode nomenclature is according to the 10–5 electrode system [35]) selected for the maximum voltage field.

19.3.5.1 Functional Source Position and Evoked Behaviour

Localization of FSS with both temporal and spectral constraints was very consistent. The ECD demonstrated very precise spatial co-localization (Fig. 19.5). This level of overlap supports the idea that the VEP and GBA are generated by spatially concordant neuronal populations. In particular, the comparison of the waveforms and localization

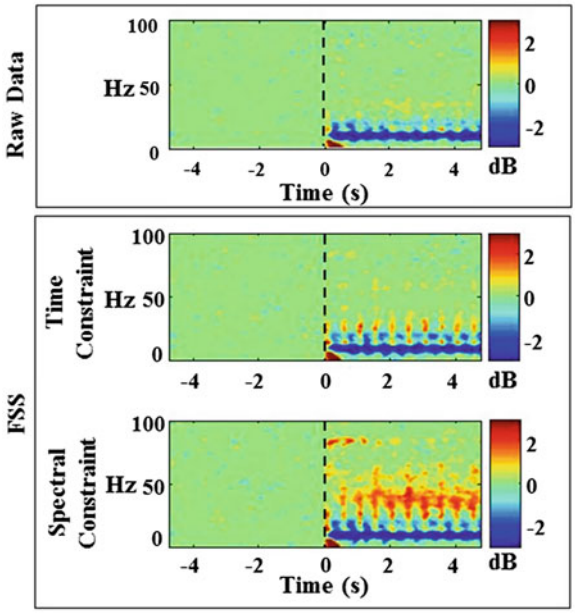


Fig. 19.6 Grand Average of ERSP for Raw Data (*first row*) and for each method EEG_rec$_{FSP100}$ (second row) and EEG_rec$_{FS\gamma}$ (*third row*). The points after 0 (from 0 to 5 s) correspond to the time when a reversal checkerboard stimulus was presented on the screen (Stimulus—Task), and the points before 0 (from −5 to 0 s) correspond to the time in which no reversal checkerboard stimulus was present (No-Stimulus—Rest)

between the two extracted sources revealed a high degree of symmetry: maximising the signal extraction for the low frequency VEP leads to a source containing strong GBA. At the same time, maximising the GBA resulted in a source with a clear VEP.

19.3.5.2 Time-Frequency Dynamics

Time-frequency analysis was performed using a short-time Fourier analysis using Fast Fourier Transforms (FFTs) with a moving windows size of 256 samples (500 ms) wide as implemented in EEGLAB [15]. Event-related spectral perturbation (ERSP), a 2-D (frequency-by-latency) image of mean change in spectral power (in dB) from baseline [27, Concept and Terms] was computed for the POO1 electrode, and the results were compared among the methods. The time-frequency plot was thresholded at a bootstrap significance level of $p = 0.01$.

For the ERSP, the gamma activity after the stimulus presentation was more evident in the FSS methods than in the raw data (Fig. 19.6). As expected, since it is optimised to do so, the GBA was most robust in the EEG_rec$_{FS\gamma}$ data. To be noted that the gamma activity after the stimulus presentation was more evident in the semi-blind

methods than in the raw data (Fig. 19.6). As expected, since it is optimised to do so, the GBA was most robust in the EEG_rec$_{FS\gamma}$.

19.3.6 Motor Intention Network

Human beings are able to interact with the environment through several modalities, which involves neural signals, muscle activities and cutaneous/proprioceptive sensory organs. At the amputation site all afferent and efferent nerves, originally devoted to the lost segment, are interrupted. The consequent deprivation of peripheral inputs and actuators results in retrograde changes that affect not only the peripheral nervous system but also central structures including the motor and somatosensory cortices. In fact, following amputation the deafferented cortical areas become responsive to inputs from the parts of the body that are represented adjacent in the Penfield homunculus [22, 32, 41]. It has recently been shown that after amputation the ability to execute a movement with the affected limb is maintained by the amputees' brain [49] despite the absence of a peripheral effector. Similarly to healthy controls, motor imagination and execution differentially activate cerebral areas in amputees with a significantly greater activation in primary motor and sensory cortices during execution, while imagination is associated with greater parietal and occipital lobe activity [50]. Even though the nervous system is altered by the amputation, residual neural patterns appropriate for the lost limb may still be activated when suitably stimulated [31, 51]. Those pathways are therefore a possible target of central or peripheral neural implants to restore a direct and relatively 'natural' channel for data exchanging.

In the present section, we proposed if and how a direct connection between the brain and a prosthetic hand via a neural implant modifies bi-hemispheric EEG activity in primary sensorimotor cortical areas controlling movements of the lost limb. In this particular case, identifying specific neuronal pools on the bases of their functional properties instead of and independently from their spatial positions became crucial. We also had to take into account possible cerebral plastic changes altering the location of brain functions with respect to standard anatomical landmarks typical of healthy people.

To achieve this goal, the FSS functional constraint exploited the mu-rhythm (8–25 Hz) reactivity [42] that occurs in contralateral sensorimotor areas during unimanual motor tasks, by requiring maximal variation of spectral power in α [8–13 Hz] and β [14–25 Hz] bands between the period of prosthesis control and rest (Fig. 19.2, bottom right). The ad-hoc functional constraint R_{FS} was built as follows:

$$R_{FS_{mu}} = \frac{\sum_{\alpha+\beta} \mathrm{PSD}^{\mathrm{Task}}_{FS(\alpha+\beta)} - \sum_{\alpha+\beta} \mathrm{PSD}^{\mathrm{Rest}}_{FS(\alpha+\beta)}}{\sum_{\alpha+\beta} \mathrm{PSD}^{\mathrm{Rest}}_{FS(\alpha+\beta)}} \tag{19.13}$$

with the Power Spectrum Density (PSD) during *Task* estimated in the 5 s windows of each movement trial and at *Rest* in the 5 s windows preceding each trial. The $\alpha + \beta$ frequency band included 8 to 25 Hz [16].

19.3.7 FS Evaluation for the Motor Intention Network

In this case anatomical position and functional behaviour were used to evaluate the *goodness* of the functional network extracted.

19.3.7.1 Functional Source Position

In order to investigate cortical recruitment occurring during the intention of moving of the (intact) right hand and (phantom–prosthetic) left hand before (PRE) and after (POST) training with the implanted neural electrodes, the FSs identified by FSS were submitted to a source localization algorithm (sLORETA [38]) implemented in CURRY 6 (Neuroscan, Hamburg, Germany, http://www.neuroscan.com/). sLORETA was performed for each source using a regular grid with a spacing of 3 mm throughout the brain region and a four-shell spherical head model. The results were projected onto the brain template of the Montreal Neurological Institute (MNI) within CURRY.

