E.E. GÜLER

ATILIM UNIVERSITY

2022

EXAMINATION OF INDEPENDENT COMPONENT ANALYSIS IN AUDIO SOURCE SEPARATION

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

ELİF EZGİ GÜLER

A MASTER OF SCIENCE THESIS

IN

THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

JUNE 2022

EXAMINATION OF INDEPENDENT COMPONENT ANALYSIS IN AUDIO SOURCE SEPARATION

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

ELİF EZGİ GÜLER

A MASTER OF SCIENCE THESIS

IN

THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

JUNE 2022

Approval of the Graduate School of Natural and Applied Sciences, Atilim University.

Prof.Dr. Ender KESKİNKILIÇ
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of **Master of Science in Electrical and Electronics Engineering, Atilim University**.

Assoc.Prof.Dr. Kemal Efe ESELLER
Head of Department

This is to certify that we have read the thesis EXAMINATION OF INDEPENDENT COMPONENT ANALYSIS IN AUDIO SOURCE SEPARATION submitted by Elif Ezgi GÜLER and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. İbrahim Baran USLU
Advisor

Examining Committee Members:

Asst. Prof. Dr. Emre SÜMER
Computer Engineering, Başkent University

Asst. Prof. Dr. İbrahim Baran USLU
Electrical and Electronics Engineering, Atilim University

Asst. Prof. Dr. Yaser DALVEREN
Electrical and Electronics Engineering, Atilim University

Date: *10/06/2022*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Elif Ezgi Güler

Signature :

ABSTRACT

EXAMINATION OF INDEPENDENT COMPONENT ANALYSIS IN AUDIO SOURCE SEPARATION

Güler, Elif Ezgi

MSc., Department of Electrical and Electronics Engineering

Supervisor: Asst. Prof. Dr. İbrahim Baran USLU

June 2022, 53 pages

In this thesis, we examine the Independent Component Analysis (ICA) method in audio source separation. This method is a type of blind source separation where the sources observed in the mixture signals are unknown. We try to solve a cocktail party problem, by extract the independent signals which are mixed by an unknown mixing matrix. There are some sub-types of the ICA algorithm such as Gradient Ascent (ICA-GA), fastICA and Kernel-ICA. In this work, we study on ICA-GA algorithm. For this purpose, different scenarios where two or three audio sources are mixed with each other, are examined. In some of the tests carried out, we separated voice and noise signals clearly from each other. In other tests, voice signals were separated. In the experiments, we focused on the η (step-size) and the maximum iteration number parameters, also examined the value of parameters on performance of ICA-GA algorithm. We obtained that, ICA method is quiet successful in blind source separation. It was concluded that increasing the value of the maximum iteration parameter alone is not a sufficient parameter for performance. Because as the maximum number of iterations increased, the running time of the algorithm also too increased, that is, the elapsed time is not at the optimum value. We can say that, increasing the value of the step-size parameter alone has more successful results on the performance of algorithm than increasing the value of the maximum iteration parameter alone. The study recommends a solution to the ICA's ambiguity about order of output signals by using the correlation values of each source signal and each output signal.

Keywords: Blind source separation, cocktail party problem, independent component analysis, speech enhancement, noise cancellation, audio separation.

ÖZ

SES KAYNAK AYRIŞTIRMASINDA BAĞIMSIZ BİLEŞEN ANALİZİ YÖNTEMİNİN İNCELENMESİ

Güler, Elif Ezgi
Yüksek Lisans, Elekrik Elektronik Mühendisliği Bölümü
Tez Yöneticisi : Dr.Öğr.Üyesi İbrahim Baran USLU
Haziran 2022, 53 sayfa

Çalışmada, Ses kaynağı ayrıştırmada Bağımsız Bileşen Analizi metodu incelenmiştir. Bu yöntem, karışım sinyallerinde gözlenen kaynakların bilinmediği bir tür kör kaynak ayırma yöntemidir. Bilinmeyen bir karıştırma matrisi tarafından karıştırılan bağımsız sinyalleri çıkararak bir kokteyl partisi problemini çözmeye çalışıyoruz. ICA algoritmasının Gradient Ascent (ICA-GA), fastICA ve Kernel-ICA gibi bazı alt türleri vardır. Bu çalışmada ICA-GA algoritması üzerinde çalışıyoruz. Bu amaçla iki veya üç ses kaynağının birbirine karıştırıldığı farklı senaryolar incelenmiştir. Yapılan bazı testlerde ses ve gürültü sinyallerini net bir şekilde birbirinden ayırdık. Diğer testlerde ses sinyalleri ayrıldı. Deneylerde η (adım-boyutu) ve maksimum iterasyon sayısı parametreleri üzerinde duruldu, ayrıca parametrelerin ICA-GA algoritmasının performansı üzerindeki değeri de incelenmiştir. Kör kaynak ayırmada ICA yönteminin oldukça başarılı olduğunu elde ettik. Maksimum iterasyon parametresinin değerinin tek başına arttırılmasının performans için yeterli bir parametre olmadığı sonucuna varılmıştır. Çünkü maksimum iterasyon sayısı arttıkça algoritmanın çalışma süresi de arttığından geçen süre optimum değerde değildir. Tek başına adım büyüklüğü parametresinin değerini artırmanın algoritmanın performansı üzerinde maksimum yineleme parametresinin değerini tek başına artırmaya göre daha başarılı sonuçlar verdiğini söyleyebiliriz. Çalışma, her bir kaynak sinyalinin ve her bir çıkış sinyalinin korelasyon değerlerini kullanarak, ICA'nın çıkış sinyallerinin sırası hakkındaki belirsizliğine bir çözüm önermektedir.

Anahtar Kelimeler: Kör kaynak ayrıştırma, kokteyl parti problemi, bağımsız bileşen analizi, ses iyileştirme, gürültü iptali, ses ayrıştırma.

To my mother…

ACKNOWLEDGEMENTS

I would like to thank my advisor, Asst. Prof. Dr. İbrahim Baran USLU, for his always motivating, supportive, kind and understanding behavior. Moreover, I would like to thank to the jury members for their time and contributions.

I especially commemorate my mother Emine EZER with love, respect and longing for all her contributions to my education and my whole life.

I would also like to thank my father Bülent ALGEN for the importance he gave to education and the support he gave to my entire education life.

I would also like to thank my beautiful little sister Bengisu ALGEN whom I love as much as my daughter for inspiring every good thing I have done since the day she was born.

I would like to express my endless love and thanks to my dear husband, Harun GÜLER, who has supported my education life with patience and devotion since the day we met.

I would also like to thank all my extended family for their moral support and interest to me and my sister in every situation.

Last but not least, I would like to thank Assoc. Prof. Dr. Bilge ÖNAL DÖLEK, who always motivated me throughout this process, for her contributions to my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ ABBREVIATIONS

η (eta)	step-size
∇h	Change in Entropy
H	Entropy

BSS	Blind Source Separation
cdf	Cumulative Density Function
ICA	Independent Component Analysis
ICA-GA	Independent Component Analysis Gradient Ascent
LOG-MMSE	Log-Spectral Minimum Mean-Square Error
MOS	Mean Opinion Score
PESQ	Perceptual Evaluation of Speech Quality
pdf	Probability Density Function
SNR	Signal to Noise Ratio
TSE	Total Square Error

CHAPTER 1

INTRODUCTION

The fastest and most effective method for communication, which is the most basic need of human beings, is voice communication. Today, apart from face-to-face communication, there are many technological tools to establish voice communication. The basic operating component of these technological tools is signals. Signal processing is a technology which making possible for some applications like generation, transformation, subtraction of information via electronic signals is an obligatory field whose mathematical foundations date back to the 17th century. The development of electronic computers and their proliferation from the second half of the 20th century gave rise to Digital Signal Processing (DSP) that has allowed us to apply these techniques to a wide range of problems in today's world [1].

At this study, we focus on audio signal that is the source for some of the main applications of these still developing technologies. Audio Processing consists in the study of sounds recorded by one or more microphones and includes multiple objectives such as analysis of the sounds for its characterization, coding for transmission or storage through digital media, enhancement of sound, processing of musical signals etc.[1][2].

When there is not any or enough information about sources, the process for separation of sources is called as Blind Sources Separation (BSS) and in this situation, Audio Source Separation methods are not always enough for separating the components clearly. One of the most effective, easy to use and open to improving algorithm for solving the Blind Source Separation is Independent Component Analysis (ICA).

In real life, there are many environments where sounds, noises and speech are mixed, such as a video conference or a presentation room. There may be a lack of meaning or incomprehension in the sounds that are mixed with each other. In such environments, it is important to separate sound sources for both humans and machines and to draw conclusions from independent sources in order to reach the necessary information.

1.1. Scope

In this study, Independent Component Analysis-Gradient Ascent (ICA-GA) method applied for separation two and three audio sources mixture. For both two and three sources, many scenarios which includes noises, male and female speeches, music sounds have been tried. The sounds used in this study have been taken from MATLAB sound database, NOISEX-92 database and some old movies, TV shows or radio programmes. Noises that used in the study; bell sound, bird sound, factory sound and music. The female voices that used belong to the actress Türkan Şoray and Atatürk's adopted daughter Ülkü Çukurluoğlu. The male voices belong to actor Kemal Sunal and professor doctor İlber Ortaylı. And the sampling frequencies of these used sounds are 22050 and 44100 Hertz. The sampling frequencies of the sounds in the same scenario were equalized using the Praat program.

The study has been focused on the separation two and three sources with ICA-GA, examining the parameters which are the most critical on the performance of the method. It is predicted that even changing some parameters within a certain range will be effective in the performance of the ICA-GA method. Therefore, while the method is applying to separate mixed sounds, it is aimed to achieve much better results by making minor changes in some parameter values. And while doing this, also it is aimed to keep the working speed of the algorithm at the optimal range.

Also in the study, a proposal has been made about one of the ambiguity in Independent Component Analysis applications which is about detecting the order of the independent components.

1.2. Motivation

Blind Source Separation is a shape of Audio Source Separation appeared when there is not exist any information or there exists a little information about sources of signals mixtures. This situation called as “Cocktail Party Problem”. For this reason, it is more difficult to solve but there exist many successful methods developed for this subject. One of the firstly think of blind source separation method to solve cocktail party problem is Independent Component Analysis. If studied based on maximum entropy with ICA, this is called as Gradient Ascent method and this method used for application of ICA algorithm depends on many parameters. The basic scenario for cocktail party problem is that; in an environment with at least two sound sources, there should be microphones in separate locations that record environmental sound and the number of microphones must be equal to the number of sources. The simplest scenario for cocktail party problem is shown in Figure 1.1. Here the sources can be music, speech, effects, noises, etc.

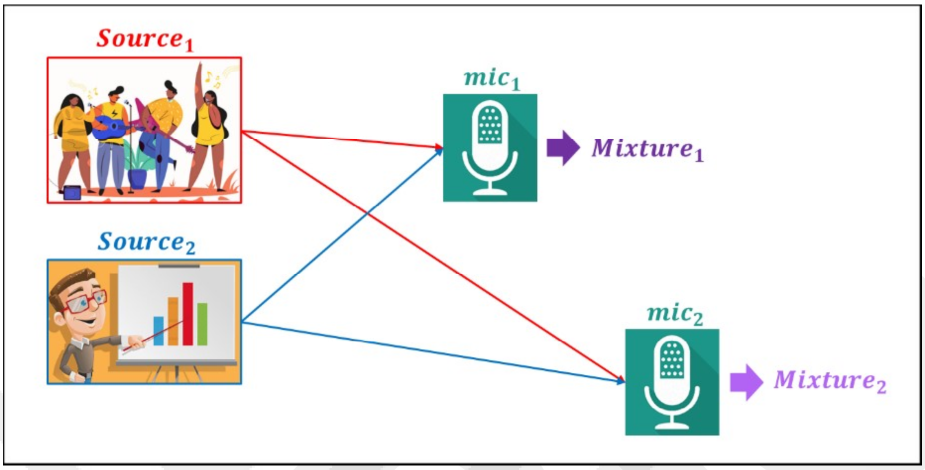


Figure 1.1 Cocktail Party Problem Scenario

In literature, there are many study about both ICA method in several fields and also in field that audio source separation. The study aimed to contribute to literature that examination of how successful and effective ICA-GA method for audio source separation and can the achievement of the method be increased or not. Basic scenario of ICA method shown in Figure 1.2.

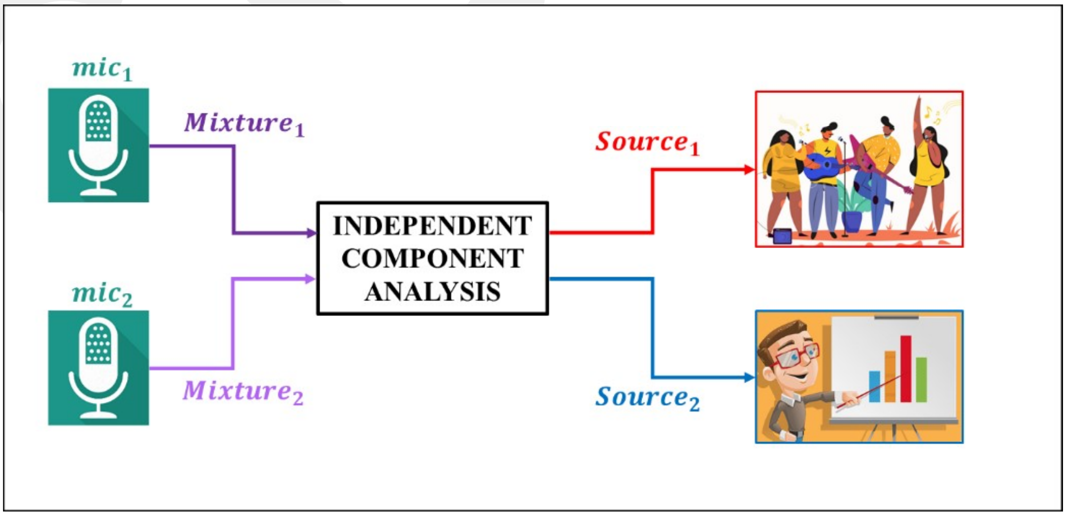


Figure 1.2 Independent Component Analysis

1.3. Definitions

In this part, some concepts that needed to known in the study are explained in this part. One of them is “statistical independence”. Statistical independence is the independency of probability density functions (pdf) of variables.

