DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

École Nationale Polytechnique

Electronic Departement
Laboratory of Communication and Photovoltaic Conversion

Master Thesis in Electronics

Blind source separation.

BENDERMEL Qasem

Supervised by

Pr. Adel BELOUCHRANI PhD. Mourad ADNANE

Jury members :

President	Mr. M. MEHENNI	Professor	ENP
Examinator	Mr. R. AKSAS	Professor	ENP
Supervisors	Mr. A. BELOUCHRANI	Professor	ENP
	Mr. M. ADNANE	PhD	ENP

ENP 2017

École Nationale Polytechnique 10, Avenue HASSEN Badi, 16200 El Harrach, Alger.

DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

École Nationale Polytechnique

Electronic Departement
Laboratory of Communication and Photovoltaic Conversion

Master Thesis in Electronics

Blind source separation.

BENDERMEL Qasem

Supervised by

Pr. Adel BELOUCHRANI PhD. Mourad ADNANE

Jury members :

President	Mr. M. MEHENNI	Professor	ENP
Examinator	Mr. R. AKSAS	Professor	ENP
Supervisors	Mr. A. BELOUCHRANI	Professor	ENP
	Mr. M. ADNANE	PhD	ENP

ENP 2017

École Nationale Polytechnique 10, Avenue HASSEN Badi, 16200 El Harrach, Alger.

Acknowledgements

My sincere thanks go to Pr. Adel Belouchrani for his supervision of this thesis, his guidance and support throughout the year were invaluable.

Also, Ph.D Mourad Adnane's assistance is most appreciated and was important to the experimentation that was undertaken in this work.

Finally, thanks to my family and all my friends for helping me get through the year.

ملخص

تفريق العناصر هو مجال احصائي لاستخراج كل عنصر لوحده بين خليط من العناصر المجهولة. هذه التقنية لا تعتمد على كثير من القواعد لذلك نجدها فعالة في كثير من الحالات. أدت البحوثات في هذا المجال إلى إيجاد نوع من الخوارزميات تدعى تحليل العناصر المستقلة. يوجد نوعان متباينان من هذه الخوارزميات: اللحظية و المتزامنة. الفرق بينها هو في طبيعة الإشارات المدروسة. في هذه الأطروحة قمنا بشرح الإطار الرياضي لكل من الحالتين مع تقديم نتائج تجريبية.

الكلمات المفتاحية: تفريق العناصر، تحليل العناصر المستقلة، معالجة الإشارة.

Résumé

La séparation de source est une approche pour tirer les signaux individuels à partir d'une mixture. la séparation de source se pose sur des faibles hypothèses sur les signaux et le processus de mixage qui rend l'application efficace dans plusieurs situations. Les recherches dans cette domaine ont donné naissance à une famille d'algorithmes connues sous le nom de "analyse des composants indépendant". Ces algorithmes se divise en deux suivant le modèle de signal. Dans cette thèse, le cadre mathématique des deux types a été donné et les résultats expérimentaux ont été présentés.

Mots clés : séparation de sources, analyse des composants indépendants, traitement de signal.

Abstract

Blind Source Separation (BSS) is a statistical approach to separating individual signals from an observed mixture of a group of signals. BSS relies on only very weak assumptions on the signals and the mixing process and this blindness enables the technique to be used in a wide variety of situations. Research in the field of Blind Source Separation has resulted in the development of a family of algorithms, known as Independent Component Analysis (ICA) algorithms, that can reliably and efficiently achieve blind separation of signals. There are two important problems that are generally considered: instantaneous BSS and convolutive BSS. The difference between these two is based on the nature of the signal mixing process. In this thesis, the mathematical foundations of both instantaneous and convolutive BSS are developed. Once this mathematical framework has been established, the emphasis of the thesis moves to experimental results obtained with ICA techniques.

Key words: blind source separation, Independent Component Analysis, signal processing.

Contents

List of Figures

List of Tables

1 Introduction **9**

2 Literature Review **11**

2.1 Development of BSS 11

2.1.1 Instantaneous BSS 11

2.1.2 Convolutive BSS 13

3 Instantaneous Blind Source Separation **14**

3.1 Mathematical Description of Source Mixing 14

3.2 Independent Component Analysis 16

3.3 BSS Assumptions 17

3.4 Preprocessing 18

3.4.1 Centering 18

3.4.2 Whitening 18

3.5 ICA Algorithms 19

3.5.1 FastICA 20

3.5.2 Bell and Sejnowski 20

3.6 Simple Illustrations of ICA 21

3.7 conclusion 24

4 Convolutive Blind Source Separation **25**

4.1 Convolutive ICA Framework 25

4.2 Time Domain Approaches to Convolutive ICA 26

4.2.1 Feedforward Time Domain Network 27

4.2.2 Feedback Time Domain Network 27

4.3 Frequency Domain Approach . . . 28
4.4 Conclusion . . . 30

5 Experimental Results for Signal Separation **31**
5.1 Instantaneous Mixtures . . . 31
5.2 Convolutive Mixtures . . . 33
5.2.1 Artificially mixed signals . . . 34
5.2.2 Separation of Signals in a Real Acoustic Environment . . . 36
5.3 conclusion . . . 37

Conclusion **38**

Bibliographie **39**

List of Figures

1.1 Diagram illustrating basic setup required for BSS. We have a number of independent sources (voices in this case) mixed at a number of different sensors (microphones). 10

3.1 Sources s_i have been linearly mixed by the unknown mixing matrix $\mathbf{A}$ and we estimate the sources by estimating the unmixing matrix $\mathbf{W}$ 15
3.2 Independent components s_1 and s_2 22
3.3 Observed signals, x_1 and x_2, from an unknown linear mixture of unknown independent components 22
3.4 Estimates of independent components 23
3.5 Joint density of observed signals x_1 and x_2 obtained from an unknown linear transformation of the independent components 23
3.6 Joint density of estimates of independent components 24

4.1 Simplified diagram of signal mixing in a real acoustic environment where s_1 and s_2 are sources and x_1 and x_2 are microphones 26
4.2 Feedforward convolutive ICA network. The w_{ij} blocks are the separating filters. 27
4.3 Feedback convolutive ICA network with direct filters w_{11} and w_{22} forced to be scalars to avoid temporal whitening. 28
4.4 Frequency domain ICA, the STFT bank transforms the mixed signals into the frequency domain where they are separated using instantaneous ICA and then converted back into the time domain using the ISTFT bank. 29

5.1 Time domain plots (5 seconds total) of instantaneous ICA test signals. 32
5.2 Spectrograms of original source signals used in algorithm analysis. Source (a) is an American male newsreader, and (b) is a female voice speaking in a foreign language. 34

5.3 Comparison of separation of convolutively mixed signals. The spectrogram of the original source is shown in (a), and the spectrograms in (b) and (c) correspond to instantaneous and convolutive ICA respectively. 35
5.4 Comparison of spectrograms for real audio mixtures, (a) shows the spectrogram of one of the two original recordings in which both voices are clearly audible, (b) shows one of the spectrograms obtained using instantaneous ICA, and (c) and (d) show the two signals resulting from convolutive ICA. 37

List of Tables

5.1 Comparison of mean crosstalk error using FastICA with different non-linear functions. Results were obtained by averaging results from 20 runs of the algorithm. 33

5.2 Comparison of mean crosstalk error using different learning rates and with $g(u) = tanh(u)$ in each case. Results were averaged for 20 runs of BS for each learning rate. 33

Chapter 1

Introduction

There are zmany situations in the real world in which a number of independent signals are mixed and interfere with one another. One obvious situation is in a noisy room, such as at a cocktail party with many people talking, music playing in the background, glasses tinkling and so on and all these acoustic signals arrive as a single waveform at a person's ear (separating voices from an environment such as this is called the "Cocktail Party Problem" and is discussed further in chapter 2). Acoustic signal mixing is familiar in our day to day lives but there are many other situations, some less familiar, in which it is important to separate independent signals from mixtures. In all of these situations, Blind Source Separation is an important signal processing technique (for brevity, Blind Source Separation will henceforth be referred to as BSS).

This thesis examines both the mathematical foundations of BSS techniques as well as applications of the techniques to real world signal processing problems. BSS is a very widely applicable technique and although it was developed only relatively recently, it has grown into an important branch of signal processing research. Figure 1.1 gives an indication of the "Cocktail Party Problem" situation that was a prime motivator research into the field of BSS. From the figure it is clear that we obtain different mixtures of the independent sources (in this case the voices of different speakers) at the microphones. The signal mixing model represented in Figure 1.1 is analogous to signal mixing in many other contexts and as a result BSS is relevant to many signal separation problems.

A major goal of this thesis was to apply BSS in practical situations, and as a result of this goal BSS was used to obtain experimental results for acoustic signal separation. Before discussing experimental results, however, the concepts of BSS are discussed in some detail in the earlier chapters. In the section below, an overview of this thesis is given.