Motor intention of the healthy right hand: Right hand movement intention recruited well-segregated contra-lateral left sensorimotor areas (Fig. 19.7—first row). In particular, source localization shows a clear activation of the primary sensory and motor areas (left Postcentral Gyrus—BA2, BA3 and Precentral Gyrus—BA4).

Motor intention of the left cybernetic hand prosthesis: Delivery of motor commands to the phantom of the left amputated limb before the Longitudinal Intra-Fascicular Electrodes (LIFEs) implant [59] recruited areas in the ipsilateral primary sensory and motor areas (left Postcentral Gyrus—BA2, BA3 and Precentral Gyrus—BA4, Fig. 19.7—second row top) and bilateral premotor and supplementary motor cortex (left and right BA6, Fig. 19.7—second row bottom). At this stage no contralateral primary motor cortex activity was found. The cerebral recruitment during the intent to move the phantom of the left amputated limb changed markedly after the four weeks of prosthesis motor control training with implanted LIFEs. Cortical recruitment became almost symmetrical with respect to right hand movements, with selective involvement of the contralateral sensorimotor cortex (right Postcentral Gyrus— BA2, BA3 and Precentral Gyrus—BA4, Fig. 19.7—third row).

19.3.7.2 Functional Source Behaviour

The cortical activation of the areas devoted to intentional control were evaluated by time-frequency spectral modulations during the motor task (intention of movement)

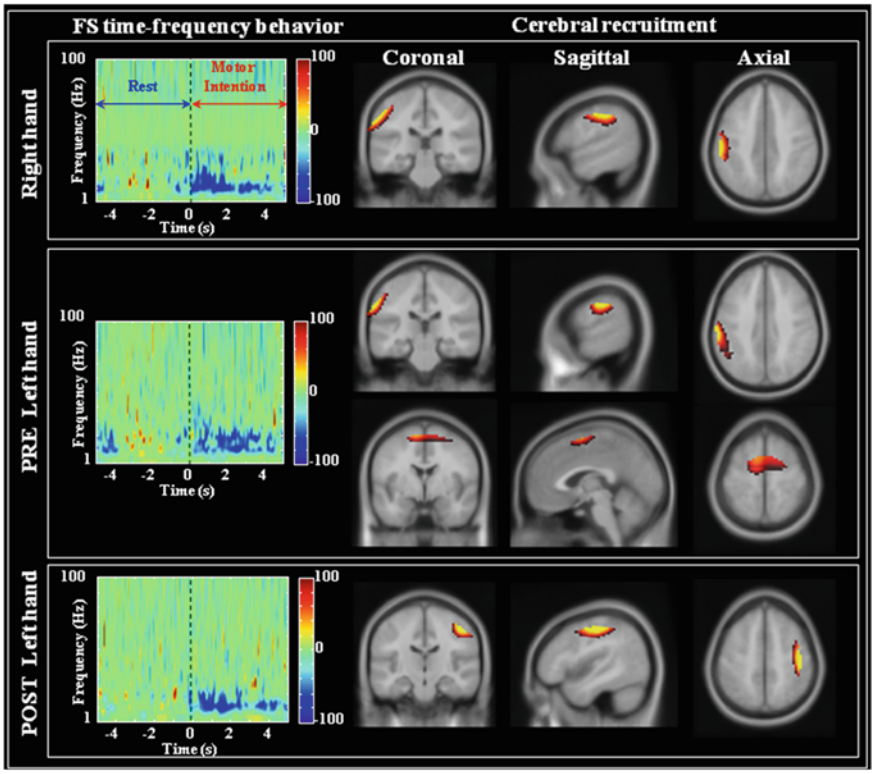


Fig. 19.7 *Left* Time frequency representation of each FS activity. The colour code represents significant changes in average power (across epochs) as a function of time and frequency. *Right* solutions of the sLORETA model for the FSs related to right hand control (*Top*), with the missing limb phantom PRE (*Middle*) and POST (*Bottom*) shown superimposed on the MNI brain template

and compared to the resting state baseline (resting with open eyes). Twenty single trials from the source were chosen in the time interval of −5 to 5 s, where time 0 is the intention movement go signal. They were convolved by a Morlet wavelet (setting the constant equal to 7, which defines the compromise between time and frequency resolution), and the squared absolute values of the convolution over trials were averaged. For each frequency, the time course of power modulation was represented as the percentage of the mean of the baseline period (−5 to 0 s). Moreover, the cortical recruitments and the time frequency behaviours were compared before and after neural implant. Significant power changes from *Rest* to *Task* periods were assessed using a resampling bootstrap technique thresholded at $P = 0.05$, while non-significant changes were set to 0. The above procedure was applied to the PRE (pre-implantation period), POST (post-implantation period) and on the right hand intention movement.

Motor intention of the healthy right hand: The analysis of time frequency behaviour exhibited a clear response that was stronger in the first 2 s and covered the whole mu-band including both α and β frequencies, becoming concentrated in the α band for the 2–5 s period.

Motor intention of the left cybernetic hand prosthesis: The analysis of time frequency behaviour evidenced activity in α and β bands lasting the entire task duration with no time-specificity. As in the case with anatomical position, the functional behaviour drastically changed after the 4-week training period. The time frequency behaviour regained evolving properties similar to those for right hand control: stronger in the first 2 s involving the entire α and β frequencies, while weaker and more concentrated in the α band in the 2–5 s period (Fig. 19.7—third row right).

19.4 FSS and Simultaneous EEG-fMRI Recordings

The simultaneous measurement of EEG and functional magnetic resonance imaging (fMRI) is an attractive, non-invasive technique for the investigation of human brain function, with the potential to offer a higher spatiotemporal resolution than either method alone. It is increasingly widely used as a tool in cognitive and sensory neuroscience (e.g. [7, 12, 17, 37, 46]) and can also shed light on the properties of the underlying neurovascular coupling which, particularly at the macroscopic level where scalp EEG and whole brain fMRI are measured, are not fully understood [61]. However, if the potential strengths of EEG-fMRI are to be fully realized, and new methods for data integration developed and exploited, it is vital that good quality EEG and fMRI data are recovered from recorded signals.

In particular, EEG data acquired in the MRI scanner are strongly contaminated by artefacts of biological and non-biological origin that may prevent the correct determination of the characteristics of the brain signals that are of primary interest.