The “cumulative density function” (cdf) is a method for describing the distribution of random variables which contains the probabilities that X is less than or equal to X. It denoted by g.

$$F_X(x) = g(x) = P(X \leq x) \tag{1.1}$$

This should be noted; uncorrelated has not the same meaning with independence. If two random variables are uncorrelated, this means that their correlation coefficient (ρ) is zero, namely they have zero covariance. And it is expressed as;

$$\rho(X, Y) = 0,$$

$$cov[X, Y] = 0 \tag{1.2}$$

Another concept is “Gaussian distribution”; is the most frequently encountered distribution in nature. It is also called as normal distribution.

“Gradient ascent” is maximizing of the function for obtain the better optimization. In independent component analysis, gradient ascent method works maximizing the entropy.

SNR or S/N is the “signal to noise ratio” that obtained with dividing the signal we want to measure the noise level by the noise affecting this signal. SNR value can be negative, zero or positive and its expressed in decibels (dB) It is would that the value be as high as possible, because its means that the noise is well cleaned up from the signal.

Additionally, in this study, some definitions which mostly encountered in all studies on this field are explained. While explaining the definitions, it has been tried to proceed from the simple to the complex. In this study, the definitions are discussed in order of as follows; in its simplest form, the concepts of “signal” and “signal processing”, “audio signal processing” and “speech enhancement” as a sub-application area of audio signal processing, “noise types” and some “noise-speech databases”, “Audio

Source Separation", "Blind Source Separation" and "Cocktail Party Problem" and finally "Independent Component Analysis-Gradient Ascent Method".

In the literature review, there are many useful studies about the ICA method which is one of the most preferred methods for blind source separation. These studies contain a lot of information such as the purpose of the method, its mathematical history, areas of use, and performance values. However, it is not mentioned about the parameters that are most effective in the performance of the ICA method and whether these parameters must have an optimum value range or these must constant for all sources. In this study, firstly, the most effective parameters were determined for the method to give the best performance. Then it was decided whether these parameters should be constant or variable for different sources. And it was concluded that parameter values should be taken at different optimum values for different sources. Finally, the optimum values of each parameter were found in different scenarios and thus the obtained result was proved. The innovative aspect of the study is, together with the method the examination of the most effective parameters and to determine what should be their values for the best performance.

CHAPTER 2

LITERATURE REVIEW

2.1. Speech Enhancement

Signal defines a wave which carries information or an electrical impulse. Signal Processing scopes any kind of changes on a signal such as modification, synthesizing or analysis [3]. If the used signal is speech signal, then the application becomes speech processing. Namely, speech processing is a subtitle of signal processing. Speech signal is usually sampled by 16-44 kHz. One of the most popular and required applications field of speech processing is speech enhancement applications.

In speech enhancement applications, to enhance the quality and intelligibility of speech degraded by noise is the object [4]. These applications provides increasing the ease of listening and the amount of information [5]. Speech enhancement algorithms reduce or suppress the background noise to some extent. To better understand the effects of noise on the speech signal, needed to know better the types of noise. Basic process for speech enhancement is shown in Figure 2.1.

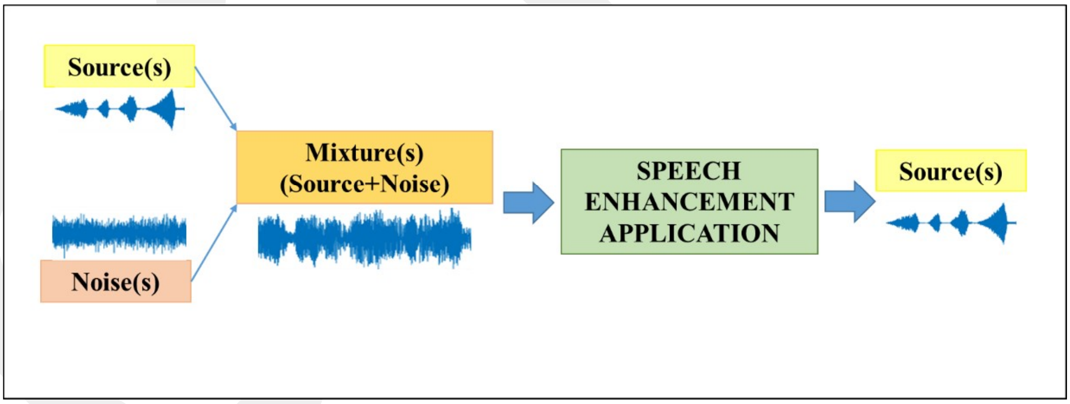


Figure 2.1 Basic Speech Enhancement Process

Numerous theses and articles have been published recently about the subject. Studies that contributed to the literature were also references for this study. The results and methods of the studies on the subject between the years 2010-2022 were examined. Since it is not possible to examine all of them, the most current studies in the field have been reviewed.

In 2010 studied on a paper by M. Shujau, C. H. Ritz and I. S. Burnett. The study measured the performance of ICA algorithm in speech enhancement for AVS (Acoustic Vector Sensor) recordings. And the results evaluated by PESQ (Perceptual Evaluation of Speech Quality). In order to measuring performance of method, compared using simulated and real recordings from various types of microphone arrays. In study, some female and male sentences used and taken from IEEE speech corpus. The lengths of taken sentences are 10 s periods and each period have 1 s of silence at the start and at the end. The lengths of noises are equal to length of sentences and include the recording of babble, factory, vehicle, white and pink noises. These sentences and the noises were used as the test database. Two scenarios for sources consisted of one source-one interferer and one source-diffuse noise (synthesized using four interferers). Anechoic recordings were processed using FastICA while reverberant recordings were processed using a convolutive FastICA algorithm. The results based on experiments show a significant improvement in speech enhancement using ICA applied to an AVS compared to ICA applied to a traditional linear microphone array, in both anechoic and reverberant environments [6].

Another study published in 2010 and prepared by Li Hongyan and Ren Guanglong, proposed to independent component analysis when the measured signals are contaminated by additive noise. Two speech signals were used for simulations, first single channel ICA speech enhancement algorithm has been used to de-noise for each mixed noisy speech and then the FASTICA algorithm has been used to separate the de-noised speech signals. The performance of study measured by SNR (signal to noise ratio) values of results [7].

Another study in this field which prepared by K. Mohanaprasad and P. Arulmozhivarman. The contribution combining two ICA algorithm shown that the method can renew the original speech signals effectively. This work was designed for compare to Gradient based negentropy algorithm and Fast ICA based negentropy algorithm efficiency in separating speech signals. Results of applied with using MATLAB shown that Fast ICA needs less execution time as compared to gradient based negentropy with minimum number of iterations [8].

The study was included in the speech enhancement studies in the literature in 2012 published by R. İrem Bor. A real-time implementable noise cancellation algorithm is proposed with separating speech and noise signals. In order to separation part a combination of independent component analysis (ICA) and particle swarm optimization (PSO) algorithms is used. To overcome the ambiguity of ICA that it is not possible to know which one of the separated signals is speech or noise, a pitch extraction (PE) algorithm is developed and combined with ICA-PSO. The ICA-PSO-PE algorithm is implemented in MATLAB. Before application signals are mixed and provided in frames of 40 ms. In order to fasten the process pre-processing steps for ICA, except centering, were not implemented. And a learning period is introduced for increasing accuracy of separation. According to the experiments ICA-PSO-PE is a robust way to real-time noise cancellation algorithm in the sense that it is computationally efficient, accurately extracts speech signal from its mixtures, even with very low SNR levels. Lastly the proposed algorithm is compared with FastICA and the subtraction method. The algorithm outperforms FastICA in the sense of real-time implementability and outperforms subtraction method in the sense of robustness [9].

One of the studies that have recently joined the literature is published by Alyaa Mahdi in 2017. The aim of study is compare the performance of three popular Independent Component analysis (ICA) and Independent Vector Analysis (IVA) algorithms to solve BSS problems. The used algorithms are Fast-ICA, Kernel-ICA and Fast-IVA. The performances of algorithms were compared with Source-to-Artifact Ratio, Source-to- Distortion Ratio, Source-to-Noise Ratio metrics. As a major result of implementation that Fast – IVA method has better performance than Fast-ICA and Kernel-ICA [10].

The study published in 2010 by Cecelioğlu S. proposed, the performance of a short-time noise reduction method based on a peak picking algorithm has been analyzed for noisy speech signals. The noisy speech signals have been degraded by additive white Gaussian noise at different signal-to-noise ratios (SNR) and different types of additive background noise. MATLAB simulation program was used for application. The HRNR method gives better results in preserving the harmonics of speech compared to the TSNR method [11].

One of the useful studies in the literature was published by F. Alim in 2011 and aimed at real-time speech enhancement. The developed algorithm tested on several combinations from a speech and noise database which consist of Atatürk's address to the Turkish youth, several news contents, a speech related to old ages, soccer games with heavy “Vuvuzela” noise, sounds of pilots in an airplane cockpit, white noise, pink noise, speech noise, restaurant ambiance noise, and airport noise. In order to obtain the best performance in real time, two integrated algorithms including speech enhancement algorithm, newly developed fusion noise estimation algorithm and other signal processing techniques were developed and their effectiveness were compared. The used main algorithms during the study are; Log-spectral amplitude estimator (LOG-MMSE) and Wiener-SNR algorithms. Wiener-SNR algorithm together with recursive-averaging noise estimation algorithm has produced the best results among others for all noise types that are experimented with [4].

2.1.1. Speech Enhancement Techniques

Speech enhancement techniques can be defined by two basic categories, these are “single channel” and “multiple channels (array processing)”. Single channel locution means source acquired from single microphone and multi-channel means source acquired from multiple microphone sources.

The three techniques of single-channel speech enhancement can be listed as; Spectral Subtraction (SS) Method, Spectral Subtraction with Over subtraction Model (SSOM) and Non-Linear Spectral Subtraction (NSS) [12].

Spectral subtraction method is very simple and easy method in terms of the implementation, it based on the principle that obtained an estimate of the original signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. SSOM procedure applies in order to reduce percept of the musical noise effect. This Method consists of subtracting an overestimate of the noise power spectrum and presenting the resultant spectral components from going below a preset minimum spectral floor value. Finally, NSS is based on connecting two different ideas that the use of an extended noise and an over subtraction model and non-linear implementation of the subtraction process, taking into account that the subtraction process must depend on the SNR of the frame, in order to apply less subtraction with high SNRs or vice versa. [13][14].

Multi-channel enhancement techniques can be also listed as; Adaptive Noise Cancellation and Multisensor Beamforming. Adaptive noise cancellation is an effective and popular method for multi-channel speech enhancement and briefly, based in the availability of an auxiliary channel, known as reference path, where a correlated sample or reference of the contaminating noise is present. Beamforming is a multiple-input and single-output (MISO) application and consists of multi-channel advanced multidimensional filtering techniques that enhance the desired signal as well as suppress the noise signal [12].

On the other hand; when said speech enhancement, removal of noise from speech signal is what one of the firstly think of. So techniques are generally focus on noise removal. Noise which is distorts the speech signal, can be periodic noise, wide band noise, and interfering speech [14]. Speech enhancement techniques according to noise types briefly explained below.

The periodic noise can be removed using three basic filters; stationary filters, adaptive filters, or transform domain filters. Spectral Subtraction method (SS) and adaptive cancellation can be used for the wide band noise. In spectral subtraction method, estimated noise spectrum is subtracted from the spectrum of the noisy speech. The noise correlated with signal can be removed using adaptive cancellation. Speech enhancement techniques are not useful when two speech signals are interfering. If the different pitches can be identified, then the voices of different speakers can be isolated. Speech enhancement techniques are not useful when two speech signals are interfering. If the different pitches can be identified, then the voices of different speakers can be isolated [14].

Speech enhancement techniques for interfering speech called Audio Source Separation. The audio source separation will be explained in more detail as a sub-title in the following sections. In an audio source separation application, when there exist blind sources, need to some special Blind Source Separation Techniques. The blind here means that the mixed components are latent or unobservable, actually only the mixed signals are observable [15].

Methods for Blind Source Separation can be listed as; Projection Pursuit, Complexity Pursuit, Gradient Ascent, Principle Component Analysis and Factor Analysis, Independent Component Analysis [16]. Several approaches have been proposed for the Blind Source Separation and development of approaches is currently in progress. One of the most successful and open to development method is Independent Component Analysis. It is also the main subject of this study, so with its application results will be discussed in detail.

2.1.2. Noise Types

If we want to define noise in its simplest form, we can say that it's any background audio that's surplus to requirements. And since this is undesirable, we mostly want to eliminate the noise.

Some environmental noise examples in daily life;

- Aircrafts,
- Alarms,
- Animals,
- Construction sites,
- Exhibition hall,
- Factories,
- Music and parties,
- Rail systems,
- Traffic,
- Vehicles and vehicle alarms etc.

In signal processing, noise is included the unwanted and unknown modifications that a signal may suffer during capture, storage, transmission, processing, or conversion. Also the noise term in signal processing is used to mean signals that are random and carry no useful information [17][18].

Signal processing noise can be classified by its statistical properties and by how it modifies the intended signal. Statistical properties of noise generally can be called as color of noise [17]. The most popular ones in studies are white noise, pink noise, brown noise and Gaussian noise.

White noise is random noise, it has a flat spectral density and also includes all frequencies. It means, the white noise has the same amplitude, or intensity, throughout the audible frequency range (20 to 20,000 Hz). Since it includes all audible frequencies, white noise can masked other sounds [19]. Gaussian noise is a statistical noise which has probability density function (pdf) equal to the normal distribution. As known, the normal distribution also called as Gaussian distribution. In other words, the values that the noise can take on are Gaussian-distributed. Additive White Gaussian Noise (AWGN) is as the name suggest, additive noise to another audio signal that has the characteristics of both white and gaussian noise.

2.1.3. Speech and Noise Databases

There are some speech databases which includes different sound recordings from people of different genders and several languages, additionally there are some noise databases which includes readily available for the most common types of noise used in speech for studies and facilitate studies.