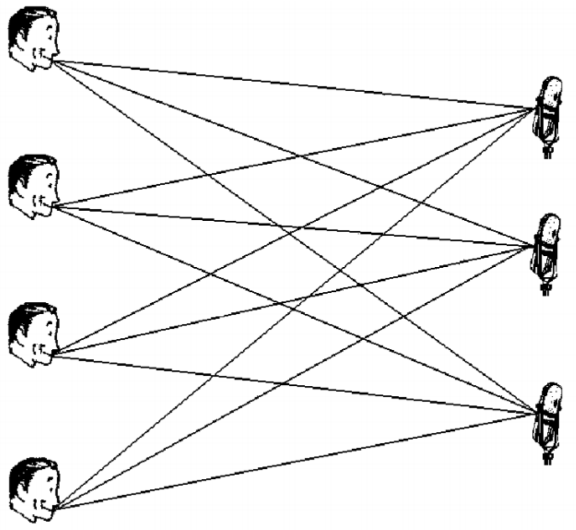


Figure 1.1: Diagram illustrating basic setup required for BSS. We have a number of independent sources (voices in this case) mixed at a number of different sensors (microphones).

Chapter 2

Literature Review

Since its first development nearly thirty years ago, Blind Source Separation has matured into an important signal processing technique. The seminal research papers in the development of Blind Source Separation are chronicled in this chapter and in discussing the significance of these papers a broad introduction to important concepts in Blind Source Separation is given.

2.1 Development of BSS

A critical distinction is drawn in BSS research between *instantaneous* and *convolutive* BSS algorithms. The distinction is based primarily on the type of signal mixing being considered – if mixing does not involve time delays then instantaneous BSS is used, otherwise convolutive BSS is necessary. Instantaneous BSS is the more well-developed and widely applicable of the two and the historical development of BSS as a research field began with the instantaneous BSS. Convolutive BSS is an extension of instantaneous BSS that was developed more recently to address some specific applications.

2.1.1 Instantaneous BSS

Blind Source Separation is a recent, and remarkable, chapter in the development of signal processing. The field began at a neural networks conference in Utah in 1986 where Jutten and Herault presented a paper entitled "Space or time adaptive signal processing by neural network models" [1].

Many algorithms have been developed to perform separation, but prior to BSS major assumptions were always required on the nature of the sources. Jutten and Herault's technique was revolutionary in that it did not require assumptions on the nature of the signals being separated. Despite this, however, their technique did not initially attract much attention. This is primarily because in the 1980s, neural network research focused on Hopfield networks and Jutten and Herault's work went largely unnoticed. It was only with a much clearer formulation of BSS by Comon in 1994 [2] that BSS became a mainstream topic of research.

Algorithms developed to perform blind separation of sources were given the name Independent Component Analysis (ICA) algorithms wheras BSS refers to the entire body of knowledge relevant to blindly separating signals. Separation techniques were named ICA to highlight the fact that independent components were being separated from mixtures of signals, but also to emphasise a close link with the classical signal processing technique of Principal Component Analysis (PCA). PCA can be used to separated mixtures of signals using decorrelation. A well-known fact from elementary statistics is that for non-Gaussian signals, uncorrelated signals are not necessarily independent. To decorrelate, it is only necessary to consider second-order statistics, whereas independence requires higher order statistics. As a result, it is common to consider ICA to be an extension of PCA that is able to separate non-Gaussian signals. This partly explains the late development of ICA as until fairly recently, Gaussian sources were assumed in most signal processing research. Despite being limited to second-order statistics, PCA is still a powerful technique and has many uses, including feature extraction and data compression [3].

Following Comon's seminal paper, there was a rapid proliferation of ICA algorithms. Algorithms were formulated based on a wide variety of principles, including mutual information, maximum likelihood and higher order statistics, to name just a few of the more popular approaches. Despite such wide variety, all ICA algorithms are fundamentally similar. ICA algorithms invariably obtain estimates of the independent signals by adopting a numerical approach (e.g. gradient descent) to maximizing an “independence metric”, i.e. a measure of the signals' independence. The main difference between different ICA algorithms is in the metric that is used.

On publication of their algorithm in 1995, Bell and Sejnowski's [4] approach to ICA became the most popular choice due to its simplicity and its favourable convergence properties. However, the algorithm involved matrix inversion which significantly hindered efficiency. Amari discovered an important improvement (using “natural” gradient descent, see [5]) to the algorithm of Bell and Sejnowski which eliminated the matrix inversion. This gave a significant performance improvement and made ICA more practical for real world problems, especially in separating large numbers of sources.

Another important ICA algorithm, called FastICA [6], was developed in 1997 by Oja and Hyvärinen of the Helsinki University of Technology. It was shown to be a very good alternative to Bell and Sejnowski's algorithm, and is probably currently the most widely used ICA algorithm.

2.1.2 Convolutive BSS

The approach to BSS discussed in Section 2.1.1 is based on an assumption that there are no time delays involved in the mixing of independent signals. In some situations, however, the instantaneous mixing model is inadequate. An obvious case in which it is necessary to account for time delays in signal mixing is in the separation of speech signals recorded in a real environment. Sound wave propagation is relatively slow and the waves are subject to reflections so that time delays are introduced in the mixing process. To separate sources mixed in this way a technique called *convolutive* Blind Source Separation was developed which extends the standard instantaneous BSS model by treating signal mixing as a convolution, allowing time delays to be accounted for.

The primary motivation for the development of convolutive ICA algorithms was to treat the speech separation problem . This separation problem has been named the "Cocktail Party Problem". The name itself conjures the familiar image of a crowded, noisy room, but a room in which people can still communicate since the human brain is effective at isolating signals. It has been an intriguing problem in signal processing and artificial intelligence to try to develop algorithms to simulate this ability of the human brain. Convolutive BSS is one approach that has shown some promise.

The first solution to the convolutive BSS problem was developed by Bell and Sejnowski [4], who proposed a feedforward neural network structure using FIR filters. This approach was improved by Torkkola who employed a similar network, but with feedback structure [7]. Both of these approaches are limited in their practicality, primarily because they involve computationally expensive convolutions of long filters.

To overcome the shortcomings of these time domain algorithms, Smaragdis proposed moving into the frequency domain [8]. Smaragdis realized that by moving to the frequency domain, the problem of convolutive mixing simplifies to instantaneous mixing allowing standard instantaneous ICA algorithms to be employed. It is clear that the frequency domain approach is superior to the time domain algorithms proposed initially.

Convolutive ICA is generally required to separate audio mixtures recorded in real acoustic environments. The promise of convolutive ICA has been demonstrated as an approach to solving the real-world "Cocktail Party Problem". The success of convolutive ICA in audio separation has resulted in a number of practical applications of the technique, such as in voice recognition systems [9] or in improving hearing aid technology [10]. In both of these applications, speech signals are separated from contaminating noise sources to improve system performance.

Chapter 3

Instantaneous Blind Source Separation

Consider a situation in which we have a number of sources emitting signals which are interfering with one another. A familiar situation in which this occurs is a crowded room with many people speaking at the same time. In this kind of situations, the mixed signals are often incomprehensible and it is of interest to separate the individual signals which is the goal of Blind Source Separation.

3.1 Mathematical Description of Source Mixing

The first step in deriving a solution to the source separation problem is to adequately model source mixing. BSS can be applied to a collection of statistically independent sources which are emitting signals that interfere with each other and the interfering signals are recorded using a number of spatially separated sensors. For the purpose of clarity, the simplest case where the number of sources is equal to the number of sensors is considered.

Suppose we have N statistically independent signals, $s_i(t)$, $i = 1, ..., N$. We assume that the sources themselves cannot be directly observed and that each signal, $s_i(t)$, is a realization of some fixed probability distribution at each time point t. Also, suppose we observe these signals using N sensors, then we obtain a set of N observation signals $x_i(t)$, $i = 1, ..., N$ that are mixtures of the sources. A fundamental aspect of the mixing process is that the sensors must be spatially separated so that each sensor records a different mixture of the sources. With this spatial separation assumption in mind, we can model the mixing process with matrix multiplication as follows:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{3.1}$$

where $\mathbf{A} \in \Re^{N \times N}$ is an unknown matrix called the mixing matrix and $\mathbf{x}(t), \mathbf{s}(t) \in \Re^N$ are the two vectors representing the observed signals and source signals respectively. Incidentally, the justification for the description of this signal processing technique as *blind* is that we have no information on the mixing matrix, or even on the sources themselves.

The objective is to recover the original signals, $s_i(t)$, from only the observed vector $\mathbf{x}(t)$. We obtain estimates for the sources by first obtaining the "unmixing matrix" $\mathbf{W}$, where:

$$\mathbf{W} = \mathbf{A}^{-1} \tag{3.2}$$

This enables an estimate, $\mathbf{y}(t)$, of the independent sources to be obtained:

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{3.3}$$

where the time index t has been omitted for notational simplicity.

The diagram in Figure 3.1 illustrates both the mixing and unmixing process involved in BSS. The independent sources are mixed by the matrix $\mathbf{A}$ (which is unknown in this case). We seek to obtain a vector $\mathbf{y}$ that approximates $\mathbf{s}$ by estimating the unmixing matrix $\mathbf{W}$. If the estimate of the unmixing matrix is accurate, we obtain a good approximation of the sources. It is important to note that the formulation in Eq.(3.1) assumes instantaneous mixing; there are no time delays involved in the mixing process. This model is inadequate in many situations and a more elaborate model which can handle non-instantaneous mixing is presented in Chapter 4.