There are several artefacts that contaminate the measurement of neurophysiological EEG and that need to be removed from the recordings before further analysis. Specific to the MRI environment are gradient artefacts (GA) and ballistocardiogram artefacts (BCG), while ocular artefacts (OA) and electrode artefacts (EA) are present in the EEG acquired inside and outside of the scanner. The most widely used techniques to reduce the effects of GA and BCG are variations of template averaging approaches [1, 2], with ICA often used as an alternative or secondary step [13, 29]. However these massive preprocessing steps are often not sufficient, particularly when the focus is on using ST variability to integrate the two data sets. Recently, FSS has been demonstrated to reliably improve single-trial EEG data recorded during simultaneous EEG-fMRI [46].

In this particular case, a reversing checkerboard stimulus was used to generate VEPs in healthy control subjects and an ad-hoc functional constraint was maximised around the principal peak (P100) of the VEP (see Eq. 19.11 and Fig. 19.2, top right).

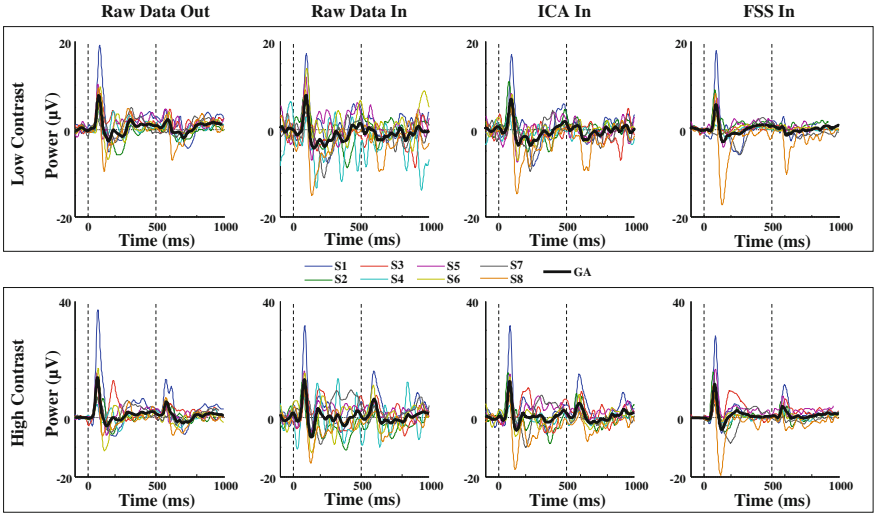


Fig. 19.8 Time course of the stimulus-averaged VEP in the [−100 1000] ms time period following low (*top row*) and high (*bottom row*) contrast stimulation. Average VEPs are shown for each subject (*coloured lines*) along with the grand average across subjects (*thick black line*). The first *vertical dashed line* indicates stimulus onset, while the second indicates the reversal of the checkerboard

19.4.1 FS Evaluation During Simultaneous EEG-fMRI Recordings

To evaluate the quality of the data following FSS approach, two criteria were used: the functional behaviour and correlation between electrophysiological and hemodynamic response functions (HRFs). To further evaluate the performance of FSS, these metrics were also applied to the raw data (i.e., only GA and BCG were removed using standard techniques) and to that recorded before the MRI scanning session. In order to facilitate comparison, the analysed data were taken from a single electrode (right occipital electrode—channel O2) that showed the highest amplitude P100 in the average VEP.

19.4.1.1 Functional Source Behaviour

The VEPs for high contrast (HC—100 % black and white contrast) and low contrast (LC—25 % grey and white contrast) stimuli were calculated for all subjects, and the grand average across subjects was calculated. Comparisons were made among the raw data before and after scanning, and the data in the scanner was pre-processed by ICA and FSS.

For each individual subject Fig. 19.8 shows the average VEPs (coloured lines) and the grand average over all subjects (thick black). The data for the individual subjects

Table 19.1 Signal-to-noise ratio comparisons

	Low contrast			High contrast		
	Raw Out	Raw In	FSS In	Raw Out	Raw In	FSS In
S1	31.7	23.3	47.5	51.6	22.0	62.1
S2	16.4	19.5	33.0	39.0	7.9	58.2
S3	21.4	4.2	18.4	28.9	30.1	38.1
S4	39.6	1.0	20.7	31.1	11.2	46.8
S5	12.2	1.7	16.1	38.5	17.0	50.1
S6	40.4	14.2	17.9	50.6	7.4	59.6
S7	17.7	8.0	30.9	38.2	34.6	44.1
S8	27.9	23.8	53.4	43.7	19.9	50.6
Mean	25.9	12.0	29.7	40.2	18.8	51.2
SD	10.7	9.5	14.3	8.2	10.0	8.3

The signal-to-noise ratio of the average VEP was calculated for each subject and for FSS method and raw data. Mean and standard deviation (SD) over subjects are also given

show a considerable amount of variability, which is most evident in the raw data recorded inside the scanner. Comparing the raw data acquired inside and outside of the MRI scanner shows that there is much more variability when recording within the MRI environment. Given that these are the same subjects undergoing the same stimulation paradigm, it is evident that the primary cause of this increased variability is the reduction in EEG data quality caused by the various MRI artefacts. This inter-subject variability is improved with ICA, but the most obvious improvement is between ICA and FSS, with FSS showing similar, or less, variability, than the data recorded outside of the scanner. It is also worth noting that the grand average VEPs are very similar for the different methods, indicating that the grand average VEP is not a good measure of the underlying data quality.

The signal-to-noise ratio (SNR) of the average VEPs was calculated, with the noise level calculated between −100 and 0 ms and the signals between 1 and 1000 ms (Table 19.1). The SNR in Table 19.1 is higher for FSS than for raw data and comparable with the data acquired outside the scanner.

19.4.1.2 Correlation Between Electrophysiological and Hemodynamic Responses

In order to address the issue of whether improving EEG data quality affects the correlation of the EEG and fMRI data (Fig. 19.9), haemodynamic response functions (HRF) were extracted from spherical volumes of interest (VOI, radius 5 mm) centred on the maximally responsive voxel in the fMRI statistical map.