TED-LIUM and TED-LIUMv2 is the English speech recognition training corpus from only English language TED talks, with transcriptions, sampled at 16 kHz. It contains about 118 hours of speech [20][21]. LibriSpeech is another corpus of approximately 1000 hours of 16 kHz read English speech, prepared by Vassil Panayotov with the assistance of Daniel Povey. The data is derived from read audiobooks, and has been carefully segmented and aligned [20][22]. RWCP Sound Scene Database is also another speech database. The data includes non-speech sounds recorded in an anechoic room, reconstructed signals in various rooms, impulse responses for a microphone array, speech data recorded with the same array, and recordings of background noises. It is intended for use when simulating sound scenes. It was developed by the Real Acoustic Environments Working Group of the Real World Computing Partnership (RWCP). The data was recorded from 1998 to 2000 [20][23]. The AMI Meeting Corpus consists of 100 hours of meeting recordings. The recordings use a range of signals synchronized to a common timeline. These include close-talking and far-field microphones, individual and room-view video cameras, and output from a slide projector and an electronic whiteboard. During the meetings, the participants also have unsynchronized pens available to them that record what is written. The meetings were recorded in English using three different rooms with different acoustic properties, and include mostly non-native speakers [20][24][25]. MUSAN is also a corpus of music, speech, and noise recordings. This work was supported by the National Science Foundation Graduate Research Fellowship [20][26]. The common aspect of all speech databases is that they contain data in English language.

NOIZEUS is a noisy speech corpus for evaluation of speech enhancement algorithms. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by 8 different real-world noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train-station noise. This database is available to free of charge [27]. NOISEX-92 is another database of recording of various noises on 2 CDROMs available free. It includes these noise samples; Voice babble, Factory noise, HF radio channel noise, pink noise, white noise, Various military noises; fighter jets (Buccaneer, F16), destroyer noises (engine room, operations room), tank noise (Leopard, M109), machine gun and Volvo 340 [28].

The common property of all databases both speech and noise is; free access.

2.2. Audio Source Separation

The term “audio source” is refer to an individual physical source or to an entity that humans perceive individually. While humans have advanced skills in “hearing out” individual sources from complex mixtures even in noisy conditions; the computer-based modeling of this process is a very difficult problem [29].

The process of estimating the individual sources from the mixture signal is called sound source separation. Audio source separation has many applications including music transcription, remixing, chord estimation, pitch modification, rendering of stereo CDs on multi-channel devices, robust speech recognition, speaker separation from recorded meeting and video conferencing, “Cocktail party” problem. When the number of observations is less than the number of sources, the problem is called under-determined. The single-channel source separation problem is the most difficult case among the under-determined separation problems in which only a single observation exists. Moreover, according to the information used, the sound source separation methods are referred to as blind when any prior information about the sources and the mixing matrix is not available [29].

2.2.1. Single-Channel and Multi-Channel Speech Enhancement

By channel, we mean the output of the microphone in the case when the observed signal has been recorded by one or more microphones, or the input of the one loudspeaker in the case when it is destined to be played back on one or more loudspeakers. This is the usual meaning of "channel" in the field of telecommunications and in some speech enhancement studies "channel" refers to the distortions (e.g., noise and reverberation) occurring when transmitting a signal instead [30].

When the only available signal is the degraded speech signal, the procedure is commonly referred as ‘single channel speech enhancement’. When there are multiple input signals (from different microphones etc.) the process is named accordingly, for instance ‘dual-microphone speech enhancement’ etc. [30].

Single-channel audio source separation problem occurs when a single observation of the mixture of a number of sources is available. In this type of problems, the aim is to estimate the individual sources constituting the mixture. In practice, it is very common to have multiple sources being active at the same time in a single recording. In such a situation, human listener has the ability to keep the attention to a single audio source in an adverse acoustical condition. However, the problem of automatically estimating several sources from one input signal is an under-determined and ill-posed challenging problem for the researchers [29].

The diagrams for single and multi-channel speech enhancement are as follows in Figure 2.2.

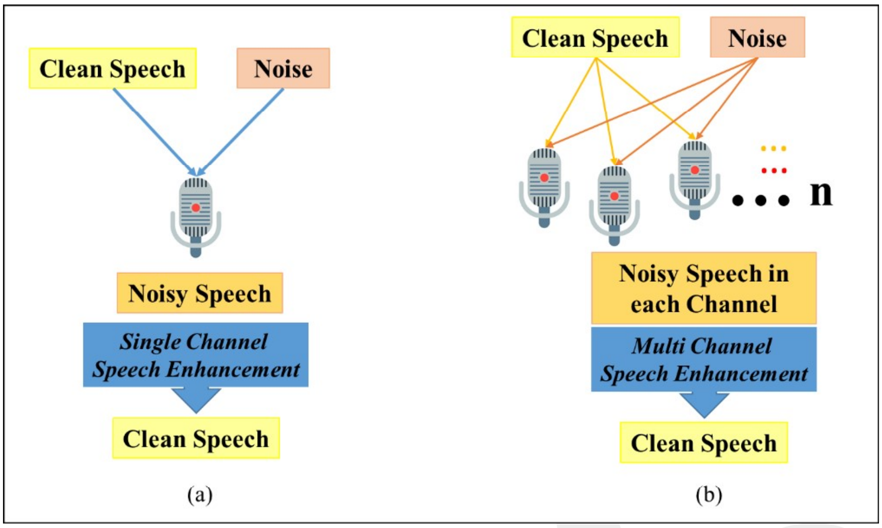


Figure 2.2 (a) Single Channel Speech Enhancement, (b) Multi-Channel Speech Enhancement

2.2.2. Audio Source Separation Techniques

Audio source separation can be also a speech enhancement method. This method can be used for separate clearly the source signal from their mixtures and also can be used for speech recognition in noisy environment. Several methods have emerged and still emerging for audio source separation.

Many algorithms have been proposed for solving the source separation problem. These algorithms can be grouped into three main categories: the methods based on Computational Auditory Scene Analysis (CASA), statistical spatial methods and statistical spectral methods [29].

Statistical spatial methods commonly use simple probabilistic source models for extracting the underlying sources of multi-channel recordings. Independent Component Analysis (ICA), which remains in this group, has been successfully used to solve source separation problems in several application areas [29].

Audio source separation is one of the basic applications of independent component analysis which is the main subject of this thesis. Since the noise outside can be considered as an independent component, ICA algorithm can be used to reduce noise or increase SNR, thus can be used for speech enhancement [15].

Audio source separation with ICA can also be used for increasing the performance of speech recognition algorithms especially in noisy environments [15].

ICA algorithms are currently in use for electroencephalographic (EEG) and magnetoencephalographic (MEG) data to separate certain source signals that are artifacts or noise sources not corresponding to brain activity. Another important application area of ICA is feature extraction which is used in data compression and pattern recognition. ICA is also used for data analysis in areas as economics, psychology or social sciences [15].

2.3. Blind Source Separation and Cocktail Party Problem

Blind source separation (BSS) is an emerging technique of multi-channel signal processing and data analysis. BSS is the process of recovering a set of signals, which are called source signals in audio source separation, from a set of mixture of those signals, which are called the observations. The term 'blind' refers to that neither the characteristics of the source nor the mixing process is known. The "blind" here means that the mixed components are latent or unobservable, actually only the mixed signals are observable. Typically, only some weak assumptions are made about the sources and mixing process. The BSS problem can be defined as a decomposition problem where the decomposed factors are unknown [16][32].

The blind audio source separation, recovering audio source signals from their linear mixtures. This issue is also referred to as 'cocktail party problem' that was first proposed by Colin Cherry in 1953. Many efforts have been dedicated to this problem in variety of science fields such as physiology, neurobiology, psychophysiology, cognitive psychology, biophysics, computer science, and engineering. [10][31]. Assume that you are in a room with lots of people talking simultaneously and you are trying to focus on one of the speakers [15]. The human hearing system has to separate those mixtures in order to follow particular speaker in the room. Even though this problem is handled easily by human brain, this is not the case in automatic source separation by applying a digital signal processing method. From the digital signal processing point of view the speakers can be considered as different simultaneously active audio sources and transmission through the room as the mixing process and the recordings made by microphones, placed in different spatial locations in the room, as the observations (mixtures) [31].

Several techniques were proposed to solve BSS problem specifically the Blind Audio Source separation problem was firstly addressed by Herault and Jutten in 1985. In their work, the sound is directly transmitted to the microphones without any delay which is known as standard blind source separation. Then in 1995, Bell and Sejnowski developed the Independent Component Analysis (ICA) method to separate the sources when they are mixed simultaneously [10]. Until today, many representational algorithms of ICA have been proposed and many publications have been made on these algorithms. There is still ongoing work on ICA algorithms, which is one of the most popular and effective methods for solving the BSS problem.

The number of sources and microphones determine the BSS problem model. Thus, when the number of microphone is more than number of sources this is known as over-determined BSS. The underdetermined BSS expression refers to the situation when the number of microphones is less than the number of sources, and classical BSS when the number of sources and microphone are equal [10]. Principles of BSS illustrated as below in Figure 2.3.

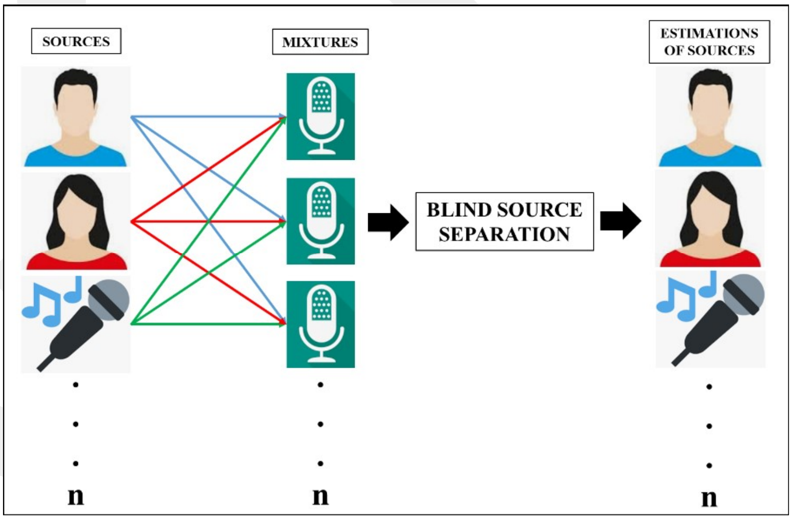


Figure 2.3 Blind Source Separation

The last thing needed to known about BSS is that; each of BSS model can occur with two different mixture model. These mixture models are; “Instantaneous Mature

Model” which the sources signals reach the microphones at the same time without any delay in time and “Convolutive Mixture Model” which refers to the mixing process that happened in a real room. The difference between the models is that; in convolutive model which more complicated one occurs the reflection caused by the room walls and the furniture. Therefore, the source signals can not arrive at the microphone at the same time.

2.4. Independent Component Analysis (ICA)

Before we deal with the “Independent Component Analysis” algorithm and the ICA-Gradient Ascent method in more detail, it may be useful to remember the concepts have encountered so far in the study. As explained before the main object of this study is to focus on an application for blind source separation. BSS is the sub application area for speech enhancement, but its scope is too wide. Therefore, there exist various type of BSS according to used sources. The best and perceptible example for BSS is using audio sources; called as “Cocktail Party Problem”. The most popular application with its good results, easy mathematical model and practical applications for cocktail party problem is “Independent Component Analysis”.

The cocktail party problem is one of the most encountered situation in daily life, so easy to understand of the problem root. This problem occurs when the various sounds are overlapped in an environmental, in this situation sources become incomprehensible and it becomes difficult to understand the desired source. For decades, the focus has been on the solution and many algorithms have been proposed for this. ICA is one of them that developing for separate the mixing signals. This is a usage field of ICA and this study interested in this, but there exist many different field for usage of ICA. PCA is another popular method which interested in this problem. Where ICA finds a set of independent source signals, PCA (Principal Component Analysis) finds a set of signals which are uncorrelated with each other.

ICA is of interest to a wide variety of scientists and engineers. Some of the applications where ICA can be used are: firing of a set of neurons, mobile phone signals (finding out their sources), brain images (e.g., functional magnetic resonance imaging, fMRI), stock prices, and as in this thesis finding the source voices [16].

ICA is widely used in various engineering disciplines like multidimensional data processing, wireless communication, speech processing, biomedical signal processing, vibration analysis and machinery fault diagnoses. In all these applications multidimensional mixed data is first recorded and then processed through ICA algorithms [32]. The applicability of ICA based on many parameters but the first requirement is; being equal number of sources and suitable recorder for the type of source in environment. Basic principle of ICA is show in Figure 2.4.

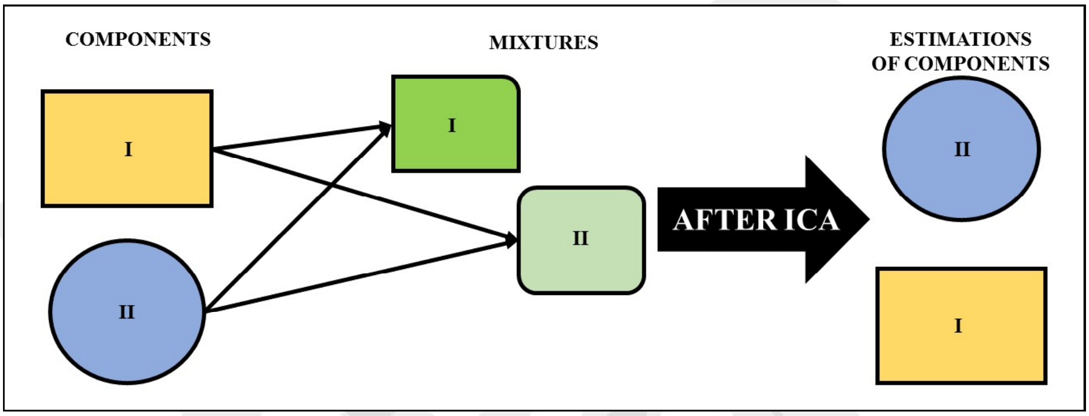


Figure 2.4 Principle of ICA

In the applications, there must be at least as many different mixtures as there are source signals. For the example of speech signals this implies that there must be at least as many microphones as there are voices.

In practice, the number of signal mixtures is often larger than the number of source signals. For example, with electroencephalography (EEG) the number of different signal mixtures of a single set of source signals is equal to the number of electrodes on the head (usually greater than 10), and the number of sources is typically less than 10. If the number of source signals is known to be less than the number of signal mixtures then the number of signals extracted by ICA can be reduced either by preprocessing signal mixtures using principal component analysis or by specifying the exact number of source signals to be extracted [16].