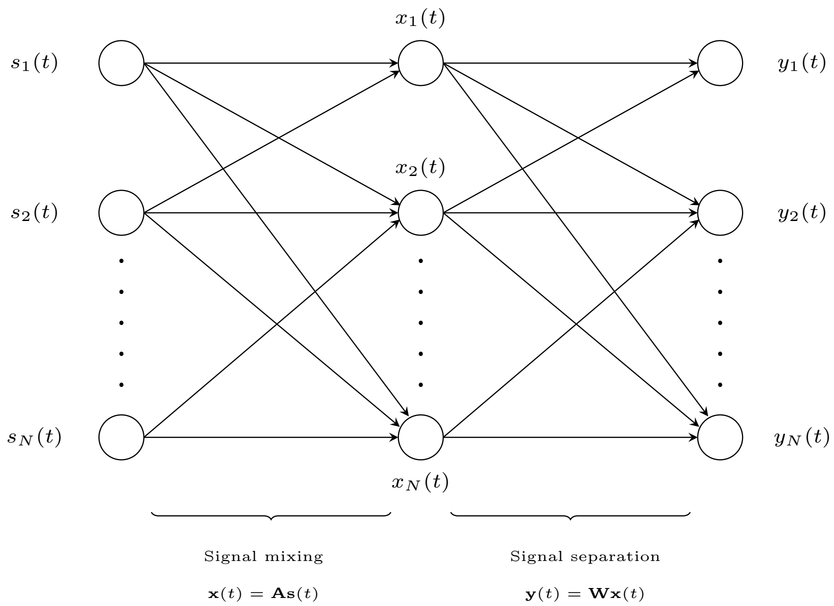


Figure 3.1: Sources s_i have been linearly mixed by the unknown mixing matrix $\mathbf{A}$ and we estimate the sources by estimating the unmixing matrix $\mathbf{W}$

3.2 Independent Component Analysis

Algorithms that perform Blind Source Separation are known as Independent Component Analysis (ICA) algorithms. Intuitively, ICA involves estimating the linear transformation that maximizes the independence of the signals. This linear transform is referred to as the unmixing matrix, $\mathbf{W}$. Since the original sources, $s_i(t)$, were assumed to be independent, we know that maximising the independence of the components of $\mathbf{y}$ from Eq.(3.3) we will obtain estimates of the original sources.

Suppose the observation vector $\mathbf{x}$ is formed according to Eq.(3.1), that is, $\mathbf{x}$ is a linear combination of independent components. To estimate one of the independent components we consider a linear combination of the x_i terms. Let y_i be one of the estimates, then we have:

$$y_i = \mathbf{w}^T \mathbf{x} \tag{3.4}$$

The crucial point is that if $\mathbf{w}$ was one of the rows of the inverse of the mixing matrix, then y_i would actually be one of the original independent components. At this point, it is not clear how we can determine such a vector $\mathbf{w}$ when we have no information about the mixing matrix $\mathbf{A}$. Independent Component Analysis has been developed to solve this problem.

ICA depends fundamentally on the independence of the original sources, $s_i(t)$. The following equation, derived from Eq.(3.4), illustrates this link:

$$y_i = \mathbf{w}^T \mathbf{A}\, \mathbf{s} = \mathbf{z}^T \mathbf{s} \tag{3.5}$$

From this equation it is clear that y_i is a linear combination of the independent components, $s_i(t)$.

The approach to Independent Component Analysis is to consider the "Gaussianity" of the signals y_i of Eq.(3.4). From the Central Limit Theorem we know that the distribution of a sum of independent random variables approaches a Gaussian distribution [11]. From Eq.(3.4) it is clear that the estimate y_i is a sum of independent random variables, thus we expect its distribution to be "more Gaussian" than the distribution of the independent components s_i. Assuming none of the independent components are Gaussian distributed, we obtain the most "non-Gaussian" distribution for y_i when it is exactly one of the independent components s_i. As a result, by choosing the vector $\mathbf{w}$ to maximize the non-Gaussianity of y_i, we will obtain an estimate of one of the independent components s_i. To proceed further we require a method of measuring non-Gaussianity.

Perhaps the most straightforward approach to measuring non-Gaussianity is kurtosis. Kurtosis is the fourth order cumulant [12] defined by:

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{3.6}$$

It is discussed here because of its simplicity and importance in independent component analysis. The kurtosis of a Gaussian random variable is zero, and is non-zero for (almost) any non-Gaussian random variable. As a result, by maximizing the absolute value of the kurtosis we can maximize non-Gaussianity. While kurtosis offers a simple approach to measuring non-Gaussianity, it has been established that it is not a robust mesure [12].

A more practical approach is to measure Gaussianity by calculating the entropy of the sources since a Gaussian random variable has the greatest entropy of all random variables of equal variance [12]. Entropy (or negentropy to be precise) is the fundamental independence metric used in deriving the FastICA algorithm (see Section 3.5.1). Once an appropriate measure of non-Gaussianity has been established (kurtosis and entropy are two possibilities) then this measure can be used to define a cost function which when minimised yields the independent components, $s_i(t)$.

3.3 BSS Assumptions

Blind Source Separation is distinguished from other approaches to source separation in that it requires relatively few assumptions on the sources and on the mixing process. The assumptions essential to BSS are discussed here [12]:

1. The sources being considered are statistically independent
2. The independent components have non-Gaussian distribution
3. The mixing matrix is invertible

The first assumption is fundamental to ICA. As discussed in Section 3.2, statistical independence is the key feature that enables estimation of the independent components $y_i(t)$ from the observations $x_i(t)$.

The second assumption is necessary because of the close link between Gaussianity and independence. It is impossible to separate Gaussian sources using the 3.2 framework described in Section 3.2 because the sum of two or more Gaussian random variables is itself Gaussian [12], and for this reason Gaussian sources are forbidden. This is not an overly restrictive assumption as in practice most sources of interest are non-Gaussian.

The third assumption is straightforward. If the mixing matrix is not invertible then clearly the unmixing matrix we seek to estimate does not even exist. If these three assumptions are satisfied, then it is possible to estimate the independent components modulo some trivial ambiguities. It is clear that these assumptions are not particularly restrictive and as a result we need only very little information about the mixing process and about the sources themselves.

3.4 Preprocessing

Before examining specific ICA algorithms, it is instructive to discuss preprocessing steps that are generally carried out before ICA.

3.4.1 Centering

A simple preprocessing step that is commonly performed is to "center" the observation vector $\mathbf{x}$ by subtracting its mean vector $\mathbf{m} = E\{\mathbf{x}\}$. That is then we obtain the centered observation vector, $\mathbf{x}_c$, as follows:

$$\mathbf{x}_c = \mathbf{x} - \mathbf{m} \tag{3.7}$$

This step simplifies ICA algorithms by allowing us to assume a zero mean.

Once the unmixing matrix has been estimated using the centered data, we can obtain the actual estimates of the independent components as follows:

$$\mathbf{y} = \mathbf{A}^{-1}(\mathbf{x}_c + \mathbf{m}) \tag{3.8}$$

From this point on, all observation vectors will be assumed centered.

3.4.2 Whitening

Another step which is very useful in practice is to pre-whiten the observation vector $\mathbf{x}$. Whitening involves linearly transforming the observation vector such that its components are uncorrelated and have unit variance [12]. Let $\mathbf{x}_w$ denote the whitened vector, then it satisfies the following equation:

$$E\{\mathbf{x}_w\mathbf{x}_w^T\} = \mathbf{I} \tag{3.9}$$

where $E\{\mathbf{x}_w\mathbf{x}_w^T\}$ is the covariance matrix of $\mathbf{x}_w$. Also, since the ICA framework is insensitive to the variances of the independent components, we can assume without loss of generality that the source vector, $\mathbf{s}$, is white, i.e. $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{I}$.

A simple method to perform the whitening transformation is to use the eigenvalue decomposition (EVD) [12] of **x**. That is, we decompose the covariance matrix of **x** as follows:

$$E\{\mathbf{x}\mathbf{x}T\} = \mathbf{V}\mathbf{D}\mathbf{V}^T \tag{3.10}$$

where **V** is the matrix of eigenvectors of $E\{\mathbf{x}\mathbf{x}T\}$, and **D** is the diagonal matrix of eigenvalues, i.e. $\mathbf{D} = \mathrm{diag}\{\lambda_1, \lambda_2, ..., \lambda_n\}$.

The observation vector can be whitened by the following transformation:

$$E\{\mathbf{x}_w\} = \mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{x} \tag{3.11}$$

We can confirm that this yields a whitened vector $\mathbf{x}_w$:

$$\begin{aligned} E\{\mathbf{x}_w\} &= \mathbf{D}^{-1/2}\mathbf{V}^T(E\{\mathbf{x}\mathbf{x}T\})\mathbf{V}\mathbf{D}^{-1/2} \\ &= \mathbf{D}^{-1/2}\mathbf{V}^T(\mathbf{V}\mathbf{D}\mathbf{V}^T)\mathbf{V}\mathbf{D}^{-1/2} \\ &= \mathbf{I} \end{aligned} \tag{3.12}$$

since $\mathbf{V}\mathbf{V}^T = I$ as the matrix of eigenvectors is orthogonal.