The averaged HRFs were compared with the data extracted using the different EEG preprocessing techniques. Correlation analysis was performed between the normalized area of the VEP over the same time interval (calculated as the sum and normalized with respect to the window length) and the normalized area of the HRF

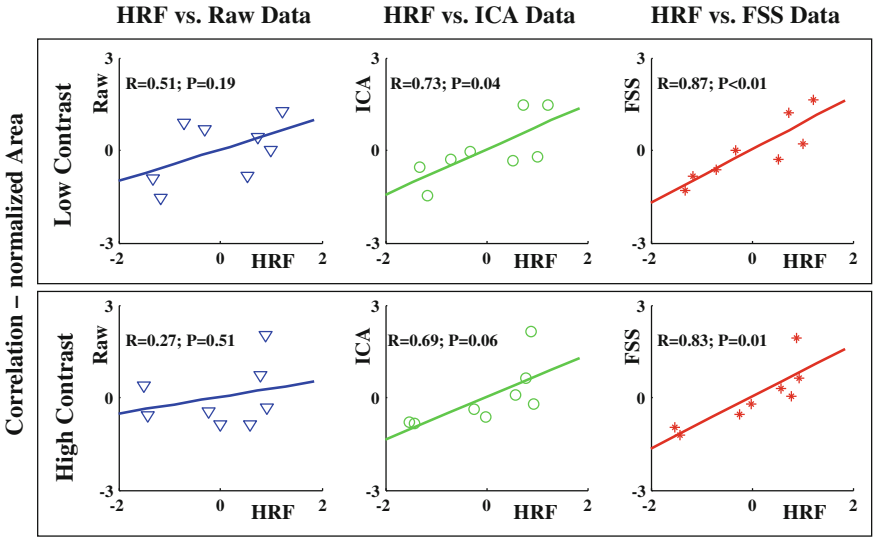


Fig. 19.9 Correlation between VEPs and HRFs across Subjects. Pearson correlation (two tailed) between EEG and fMRI data was calculated across subjects for the areas of the evoked responses. The data was centred and scaled to facilitate the comparison across the methods

(also centred on the maximum peak and normalized with respect to the window length). Clearly, the correlation analysis demonstrates a much improved correlation between the area of the EEG and fMRI evoked responses when using the EEG data processed with FSS.

19.5 FSS and Single Trial Behaviour

The advantages of source separation techniques are most evident when dealing with trial-by-trial variations of electrophysiological signals [27, 28, 46, 47], or with other low amplitude and noise aspects of signals reaching the extra-cranial sensors such as oscillations in the γ band (30–90 Hz [4, 19, 34]). In this section, we show how FSS enables the investigation of the single trial (ST) behaviour of the source of interest in two exemplificative cases, analysing GBA in V1 and simultaneous EEG-fMRI.

19.5.1 Single Trial γ Band Activity (GBA) Investigated by FSS

GBA is hardly detectable from outside the scalp. We applied an FSS method to dense array EEG data recorded during full checkerboard stimulation, comparatively

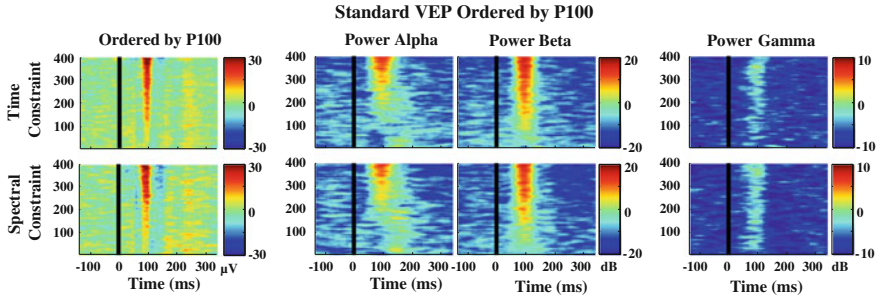


Fig. 19.10 ERPImage of the single trial P100 ordered by the area under the P100 curve (i.e. ∼80 ms and ∼120 ms) x-axis indicate the number of the trials and y-axis indicate the time in milliseconds

investigating neuronal pools generating the main peak around 100 ms of the VEP (temporal constraint VEP_{100}, Fig. 19.2, top right) and the GBA (spectral constraint, Fig. 19.2, bottom middle). Our objective was to determine if this neural activity is generated by the same neuronal pools.

A single trial (ST see ERPImage [27], Concept and Terms) analysis compared VEP_{100} and GBA in response to checkerboard stimulation in time and in different frequency bands. The single trails of VEP_{100} and GBA were ordered by the area under the P100 peak (between ∼80 and ∼120 ms). Furthermore, the event-related spectral perturbation (ERSP) for each trial was calculated and averaged within frequency ranges of 8–13 Hz (α band), 14–30 Hz (β band) and 31–90 Hz (γ band) with baseline correction using the interval −100 to 0 ms. The trial ordering based on the P100 area was also then applied to the ERSP data in order to investigate, for example, whether trials with a large P100 also had a large ERSP in each of the specific frequency bands: Figure 19.10 shows that this is indeed the case. The consistency of the behaviours of VEP_{100} and GBA, in terms of single trial responsiveness and time-frequency power changes, documented that the same neuronal pools generate the evoked activity and gamma band modulations.

19.5.2 FSS Improves the Quality of Single Trial Analysis in Simultaneous EEG-fMRI Recordings

While it is at least conceptually clear that EEG-fMRI can improve spatiotemporal resolution compared to each method alone, it is less obvious how the data should be combined in order to achieve this goal. Often, data integration relies upon the use of some single trial features of the EEG data, either properties of evoked potentials or spectral power variations [7, 12, 13, 17, 25], which are then used to form regressors for a standard general linear model analysis of the fMRI data. The underlying neurophysiology and biophysics relating the macroscopic measures of EEG and fMRI

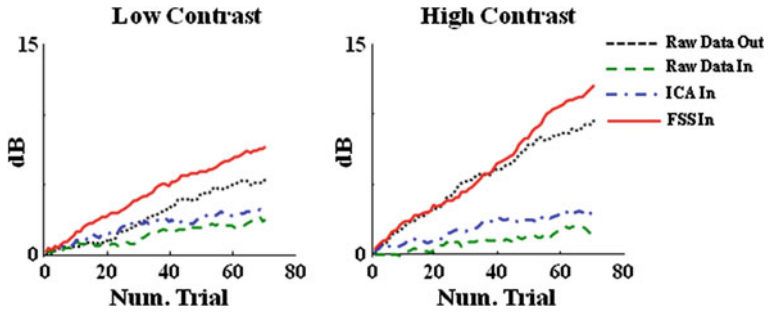


Fig. 19.11 The normalized cumulative signal-to-noise ratio (ncSNR) is shown for both low (*left*) and high (*right*) contrast, averaged over all subjects. The ncSNR is shown for raw data recorded outside of the scanner (*dotted black line*), raw data recorded inside of the scanner (*dashed green line*), data recorded inside the scanner after preprocessing by ICA (*dash-dot green line*) and after FSS (*solid red line*)

are not currently sufficiently developed to allow a principled identification of the features which best link the two. This type of approach is further complicated by the reduction in data quality caused by simultaneous recording especially in the single trial case. The development of additional methods to improve the quality of EEG data acquired in the MRI scanner is therefore an ongoing area of research. Towards this research, in this section we propose the FSS method as a possible candidate to improve the single trial quality of the EEG data recorded in the MRI environment.