Resources with different physical properties will be statistically independent of each other. ICA profits from independence when separating mixtures of these resources. However, while the source components are independent of each other, their mixtures are not. Because each mixture consists of the same sources components.

Let consider the n mixtures from sources; x_1, x_2, …, x_n. It is a requirement that the number of sources equal to number of mixtures and the sources independent each other; s_1, s_2, …, s_n. If there exist mixtures matrix, there should be a mixing matrix. Namely; mixtures are consisting of sources signals multiplying by the mixture matrix A. The ICA model expressed as;

$$x = As \ ,$$

$$x = [x_1, x_2, \dots, x_n]$$

$$s = [s_1, s_2, \dots, s_n]$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \tag{2.1}$$

These mixtures and independent source components are random variables so the time index can be ignored for ICA.

CHAPTER 3
METHODOLOGY

3.1. Definition of ICA

As previously defined, ICA is a separation method for blind mixtures and the method use to independency between sources signals. In the study, we focused on the ICA method and its performance. There are many different ways to use ICA method for the BSS problem. We chose the ICA gradient ascent method because of its openness to development, ease of application and easy-to-understand mathematical principle. The block diagram of the study is shown in Figure 3.1.

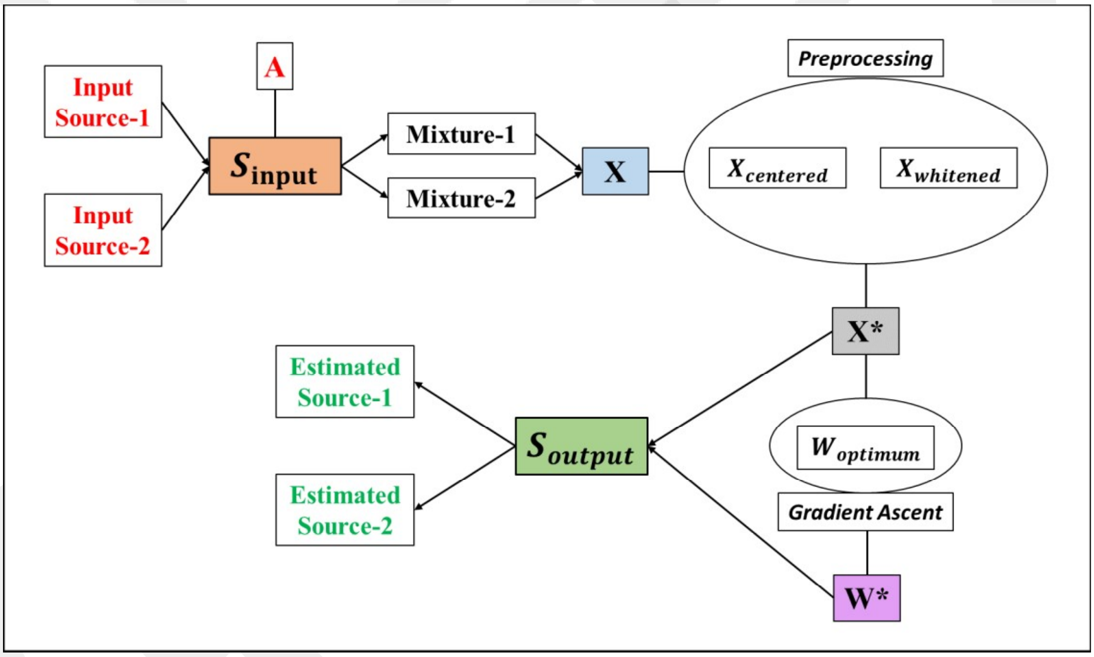


Figure 3. 1 Block Diagram of the ICA-GA Method

3.1.1. Illustration of ICA

- **How ICA Works:**

If a set of source signals are mixed to make a corresponding set of signal mixtures, then we can observe changes in three parameters:

• *independence;* the signal mixtures are no more independent as the source signals.

• *non-gaussianity;* signal mixtures tend to have a gaussian histogram while the histogram of each source signal is more non-gaussian (e.g., peaky).

• *complexity;* the complexity of the signal mixtures will be more than (or equal to) the complexity of the individual source signals.

Actually, the independent signals show a pattern as shown in part of the Figure 3.2 given below. But, when the signals are mixed the pattern evolves into the one shown in part b of the same Figure 3.2.

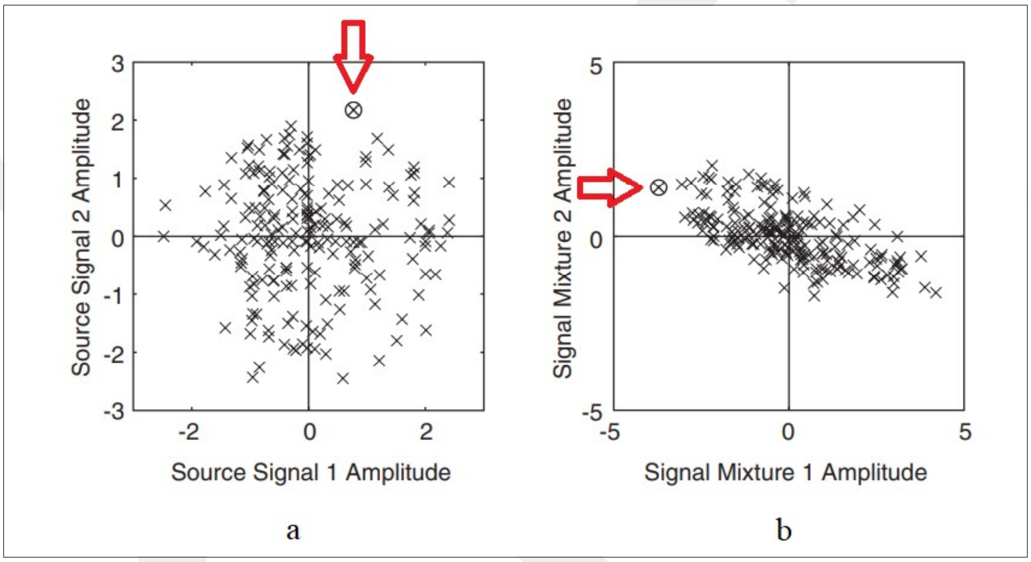


Figure 3.2 (a) Pattern of the Independent Signals (b) Pattern of when the Signals are Mixed [16]

Geometrically, we can observe one important thing; when the signals are mixed, a point shown with the arrow on the left in Figure 3.1, will correspond to the point shown again with the arrow.

It will be easier to understand its mathematical principle if worked through the simplest scenario of ICA, the separation of the two-source mix.

There exist two different sound sources in an environment, then there exist also two microphones in separate locations to record sounds from sources. An equal number of sources and receivers means that the mixing matrix is square and invertible. The mixing matrix A depends on the microphone location (distance from source) and the source signals. Namely, mixing matrix components are distances between the sources and the microphones. Then the mixtures are obtaining from sources to each receivers. The model of this scenario illustrated as follow in Figure 3.3;

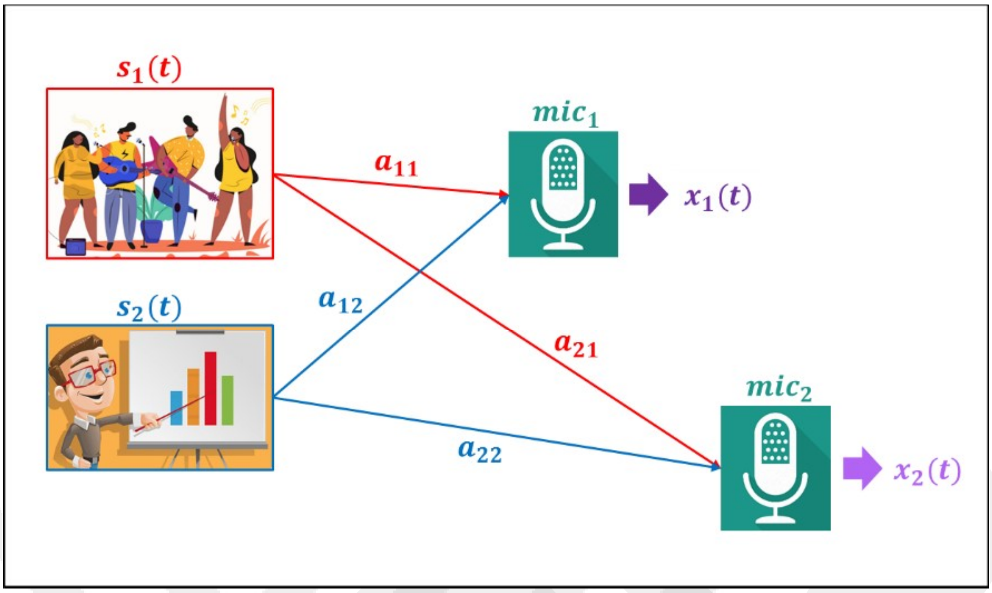


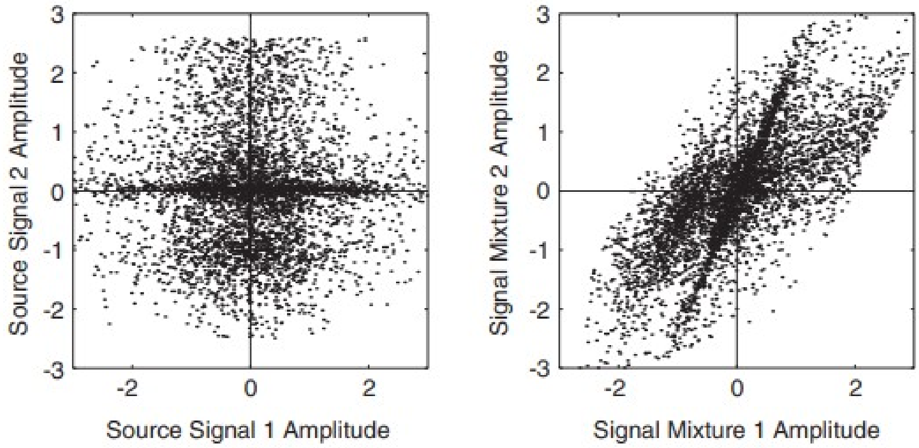
Figure 3.3 Formation of the Mixing Matrices; x_1 and x_2

- The inputs $s_1(t)$ and $s_2(t)$ are the sources signals,
- a_{11}, a_{21}, a_{12} and a_{22} are mixing coefficients,
- The outputs $x_1(t)$ and $x_2(t)$ are the mixtures, the formula expressed as;

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \tag{3.1}$$

ICA assumes that $s_1(t)$ and $s_2(t)$ are the independent but $x_1(t)$ and $x_2(t)$ are not. This independency situation is shown in Figure 3.4 with amplitude values of signals and their mixtures;



(a) (b)

Figure 3.4 Amplitude Distribution of the Mixing Matrices; x_1 and x_2 [16]

The aim of the ICA model is that find the best estimation of the sources signals. The challenge during the estimation process is, there is not any information or there is minimum information about the sources and the mixing matrix A. The process of obtained the estimations of sources with ICA illustrated in Figure 3.5.

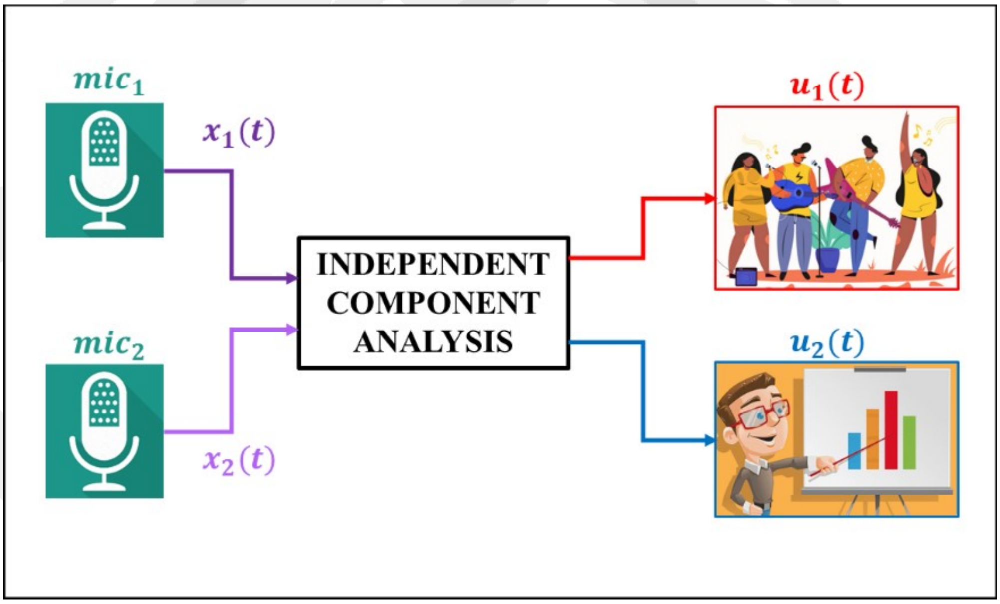


Figure 3.5 Formation of Estimations; u_1 and u_2

- The inputs $x_1(t)$ and $x_2(t)$ are the mixtures,
- The outputs $u_1(t)$ and $u_2(t)$ are the estimations of sources.

For simplicity, in order to reach estimations of sources, needed to know inverse of mixing matrix A. The inverse of mixing matrix is the unmixing matrix W. The mixing matrix A does a mapping from signal space to mixture space, and on the contrary, the unmixing matrix W implements a geometric transformation from signal mixture space to source signal space. W consists of weight vectors which each one extracts a different source signal. And each weight vector consists of a pair of unmixing coefficients. It defines as;

$$s = (s_1 , s_2)^T$$

$$s_1 = \alpha x_1 + \beta x_2 = {w_1}^T x$$

$$s_2 = \gamma x_1 + \delta x_2 = {w_2}^T x \tag{3.2}$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$x_1 = a_{11} s_1 + a_{12} s_2$$

$$x_2 = a_{21} s_1 + a_{22} s_2$$

$$x = (x_1 , x_2)^T \tag{3.3}$$

$$W = (w_1 , w_2)^T$$

$${w_1}^T = (\alpha, \beta)^T$$

$${w_2}^T = (\gamma, \delta)^T \tag{3.4}$$

- s is a vector which consist of the source signals; s_1 and s_2 ,
- A is the mixing matrix,
- x is a vector which consist of the mixtures; x_1 and x_2 ,
- W is the unmixing matrix which consist of unmixing vectors; w_1 and w_2 ,
- $\alpha,\ \beta, \gamma$ and δ are unmixing coefficients generating source signals.