By pre-whitening the data, we transform the mixing matrix **A** into a new matrix $\mathbf{A}_w$. The usefulness of whitening is that the new mixing matrix is orthogonal, as shown by the following chain of equations:

$$I = E\{\mathbf{x}_w\mathbf{x}_w^T\} = \mathbf{A}_w E\{\mathbf{s}\mathbf{s}^T\}\mathbf{A}_w^T = \mathbf{A}_w I \mathbf{A}_w^T = \mathbf{A}_w\mathbf{A}_w^T \tag{3.13}$$

An orthogonal matrix contains only $n(n-1)/2$ degrees of freedom compared to n^2 degrees of freedom for an unconstrained matrix. As a result, pre-whitening the vector **x** effectively reduces the number of parameters that need to be estimated by ICA by about half. This is a very useful step as whitening is a simple and efficient process that significantly reduces the computational complexity of ICA.

From this point on all observation vectors will be assumed centered and whitened.

3.5 ICA Algorithms

In this section, two of the most important ICA algorithms are presented in some detail. Oja and Hyvärinen's FastICA algorithm [6] is presented in Section 3.5.1 and Bell and Sejnowski's information maximization algorithm in Section 3.5.2 These two algorithms are probably the most widely used and they each illustrate the important principles of ICA.

3.5.1 FastICA

This algorithm is based on using "non-Gaussianity" as a metric for independence. FastICA is based on using entropy as a measure of non-Gaussianity. A fundamental result of information theory is that a Gaussian random variable has the greatest entropy of all random variables of equal variance. As a result, entropy can be used as a measure of non-Gaussianity. To be precise, FastICA is not based on entropy, but rather on negentropy. Negentropy is a related concept defined by:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{3.14}$$

where $H(\cdot)$ denotes the entropy of a random variable, $J(\cdot)$ denotes negentropy and $\mathbf{y}_{gauss}$ is a Gaussian random vector with the same covariance matrix as $\mathbf{y}$. Negentropy is always non-negative since $H(\mathbf{y}_{gauss}) \geq H(\mathbf{y})$.

FastICA maximizes negentropy using Newton's iterative method, for details of the derivation of FastICA refer to [6]. The algorithm determines the unmixing matrix one column at a time, with the update rule for each column defined by:

$$\mathbf{w}^{+} = E\{\mathbf{x}g(\mathbf{w}^T\mathbf{x})\} - E\{g'(\mathbf{w}^T\mathbf{x})\}\mathbf{w} \tag{3.15}$$

The function g can be almost any non-quadratic function, but hyperbolic tangent functions have been shown to behave well in practice [6].

3.5.2 Bell and Sejnowski

Bell and Sejnowski's algorithm (henceforth referred to as the BS algorithm) is based on the neural network principle of Information Maximization [4] but is essentially one of the family of maximum-likelihood (ML) ICA algorithms. That is, independence is maximised by estimating the unmixing matrix that maximises the probability of the observation vector $\mathbf{x}$.

Based on Eq.(3.1), we can find the probability density of the observation vector, $\mathbf{x}$, in terms of the density of $\mathbf{s}$:

$$p_x(x) = |det\mathbf{W}|p_s(\mathbf{s}) = |det\mathbf{W}|\prod_i p_i(s_i) \tag{3.16}$$

where $\mathbf{W} = \mathbf{A}^{-1}$, and $p_i(s_i)$ denotes the density of the i^{th} component of $\mathbf{s}$. The second equality in Eq.(3.16) follows because of the independence of the components of $\mathbf{s}$.

Let us assume we have a vector $\mathbf{w}_i$ that satisfies $\mathbf{w}_i^T\mathbf{x} = s_i$. We can then rewrite Eq.(3.16) as:

$$p_x(\mathbf{x}) = |det\mathbf{W}| \prod_i p_i(\mathbf{w}_i^T \mathbf{x}) \tag{3.17}$$

This expression can be used to define a likelihood for the unmixing matrix **W** since for a given observation vector **x** we can determine the vector $\mathbf{w}_i$ that maximises the likelihood that **x** would be observed. Suppose we have T observations of **x**, then we can estimate the likelihood of the unmixing matrix **W** with the log-likelihood expression:

$$\log L(\mathbf{W}) = \sum_{t=1}^{T} \sum_{i=1}^{n} \log p_i(\mathbf{w}_i^T \mathbf{w}(t)) + T \log |\det \mathbf{B}| \tag{3.18}$$

Bell and Sejnowski's algorithm maximizes this likelihood expression by performing gradient ascent. For details of the derivation of the BS algorithm see [4], but the actual algorithm is given here:

$$\mathbf{\Delta W} \propto \left[\mathbf{W}^T\right]^{-1} + E\{g(\mathbf{W}\mathbf{x})\mathbf{x}^T\} \tag{3.19}$$

Technically this algorithm is only semi-blind because the derivation requires that the non-linear activation function **g** must approximate the Cumulative distribution function (CDF) of **s**, thus some assumptions must be made about the distribution of each independent component, s_i. It is fortunate, however, that this algorithm is quite insensitive to the accuracy of this approximation and that in most cases of interest the independent components are Gaussian-like in nature. Hyperbolic tangent and logistic functions are good approximations of the CDF of Gaussian-like random variables and either can be used effectively for g_i.

3.6 Simple Illustrations of ICA

To clarify the concepts discussed in the preceding sections two simple illustrations of ICA are presented here. The results presented below were obtained using the FastICA algorithm of Section 3.5.1 but could equally well have been obtained from any of the numerous ICA algorithms that have been published in the literature.

Separation of Two Signals

In this illustration two independent signals, s_1 and s_2, are generated. These signals are shown in Figure 3.2. The independent components are then mixed according to Eq.(3.1) using an arbitrarily chosen mixing matrix **A**, where

$$\mathbf{A} = \begin{bmatrix} -0.3784 & 0.8537 \\ 0.8600 & 0.5936 \end{bmatrix} \tag{3.20}$$

The resulting signals from this mixing are shown in Figure 3.2. Finally, the mixtures x_1 and x_2 are separated using ICA to obtain y_1 and y_2, shown in Figure 3.4. By comparing Figure 3.4 to Figure 3.2 it is clear that the independent components have been estimated accurately and that the independent components have been estimated without any knowledge of the components themselves or the mixing process. This example also provides a clear illustration of the scaling and permutation ambiguities. The amplitudes of the corresponding waveforms in Figures 3.2 and 3.4 are different and the sawtooth waveform in Figure 3.4 has been reflected vertically with respect to the sawtooth waveform in Figure 3.2. Thus the estimates of the independent components are some multiple of the independent components of Figure 3.2, and in the case of s_1, the scaling factor is negative. The permutation ambiguity is also demonstrated as the order of the independent components has been reversed between Figure 3.2 and Figure 3.4.

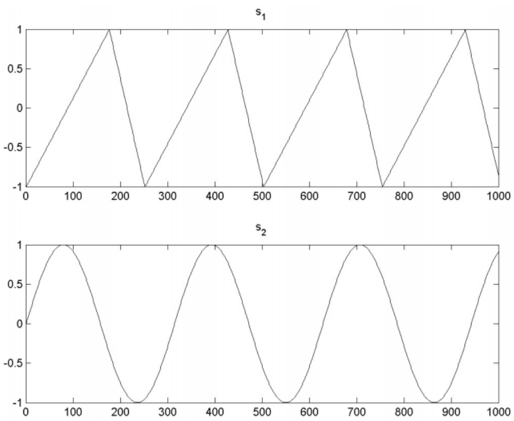


Figure 3.2: Independent components s_1 and s_2

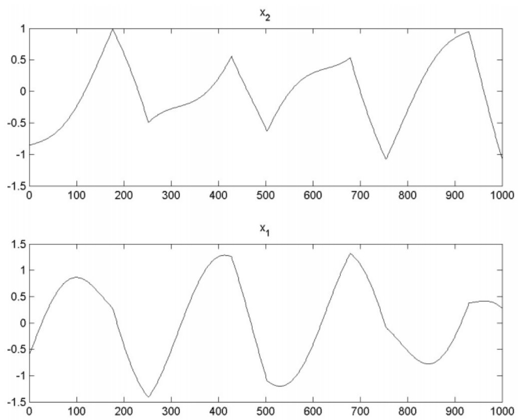


Figure 3.3: Observed signals, x_1 and x_2, from an unknown linear mixture of unknown independent components

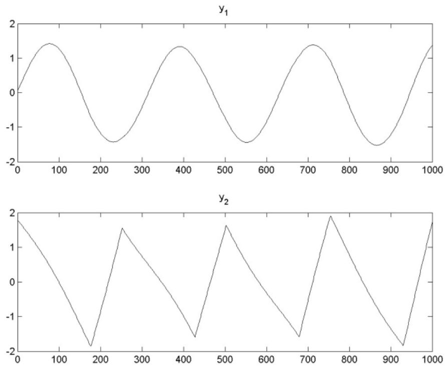


Figure 3.4: Estimates of independent components

Illustration of Statistical Independence in ICA

The previous example was a simple illustration of how ICA is used; we start with mixtures of signals and use ICA to separate them. However, this gives no insight into the mechanics of ICA and the close link with statistical independence. The statistical basis of ICA is illustrated more clearly in this example.

We assume that the independent components can be modeled as realizations of some underlying statistical distribution at each time instant (e.g. a speech signal can be accurately modeled as having a Laplacian distribution [13]). One way of visualizing ICA is that it estimates the optimal linear transform to maximise the independence of the joint distribution of the signals x_i.