To evaluate the ST performance among the methods, normalized cumulative signal-to-noise ratio (ncSNR), localization and single trial variability were calculated.

19.5.2.1 Normalized Cumulative Signal-to-Noise Ratio

The ncSNR for each trial was calculated, as in the case of the average VEPs (Sect. 19.4.1.2); Fig. 19.11 shows the ncSNR. This measure summarizes the quality of the data at the level of the individual trials and is of primary importance for the application to ST EEG-fMRI. Examination of Fig. 19.11 leads to a similar conclusion to that based on the average VEP (Table 19.1); i.e., the FSS data are of similar quality to that recorded outside of the scanner and considerably better than the ICA data, while the ICA data are an improvement on the raw data.

19.5.2.2 Localization and Single Trial Variability

In this section we show the comparison of single trial (ERPImage [27], Concept and Terms) plots and source localization of average VEPs (low Contrast, Fig. 19.12 and High Contrast, Fig. 19.13). The source localization of the average VEPs appears to be relatively robust, with similar localization for different preprocessing methods.

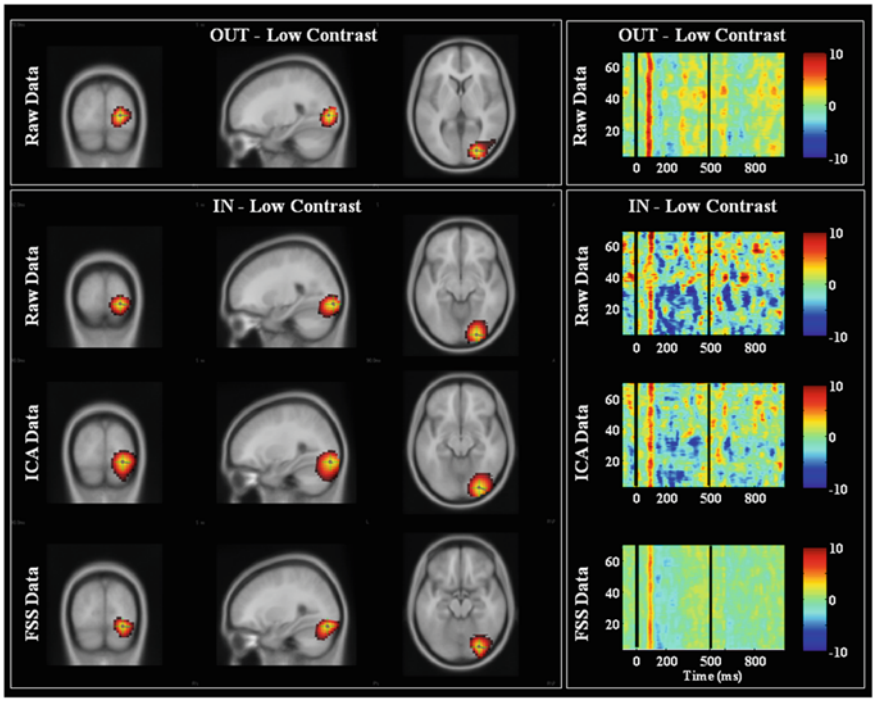


Fig. 19.12 (Low Contrast). sLORETA localization and ERPImage plots are shown for the different data sets and preprocessing strategies. sLORETA results are shown superimposed on the MNI brain template

However, the level of single trial variability as assessed by the ERPImage plots is clearly very dependent on the data preprocessing (Figs. 19.12 and 19.13, last columns). In both cases, the raw data (inside the scanner) are clearly very noisy, with little obvious P100 in the ST data. ICA does not improve this very much but, consistent with the other measurements that have been shown, the difference between ICA and FSS is obvious. This is even clearer for the low contrast data (Fig. 19.12). The FSS ST plot, however, clearly shows a consistent P100 across trials. The low contrast data also demonstrates that improving the ST data quality can affect the localization of the average VEP, since the localization of the FSS data is must more realistic than that of the raw or ICA data.