For the unmixing operation, the weight vectors that we are searching for must satisfy the condition that the weight vectors must be orthogonal to each other and to the axis of the other source components.

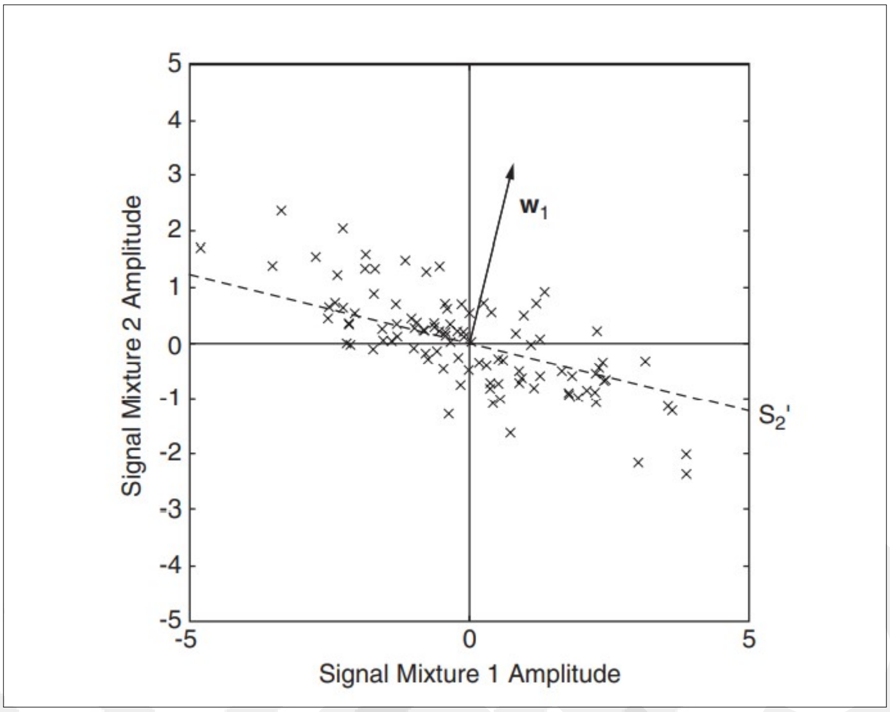


Figure 3.6 Illustrating Condition of Weight Vector w_1 according to the Axis of the Source Component [16]

The most general form of Equation set 3.2 is;

$$u = Wx = A^{-1}x$$

$$x = As$$

$$u = A^{-1}(As) = s \tag{3.5}$$

The expected result after these transactions is that the estimation signals are equal to the source signals. Summary, in ICA model, the processes proceed in Figure 3.7 as follows;

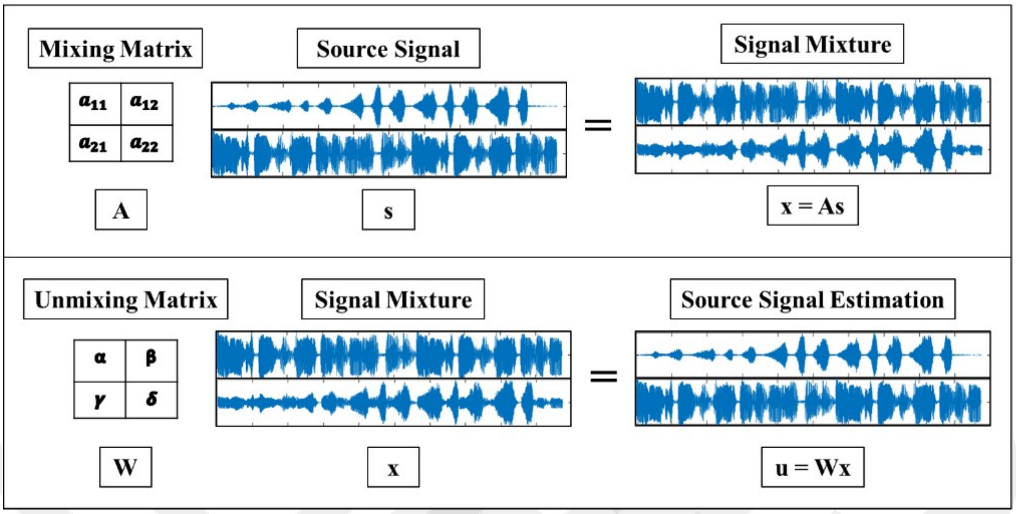


Figure 3.7 ICA Model in terms of Mixing and Unmixing Matrices

3.2. Assumptions for ICA

The first assumption for ICA is that; the sources signals are statistically independent because ICA maximizes independence of components for separate them. Independence means as a word, when there exist two components y_1 and y_2 which have some variables, the variables of y_1 does not carry any information about variables of y_2. The reverse of this is also true. Speech signals are personal due to their physical characteristics. Namely, each speech signal is statistically independent from each other.

On the other hand, technical means of statistically independent is about probability densities (pdf). y_1 and y_2 statistically independent means that their *joint pdf* is equal to product of each components' pdf and defined as Equation 3.6;

$$pdf(y_1, y_2) = pdf(y_1).pdf(y_2) \tag{3.6}$$

The second assumption in ICA model is, for simplicity mixing matrix A is invertible and square. Thus unmixing matrix W can be obtained and source estimations can be found.

The last assumption is, source signals have not Gaussian or normal distribution. ICA can not separate them. Because the Gaussian distribution signal seems like a linear combination of many independent signals. For this reason, the ICA model cannot perceive sources which have Gaussian distribution as a single source.

3.2.1. Non-Gaussianity Estimation

There exist some methods for measuring the non-Gaussianity of source.

Kurtosis is a measurement method for non-Gaussianity estimation. The moment of a signal s used for characterize the pdf of signal and Kurtosis is the fourth moment. The value of kurtosis can be negative or positive and expressed by Equation 3.7 follows for zero mean signals.

$$kurt(s) = E[s^4] = E\{s^4\} - 3(E\{s^4\})^2 \tag{3.7}$$

where E means extracted.

Entropy is the measurement of uniformity of any random system. The more unpredictable the system, the higher the entropy. The entropy of a random variable is also defining as the degree of information that taken from the observation of the variable [33]. The random variables which have Gaussian distribution has the largest entropy within the all random variables of equal variance. The entropy H definition of the signals is;

$$H(s) = -\int P(s)logP(s)ds \tag{3.8}$$

where P is probability.

Negentropy J is a shape of normalized the entropy measurement for simplicity. It can also be described as a slighlty modified version of differential entropy. Negentropy is zero for Gassian variables and in other case always positive. It can be defined as follow:

$$J(s) = H(s_{gauss}) - H(s) \tag{3.9}$$

where s_{gauss} is a Gaussian random variable of the same covariance matrix as s.

3.3. Preprocessing in ICA

Before the ICA algorithm applied to mixture signals, some preprocessing are needed and very useful for results. These preprocessing have very positive effect on performance.

3.3.1. Centering

The first preprocessing is centering applied on the data matrix x which is the observed mixtures. The aim of these process is obtaining zero-mean by subtracting the mean vector of data. The mean vector of x shown in Equation 3.9;

$$\mu = E(x) \tag{3.9}$$

and the centered vector is defined as Equation 3.10;

$$x_{centered} = x - \mu \tag{3.10}$$

This is using while independent components are estimating. Namely, while unmixing matrix estimation, profits form centered data.

$$s = A^{-1}(x_{centered} + \mu\ ) \tag{3.11}$$

3.3.2. Whitening

Other useful and effective preprocessing is whitening for ICA and aim of this preprocessing is a linearly transforming for the observed mixtures such that its components are uncorrelated and have unit variance. The eigenvalue decomposition (EVD) is a simple and frequently used method for whitening. And this is defined as;

$$E\{xx^T\} = VDV^T \tag{3.12}$$

- $E\{xx^T\}$ is the covariance matrix of x,
- V is the eigenvectors of x,
- $D = diag(d_1, d_2, \ldots, d_n)$ is the diagonal matrix of eigenvalues.

Whitening can be defined as;

$$x_{whitened} = VD^{-1/2}V^T x \tag{3.13}$$

As a result of this process, reducing the number of parameters to be estimated with whitening [34].

3.4. Ambiguities of ICA

In ICA model there are two main ambiguities which easy to understand because of principle of algorithm is known.

First ambiguity is that we can not know the order of independent components namely estimation sources. This is because source signals s and mixing matrix A are unknown, so order of terms in the ICA can change freely. This ambiguity can be expressed by the Equation 3.14;

$$x = AP^{-1}P_s$$

$$x = A's' \tag{3.14}$$

P is the permutation matrix and the elements of P_s are the initial input sources s, but ordered differently. The expression AP^{-1} in the equation is new unknown mixing matrix within the ICA.

The second ambiguity is that uncertainty of variances (energies) of independent components. Because of can not known sources s and mixing matrix A, any scalar multiplier of a source component s_j can be cancelled by dividing the corresponding column a_j of A with the same scalar. Consequently, magnitudes of the independent components can be fixed by assuming that each has unit variance: $E = \{s_i^2\}$. This ambiguity can define these equations;

$$x = As$$

$$x = \sum_{j=1}^{N} a_j\, s_j \tag{3.15}$$

where, a_j is the j-th column of the A.

$$x = \sum_{j=1}^{N} \left(\frac{1}{a_j}\, a_j\right)\left(a_j s_j\right) \tag{3.16}$$

3.5. Gradient Ascent ICA (ICA-GA)

When there is a function that determines the agreement between data and the fitting model for a particular choice of the parameters, gradient ascent can be a method for finding parameter values that maximize that function. It can also be said that for this study, finding the highest point on amount of kurtosis where each point on the different values of the two unmixing coefficients corresponds to a specific set of parameter values, and height corresponds to the function value. A set of optimal unmixing coefficients are obtained by growingly changing the values of the unmixing coefficients in order to increase the kurtosis of an extracted signal which we assume to be a source signal.

ICA gradient ascent method (ICA-GA) algorithm works by maximizing the entropy of the estimated components.

Assume that the sources in ICA model have a common cumulative density function (cdf) g and probability density function (pdf) p_s. The entropy H of the components of $U = g(u)$ defined as;

$$H(U) = H(x) + E\left\{\sum_{i=1}^{n} ln p_s(u_i)\right\} + ln|W| \tag{3.17}$$

- H is entropy,
- U is the components of cdf,
- H(U) is entropy of cdf components,
- x is mixtures in the ICA model,
- H(x) is entropy of mixtures,
- u_i is extracted components in ICA model,
- W is unmixing matrix in ICA model.

The pdf of a variable is equal to derivative of cdf ; $p_s(u_i) = g'(u_i)$**.** Then the Equation 3.17 can rewrite as;

$$H(U) = H(x) + E\left\{\sum_{i=1}^{n} ln g'(u_i)\right\} + ln|W| \tag{3.18}$$

The contribution of H(x) to H(U) is constant, and can therefore be ignored because of the entropy H(x) is unaffected by W. The object of this process is find an unmixing matrix W that maximizes the entropy of U.

The change in entropy associated with the mapping from x to U is defined as;

$$h(U) = E\left\{\sum_{i=1}^{n} lng'(u_i)\right\} + ln|W| \tag{3.19}$$

With the help of Equation 3.19 the optimal W* can find using gradient ascent on h. In order to apply gradient ascent efficiently, an expression is necessary for the gradient of h with respect to the matrix W. It is possible to proceed as finding the partial derivative of h with respect to one scalar element W_{ij} of W. W_{ij} refers that the i-th row and j-th column of W. The weight of W_{ij} refers to the proportion of the x_j in the i-th extracted component u_i.

Every component in estimated sources has the same pdf p_s or g'. The partial derivative of change in entropy in W is;

$$\frac{\partial}{\partial W_{ij}} h(U) = E\left\{\sum_{i=1}^{n} \varphi(u_i)x_j\right\} + [W^{-T}]_{ij} \tag{3.20}$$

Consider all the element of W, then the equation modified as;

$$\nabla h = W^{-T} + E\,\{\varphi(u)x^T\} \tag{3.21}$$

where ∇h is a Jacobian matrix of derivatives. These derivatives are the ijth element in $\partial h/\, \partial W_{ij}$.

When there is a finite sample N for observed mixture values of xk; k = 1,2, …, N, unmixing matrix W can be estimated as:

$$E\{\varphi(u)x^T\} = \frac{1}{N}\sum_{k=1}^{N} \varphi(u^k)[x^k]^T \tag{3.22}$$

where $\boldsymbol{u^k} = \boldsymbol{Wx^k}$,

Thus in its most general form of the gradient ascent rule can be defined as;

$$W_{new} = W_{old} + \eta \nabla h \tag{3.23}$$

where η is a small constant.

Consequently, the new rule for updating W for maximizing the entropy of $U = g(u)$ is defined as;

$$W_{new} = W_{old} + \eta \left(W^{-T} - \frac{2}{N} \sum_{k=1}^{N} \tanh(u^k)[x^k]^T \right) \tag{3.24}$$

CHAPTER 4

TESTS AND RESULTS

For the application of this study, an environment was simulated in the MATLAB program. Sound mixes in different scenarios were designed with some noise sounds, music and male/female voices. Sounds used in the application are from Noisex-92 noise database, MATLAB sound database; the sounds used were obtained in sections from movie lines and interviews of some old actors. Scenarios were prepared with two and three sources.

The scenarios prepared with two sources are given below;

Scenario 1: A male voice - noise(bird),

Scenario 2: Music - noise(factory),

Scenario 3: A male voice - a female voice.

The scenario with three source are;

Scenario 4: A male voice – noise(bird) – noise(factory).

The sampling frequencies of the given source sounds are 22050 Hz and 44100 Hz. An audio analysis program; Praat, was used to equalize the sampling frequencies and signal lengths of the source sounds.