Suppose we have the joint probability distribution for the observed signals x_1 and x_2 shown in figure 3.5. From the figure it is clear that the two signals are not statistically independent because, for example, if x_1 = 0 or 3 then x_2 is totally determined. By applying ICA, we seek to transform the data such that we obtain two independent components.

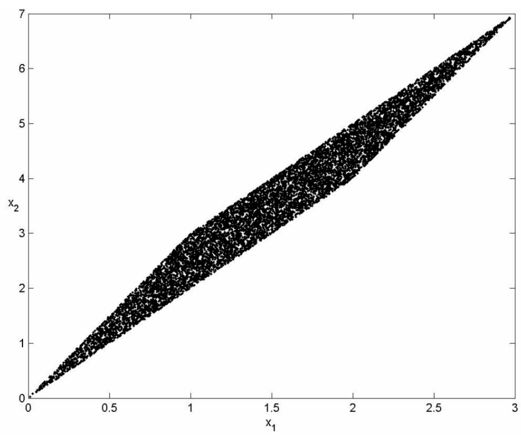


Figure 3.5: Joint density of observed signals x_1 and x_2 obtained from an unknown linear transformation of the independent components

The joint distribution resulting from applying ICA to x_1 and x_2 is shown in figure 3.6. This is clearly the joint distribution of two independent, uniformly distributed random variables. Independence can be intuitively confirmed as each random variable is unconstrained regardless of the value of the other random variable (this is not the case for x_1 and x_2). The uniformly distributed random variables in figure 3.6 take values between 0 and -4, but due to the scaling ambiguity, we do not know the range of the original independent components.

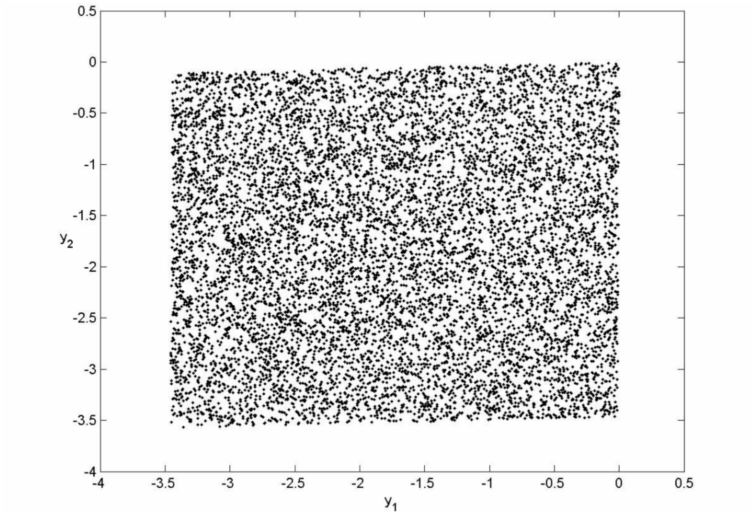


Figure 3.6: Joint density of estimates of independent components

The two examples in this section are simple but they illustrate both how ICA is used and the statistical underpinnings of the process. The power of ICA is that an identical approach can be used to address problems of much greater complexity.

3.7 conclusion

This chapter has introduced the fundamentals of Blind Source Separation. The mathematical framework of the source mixing problem that BSS addresses was examined in some detail, as was the general approach to solving BSS. As part of this discussion, some inherent ambiguities of the BSS framework were examined as well as the two important preprocessing steps of centering and whitening. Finally, specific details of the approach to solving the mixing problem were presented and two important ICA algorithms were discussed in detail. The material covered in this chapter is important not only to understand the algorithms used in this project to perform BSS, but it also provides the necessary background.

Chapter 4

Convolutive Blind Source Separation

The approaches to BSS presented in the previous chapter are robust and efficient enough to use in real world applications. However, they are limited to treating situations in which signals are mixed with no time delays. This chapter presents an extension to the instantaneous ICA framework of Chapter 3, called convolutive ICA, which enables separation of signals mixed with time delays. Convolutive ICA is especially important when separating audio signals. In the real world, it is rare to find instantaneous mixtures of audio signals because of reflections. As an illustration, consider the situation pictured in figure 4.1 which shows two audio sources and two microphones in a room. The microphones record both the direct sounds and the reflections off the walls and each will involve a different time delay and level of attenuation. This mixing situation is described using a convolution.

4.1 Convolutive ICA Framework

The equation that describes the mixing process of figure 4.1 is the convolution:

$$x_i(t) = \sum_{j=1}^{N} \sum_{k=0}^{M} s_j(t-k) a_{ij}(k) \tag{4.1}$$

The above equation is the summation of the convolution of sources, s_j with filters a_i. The filters are generally assumed to be Finite Impluse Response (FIR) filter and they model the acoustic environment for each source/sensor pair. This mixing process is more complicated than the one considered in Chapter 3 and requires a more advanced approach to achieve separation.

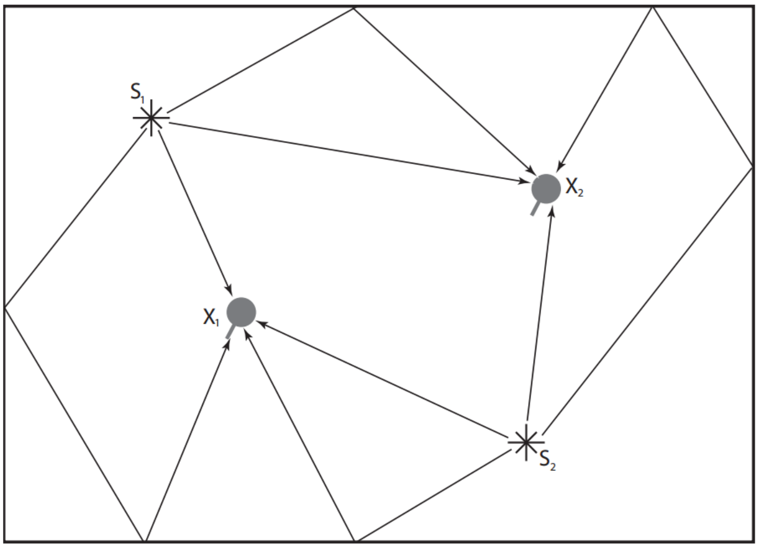


Figure 4.1: Simplified diagram of signal mixing in a real acoustic environment where s_1 and s_2 are sources and x_1 and x_2 are microphones

Using FIR linear algebra notation Eq.(4.1) can be rewritten more elegantly:

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t) \tag{4.2}$$

The goal of convolutive BSS is to determine an unmixing FIR matrix, $\mathbf{W}$, that will invert the mixing process of Eq.(4.2). Considering only the case of two sources for simplicity, A can be inverted analogously to standard linear algebra to yield the optimal unmixing FIR matrix $\mathbf{W}$:

$$\mathbf{W} = \frac{1}{a_{11} * a_{22} - a_{12} * a_{21}} \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix} \tag{4.3}$$

and the estimate $\mathbf{y}$ of the separated sources is given by:

$$y = \mathbf{W} \cdot \mathbf{x} \tag{4.4}$$

As in Chapter 3, the mixing FIR matrix is unknown so that statistical techniques must be employed to estimate the unmixing FIR matrix.

4.2 Time Domain Approaches to Convolutive ICA

The first approaches to addressing the convolutive BSS problem were algorithms that worked purely in the time domain. Two important time domain approaches to convolutive ICA are the feedforward network approach proposed by Bell and Sejnowski [4] and the feedback approach of Torkkola [7]. Both approaches use networks with unmixing filters and the taps of the filters are adapted using learning rules in order to approximate the ideal solution of Eq.(4.3).

4.2.1 Feedforward Time Domain Network

Bell and Sejnowski's feedforward network for the two-source convolutive BSS problem is shown in Figure 4.2. The network contains four filters which are used to separate the signals x_1 and x_2. The estimates of the separated signals are y_1 and y_2 and the non-linear output function g is analogous to the CDF approximating non-linear function in Bell and Sejnowski's instantaneous ICA algorithm.

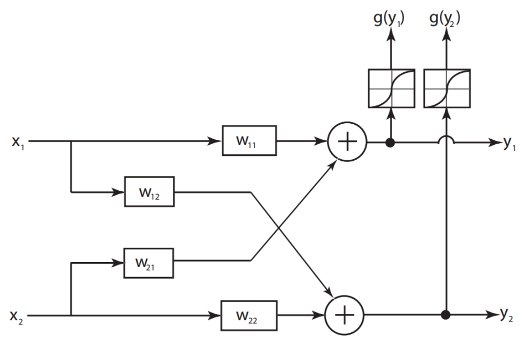


Figure 4.2: Feedforward convolutive ICA network. The w_{ij} blocks are the separating filters.

The network structure of Figure 4.2 is suggested by Eq.(4.3) and Eq.(4.4) since we will achieve separation if the filters in Figure 4.2 match the filters in Eq.(4.3).