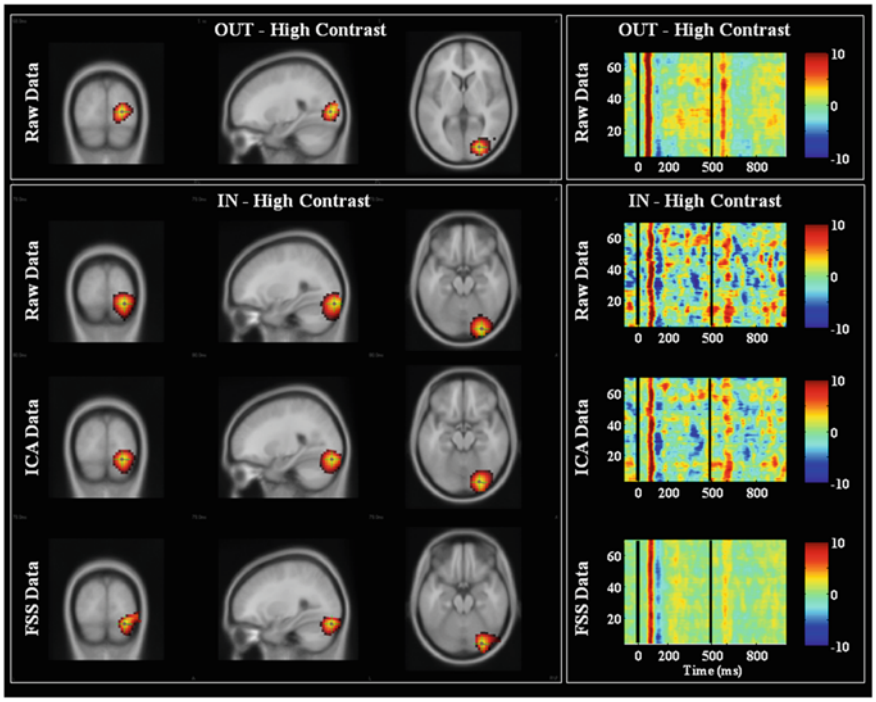


Fig. 19.13 (High Contrast). As in Fig. 19.12

19.6 Discussion and Conclusion

19.6.1 FSS Strengths

The main difference between FSS and other source extraction methods, ranging from inverse problem solving algorithms to spatial filters like beam-forming [6], is that FSS requires no information about the physical relationship between cerebral source generators and their field distributions. This means that FSs are not identified exploiting their positions and intensities with respect to the recording sensors/channels. Once separated, FSs provide the source activities in time and the spatial distribution of the electric field they generate, from which appropriate modelling is used to solve the inverse problem to identify their spatial positions. The solution of the inverse problem theoretically provides in one go both the source position and its time evolution. Unfortunately, on one side it is ill-posed and requires adjunctive information to be selectively added one at a time. On the other side, the solution is based on the relationship between the current distribution and that of the generated field. The spatial distribution of magnetic or electric filed generated by a neuronal current depends on the physical properties of the extra-cerebral tissues, namely shape and thickness

of the head and of all structures including scalp, skull, meninges, cerebral fluid and specific brain regions. All of them can be nowadays assessed anatomically by high-resolution MRIs, but this information can be integrated with MEG/EEG investigation via integration procedures which are theoretically simple but affected by relevant error in technological practical settings. Furthermore, all these tissues have specific conductance and impedances which are not known and change individually both in physiological and pathological conditions. Finally, the current distribution itself can be modelled only posing schematization of real neuronal/glia features generating electric inhomogeneity changing in time. As a consequence, the inverse problem solution is based on less accurate information that is provided by electrophysiological techniques, while FSS algorithms solve the source identification problem using the most accurate information, such as the statistical temporal frequency properties of the signal. In many cases, scientific interest is in the morphological and temporal characteristics of the signal and its modulation in the different experimental conditions. Whenever the investigation requires to focus on spatial location, FSS allows the use of localization algorithms having isolated the field distribution generated solely by the sources of interest.

Even if a source is extracted by exploiting a functional constraint related to a specific time portion of the experiment, the estimated signal could be studied along the entire length of the session.

19.6.1.1 FSS Limitations

As the method is developed to exploit a well-known a priori 'functional' property to identify a pre-defined source of interest, it cannot be used the extract unexpected or unknown brain activities. A second limit is that, the quantity maximised by FSS (i.e., the functional constraint) needs the corresponding experimental paradigm. In other words, an 'ad-hoc' task is required in the recording session to 'activate' the property exploited in the cost function: in examples, FSS needs to have stimulated the median nerve to identify primary sensory hand areas, it needs to have executed an isometric contraction to identify primary motor areas, it needs have presented a reverse pattern or another visual stimuli to identify primary visual areas. This has the inherent implication of lengthening the experimental session. Moreover, FSS is not effective in the absence of a 'distinct' activation property of the area of interest, as it is often the case for associative areas involved in complex cognitive functions.

19.6.2 Conclusive Considerations

FSS methods have access to interesting properties of brain organization and allow the separation of distant regions (about 5–15 mm, which is the typical resolution of non-invasive EEG/MEG techniques) through ecological experimental paradigms. Once identified, the FSS can be localised and studied in all experimental situations

of interest as for all sources extracted by BSS. In fact, FSS introduces a new measure of within-area intra-cortical connectivity in S1, displaying the properties of a new dexterity code to complement the typical 'magnification principle' of cortical organization [55, 57]. In this study, we provided a measurable index of the efficiency of sensorimotor feedback while controlling a simple hand movement [58], which can be potentiated through attention even during a passive movement [40] and is developed through emphatic sharing [8]. It is also sensitive to the dynamic alterations of brain functional organization in dystonic people [30]. We believe that the proposed tool is particularly efficient in investigating the inter-hemispheric balances between functionally homologous areas [11, 39], even in pathological conditions where cortical plasticity phenomena can occur [53, 56].

Altogether, these applications emphasise that FSS derives its power from the identification standing on the source behaviour. In fact, FSS provides a reliable tool to: 1. Discriminate between very closely located cerebral regions, since neuronal structures within the same cortical patch can display definitely different behaviours; 2. Assess hemispheric homologous areas, in which balances are crucial features of brain network functionality [14], in terms of 'functional homology' instead of 'spatial homology', two concepts which typically coincide in physiological conditions but can be un-coupled when disease/damage related plasticity occurs. Furthermore, FSS can assess plastic changes linked with motor improvements through an experimental procedure fully independent of patient compliance [16, 39].

References

1. Allen, P.J., Polizzi, G., Krakow, K., Fish, D.R., Lemieux, L.: Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. Neuroimage **8**, 229–239 (1998)
2. Allen, P.J., Josephs, O., Turner, R.: A method for removing imaging artifact from continuous EEG recorded during functional MRI. Neuroimage **12**, 230–239 (2000)
3. Allison, T., McCarthy, G., Wood, C.C., Jones, S.J.: Potentials evoked in human and monkey cerebral cortex by stimulation of the median nerve: a review of scalp and intracranial recordings. Brain **114**, 2465–2503 (1991)
4. Barbati, G., Porcaro, C., Hadjipapas, A., Adjamian, P., Pizzella, V., Romani, G.-L., Seri, S., Tecchio, F., Barnes, G.R.: Functional source separation applied to induced visual gamma activity. Hum. Brain Mapp. **29**, 131–141 (2008)
5. Barbati, G., Sigismondi, R., Zappasodi, F., Porcaro, C., Graziadio, S., Valente, G., Balsi, M., Rossini, P.M., Tecchio, F.: Functional source separation from magnetoencephalographic signals. Hum. Brain Mapp. **27**, 925–934 (2006)
6. Barnes, G.R., Hillebrand, A.: Statistical flattening of MEG beamformer images. Hum. Brain Mapp. **18**, 1–12 (2003)
7. Bénar, C.G., Schön, D., Grimault, S., Nazarian, B., Burle, B., Roth, M., Badier, J.M., Marquis, P., Liegeois-Chauvel, C., Anton, J.L.: Single-trial analysis of oddball event-related potentials in simultaneous EEG-fMRI. Hum. Brain Mapp. **28**, 602–613 (2007)
8. Betti, V., Zappasodi, F., Rossini, P.M., Aglioti, S., Tecchio, F.: Synchronous with your feeelings: sensorimotor gamma-band and empathy for pain. J. Neurosci. **29**, 12384–12392 (2009)
9. Brown, P., Salenius, S., Rothwell, J.C., Hari, R.: Cortical correlate of the Piper rhythm in humans. J. Neurophysiol. **80**, 2911–2917 (1998)