A: mixing matrix; is one of the most important parameter in ICA method. When we selected the mixing matrix randomly, although the source separation with the ICA-GA method was successful, efficient mixing could not be obtained in all microphones in mixed signals. In one of the microphones, the action of one source was more dominant, while the action of the other sources was minimal. This can be ignored assuming the microphone is closer to the dominant source. However, since we wanted to see the effect of the method in more homogeneous mixtures, we chose the mixing coefficients of the mixing matrix as 0.4 and 0.6 for each microphone. This means; in the first microphone, while the effect of the first source is 60 percent, the effect of the second source is 40 percent, and the opposite in the second microphone. For three-source mixes, we determined the mixing coefficient values for each microphone, which determine the effect of each source, as 0.33.

The sources in Scenario 1 are illustrated in Figure 4.1 (a) and Figure 4.1 (b) as follows;

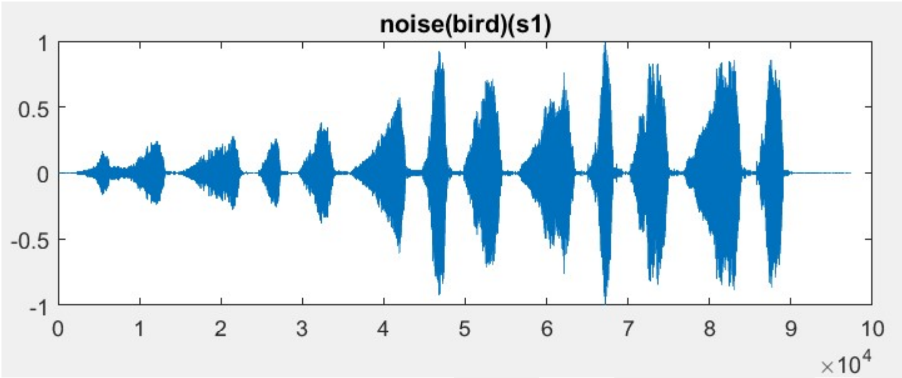


Figure 4.1 (a) Noise (bird)

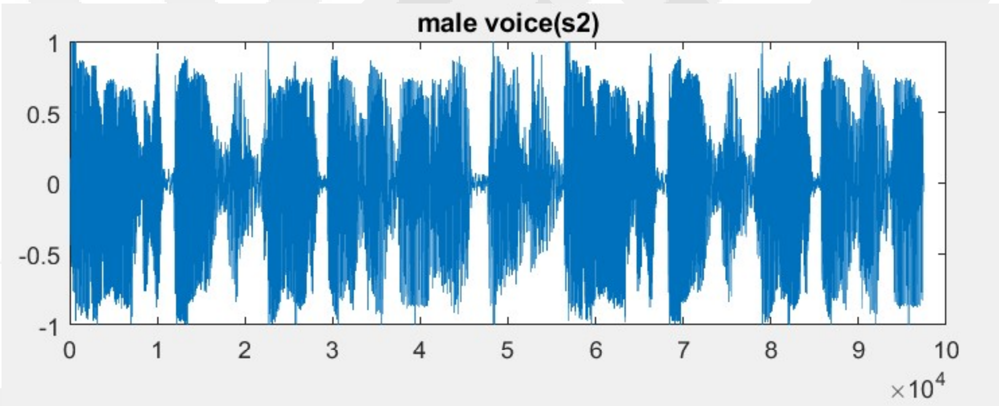


Figure 4.1 (b) Male Voice

Thus, we have proven that the ICA-GA method works even when there is a full homogeneous mixture in each microphone. But, achievement of algorithm in when A is random is better than homogeneous mixture. Figure 4.2 contains the illustration of Scenario 1. And in the Figure 4.3 below, the effect of the mixing matrix on the mixtures in the scenario involving two sources, is graphically illustrated.

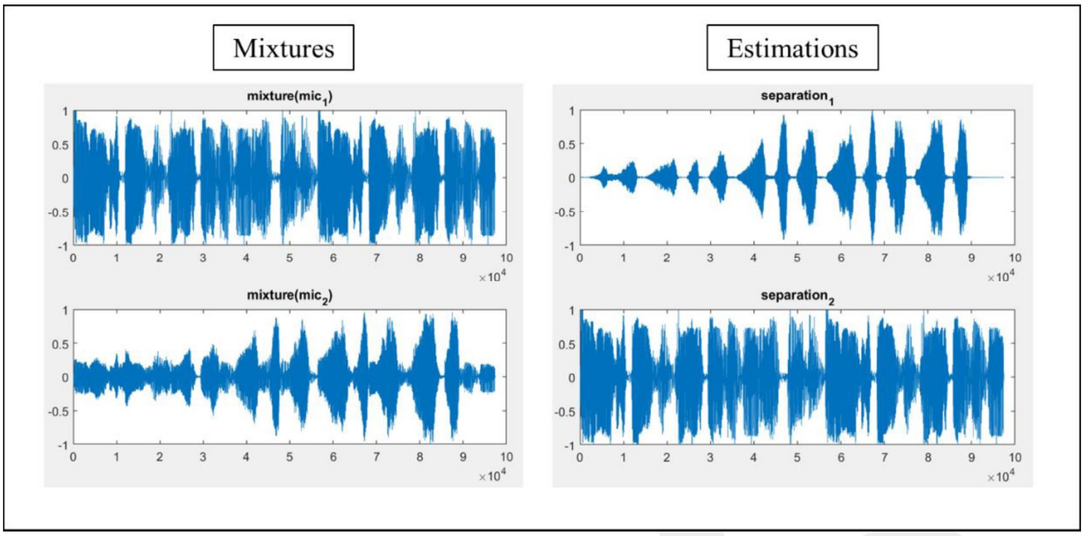


Figure 4.2 Mixtures and Estimated Sources when A is Random

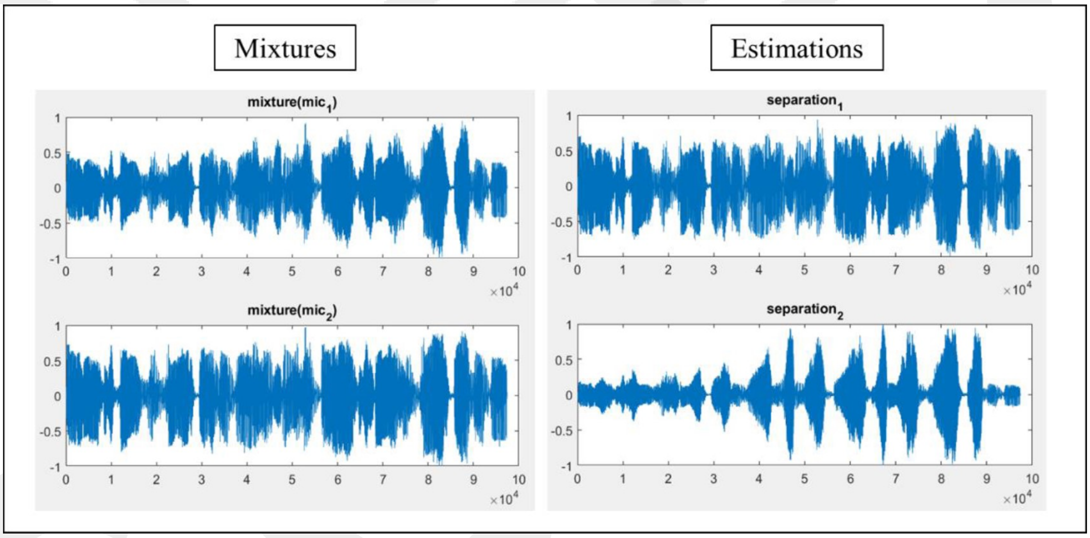


Figure 4.3 Homogeneous Mixtures and Estimated Sources

Separating full homogeneous mixtures is a more difficult and complex problem for the ICA-GA method. To overcome this challenge and perfectly separate even full homogeneous mixtures, we need to examine the effect of other important parameters on the method. While the effect of any parameter on the result was observed, all other parameters were kept constant, only the value of parameter which want to be observed was changed within a certain range and the effect of each given value on the result was examined. In order to accept a parameter value, it is not enough that it has done the separation clearly, also it must have done it in the optimum time. That is, the time index is one of the important performance criteria of an algorithm.

The maximum number of iterations; is the number of cycles processed to get unmixing matrix, W. In other words, it is the number of times the unmixing matrix W, is updated to find its optimum value. When only maximum iteration value changes by increasing in a wide range, the high values gave good results for separation. But, as the maximum number of iterations increased, the running time of the algorithm also too increased. The elapsed time is not at the optimum value. For this reason, it has been observed that only increasing the value of the maximum iteration parameter alone is not a sufficient parameter for performance.

The step-size value for gradient ascent; is eta (η) parameter is also very important in terms of performance in the ICA-GA algorithm. The eta parameter weights the ∇h value which defined in Equation 3.21 and thus determines the unmixing matrix W. h is change in entropy. According to the methodology, the eta parameter is expected to be a small constant. However, in experimental studies, a positive contribution to the result was observed when the step-size value was gradually increasing in the separation of homogeneous mixtures. In addition, giving high values to the step-size parameter did not take out the running time of the algorithm from the acceptable range.

In ICA-GA, step sizes decrease as the distance to the maximum decreases. Because, the gradient magnitude of any function decreases as its closer to maximum. Namely, the step-size is proportional to the magnitude of the gradient. Finally, based on this information, tried to get the maximum performance with more optimum (smaller) values by bringing the maximum iteration number and step-size values closer each other. In this way, both clean resources were obtained as a result of separation and the elapsed time for running time of the algorithm was kept at an optimum level. For this reason, it has been decided that this is the most effective way to get the best efficiency for separating full homogeneous mixtures with ICA-GA method.

In the application, before the estimated source signals enter the gradient ascent cycle, the correlation coefficients between the estimated signals and the original signals were calculated and expressed by $r_{initial}$. After the gradient ascent process, again the correlation coefficients between the new output signals and the original signals were calculated and expressed by r_{final}. Thus, the contribution of the gradient ascent process to the simple ICA algorithm is expressed.

In addition, the correlation values R_{xy}'s of each source signal and each output signal were calculated, and it was determined that which source signal takes place in the result in what order with the maximum value of correlation. This process can be recommended as a solution to the ICA's ambiguity about order of output signals.

Finally, the performance of the ICA-GA algorithm was expressed mathematically by calculating the total squared error (TSE) between the output signal and the original signals. In addition to this, the performance of algorithm was observed by listening to the output signals and visually with the graphic comparison.

The results were obtained with the most optimum values possible, measurement values of performance and their graphics and representations are given below.

Table 4.1 and Figure 4.4, 4.5 contain the result for Scenario 1;

Table 4.1 Parameter and Results Value for Scenario 1

Maximum number of iterations	1e4
Step-size parameter value (η)	1e2
$r_{initial}$	0.1843 0.8917 0.9828 0.4528
r_{final}	0.0058 1.0000 1.0000 0.0081
TSE	1.1621 0.2410
Elapsed Time	0 min., 43.244639 sec.

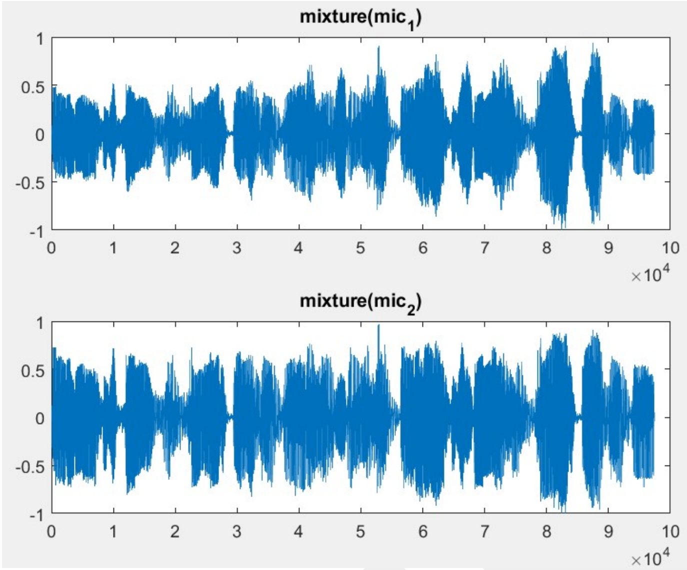


Figure 4.4 Mixtures for Scenario 1

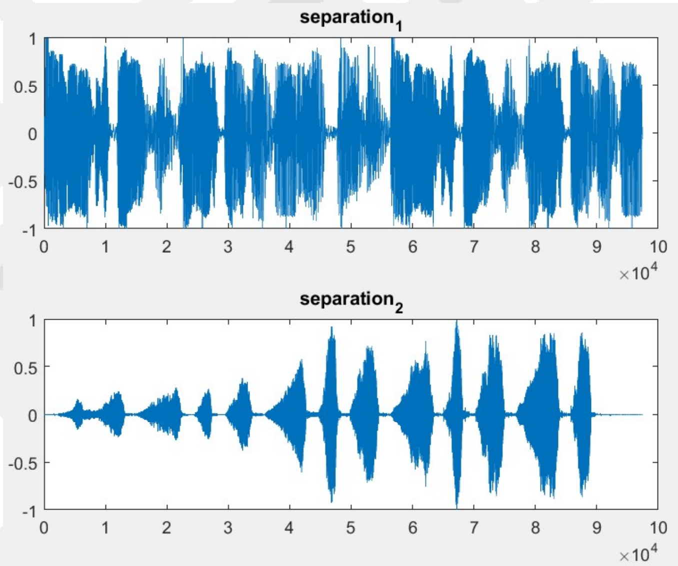


Figure 4.5 Estimations for Scenario 1

Table 4.2 and Figure 4.6, 4.7, 4.8 contain the result for Scenario 2;

Table 4.2 Parameter and Results Value for Scenario 2

Maximum number of iterations	1e2
Step-size parameter value (η)	0.1
$r_{initial}$	0.0347 0.9495 0.9997 0.3024

r_{final}	0.1788 0.9843 0.9859 0.1646
TSE	0.0013 0.0009
Elapsed Time	0 min., 0.401321 sec.