The filter taps are updated individually using the learning rule from Bell and Sejnowski's instantaneous BSS algorithm in order to maximise the independence of the outputs y_1 and y_2 based on the values of the non-linear functions $g(y_1)$ and $g(y_2)$

4.2.2 Feedback Time Domain Network

Torkkola [7] showed that the feedforward network architecture of Section 4.2.1 is not ideal because the w_{11} and w_{22} filters enable temporal whitening of the input signals. Temporal whitening is encouraged by the feedforward network structure because it increases signal independence, but it is undesirable because it tends to highpass filter the signals and distort their spectral quality [7]. To avoid this problem, Torkkola proposed the feedback architecture shown in Figure 4.3 which only has cross filters (w_{12} and w_{21}) and forces the direct filters w_{11} and w_{22} to scalar values so that it is impossible for the network to temporally whiten the signals.

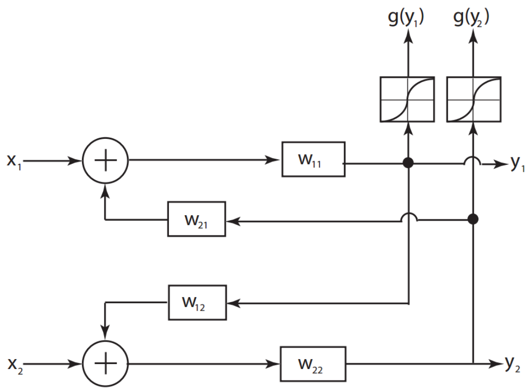


Figure 4.3: Feedback convolutive ICA network with direct filters w_{11} and w_{22} forced to be scalars to avoid temporal whitening.

The same update algorithm was employed by Torkkola as by Bell and Sejnowski, but the feedback network structure resulted in superior separation and avoided distortion due to temporal whitening.

4.3 Frequency Domain Approach

The time domain algorithms of the previous section result in successful separation of convolutively mixed signals. However, they are very inefficient because they involve convolutions of long filters (for audio separation problems filters may routinely need around 1000 taps to model the acoustic environment). Smaragdis [8] proposed a more efficient approach by working in the frequency domain. By transforming the convolutive mixing problem into the frequency domain, Smaragdis realised that it would reduce to an instantaneous mixing BSS problem which and treated using the methods of Chapter 3.

Smaragdis' approach was to first transform the observed mixtures into the frequency domain using a Short-Time Fourier Transform (STFT) yielding a set of frequency bins for each signal. Let $u_k(\omega, t)$ denote time point t in frequency bin ω for the signal $x_k(t)$. The frequency bins at each frequency ω are instantaneously mixed and an instantaneous ICA algorithm is used to find the unmixing matrix $\mathbf{W}$ for each bin. Once separated, the STFT is inverted to reconstruct the separated signals in the time domain. Smaragdis' procedure for frequency domain ICA is illustrated in figure 4.4.

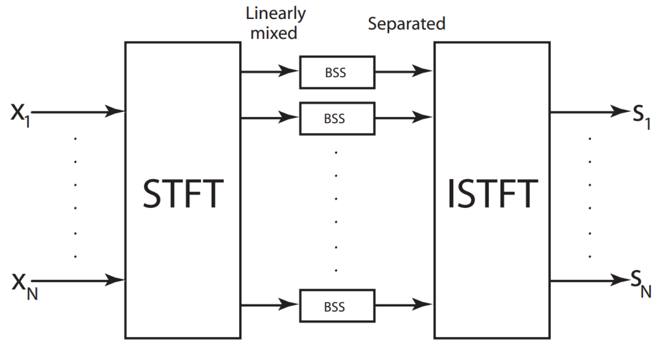


Figure 4.4: Frequency domain ICA, the STFT bank transforms the mixed signals into the frequency domain where they are separated using instantaneous ICA and then converted back into the time domain using the ISTFT bank.

There are two major complications in implementing this convolutive framework:(1) instantaneous ICA algorithms must be modified to work with complex numbers and; (2) we must ensure that each frequency bin separates to the same permutation. The first issue is addressed in a paper by Bingham and Hyvärinen [14] that extends the FastICA algorithm to work with complex numbers. The second problem arises from the permutation ambiguity in BSS and as yet has no robust solution. In instantaneous ICA the permutation ambiguity was of minor importance. In Smaragdis' approach, it is absolutely essential to keep the same permutation in frequency bin as otherwise the signals remain mixed.

A simple approach to avoiding the permutation problem was suggested by Smaragdis: work sequentially through the frequency bins using the converged value of $\mathbf{W}$ from the i^{th} bin as the initial value for the $(i+1)^{th}$ bin. This heuristic encourages the unmixing matrices to converge to the same permutation solution in each bin, but in general it is unreliable.

Davies and Mitianoudis [15] proposed a more effective approach to solving the permutation problem which relied on using time-domain information to identify the correct permutations. Specifically, Davies and Mitianoudis proposed using time domain data to determine the most likely permutation of the instantaneous unmixing matrix $\mathbf{W}$. In the 2×2 case, the technique compares the likelihood of the unmixing matrix $\mathbf{W}$ with $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{W}$ at frequency bin ω to determine which permutation matches the rest of the frequency bins more closely. The likelihood of each permutation is determined by calculating the *likelihood ratio* (LR), given by:

$$LR = \frac{\gamma_{12}\gamma_{21}}{\gamma_{11}\gamma_{22}} \tag{4.5}$$

where

$$\gamma_{ij} = \sum_{t=1}^{T} \frac{|u_i(t)|}{\beta_j(t)} \tag{4.6}$$

The term $\beta_j(t)$ in γ_{ij} provides the time domain information that is crucial in eliminating the permutation problem, it is effectively a time average over frequency and is given by:

$$\beta_j(t) = \frac{1}{N} \sum_{\omega} |u_j(\omega, t)| \tag{4.7}$$

This technique enables efficient separation, but unfortunately it is not easily extensible to more than two signals.

4.4 Conclusion

This chapter introduced the concept of convolutive BSS. The most significant application of convolutive BSS is in separation of acoustic signals recorded in a real environment. Two time domain approaches to convolutive BSS were presented in Section 4.2 but due to the long convolutions involved in these procedures they are too inefficient for practical use. A more efficient frequency domain algorithm was considered in Section 4.3.

Chapter 5

Experimental Results for Signal Separation

In this chapter experimental results from both instantaneous and convolutive audio signal separation using the ICA algorithms discussed in Chapters 3 and 4 are presented. The objectives of this experimentation were to compare the effectiveness of Bell and Sejnowski's instantaneous ICA algorithm with FastICA, show that Davies' convolutive ICA algorithm is effective in separating convolutive audio mixtures and demonstrate the inadequacy of instantaneous BSS in dealing with convolutive audio mixtures.

5.1 Instantaneous Mixtures

In practical applications of instantaneous ICA the mixing matrix is unknown, but for the purposes of algorithm assessment, in this section a number of signals were mixed using a known mixing matrix, **A**. When the mixing matrix is known it is straightforward to assess the effectiveness of instantaneous ICA algorithms since for ideal separation the product matrix $\mathbf{Q} = \mathbf{WA}$ (where **W** is the unmixing matrix found through BSS) should be a permutation matrix. By calculating how far **Q** deviates from being a permutation matrix it is possible to measure the effectiveness of the ICA algorithm under question.

Mutihac and Van Hulle [16] proposed the cross-talk error (CTE) as an index formeasuring ICA separation based on comparing **Q** to a permutation matrix. CTE is defined as:

$$CTE = \sum_{i=1}^{N}\left(\sum_{j=1}^{N}\frac{|Q_{ij}|}{\max|Q_i|} - 1\right) + \sum_{j=1}^{N}\left(\sum_{i=1}^{N}\frac{|Q_{ij}|}{\max|Q_j|} - 1\right) \tag{5.1}$$

CTE is used in to compare the effectiveness of FastICA and BS in performing blind signal separation.

A set of audio test signals were used in assessing the FastICA and BS algorithms. The signals are described and plotted in the time domain in Figure 5.1.

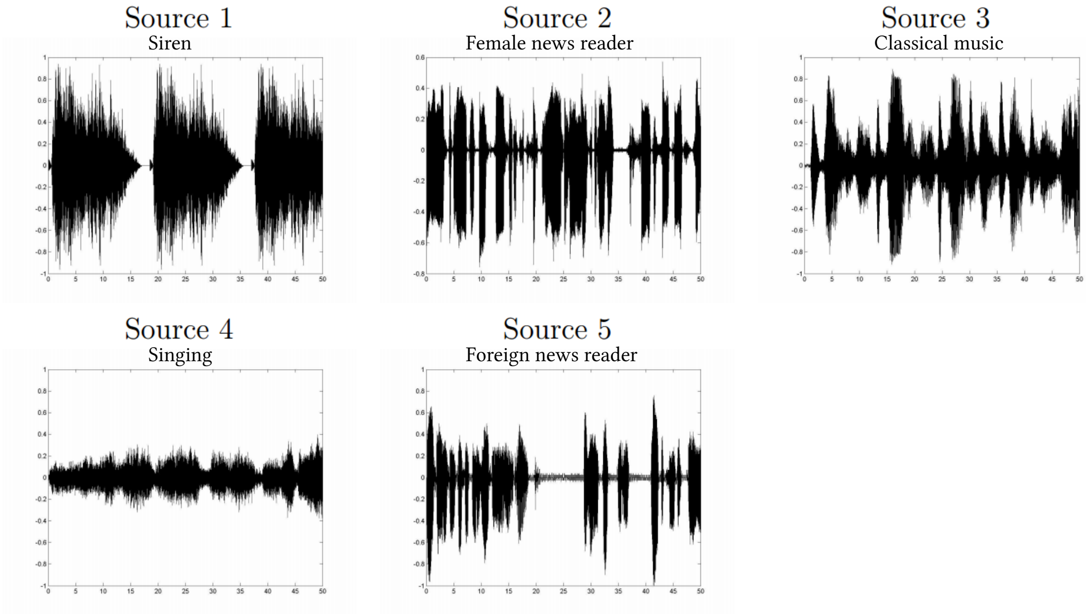


Figure 5.1: Time domain plots (5 seconds total) of instantaneous ICA test signals.