10. Brown, P., Farmer, S.F., Halliday, D.M., Marsden, J., Rosenberg, J.R.: Coherent cortical and muscle discharge in cortical myoclonus. Brain **122**, 461–472 (1999)
11. Cottone, C., Tomasevic, L., Porcaro, C., Filligoi, G., Tecchio, F.: Physiological aging impacts the hemispheric balances of resting state primary somatosensory activities. Brain Topogr. **26**, 186–199 (2013)
12. Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D.Y., Engel, A.K.: Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identified the dynamics of performance monitoring. J. Neurosci. **25**, 11730–11737 (2005)
13. Debener, S., Ullsperger, M., Siegel, M., Engel, A.K.: Single-trial EEG-fMRI reveals the dynamics of cognitive function. Trends Cogn. Sci. **10**, 558–563 (2006)
14. Deco, G., Corbetta, M.: The dynamical balance of the brain at rest. Neuroscientist **17**, 107–123 (2011)
15. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods **134**, 9–21 (2004)
16. Di Pino, G., Porcaro, C., Tombini, M., Assenza, G., Pellegrino, G., Tecchio, F., Rossini, P.M.: A neurally-interfaced hand prosthesis tuned inter-hemispheric communication. Restor. Neurol. Neurosci. **30**, 407–418 (2012)
17. Eichele, T., Specht, K., Moosmann, M., Jongsma, M.L.A., Nordby, H., Hugdahl, K.: Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. Proc. Natl. Acad. Sci. USA **49**, 17798–17803 (2005)
18. Gross, J., Tass, P.A., Salenius, S., Hari, R., Freund, H., Schnitzler, A.: Cortico-muscular synchronization during isometric muscle contraction in humans as revealed by magnetoencephalography. J. Physiol. **527**, 623–631 (2000)
19. Hadjipapas, A., Adjamian, P., Swettenham, J.B., Holliday, I.E., Barnes, G.R.: Stimuli of varying spatial scale induce gamma activity with distinct temporal characteristics in human visual cortex. Neuroimage **35**, 518–530 (2007)
20. Hari, R., Kaukoranta, E.: Neuromagnetic studies of somatosensory system: principles and examples. Prog. Neurobiol. **24**, 233–256 (1985)
21. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
22. Kaas, J.H.: Plasticity of sensory and motor maps in adult mammals. Annu. Rev. Neurosci. **14**, 137–167 (1991)
23. Khan, O.I., Farooq, F., Akram, F., Choi, M.T., Han, S.M., Kim, T.S.: Robust extraction of P300 using constrained ICA for BCI applications. Med. Biol. Eng. Comput. **50**, 231–241 (2012)
24. Kirkpatrick, S., Gelatt Jr, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**, 671–680 (1983)
25. Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., Kleinschmidt, A.: Electroencephalographic signatures of attentional and cognitive default moes in spontaneous activity fluctuations at rest. Proc. Natl. Acad. Sci. USA **100**, 11053–11058 (2003)
26. Lu, W., Rajapakse, J.C.: Approach and applications of constrained ICA. IEEE Trans. Neural Netw. **16**, 203–212 (2005)
27. Makeig, S., Debener, S., Onton, J., Delorme, A.: Mining event-related brain dynamics. Trends Cogn. Sci. **8**, 204–210 (2004a)
28. Makeig, S., Delorme, A., Westerfield, M., Jung, T.P., Townsend, J., Courchesne, E., Sejnowski, T.J.: Electroencephalographic brain dynamics following manually responded visual targets. PLoS Biol. **6**, 0747–0762 (2004b)
29. Mantini, D., Perrucci, M.G., Cugini, S., Ferretti, A., Romani, G.L., Del Gratta, C.: Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis. Neuroimage **34**, 598–607 (2007)
30. Melgari, J.M., Zappasodi, F., Porcaro, C., Tomasevic, L., Cassetta, E., Rossini, P.M., Tecchio, F.: Movement-induced uncoupling of primary sensory and motor areas in focal task-specific hand dystonia. Neuroscience **250**, 434–445 (2013)
31. Mercier, C., Reilly, K.T., Vargas, C.D., Aballea, A., Sirigu, A.: Mapping phantom movement representations in the motor cortex of amputees. Brain **129**(Pt 8), 2202–2210 (2006)