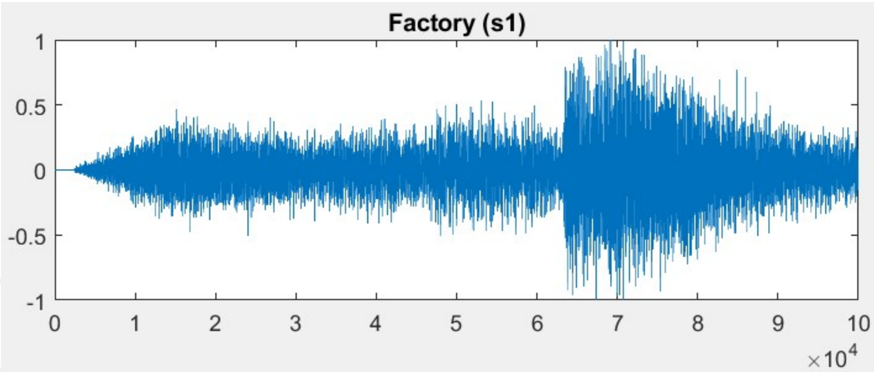


Figure 4.6 (a) Noise (factory)

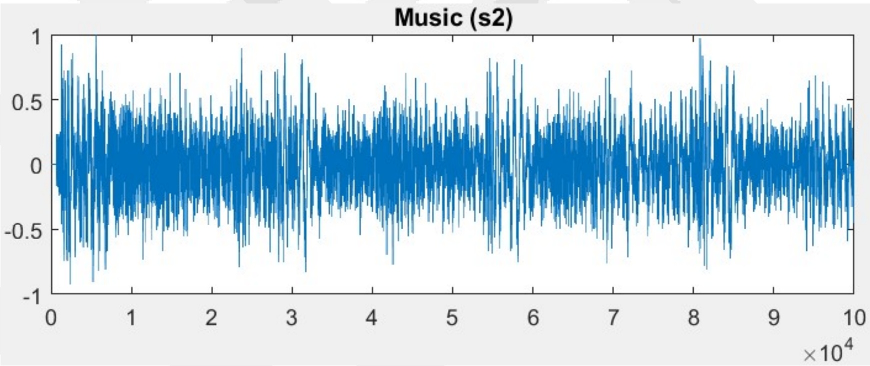


Figure 4.6 (b) Music

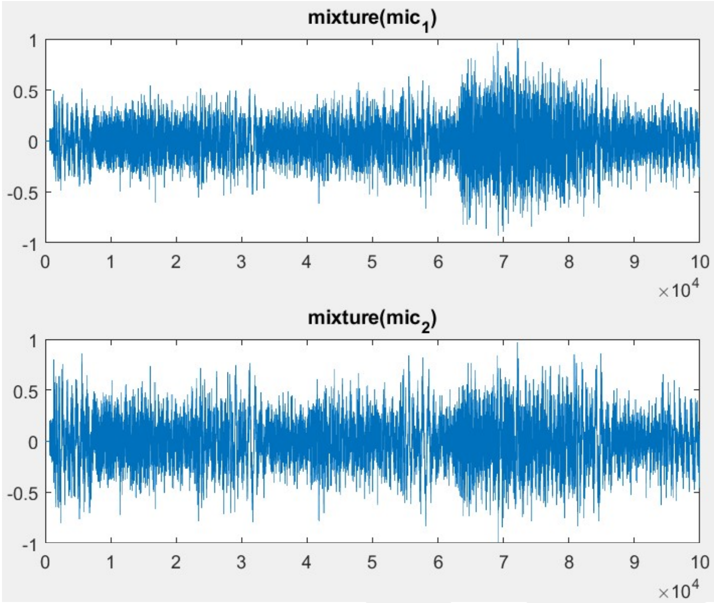


Figure 4.7 Mixtures for Scenario 2

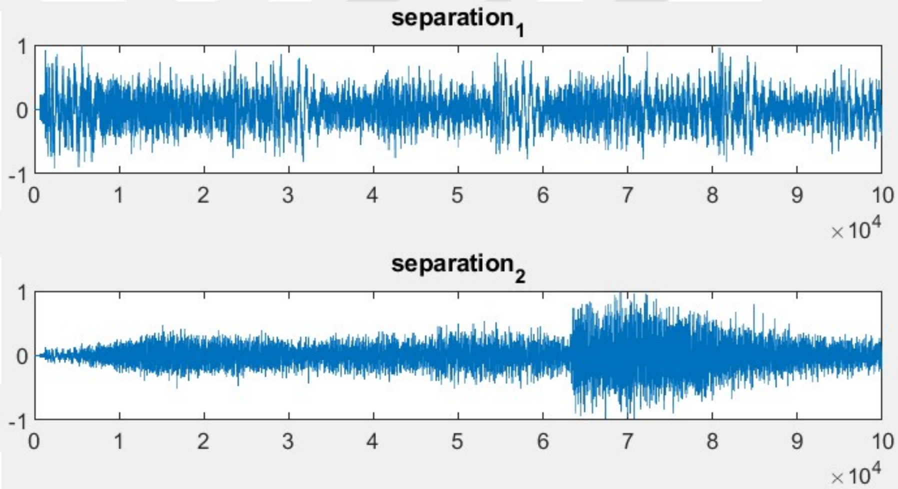


Figure 4.8 Estimations for Scenario 2

Table 4.3 and Figure 4.9, 4.10, 4.11 contain the result for scenario 3;

Table 4.3 Parameter and Results Value for Scenario 3

Maximum number of iterations	1e2
Step-size parameter value (η)	0.1
$r_{initial}$	0.1435 0.9902 0.9893 0.1373

r_{final}	0.0011 1.0000 1.0000 0.0041
TSE	0.1685 0.0697
Elapsed Time	0 min., 3.294695 sec.

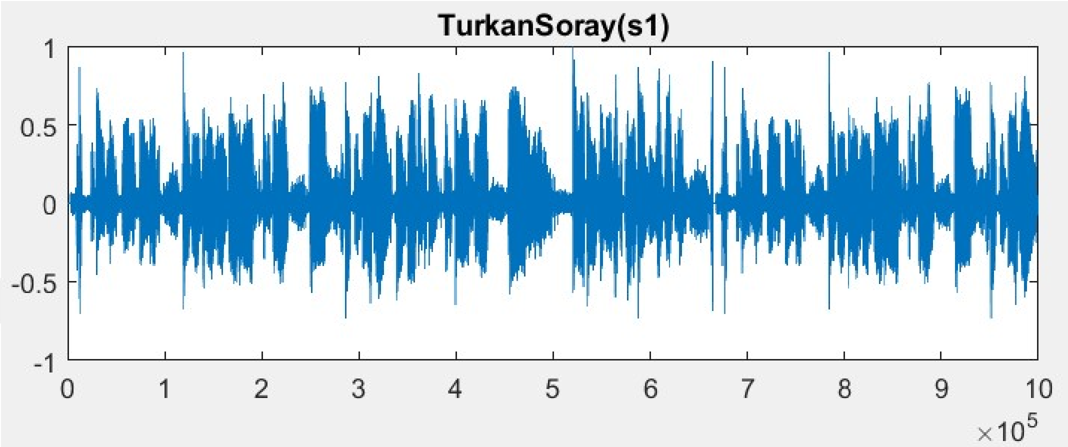


Figure 4.9 (a) Female Voice

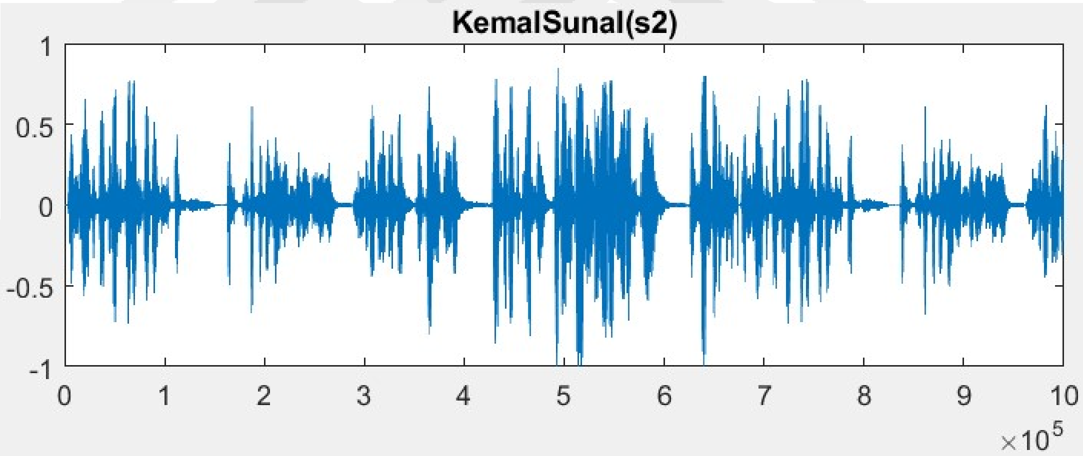


Figure 4.9 (b) Male voice

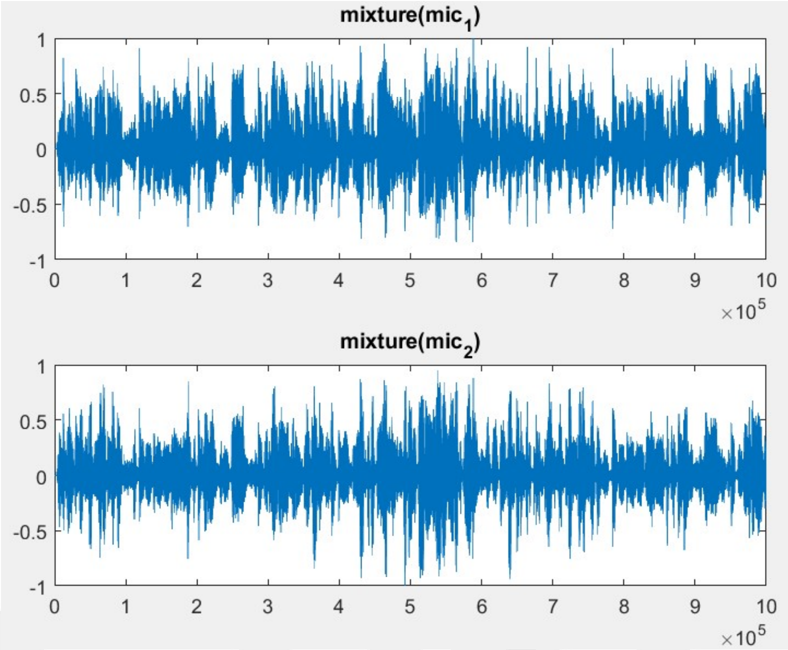


Figure 4.10 Mixtures for Scenario 3

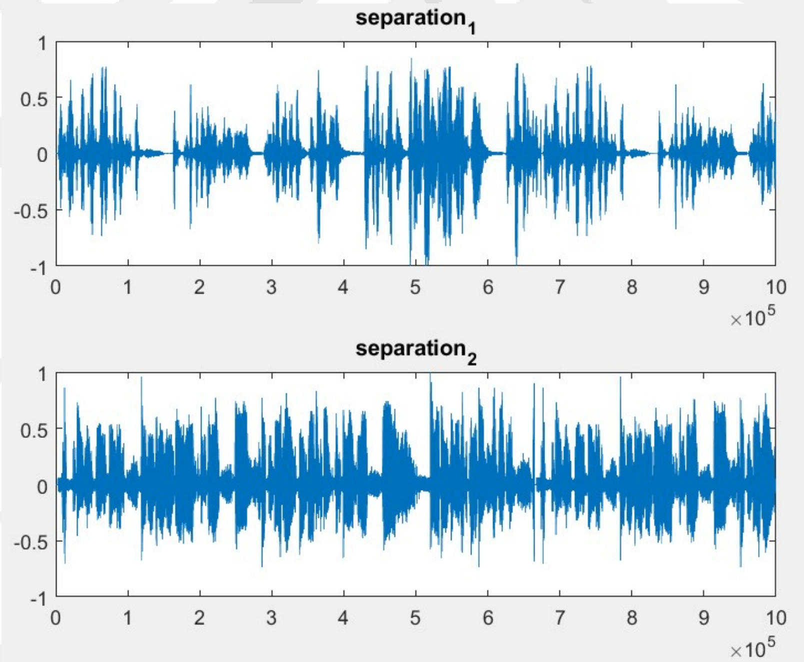


Figure 4.11 Estimations for scenario 3

Table 4.3 and Figure 4.12, 4.13, 4.14 contain the result for scenario 4;

Table 4.4 Parameter and Results Value for Scenario 4

Maximum number of iterations	1e4
Step-size parameter value (η)	1e2
$r_{initial}$	0.1907 0.1742 0.5550 0.9442 0.6786 0.8239 0.2396 0.6922 0.0902
r_{final}	0.0177 0.0052 0.9997 0.9984 0.0572 0.0242 0.0233 0.9996 0.0058
TSE	0.1660 0.0691 0.1872
Elapsed Time	2 min., 11.260485 sec.