In order to reliably determine the separation effectiveness of FastICA and Bell and Sejnowski's maximum likelihood algorithm (henceforth referred to as the BS algorithm), separation was performed 20 times for each algorithm and the CTE results were averaged. In each case the signals of Figure 5.1 were mixed using a randomly generated 5×5 matrix. Figure 5.2 shows a typical set of mixtures.

Tables 6.2 and 6.3 summarise the results from applying FastICA and BS respectively to the test mixtures. In order to use the BS algorithm a non-linear output function must be specified. The BS algorithm is somewhat restricted in the nonlinear output functions that can be used since, as explained in Section 3.6.2, the algorithm requires that the function approximately match the CDF of the data being separated – a property sufficiently satisfied by the function $g(u) = tanh(u)$ which was used throughout this experimentation. There is, however, greater flexibility with the learning rate parameter, L, in the BS algorithm and the results of Table 6.3 compares the signal separation using BS with different learning rate parameters. Note that in order to make a meaningful comparison of algorithmic performance when varying learning rate, the number of iterations was kept constant. In each case, BS was performed with 500,000 iterations (or 10 complete passes through the data) which, based on experience, is sufficient to enable reliable convergence to a solution.

Non-linearity	Mean CTE
$g1(u) = tanh(u)$	0.274
$g2(u) = u^3$	0.326
$g3(u) = u \exp(-u^2/2)$	0.315

Table 5.1: Comparison of mean crosstalk error using FastICA with different non-linear functions. Results were obtained by averaging results from 20 runs of the algorithm.

Learning rate	Mean CTE
5.0×10^{-5}	0.565
1.0×10^{-4}	0.282
5.0×10^{-4}	0.463
1.0×10^{-3}	0.635

Table 5.2: Comparison of mean crosstalk error using different learning rates and with $g(u) = tanh(u)$ in each case. Results were averaged for 20 runs of BS for each learning rate.

From the results in Tables 5.1 and 5.2, it is clear that varying the non-linearity in FastICA and the learning rate in BS has a significant impact on the quality of signal separation. Unlike in BS there are no stringent requirements on the output function g that is used in FastICA, except that is must be non-quadratic. As a result, a variety of functions can be used with FastICA and the results from separating with FastICA with three of the most commonly used functions, $g(u) = tanh(u)$, $g(u) = u^3$ and $g(u) = u \exp(-u^2/2)$ are given in Table 5.2. In this case the best separation results (CTE of 0.274) were obtained using the hyperbolic tangent function.

The results for the BS algorithm show a strong dependence on the learning rate parameter. This is to be expected as L has a significant effect on the rate of convergence. With L too large we get overshoot and poor convergence, and with L too small it takes a long time to converge to a solution. From the results, we can see that an approximately optimal learning rate parameter was $L = 1.0 \times 10^{-4}$ for the separation problem considered in this section.

The results of this section demonstrated the effectiveness of ICA in separating instantaneously mixed signals and showed there is only a small difference in performance between FastICA and BS, with FastICA shown to perform slightly better.

5.2 Convolutive Mixtures

Convolutive ICA was performed using Davies' frequency domain approach. It is more difficult to quantify the performance of frequency domain ICA algorithms as we do not have a simple analogue of CTE for the convolutive case. The results presented here are qualitative and based on comparisons of spectrograms.

5.2.1 Artificially mixed signals

Two 6.25 second speech signals, sampled at 8kHz, were used as test signals for analysis of algorithm performance. The spectrograms of these signals are shown in figure 5.2. The signals of figure 5.2 were convolutively mixed using the following mixing filters:

$$\begin{aligned}
A_{11}(z) &= 1 - 0.4z^{-25} - 0.2z^{-45} \\
A_{12}(z) &= 0.4z^{-20} - 0.2z^{-28} + 0.1z^{-36} \\
A_{21}(z) &= 0.3z^{-10} + 0.3z^{-22} + 0.1z^{-34} \\
A_{22}(z) &= 1 - 0.3z^{-20} + 0.2z^{-38} \\
& 1
\end{aligned} \tag{5.2}$$

The maximum delay introduced by these filters was due to the z^{-45} filter tap, which at 8kHz is equivalent to 5.625ms. This delay is imperceptible to the ear, but is significant when performing separation as the mixing process can no longer be modeled as matrix multiplication, as in Eq.(3.1).

(a) (b)

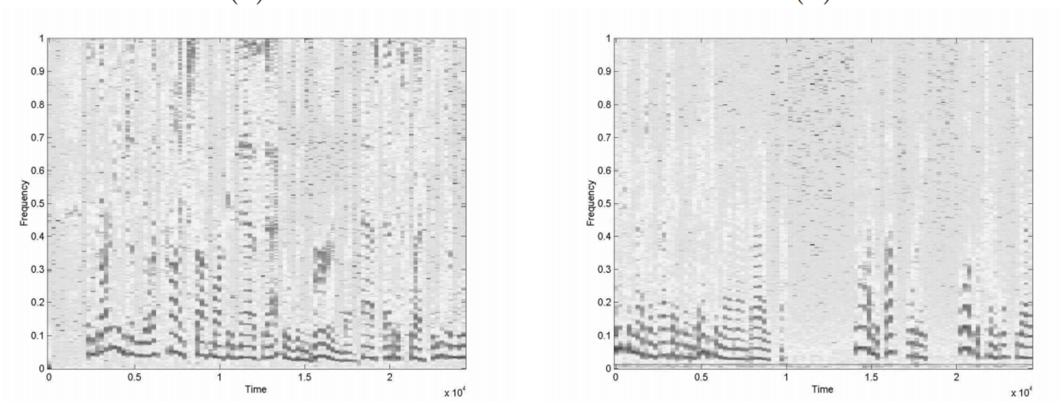


Figure 5.2: Spectrograms of original source signals used in algorithm analysis. Source (a) is an American male newsreader, and (b) is a female voice speaking in a foreign language.

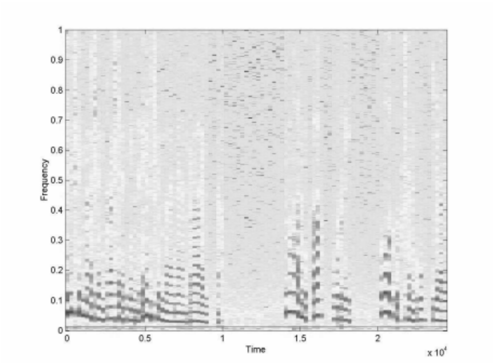
The effect of the non-linear function in FastICA has already been considered earlier in this chapter, and for the purpose of this section, we will assume that $g(u) = tanh(u)$ (a routine assumption in the literature). Another parameter is the number of iterations, the number of iterations will be held constant at 100 (which, based on experience, is ample to allow convergence).

Comparing Instantaneous and Convolutive ICA for Artificial Mixture

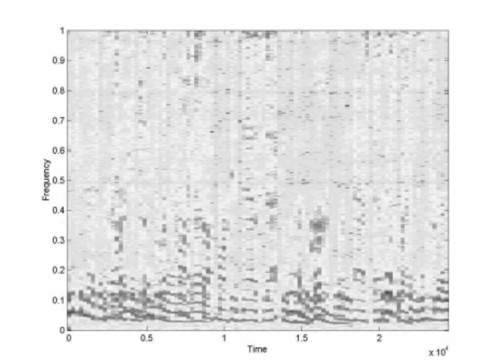
The results of separation are shown in the spectrograms of figure 5.3. FastICA yielded poor signal separation (as expected since instantaneous ICA cannot handle time delays due to convolutive mixing) and on listening to the output signals of the FastICA algorithm, significant crosstalk remained. On the other hand, with Davies' convolutive ICA algorithm good separation was achieved and crosstalk was almost inaudible.

Figure 5.3 confirms that the signal estimate using convolutive ICA matches the original source signal much more closely than the estimate using FastICA. The spectrogram in Figure 5.3.(c), obtained using Davies' algorithm, is a superior estimate of the original source. In figure 5.3.(b) there is greater signal due to interference at high-frequency and between 1.0s and 1.5s when there is a lull in the source signal.

(a) Source signal



(b) Estimate using instantaneous ICA



(c) Estimate using convolutive ICA

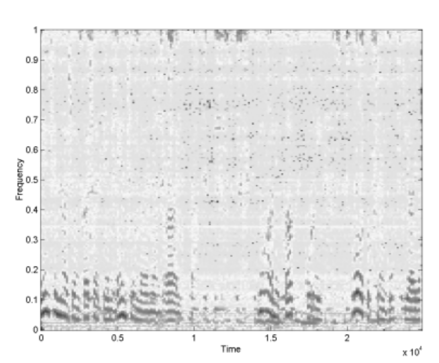


Figure 5.3: Comparison of separation of convolutively mixed signals. The spectrogram of the original source is shown in (a), and the spectrograms in (b) and (c) correspond to instantaneous and convolutive ICA respectively.