32. Merzenich, M.M., Nelson, R.J., Stryker, M.P., Cynader, M.S., Schoppmann, A., Zook, J.M.: Somatosensory cortical map changes following digit amputation in adult monkeys. J. Comp. Neurol. **224**, 591–605 (1984)
33. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087 (1953)
34. Muthukumaraswamy, S.D., Singh, K.D., Swettenham, J.B., Jones, D.K.: Visual gamma oscillations and evoked responses: variability, repeatability and structural MRI correlates. Neuroimage **49**, 3349–3357 (2010)
35. Oostenveld, R., Praamstra, P.: The five percent electrode system for high resolution EEG and ERP measurements. Clin. Neurophysiol. **112**, 713–719 (2001)
36. Oostenveld, R., Oostendorp, T.F.: Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. Hum. Brain Mapp. **17**, 179–192 (2002)
37. Ostwald, D., Porcaro, C., Bagshaw, A.P.: An information theoretic approach to EEG-fMRI integration of visually evoked responses. Neuroimage **49**, 498–516 (2010)
38. Pascual-Marqui, R.D.: Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find. Exp. Clin. Pharmacol. **24**(Suppl D), 5–12 (2002)
39. Pellegrino, G., Tomasevic, L., Tombini, M., Assenza, G., Bravi, M., Sterzi, S., Giacobbe, V., Zollo, L., Guglielmelli, E., Cavallo, G., Vernieri, F., Tecchio, F.: Inter-hemispheric coupling changes associate with motor improvements after robotic stroke rehabilitation. Restor. Neurol. Neurosci. **30**, 497–510 (2012)
40. Pittaccio, S., Zappasodi, F., Viscuso, S., Mastrolilli, F., Ercolani, M., Passarelli, F., Molteni, F., Besseghini, S., Rossini, P.M., Tecchio, F.: Primary sensory and motor cortex activities during voluntary and passive ankle mobilization by the SHADE orthosis. Hum. Brain Mapp. **32**, 60–70 (2011)
41. Pons, T.P., Garraghty, P.E., Ommaya, A.K., Kaas, J.H., Taub, E., Mishkin, M.: Massive cortical reorganization after sensory deafferentation in adult macaques. Science **252**, 1857–1860 (1991)
42. Pfurtscheller, G., Lopes da Silva, F.H.: Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin. Neurophysiol. **110**, 1842–1857 (1999)
43. Porcaro, C., Barbati, G., Zappasodi, F., Rossini, P.M., Tecchio, F.: Hand sensory-motor cortical network assessed by functional source separation. Hum. Brain Mapp. **29**, 70–81 (2008)
44. Porcaro, C., Coppola, G., Di Lorenzo, G., Zappasodi, F., Siracusano, A., Pierelli, F., Rossini, P.M., Tecchio, F., Seri, S.: Hand somatosensory subcortical and cortical sources assessed by functional source separation: an EEG study. Hum. Brain Mapp. **30**, 660–674 (2009a)
45. Porcaro, C., Coppola, G., Pierelli, F., Seri, S., Di Lorenzo, G., Tomasevic, L., Salustri, C., Tecchio, F.: Multiple frequency functional connectivity in the hand somatosensory network: an EEG study. Clin. Neurophysiol. **124**, 1216–1224 (2013)
46. Porcaro, C., Ostwald, D., Bagshaw, A.P.: Functional source separation improves the quality of single trial visual evoked potentials recorded during concurrent EEG-fMRI. Neuroimage **50**, 112–123 (2010)
47. Porcaro, C., Ostwald, D., Hadjipapas, A., Barnes, G.R., Bagshaw, A.P.: The relationship between the visual evoked potential and the gamma band investigated by blind and semi-blind methods. Neuroimage **56**, 1059–1071 (2011)
48. Porcaro, C., Zappasodi, F., Rossini, P.M., Tecchio, F.: Choice of multivariate autoregressive model order affecting real network functional connectivity estimate. Clin. Neurophysiol. **120**, 436–448 (2009b)
49. Raffin, E., Giraux, P., Reilly, K.T.: The moving phantom: motor execution or motor imagery? Cortex **48**, 746–757 (2011)
50. Raffin, E., Mattout, J., Reilly, K.T., Giraux, P.: Disentangling motor execution from motor imagery with the phantom limb. Brain **135**(Pt 2), 582–595 (2012)
51. Reilly, K.T., Mercier, C., Schieber, M.H., Sirigu, A.: Persistent hand motor commands in the amputees' brain. Brain **129**(Pt 8), 2211–2223 (2006)
52. Siegel, M., Donner, T.H., Engel, A.K.: Spectral fingerprints of large-scale neuronal interactions. Nat. Rev. Neurosci. **13**, 121–134 (2012)

53. Tecchio, F., Zappasodi, F., Tombini, M., Oliviero, A., Pasqualetti, P., Vernieri, F., Ercolani, M., Pizzella, V., Rossini, P.M.: Brain plasticity in recovery from stroke: an MEG assessment. Neuroimage **32**, 1326–1334 (2006)
54. Tecchio, F., Porcaro, C., Barbati, G., Zappasodi, F.: Functional source separation and hand cortical representation for BCI feature extraction. J. Physiol. (Review) **580**, 703–721 (2007a)
55. Tecchio, F., Graziadio, S., Barbati, G., Sigismondi, R., Zappasodi, F., Porcaro, C., Valente, G., Balsi, M., Rossini, P.M.: Somatosensory dynamic gamma-band synchrony: a neural code of sensorimotor dexterity. Neuroimage **35**, 185–193 (2007b)
56. Tecchio, F., Zappasodi, F., Tombini, M., Caulo, M., Vernieri, F., Rossini, P.M.: Interhemispheric asymmetry of primary hand representation and recovery after stroke: a MEG study. Neuroimage **36**, 1057–1064 (2007c)
57. Tecchio, F., Zito, G., Zappasodi, F., Dell'Acqua, M.L., Landi, D., Nardo, D., Lupoi, N., Rossini, P.M., Filippi, M.M.: Intra-cortical connectivity in multiple sclerosis: a neurophysiological approach. Brain **131**, 1783–1792 (2008a)
58. Tecchio, F., Zappasodi, F., Porcaro, C., Barbati, G., Assenza, G., Salustri, C., Rossini, P.M.: High-gamma band activity of primary hand cortical areas: a sensorimotor feedback efficiency index. Neuroimage **40**, 256–264 (2008b)
59. Tombini, M., Rigosa, J., Zappasodi, F., Porcaro, C., Citi, L., Carpaneto, J., Rossini, P.M., Micera, S.: Combined analysis of cortical (EEG) and nerve stump signals improves robotic hand control. Neurorehabil. Neural Repair **26**, 275–281 (2012)
60. Wang, S., James, C.J.: Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis. Comput. Intell. Neurosci. 2007, 9 (2007). doi:10.1155/2007/41468
61. Wan, X., Riera, J., Iwata, K., Takahashi, M., Wakabayashi, T., Kawashima, R.: The neural basis of the hemodynamic response nonlinearity in human primary visual cortex: implications for neurovascular coupling mechanism. Neuroimage **32**, 616–625 (2006)

Erratum to: Performance Study for Complex Independent Component Analysis

Benedikt Loesch and Bin Yang

Erratum to:
Chapter 3 in: G. R. Naik and W. Wang (eds.), *Blind Source Separation*, DOI 10.1007/978-3-642-55016-4_3

On page 70, the term "ss" should be deleted at the end of the last display equation. Hence, the equation should read as below:

$$\mathbf{R}_{\vartheta} = \left[\begin{array}{ccc:ccc:ccc} d_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{12} & 0 & b_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{13} & 0 & 0 & 0 & b_{13} & 0 & 0 \\ \hdashline 0 & b_{21} & 0 & a_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & d_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{23} & 0 & b_{23} & 0 \\ \hdashline 0 & 0 & b_{31} & 0 & 0 & 0 & a_{31} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & b_{32} & 0 & a_{32} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d_3 \end{array}\right]$$

The online version of the original chapter can be found under DOI 10.1007/978-3-642-55016-4_3

B. Loesch (✉) · B. Yang
Institute of Signal Processing and System Theory, University of Stuttgart, Stuttgart, Germany
e-mail: benedikt.loesch@iss.uni-stuttgart.de

B. Yang
e-mail: bin.yang@iss.uni-stuttgart.de

G. R. Naik and W. Wang (eds.), *Blind Source Separation*,
Signals and Communication Technology, DOI: 10.1007/978-3-642-55016-4_20,
© Springer-Verlag Berlin Heidelberg 2014