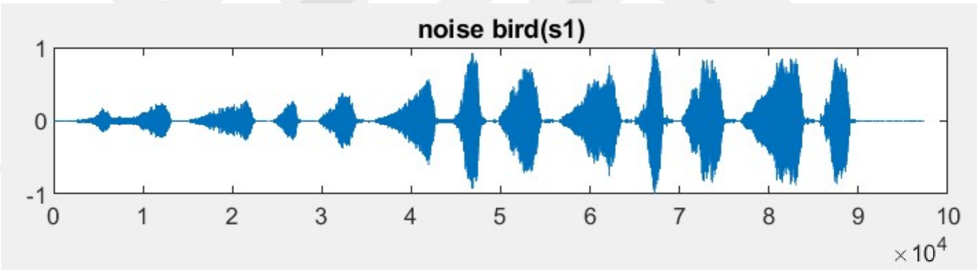


Figure 4.12 (a) Noise (bird)

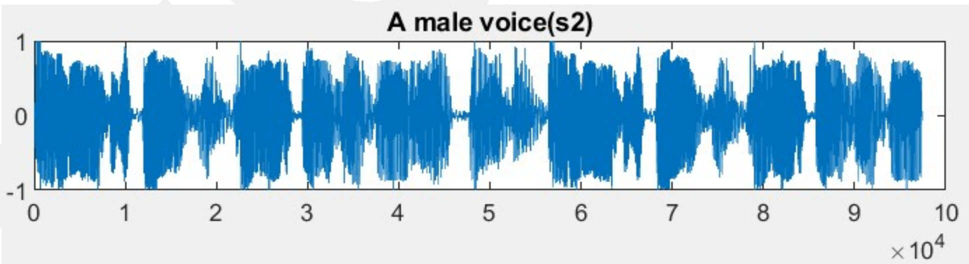


Figure 4.12 (b) Male voice

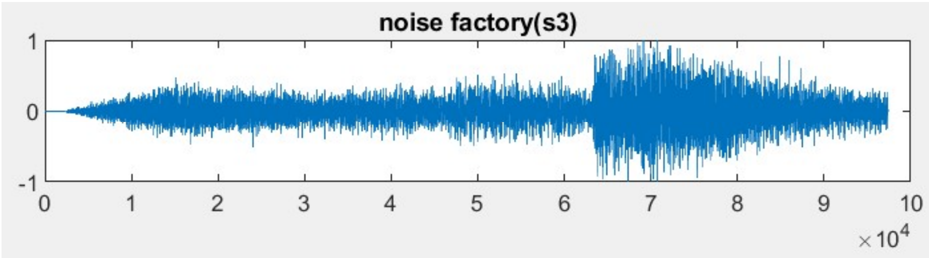


Figure 4.12 (c) Noise (factory)

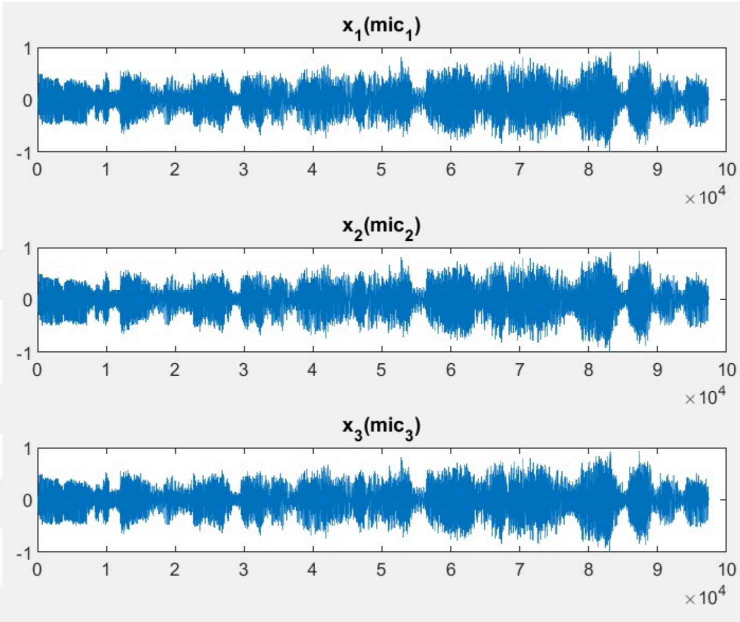


Figure 4.13 Mixtures for Scenario 4

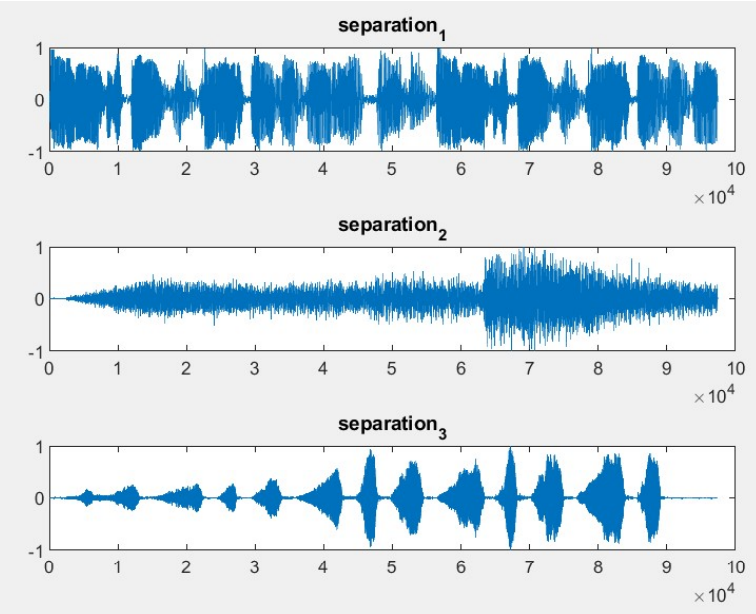


Figure 4.14 Estimations for Scenario 4

In addition to these scenarios, the method was not successful when we selected two sources from the voice of the same person in two and three-source sound separation. Because in order to be successful of ICA algorithm in separation, the sources must be independent of each other, two different voices belonging to the same person cannot be independent of each other. Therefore, we cannot expect the method to be successful in such mixtures.

CHAPTER 5

CONCLUSION

The fastest and most effective method for communication, which is the most basic need of human beings, is voice communication. Today, apart from face-to-face communication, there are many technological tools to establish voice communication. The basic operating component of these technological tools is signals. Therefore, audio signal processing applications, especially speech enhancement applications, come into play for the malfunctions in this system.

The importance of speech enhancement applications in powerful signal processing technologies is increasing. One of the situations where speech enhancement applications are most needed is the need to extract useful information from the interfering sounds. The signal processing area developed to solve this problem is audio signal separation. The need for a sub-study has occurred called as blind source separation when we have little or no knowledge about the sources of mixed sound and how they are mixed. The most familiar BSS problem is the cocktail party problem, and one of the first solutions that comes to mind for this problem is the ICA method.

In this study, we focused on the ICA methods and its performance. There are many different ways to use ICA method for the BSS problem. We chose the ICA gradient ascent method because of its openness to development, ease of application and easy-to-understand mathematical principle. The ICA-GA algorithm was applied by using the MATLAB program.

Mixing matrix, A is one of the most important parameter in ICA method. Separating full homogeneous mixtures is a more difficult and complex problem for the ICA-GA method. The ICA-GA method works even when there is a full homogeneous mixture in each microphone. But, achievement of algorithm in when A is random is better than homogeneous mixture.

Determination of the optimum value for maximum iteration has an important effect on performance of ICA-GA method. The maximum number of iterations defined as the number of cycles processed to get unmixing matrix, W. It was concluded that increasing the value of the maximum iteration parameter alone is not a sufficient parameter for performance. Because as the maximum number of iterations increased, the running time of the algorithm also too increased, that is, the elapsed time is not at the optimum value.

Another important parameter for ICA-GA algorithm is the value of step-size for gradient ascent. The parameter value defines as eta (η). According to the methodology, the step-size parameter is expected to be a small constant. However, in our study, a positive contribution to the result was observed when the step-size parameter was gradually increasing in the separation of homogeneous mixtures. In addition, giving high values to the step-size parameter did not take out the running time of the algorithm from the acceptable range. We can say that, this parameter has more successful results on the performance of algorithm than increasing the value of the maximum iteration parameter alone.

Finally, based on this information, tried to get the maximum performance with more optimum (smaller) values by bringing the maximum iteration number and step-size values closer each other. In this way, both clean resources were obtained as a result of separation and the elapsed time for running time of the algorithm was kept at an optimum level. For this reason, it has been decided that this is the most effective way to get the best efficiency for separating full homogeneous mixtures with ICA-GA method.

The contribution of the gradient ascent process to the simple ICA algorithm is expressed by comparing the values of initial and final correlation coefficients between estimated signals and original signals. The performance of algorithm was observed by listening to the output signals and visually with the graphic comparison. In addition to this, the performance of the ICA-GA algorithm was expressed mathematically by calculating the total square error (TSE) between the output signal and the original signals.

Additionally, since different voices belonging to the same person have the same physical characteristics, they are not two independent sources each other. For this reason, when solving a BSS problem with the ICA algorithm, the method was not successful when the sources were taken from the voice of same person.

The study recommended a solution to the ICA's ambiguity about order of output signals by using the correlation values R_{xy}'s of each source signal and each output signal.

Future Works:

After this study, two suggestions can be made for future studies in this field. In this study, source sounds were mixed randomly and homogeneously on MATLAB, and the mixtures were separated. In future studies, a cocktail party problem can be simulated and solved with this method in a real environment. Another issue is that objective methods such as total squared error and correlation coefficient values were used to verify the results of the study. Methods such as Mean Opinion Score (MOS), Perceptual Evaluation of Speech Quality (PESQ), Signal to Noise Ratio (SNR) values can also be added in a wider runtime.

REFERENCES

[1] Estrategic Areas, "Applications of digital audio, voice, and image processing", https://www.citsem.upm.es/en/research/strategic-areas/applications-of-digital-audio-voice-and-image-processing, [May 29, 2022].

[2] Signal Processing, "3 Reasons Why Signal Processing is the Career of the Future", https://signalprocessingsociety.org/publications-resources/blog/3-reasons-why-signal-processing-career-future, July 19, 2018 [May 29, 2022].

[3] Basic Signal Processing, "What is basic signal processing?". https://www.ssla.co.uk/basic-signal-processing/, [May 28, 2022].

[4] F. Alim, "Real-Time Audio Signal Processing for Speech Enhancement". Master thesis, University of Dokuz Eylül, İzmir, 2011.

[5] E. Mehmetçik, "Speech Enhancement Utilizing Phase Continuity Between Consecutive Analysis Windows,". Master thesis, Middle East Technical University, Ankara, 2011.

[6] M. Shujau, C. H. Ritz, and I. S. Burnett, "Speech enhancement via separation of sources from co-located microphone recordings". *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2010, pp. 137–140.

[7] H. Li and G. Ren, "Blind separation of noisy mixed speech signals based independent component analysis". *2010 First International Conference on Pervasive Computing, Signal Processing and Applications*, 2010, pp. 586–589.

[8] K. Mohanaprasad and P. Arulmozhivarman, "Comparison of fast ICA and gradient algorithms of independent component analysis for separation of speech signals". *International Journal of Engineering and Technology*, 2013, pp. 3196-3202.

[9] R. İ. Bor, “Real-time Noise Cancellation Using ICA-PSO-PE”. Master thesis, Bilkent University, Ankara, 2012.

[10] A. Mahdi, A. Elbir and F. Karabiber, “Blind audio source separation using independent component analysis and independent vector analysis”. *International Journal of Applied Mathematics, Electronics and Computers*, 4 (Special Issues), pp. 174, 2016.

[11] S. Cecelioğlu, “Tek Kanallı Toplamsal Gürültülü Konuşma Sinyal İyileştirme". Master thesis, Gazi Üniversity, Ankara, 2010.

[12] S. Dixit and Y. Mulge (2014, August), "Review on Speech Enhancement Techniques", *International Journal of Computer Science and Mobile Computing*, [Online], vol. 3, no. 8, pp. 285–290. Available: www.ijcsmc.com [May 28, 2022].

[13] B.-J. Yoon, I. Tashev2, and A. Acero, “Robust Adaptive Beamforming Algorithm Using Instantaneous Direction Of Arrival With Enhanced Noise Suppression Capability”. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2007, pp. 133–137.

[14] A. W. Bronkhorst (2015, April), “The cocktail-party problem revisited: early processing and selection of multi-talker speech”. *Attention, Perception, and Psychophysics,* [Online], vol. 77, no. 5, pp. 1465–1487. Available: https://link.springer.com/article/10.3758/s13414-015-0882-9 [May 28, 2022].

[15] D. Başaran, “Audio Source Separation with Convolutively Mixed Signals”. Master thesis, Boğaziçi University, İstanbul, 2005.

[16] J.V. Stone, *Independent Component Analysis: A Tutorial Introduction*. Learning with Kernels, 2018.

[17] Wikipedia contributors. "Noise (signal processing)". Wikipedia, The Free Encyclopedia. 11 July 2021, Web. 28 May, 2022.

[18] V. P. Tuzlukov, *Signal Processing Noise.* CRC Press, 2002, pp. 1–663.

[19] J. Castro, “What Is White Noise?”. https://www.livescience.com/38387-what-is-white-noise.html, 30 July 2013 [May 28, 2022].

[20] D. Povey, "Open Speech and Language Resources". https://www.openslr.org/resources.php, [May 28, 2022].

[21] A. Rousseau, P. Deléglise, and Y. Estève, “TED-LIUM: an Automatic Speech Recognition dedicated corpus”. http://www.ted.com, [May 28, 2022].

[22] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, “Librispeech: An Asr Corpus Based On Public Domain Audio Books”. http://www.gutenberg.org., [May 28, 2022].

[23] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, “Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition”. http://www.lrec-conf.org/proceedings/lrec2000/pdf/356.pdf, [May 28, 2022].

[24] J. Carletta, “Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus”. https://www.openslr.org/16/, [May 28, 2022].

[25] S. Renals, T. Hain, and H. Bourlard, “Recognition And Understanding of Meetings The Ami And Amida Projects”. http://sourceforge.net/projects/nite/, [May 28, 2022].

[26] D. Snyder, G. Chen, and D. Povey, “Musan: A Music, Speech, and Noise Corpus”. http://arxiv.org/abs/1510.08484, [May 28, 2022].

[27] P. Loizou, “NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms”. https://ecs.utdallas.edu/loizou/speech/noizeus/, [May 28, 2022].

[28] Y. Hu, P. C. Loizou, and S. Member, “Evaluation of Objective Quality

Measures for Speech Enhancement". IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 1, pp. 229, 2008.

[29] S. Kırbız, " Perceptual Audio Source Separation by Subspace Learning". Phd., İstanbul Technical University, İstanbul, 2013.

[30] E. Vincent, T. Virtanen and S. Gannot, *Audio Source Separation and Speech Enhancement.* John Willey and Sons Ltd., 2018, pp. 4.

[31] M. A. Keyder, "Blind Audio Source Seperation Using Nonnegative Tensor Factorization Tecniques". Master thesis, İstanbul Technical University, İstanbul, 2008.

[32] M. Altaf, A. Ahmad, and F. Alam, "Gradient Ascent Independent Component Analysis Algorithm". *Journal of Engineering. and Applications Science,* vol. 36, no. 1, pp. 125–134, 2017.

[33] A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithms and Applications". *Transactions of the ASABE*, vol. 56, no. 3, pp. 963–976, 2013.

[34] A. Mahdi, "Blind Audio Source Separation Using Independent Component Analysis And Independent Vector Analysis Methods". Master thesis, Yıldız Technical University, İstanbul, 2017.