5.2.2 Separation of Signals in a Real Acoustic Environment

In this section, speech mixtures recorded in a real acoustic environment are processed using BSS techniques. The primary goal is to demonstrate, from experimental evidence, that such signals can, for the purposes of BSS, be considered convolutive mixtures.

The speech mixture that was used to evaluate separation performance on real acoustic mixtures was a recording of two males speaking simultaneously in a regularsized office. The sample rate is 16kHz, the speakers were approximately 60cm from each other and both counted from one to ten, one in English and the other in Spanish. In this situation the original (i.e. unmixed) signals are not available for comparison. As a result, to gauge the level of separation, spectrograms of the unprocessed recordings are compared to the output signals from FastICA (instantaneous ICA) and Davies' algorithm (convolutive ICA) in figure 5.4. Both FastICA and Davies' algorithm were used with the optimal parameters: the non-linear function was $g(u) = tanh(u)$, and using 100 iterations, 512 point Hanning windows and 50 % window overlap.

The results are shown in figure 5.4. The spectrogram in figure 5.4.(a) shows one of the two original recordings in which both voices were clearly audible. Figure 5.4.(b) is very similar to the original mixture of Figure 5.4.(a) which illustrates the poor performance of FastICA in this case. The spectrograms in figure 5.4.(c) and figure 5.4.(d) show the two signals resulting from convolutive ICA. It can be seen that the spectrogram in part (a) of the figure has been decomposed into parts (c) and (d) so that superpositioning (c) and (d) would give a good approximation to the original signal. This pictorially illustrates the separation of the original mixture in figure 5.4.(a) into the two individual voices. These results indicate that instantaneous ICA is ineffective for separating audio mixtures recorded in a real acoustic environment, whereas convolutive ICA offers an effective approach which provides experimental verification that real acoustic mixtures are convolutive in nature.

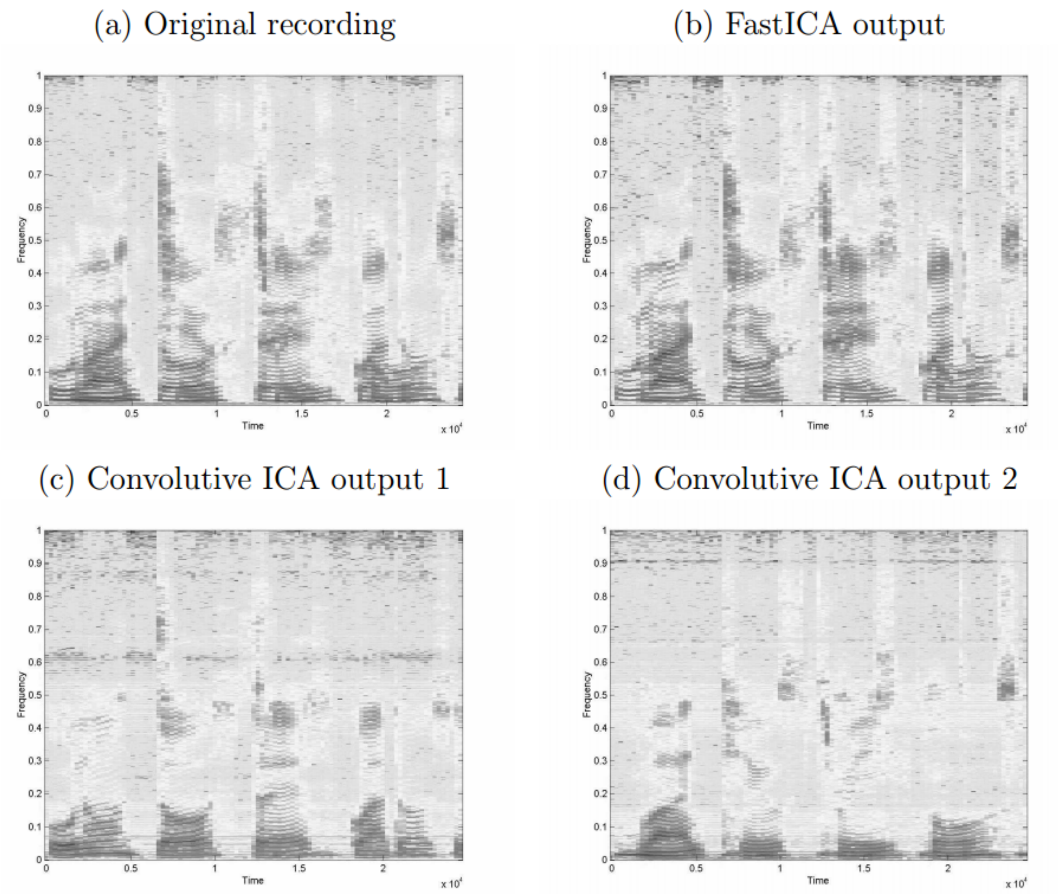


Figure 5.4: Comparison of spectrograms for real audio mixtures, (a) shows the spectrogram of one of the two original recordings in which both voices are clearly audible, (b) shows one of the spectrograms obtained using instantaneous ICA, and (c) and (d) show the two signals resulting from convolutive ICA.

5.3 conclusion

The performance of instantaneous and convolutive ICA was compared in separating both an artificially generated convolutive mixture (using known mixing filters) and a speech mixture recorded in a real acoustic environment. In each case, convolutive ICA was shown to perform significantly better. Furthermore, the ineffectiveness of instantaneous ICA in separating the real acoustic mixture provided experimental verification of the well established result that mixtures recorded in a real acoustic environment can be accurately modeled as convolutive. The efficacy of convolutive ICA in separating real acoustic mixtures suggests a number of practical applications of convolutive ICA, such as in improving hearing aid technology or in voice recognition systems.

Conclusion

This thesis has examined the topic of Blind Source Separation, focusing on both the mathematical foundations of BSS techniques as well as on some important signal processing applications.

The general mathematical framework for both instantaneous and convolutive ICA was developed in the early chapters. As well as discussing the signal mixing models and conceptual background of various algorithms, the underlying assumptions and limitations of BSS techniques were examined in some detail. The most important application of BSS considered in this thesis was acoustic signal processing.

Experimental results from instantaneous and convolutive signal separation problems were given in Chapter 5. The FastICA and BS instantaneous ICA algorithms were shown to be comparable in separating instantaneous audio mixtures (with FastICA performing slightly better). The performance of convolutive ICA in separating convolutive audio mixtures was examined by comparing separation performance using different parameters. Also, by processing convolutive audio mixtures using both instantaneous and convolutive ICA algorithms it was shown that instantaneous ICA is insufficient for sound sources. In particular, it was shown that convolutive ICA is required to separate audio mixtures recorded in real acoustic environments (i.e. the real-world “Cocktail Party Problem”).

Bibliography

[1] C. Jutten and J. Herault. "space or time adaptive signal processing by neural network models". In *AIP Conf. Proc. 151 on Neural Networks for Computing*, pages 206–211, 1986.

[2] P. Comon. Independent component analysis, a new concept? *Signal Process.*, 36(3):287–314, April 1994.

[3] K. I. Diamantaras and S. Y. Kung. *Principal Component Neural Networks: Theory and Applications*. Wiley-Interscience, 1 edition, 1996.

[4] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, 7(6):1129–1159, November 1995.

[5] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Proceedings of the 8th International Conference on Neural Information Processing Systems*, NIPS'95, pages 757–763, Cambridge, MA, USA, 1995. MIT Press.

[6] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Comput.*, 9(7):1483–1492, October 1997.

[7] K. Torkkola. Blind separation of convolved sources based on information maximization. In *Proceedings of the 1996 IEEE Signal Processing Society Workshop*, pages 423—-32, 1996.

[8] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, 22(1–3):21 – 34, 1998.

[9] R. Togneri S. Y. Low and S. Nordholm. Spatio-temporal processing for distant speech recognition. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages I–1001–4 vol.1, May 2004.

[10] J. Anemüller. Adaptive separation of acoustic sources for anechoic conditions: A constrained frequency domain approach. *Speech Commun.*, 39(1-2):79–95, January 2003.

[11] Albert Leon-Garcia. *Probability and Random Processes for Electrical Engineering (2nd Edition).* Addison-Wesley, 2 edition, 1993.

[12] E. Oja A. Hyvärinen, J. Karhunen. *Independent Component Analysis.* J. Wiley, 1 edition, 2001.

[13] S. Gazor and Wei Zhang. Speech probability distribution. *IEEE Signal Processing Letters*, 10(7):204–207, July 2003.

[14] Aapo Hyvärinen. Ica of complex valued signals: A fast and robust deflationary algorithm. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00)-Volume 3 - Volume 3*, IJCNN '00, pages 3357–, Washington, DC, USA, 2000. IEEE Computer Society.

[15] N. Mitianoudis and M. E. Davies. Audio source separation of convolutive mixtures. *IEEE Transactions on Speech and Audio Processing*, 11(5):489–497, Sept 2003.

[16] R. Mutihac and M. M. Van Hulle. Neural network implementations of independent component analysis. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 505–514, 2002.